

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

**OPTIMISATION DE LA QUALITÉ DU PROCÉDÉ DE FABRICATION DE NOURRITURE
ANIMALE PAR SPECTROSCOPIE PROCHE INFRAROUGE EN DIRECT**

**MÉMOIRE PRÉSENTÉ
COMME EXIGENCE PARTIELLE DE LA
MAÎTRISE EN INGÉNIERIE — CONCENTRATION GÉNIE MÉCANIQUE**

**PAR
JONATHAN LAPALME**

DÉCEMBRE 2025

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire, de cette thèse ou de cet essai a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire, de sa thèse ou de son essai.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire, cette thèse ou cet essai. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire, de cette thèse et de son essai requiert son autorisation.

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES
MAÎTRISE EN INGÉNIERIE — CONCENTRATION GÉNIE MÉCANIQUE (AVEC
MÉMOIRE) (1542)

Direction de recherche :

Marc-André Gaudreau Université du Québec à Trois-Rivières	Directeur de recherche
--	------------------------

Laurent Cormier Université du Québec à Trois-Rivières	Codirecteur de recherche
--	--------------------------

Jury d'évaluation

Marc-André Gaudreau Université du Québec à Trois-Rivières	Directeur de recherche
--	------------------------

Pascal Forget Université du Québec à Trois-Rivières	Évaluateur interne
--	--------------------

Tasseda Boukherroub École de Technologie Supérieure	Évaluatrice externe
--	---------------------

REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude à mes directeurs de recherche, M. Marc-André Gaudreau et M. Laurent Cormier, pour leur accompagnement tout au long de ce projet. M. Gaudreau, par son énergie inépuisable, son enthousiasme contagieux et sa passion pour l'ingénierie et la recherche, a su créer un environnement stimulant et motivant. Sa vision positive et son engagement constant ont grandement contribué à mon développement personnel et professionnel. M. Cormier, tout aussi passionné par la recherche, s'est démarqué par sa rigueur exemplaire, son esprit d'analyse remarquable et la profondeur de ses connaissances. Ses conseils avisés et sa minutie ont apporté une grande valeur aux travaux de recherche. Ensemble, ils m'ont transmis une rigueur scientifique qui continuera de guider ma démarche intellectuelle, et m'ont offert un cadre d'apprentissage riche et formateur pour lequel je suis sincèrement reconnaissant.

Je souhaite également remercier chaleureusement l'entreprise partenaire, dont l'identification doit rester confidentielle, pour sa précieuse collaboration tout au long de ce projet. J'ai eu la chance de bénéficier d'une grande confiance de la part de l'ensemble du personnel, qui m'a accueilli comme un membre à part entière de l'équipe. Ce climat de confiance et de respect mutuel a favorisé un environnement de travail propice au développement du projet de recherche. La compréhension des enjeux scientifiques démontrée par les membres de l'entreprise a permis des échanges riches et constructifs, nous poussant collectivement vers l'excellence. Leur engagement envers l'innovation et la rigueur a été une source d'inspiration tout au long de cette démarche.

Finalement, je remercie le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) ainsi que Mitacs pour leur soutien financier dans ce projet.

RÉSUMÉ

Le secteur agroalimentaire fait face à une pression croissante quant à la qualité des produits imposée tant par les réglementations en vigueur que par les attentes des consommateurs. Dans ce contexte compétitif et mondialisé, bien que les procédés de fabrication soient soumis à une variabilité importante liée à l'origine organique du produit et aux conditions de transformation, le contrôle de la qualité devient essentiel. Cette réalité souligne le besoin de solutions analytiques rapides, non destructives et fiables, capables d'assurer la maîtrise des procédés en temps réel.

Les entreprises désirant opérer un changement de paradigme envers leur système de contrôle de qualité ont avantage à faire appel à des technologies sophistiquées visant une mesure en direct des paramètres critiques de qualité. Or, ces techniques de mesure secondaires exigent le développement de modèles de calibration complexes dont les performances peuvent être affectées par plusieurs facteurs. Une entreprise de l'industrie de la fabrication d'aliments pour animaux de compagnie, consciente des bonnes pratiques en gestion de la qualité, a mis en place une lecture spectroscopique par infrarouge du taux d'humidité en direct en fin de ligne de production. Des efforts considérables ont été investis dans le développement d'une calibration sans toutefois atteindre des résultats satisfaisants.

Dans ce contexte, la présente étude a été réalisée afin d'identifier et de quantifier les facteurs pouvant affecter la précision et la robustesse de la lecture spectroscopique de l'humidité sur ce type de produit. Pour ce faire, une revue des facteurs d'influence identifiés dans la littérature scientifique ainsi qu'une analyse exhaustive de l'environnement du capteur et des procédures d'échantillonnage ont premièrement été réalisées. Par la suite, une approche méthodologique combinant des essais expérimentaux et diverses méthodes d'analyse statistique a été mise en œuvre, incluant notamment l'utilisation de statistiques descriptives, le développement d'une forêt aléatoire, l'analyse de la covariance ainsi que des analyses de la variance (ANOVA). Ces analyses ont utilisé un jeu de données existant de 422 échantillons de produits différents prélevés sur une durée de 1 an et 4 mois.

Une attention particulière a été portée à l'impact du déséquilibre du jeu de données, provoqué par un échantillonnage aléatoire, ainsi qu'à la présence de variables colinéaires, notamment le numéro de produit, la recette, la forme, la dimension, les teneurs en protéines et en gras ainsi que l'ajout de saveurs externes. Ces variables, souvent interdépendantes, ont complexifié l'interprétation des effets individuels. Par ailleurs, le taux de protéine s'est révélé exercer un effet confondant majeur sur la mesure de l'humidité, ce qui a nécessité le retrait des groupes de données associées pour isoler l'influence réelle des autres facteurs.

Les résultats obtenus ont permis de mettre en évidence l'impact significatif du taux de protéine, de la dimension du produit, de la température du produit, de la température ambiante, de la masse du sous-échantillon utilisé pour les mesures de référence et des conditions de stockage. La plupart de ces facteurs produisent des effets non linéaires sur l'écart de mesure, ce qui complique le développement d'un modèle de calibration basé sur les méthodes traditionnelles de régressions linéaires. À l'inverse, d'autres facteurs initialement suspectés, tels que la forme du produit, la recette, l'ajout de saveurs externes et la teneur en gras, se sont révélés avoir un effet négligeable sur la mesure d'humidité. De plus, l'étude a confirmé l'efficacité de la combinaison des méthodes de prétraitement SNV, *Detrend* et Savitzky-Golay. Celles-ci sont déployées pour minimiser les effets des facteurs environnementaux tels que la lumière ambiante, le spectre du convoyeur ainsi que la température interne des composants du capteur.

En somme, les travaux réalisés ont contribué à une meilleure compréhension des sources de variabilité associées à la mesure en direct du taux d'humidité par spectroscopie proche infrarouge appliquée aux aliments pour animaux de compagnie. Ils ont également permis d'identifier des pistes concrètes d'amélioration, tant sur le plan méthodologique que technologique, afin d'optimiser le développement de calibrations futures. Ces connaissances visent à soutenir l'essor de méthodes de contrôle de qualité en temps réel plus performantes, contribuant ainsi à accélérer l'adoption de technologies analytiques avancées au sein de l'industrie agroalimentaire.

Mots-clés : Contrôle qualité, Spectroscopie proche infrarouge, nourriture animale, taux d'humidité, mesure en direct, analyses statistiques.

TABLE DES MATIÈRES

REMERCIEMENTS	iv
RÉSUMÉ.....	v
TABLE DES MATIÈRES.....	vii
LISTE DES TABLEAUX	xii
LISTE DES FIGURES	xiii
LISTE DES SIGLES ET ABRÉVIATIONS	xv
CHAPITRE 1 - INTRODUCTION.....	1
1.1. Mise en contexte	1
1.2. Problématique	3
1.3. Questions de recherche	4
1.4. Objectifs	5
1.5. Structure du mémoire.....	5
CHAPITRE 2 - REVUE DE LA LITTÉRATURE	7
2.1. Les nouveaux paradigmes en gestion de la qualité	7
2.1.1. <i>Quality by Design</i>	8
2.1.2. <i>Process Analytical Technology</i>	9
2.1.2.1. Avantages d'une approche PAT.....	11
2.2. Les modes d'implémentation d'une mesure en industrie	12
2.3. Les différents capteurs dans le secteur agroalimentaire.....	12
2.3.1. La spectroscopie dans l'ultraviolet et le visible	14
2.3.2. La spectroscopie par fluorescence.....	14
2.3.3. La spectroscopie Raman	15
2.3.4. La spectroscopie infrarouge.....	16
2.3.5. La vision assistée par ordinateur	16
2.3.6. Autres technologies	17
2.4. La spectroscopie proche infrarouge	17
2.4.1. Instrumentation	18
2.4.2. Avantages.....	19
2.4.3. Limites et obstacles.....	20

2.4.4. Liens moléculaires des produits agroalimentaires	22
2.4.5. Spectres dans le proche infrarouge	22
2.4.6. Distance de pénétration	24
2.5. Chimiométrie	25
2.5.1. Prétraitement des données	26
2.5.1.1. Multiplicative Scatter Correction (MSC).....	28
2.5.1.2. Standard Normal Variate (SNV).....	28
2.5.1.3. Detrending.....	29
2.5.1.4. Mean centering	30
2.5.1.5. Savitzky-Golay.....	31
2.5.2. Calibration multivariée	32
2.5.2.1. Sélection du lot de calibration.....	33
2.5.2.2. Mesures par méthodes standards.....	35
2.5.2.3. Mesure de l'humidité	35
2.5.2.4. Acquisition et prétraitement des spectres NIR	36
2.5.2.5. Développement du modèle et validation.....	36
2.5.2.6. Valeurs aberrantes	37
2.5.2.7. Sélection et validation du modèle.....	38
2.5.2.8. Utilisation et suivi du modèle.....	41
2.6. Facteurs d'influence d'une mesure par proche infrarouge en direct.....	41
2.6.1. Description des sources consultées	42
2.6.2. Environnement	43
2.6.3. Spectromètre	44
2.6.4. Équipements	45
2.6.5. Échantillons	45
2.6.6. Méthodes	47
2.6.7. Diagramme de Pareto	48
2.7. Conclusion.....	49
CHAPITRE 3 - MÉTHODOLOGIE	51
3.1. Identification des facteurs d'influence	51

3.1.1. Étude de l'application	53
3.1.1.1. Environnement du capteur.....	53
3.1.1.2. Caractéristiques des produits.....	54
3.1.1.3. Procédure d'échantillonnage	55
3.1.1.4. Spécifications techniques du capteur NIR DA 7440™	57
3.1.1.5. Entretiens avec les experts.....	59
3.1.2. Diagramme d'Ishikawa initial.....	63
3.1.3. Sélection des facteurs	63
3.1.3.1. Facteurs non applicables	64
3.1.3.2. Facteurs constants.....	64
3.1.3.3. Facteurs non mesurables	65
3.1.4. Diagramme d'Ishikawa final.....	66
3.2. Quantification des facteurs d'influence.....	67
3.2.1. Variable dépendante.....	67
3.2.2. Jeu de données de calibration	68
3.2.3. Analyse des facteurs du jeu de données.....	70
3.2.3.1. Forêt aléatoire	71
3.2.3.2. Matrice de covariance	71
3.2.3.3. ANOVA.....	72
3.2.3.4. Modèle de calibration	74
3.2.3.5. Taux de protéine.....	75
3.2.3.6. Taux de gras.....	75
3.2.3.7. Température du produit.....	76
3.2.3.8. Température ambiante	77
3.2.3.9. Dimensions du produit	78
3.2.3.10. Forme du produit	80
3.2.3.11. Recette des produits	80
3.2.3.12. Ajout de saveur externe	84
3.2.3.13. Délai de mesure de référence.....	85
3.2.4. Essais expérimentaux.....	86

3.2.4.1.	Quantité d'échantillons	86
3.2.4.2.	Emballage des échantillons	88
3.2.4.3.	Lumière environnante	90
3.2.4.4.	Spectre du convoyeur.....	92
3.2.4.5.	Température interne du spectromètre	92
CHAPITRE 4 -	RÉSULTATS	94
4.1.	Forêt aléatoire	94
4.2.	Matrice de covariance.....	95
4.3.	Test de normalité de la variable dépendante	96
4.4.	Analyse des facteurs du jeu de données	98
4.4.1.	Modèle de calibration	99
4.4.2.	Taux de protéine	100
4.4.3.	Taux de gras	101
4.4.4.	Température du produit	103
4.4.5.	Température ambiante	106
4.4.6.	Dimensions du produit.....	108
4.4.7.	Forme du produit	109
4.4.8.	Recette des produits	110
4.4.9.	Ajout de saveur externe.....	112
4.4.10.	Délai de mesure de référence.....	115
4.5.	Essais expérimentaux.....	116
4.5.1.	Quantité d'échantillons.....	116
4.5.2.	Emballage des échantillons.....	120
4.5.3.	Lumière environnante.....	124
4.5.4.	Spectre du convoyeur	126
4.5.5.	Température interne du spectromètre.....	127
4.6.	Résumé des analyses	131
4.6.1.	Facteurs du jeu de données de calibration	131
4.6.2.	Facteurs vérifiés par essais expérimentaux	133
CHAPITRE 5 -	ANALYSE DES RÉSULTATS ET DISCUSSIONS	134

5.1. Analyse critique des résultats	134
5.1.1. Limites méthodologiques.....	134
5.1.2. Variables colinéaires et confondantes	137
5.1.3. Facteurs influents.....	138
5.1.4. Facteurs non influents	141
5.2. Discussions	142
5.2.1. Retour sur les objectifs de recherche	142
5.2.2. Comparaison avec la littérature.....	143
5.2.3. Pistes d'amélioration futures.....	144
5.2.3.1. Améliorations méthodologiques	144
5.2.3.2. Améliorations techniques et technologiques.....	145
5.2.3.3. Généralisation des résultats.....	146
5.2.3.4. Approfondissement des connaissances scientifiques	147
CHAPITRE 6 - CONCLUSION.....	148
LISTE DES RÉFÉRENCES.....	151
ANNEXE I	161
ANNEXE II	168
ANNEXE III	174
ANNEXE IV.....	176

LISTE DES TABLEAUX

Tableau 2.1 : Distances de pénétration de la lumière proche infrarouge	25
Tableau 2.2 : Méthodes de prétraitement des spectres NIR.....	27
Tableau 2.3 : Exemples de lots de calibration	34
Tableau 2.4 : Méthodes de détermination de l'humidité dans les aliments pour animaux.....	36
Tableau 3.1 : Précision des méthodes de mesure de référence de l'humidité	57
Tableau 3.2 : Informations fusionnées avec le jeu de données de calibration	69
Tableau 3.3 : Plan des mesures d'arrière-plan	91
Tableau 4.1 : Statistiques descriptives — échantillons de 230 g et 5 g	117
Tableau 4.2 : Résumé des analyses du jeu de calibration	132
Tableau 4.3 : Quantification des facteurs influents du jeu de calibration.....	132
Tableau 4.4 : Résumé des essais expérimentaux.....	133

LISTE DES FIGURES

Figure 2.1 : Illustration des quatre composants principaux du PAT	10
Figure 2.2 : Région optique du spectre électromagnétique	13
Figure 2.3 : Différents types de spectromètre proche infrarouge	18
Figure 2.4 : Modes de mesure par spectroscopie proche infrarouge.....	19
Figure 2.5 : Interactions de la radiation proche infrarouge avec une particule solide.....	19
Figure 2.6 : Exemple de l'apparence d'un spectre NIR	23
Figure 2.7 : Caractéristiques générales d'un spectre NIR	24
Figure 2.8 : Spectre d'absorption de nourriture animale et sa première dérivée.....	32
Figure 2.9 : Classification des méthodes d'analyse multivariées.....	37
Figure 2.10 : Répartition des articles consultés	42
Figure 2.11 : Diagramme de Pareto des facteurs d'influence des mesures NIR.....	49
Figure 3.1 : Montage du capteur NIR DA 7440™.....	58
Figure 3.2 : Comparaison des références internes du capteur NIR DA 7440™	60
Figure 3.3 : Performances du modèle de calibration initial.....	62
Figure 3.4 : Diagramme d'Ishikawa initial des facteurs d'influence	63
Figure 3.5 : Diagramme d'Ishikawa final des facteurs d'influence	66
Figure 3.6 : Schéma des méthodes d'analyse du jeu de données	73
Figure 3.7 : Nombre d'échantillons estimés par modèle de calibration.....	74
Figure 3.8 : Histogramme du taux de protéine.....	75
Figure 3.9 : Histogramme du taux de gras.....	76
Figure 3.10 : Histogramme de la température du produit	77
Figure 3.11 : Histogramme de la température ambiante	78
Figure 3.12 : Nombre d'échantillons par volume des produits	79
Figure 3.13 : Nombre d'échantillons par forme des produits.....	80
Figure 3.14 : Nombre d'échantillons par produits.....	81
Figure 3.15 : Dendrogramme pour 5 groupes de produits.....	82
Figure 3.16 : Dendrogramme pour 10 groupes de produits.....	82
Figure 3.17 : Dendrogramme pour 15 groupes de produits.....	83
Figure 3.18 : Nombre d'échantillons par regroupement de produits	84
Figure 3.19 : Nombre d'échantillons par saveur externe	85
Figure 3.20 : Histogramme du délai de mesure.....	86
Figure 4.1 : Importance des facteurs selon une forêt aléatoire	95
Figure 4.2 : Matrice de covariance des variables continues.....	96
Figure 4.3 : Graphique de probabilité normale — écart d'humidité.....	97
Figure 4.4 : Graphique de probabilité normale — écart d'humidité prédit	98
Figure 4.5 : Résultats graphiques ANOVA — modèle de calibration.....	99
Figure 4.6 : Résultats graphiques ANOVA — taux de protéine.....	100
Figure 4.7 : Résultats graphiques ANOVA — taux de gras.....	101
Figure 4.8 : Distribution des échantillons à haut taux de protéine selon le taux de gras	102

Figure 4.9 : Résultats graphiques ANOVA — taux de gras sans variable confondante	102
Figure 4.10 : Résultats graphiques ANOVA — température du produit.....	103
Figure 4.11 : Distribution des échantillons à haut taux de protéine selon la température du produit	104
Figure 4.12 : Résultats graphiques ANOVA — température du produit sans variable confondante.....	105
Figure 4.13 : Tendence moyenne de l'écart par température du produit.....	105
Figure 4.14 : Résultats graphiques ANOVA — température ambiante	106
Figure 4.15 : Tendence moyenne de l'écart par température ambiante	107
Figure 4.16 : Distribution des échantillons à haut taux de protéine selon la température ambiante	107
Figure 4.17 : Résultats graphiques ANOVA — volume du produit	108
Figure 4.18 : Résultats graphiques ANOVA — forme du produit	109
Figure 4.19 : Résultats graphiques ANOVA — groupe de produit (5).....	111
Figure 4.20 : Résultats graphiques ANOVA — groupe de produit (10).....	111
Figure 4.21 : Résultats graphiques ANOVA — groupe de produit (15).....	112
Figure 4.22 : Résultats graphiques ANOVA — saveur poudre	113
Figure 4.23 : Résultats graphiques ANOVA — saveur liquide.....	114
Figure 4.24 : Résultats graphiques ANOVA — saveur gras	114
Figure 4.25 : Résultats graphiques ANOVA — délai de mesure	115
Figure 4.26 : Distributions estimées — échantillons de 230 g et 5 g.....	117
Figure 4.27 : Évolution des mesures de référence d'humidité (5 g) pour trois produits	118
Figure 4.28 : Résultats graphiques ANOVA — humidité DG920.....	119
Figure 4.29 : Résultats graphiques ANOVA — humidité DGJ901.....	119
Figure 4.30 : Résultats graphiques ANOVA — humidité CHT05	120
Figure 4.31 : Évolution de la masse — échantillon ensaché en conditions réelles	121
Figure 4.32 : Évolution de la masse — échantillon ensaché en pires conditions	123
Figure 4.33 : Spectres de l'arrière-plan à différents endroits.....	125
Figure 4.34 : Spectres du convoyeur.....	126
Figure 4.35 : Graphique temporel des températures du capteur NIR DA 7440™.....	127
Figure 4.36 : Histogrammes des températures du capteur NIR DA 7440™	128
Figure 4.37 : Diagramme de boîtes des températures internes du capteur NIR DA 7440™	129
Figure 4.38 : Diagramme de boîtes de la température ambiante	130
Figure 4.39 : Matrice de covariance des températures du capteur	130
Figure 5.1 : Matrice de corrélation du jeu de données	138

LISTE DES SIGLES ET ABRÉVIATIONS

ANOVA	<i>analyse de la variance</i>
CV	<i>coefficient of variation</i> (coefficient de variation)
DA	<i>diode array</i> (matrice de diodes)
Dt	<i>detrend</i> (suppression de la tendance)
DWT	<i>discrete wavelet transform</i> (transformée en ondelettes discrète)
EMA	<i>european medicines agency</i> (agence européenne des médicaments)
EPO	<i>external parameter orthogonalization</i> (orthogonalisation des paramètres externes)
FDA	<i>food and drug administration</i> (administration des aliments et des médicaments)
FIR	<i>far infrared</i> (infrarouge lointain)
FTNIR	<i>fourier transform near infrared</i> (spectroscopie proche infrarouge à transformée de Fourier)
ICH	<i>international conference on harmonisation</i> (conférence internationale sur l'harmonisation)
IR	<i>infrared</i> (infrarouge)
KNN	<i>k-nearest neighbor</i> (k plus proches voisins)
LOD	<i>limit of detection</i> (limite de détection)
LOQ	<i>limit of quantification</i> (limite de quantification)
QbD	<i>quality by design</i> (qualité par la conception)
MC	<i>mean centering</i> (centrage sur la moyenne)
MIR	<i>middle infrared</i> (infrarouge moyen)
MSC	<i>multiplicative scatter correction</i> (correction de diffusion multiplicative)
NIR	<i>near infrared</i> (proche infrarouge)
NIRS	<i>near infrared spectroscopy</i> (spectroscopie proche infrarouge)
PAT	<i>process analytical technology</i> (technologie analytique de procédé)
PLS	<i>partial least squares</i> (moindres carrés partiels)
RMSEC	<i>root mean square error of calibration</i> (erreur quadratique moyenne de calibration)

RMSECV	<i>root mean square error of cross-calibration</i> (erreur quadratique moyenne de calibration croisée)
RMSEP	<i>root mean square error of prediction</i> (erreur quadratique moyenne de prédiction)
RPD	<i>ratio of prediction to deviation</i> (rapport de prédiction à l'écart-type)
SEC	<i>standard error of calibration</i> (erreur-type de calibration)
SECV	<i>standard error of cross-validation</i> (erreur-type de validation croisée)
SEP	<i>standard error of prediction</i> (erreur-type de prédiction)
SG	Savitzky-Golay
SNV	<i>standard normal variate</i> (variable normale centrée réduite)
UV	ultraviolet
UV-VIS	ultraviolet-visible

CHAPITRE 1 - INTRODUCTION

1.1. Mise en contexte

L'industrie agroalimentaire représente un secteur dans lequel s'exerce une pression grandissante provenant de différentes sources en regard à la qualité des produits. En effet, les exigences imposées par les normes et réglementations gouvernementales envers les entreprises obligent celles-ci à fournir des produits à la fois sécuritaires et comblant les besoins et les attentes des consommateurs (Pop et al., 2018). Selon Cen et He (2007), ceux-ci recherchent de plus en plus des sceaux de qualité ainsi que des marques de confiance sur les produits alimentaires tout en attendant des fabricants et des détaillants qu'ils fournissent des produits satisfaisants leurs exigences élevées. Par ailleurs, en raison des pressions imposées par la mondialisation et l'augmentation correspondante de la compétition, proposer des produits de qualité demeure essentiel pour ce secteur (Maduro Dias et al., 2021).

L'industrie de la fabrication d'aliment pour animaux de compagnie n'échappe pas à cette réalité. En effet, cette branche spécifique de l'agroalimentaire fait face aux mêmes défis alors qu'il existe à l'échelle mondiale un cadre réglementaire complexe mettant en évidence le besoin de renforcer le contrôle de ce type d'aliment (Pérez-Marín et al., 2008). Alors que la population d'animaux de compagnie augmente à travers le monde, la demande constante pour un produit de meilleure qualité de la part des propriétaires rend les fabricants particulièrement attentifs à garantir des aliments nutritionnels adéquats pour la santé et le bien-être des animaux (Goi et al., 2020). De Marchi et al. (2018) soulèvent également que l'intérêt croissant des gens pour le bien-être de leurs animaux de compagnie provoque un développement spectaculaire de cette industrie. Il existe maintenant sur le marché des centaines d'aliments canins et félins répondant aux différents besoins des animaux et aux nombreuses demandes de leurs propriétaires. Conséquemment, une forte compétition s'est installée dans ce secteur imposant une efficacité accrue des procédés manufacturiers (Maslovaric et al., 2011).

Les entreprises manufacturières impliquées dans la fabrication de nourriture pour animaux œuvrent donc dans un environnement compétitif de mondialisation où la qualité doit être priorisée tout en maintenant un haut niveau de profitabilité et de rentabilité. Ces entreprises

font ainsi face à des défis de taille alors que leur procédé, utilisant des matières organiques, souffre d'un degré élevé de variation causé par plusieurs facteurs (van den Berg et al., 2013). Premièrement, les matières premières utilisées sont constituées de combinaisons complexes de substances pures. Leurs propriétés physiques, extrêmement hétérogènes et variables, dépendent de la température, de la pression, de l'humidité, de leur origine et des conditions d'entreposage. Durant les étapes du procédé de fabrication, diverses réactions complexes engendrent des transformations chimiques et physiques qui sont mal comprises par les fabricants. De plus, leur nature périssable augmente le défi de livrer un produit de qualité de façon efficiente et qui perdure dans le temps (Grassi et Alamprese, 2018 ; Hitzmann et al., 2015 ; Jerome et al., 2019).

Outre la variabilité des procédés, la gestion complexe des tests de qualité et les exigences rigoureuses qui leur sont associées représentent une particularité additionnelle relative à la production alimentaire. Dans la plupart des cas, les tests de qualité effectués en production se basent sur des méthodes conventionnelles exécutées de façon manuelle par de la main-d'œuvre qualifiée et entraînée (Garrido et al., 2001 ; Ozaki et al., 2006). Ils sont effectués sur des produits finis ou semi-finis et impliquent la modification de l'échantillon ainsi que sa destruction. Ces méthodes traditionnelles sont qualifiées de laborieuses, coûteuses et chronophages (Porep et al., 2015 ; Shenk et al., 2007). De plus, les produits chimiques utilisés peuvent être dangereux pour la santé humaine et avoir un impact négatif sur l'environnement (Maduro Dias et al., 2021). Selon Alomar et al. (2006), quoique ces méthodes en laboratoire soient bien établies, le nombre de caractéristiques nutritionnelles à déterminer, le coût, la main-d'œuvre et le temps nécessaires tendent à entraver le bon degré de contrôle de la qualité de ces aliments. Mohan et al. (2005) mentionnent également que l'implication humaine induit des problèmes d'inconsistance et augmente les coûts associés à la main-d'œuvre. En outre, bien que ces tests de qualité produisent des résultats fiables, ils ne fournissent pas d'information en temps réel (Fernández-Ahumada et al., 2008 ; Novak et Jednacak, 2014) et les résultats sont souvent reçus trop tard pour arrêter la production ou la corriger à temps (Maslovaric et al., 2011).

1.2. Problématique

Une entreprise spécialisée dans la fabrication de nourriture sèche pour animaux offre des produits naturels haut de gamme devant respecter des normes de qualité les plus strictes. Le procédé de fabrication en semi-continu est constitué de multiples étapes au travers desquelles le produit passe par différentes phases, de la pâte non cuite au produit moulé, cuit, séché et refroidi. Des tests de qualité sont effectués sur les produits semi-finis en mode hors ligne et en différé pour mesurer certaines caractéristiques critiques dont les taux de protéine, de gras et d'humidité.

Selon Figura et Teixeira (2023), l'eau est un composant important d'une matière alimentaire et joue un rôle décisif dans la détermination des propriétés physiques et chimiques d'un aliment. Un taux d'humidité trop bas réduit notamment l'appétence du produit, tandis qu'un taux élevé favorise la prolifération de moisissures et de bactéries, représentant un risque pour la santé de l'animal. Le contrôle de ce paramètre à l'intérieur des limites fixées par les critères de qualité est donc un impératif que toute entreprise fabriquant ce type de produit doit satisfaire.

Afin d'atteindre une cible de taux d'humidité dans le produit, l'excédent d'eau doit être retiré par évaporation durant l'étape de séchage. Ce procédé nécessite une consommation énergétique importante qui impacte directement les coûts de production. Ainsi, fabriquer un produit plus humide permet de réaliser des économies en utilisant moins d'énergie.

L'obtention d'un produit plus humide comporte également un avantage lors de l'étape d'emballage. En effet, puisque l'eau contribue à la masse totale du produit, le poids minimum requis pour un sac de produit fini sera atteint en utilisant moins de matière que pour un produit sec, se traduisant directement en profit pour l'entreprise.

Outre les gains en énergie et les économies réalisées, le contrôle optimal du taux d'humidité permet de stabiliser les opérations en production. En effet, un produit trop sec doit être recyclé, engendrant des pertes de temps et d'argent. À l'inverse, un produit trop humide devra être retourné au séchoir retardant la production et augmentant la consommation énergétique. Indépendamment de la gestion de ces lots non conforme, les ajustements nécessaires pendant

la production sont souvent effectués trop tard dû aux tests de qualité en différé, provoquant des perturbations et accentuant la variabilité.

Afin d'améliorer la gestion de la qualité, l'entreprise a déployé une méthode de lecture en direct des paramètres critiques de qualité. Un spectromètre proche infrarouge a été installé en fin de ligne permettant notamment de suivre le taux d'humidité du produit semi-fini. Bien que l'entreprise utilise déjà en laboratoire des spectromètres de table dont la calibration est satisfaisante, le déploiement d'une calibration pour des mesures en direct dans un environnement industriel s'est révélé plus complexe que prévu.

Cette difficulté a également été constatée par la majorité des auteurs d'articles scientifiques traitant de la spectroscopie en direct, qui mentionnent que de nombreux facteurs peuvent affecter les performances d'une telle calibration. Cependant, ces affirmations s'accompagnent rarement de données quantifiables sur l'impact réel de ces facteurs. Il devient donc difficile d'identifier clairement l'origine d'un tel perturbateur ou de les classer par ordre d'importance. Par ailleurs, l'entreprise fabrique près d'une centaine de produits aux recettes et propriétés physiques variées, ce qui introduit des sources de perturbation supplémentaires non documentées dans la littérature. Cette complexité accrue rend encore plus difficile l'obtention d'une calibration fiable.

1.3. Questions de recherche

Face aux performances insatisfaisantes d'une calibration en direct qui semble affectée par de multiples facteurs, il est nécessaire de se poser les questions suivantes :

- Quels sont les principaux facteurs ayant un impact sur la calibration d'un spectromètre proche infrarouge dans un environnement industriel ?
- Quels sont les effets quantifiables réels des facteurs d'influence sur la précision de la mesure d'humidité ?

1.4. Objectifs

Ce projet de maîtrise s'inscrit dans le cadre d'un programme de recherche dont l'objectif principal est d'optimiser le contrôle de la ligne de production de nourriture pour animaux afin de réduire la variabilité de la qualité du produit. Afin de contribuer à l'atteinte de ce but, l'étude ciblée par ce mémoire se concentre sur les sous-objectifs suivants :

- Identifier les facteurs qui influencent la calibration d'un spectromètre proche infrarouge utilisé dans un environnement industriel.
- Quantifier l'effet des facteurs principaux qui influencent la précision des mesures d'humidité.

1.5. Structure du mémoire

Ce mémoire traite d'un sujet impliquant des concepts qui appartiennent à différentes disciplines dont les principales relèvent de la gestion de la qualité, la science des matériaux, les analyses statistiques multivariées, la spectroscopie, le développement de modèles de calibration ainsi que les interactions physiques de certains facteurs sur la matière étudiée et les instruments utilisés. La structure du mémoire a été pensée de manière à guider progressivement le lecteur à travers les différentes étapes de la démarche scientifique, en facilitant la compréhension de l'état de l'art, des enjeux techniques, de la méthodologie utilisée et des résultats obtenus.

Afin d'établir un portrait clair de l'état de l'art dans ce domaine, la revue de la littérature présente de manière progressive les concepts et éléments essentiels à la compréhension des enjeux étudiés. Elle débute par une introduction à la gestion de la qualité, suivie d'un survol des principaux types de mesures utilisées dans le secteur agroalimentaire. Une attention particulière est ensuite portée à la spectroscopie proche infrarouge, la méthode la plus répandue, ainsi qu'aux procédures de calibration qui y sont associées. Enfin, la revue se conclut par une synthèse des facteurs identifiés par les auteurs étant susceptibles d'influencer la qualité de ces mesures.

Le chapitre suivant présente la méthodologie adoptée dans cette étude. Il expose d'abord les approches mises en place pour la constitution des jeux de données nécessaires aux analyses, soit la liste des facteurs pertinents à l'application ainsi que les données relatives à ces facteurs.

Ensuite, il décrit les méthodes statistiques retenues pour le traitement et l'interprétation des résultats. Cette méthodologie vise à identifier les facteurs pertinents pour l'application industrielle étudiée, à structurer les données de manière cohérente, et à sélectionner les techniques d'analyse les plus appropriées pour quantifier l'influence de ces facteurs.

Dans le chapitre traitant des résultats, chacun des facteurs identifiés fait l'objet d'analyse visant à quantifier son impact sur les mesures spectroscopiques en suivant la méthodologie. À la fin de ce chapitre, un résumé permet de visualiser les résultats d'analyse.

Le chapitre de discussion propose ensuite une analyse critique de ces résultats. Il débute par l'identification des principales limites méthodologiques de l'étude, puis offre une classification des facteurs selon leur influence observée. Ce chapitre inclut également un retour sur les objectifs de recherche, une comparaison avec les données issues de la littérature scientifique, ainsi que des pistes d'amélioration pour de futurs travaux.

Le chapitre de conclusion vise à faire le bilan des travaux réalisés dans le cadre de cette étude. Il reprend les principaux constats issus des analyses et présente la manière dont ceux-ci permettent de répondre aux objectifs de recherche. Une synthèse des contributions scientifiques et pratiques est présentée, suivie d'une réflexion sur les limites de l'étude et les perspectives de recherche futures. Cette conclusion permet ainsi de situer la portée des résultats obtenus et de souligner leur pertinence dans le contexte industriel ciblé.

CHAPITRE 2 - REVUE DE LA LITTÉRATURE

Comme présenté dans le chapitre précédent, le contexte du secteur agroalimentaire ainsi que les enjeux associés sont connus et détaillés depuis plusieurs années. Avec le temps, ce contexte a cependant subi une évolution qui a généré un changement des paradigmes dans cette industrie, particulièrement ceux reliés à la gestion de la qualité des produits. D'une autre part, des percées technologiques dans le domaine de la spectroscopie et des calibrations multivariées ont permis de mesurer des éléments physico-chimiques qui autrefois étaient impossibles ou très difficiles à quantifier.

La revue de la littérature permet de constater cette évolution tout en fournissant une meilleure compréhension des éléments critiques impliqués dans la gestion de la qualité des produits alimentaires. En effet, plusieurs articles font office de résumé de l'état de l'art décrivant l'historique, les concepts impliqués ainsi que les différentes techniques et méthodes en lien avec l'analyse des paramètres critiques de qualité. De plus, une multitude d'exemples d'applications sont présentés dans des domaines variés et des études effectuées sur un large éventail de produits différents exposent leur méthodologie et leurs résultats.

L'objectif de ce chapitre est donc de présenter la revue des articles scientifiques en lien avec l'optimisation d'un procédé de fabrication à l'aide de la spectroscopie proche infrarouge ainsi que les défis associés. La première section se concentre sur le changement des paradigmes en gestion de la qualité et décrit les nouveaux concepts et systèmes innovants qui ont généré cette transformation. Les sections suivantes décrivent les modes d'implémentation des mesures et les différentes technologies de capteurs, incluant la spectroscopie proche infrarouge faisant l'objet d'une section spécifique. La chimiométrie utilisée dans le processus de calibration multivariée est ensuite détaillée, puis les multiples facteurs affectant les calibrations NIR sont expliqués avant de terminer par un court résumé à titre de conclusion du chapitre.

2.1. Les nouveaux paradigmes en gestion de la qualité

Les procédés de fabrication de produits alimentaires ont à l'origine été pensés et conçus à partir de suivis et de contrôles inférentiels. Ainsi, afin d'assurer une qualité des constituants du produit,

les variables associées au contrôle du procédé (température, pression, débit, pH et autres) sont suivies et ajustées à partir des résultats de tests périodiques effectués en fin de ligne. Selon Hitzmann et al. (2015), ces contrôles de qualité basés sur des analyses discontinues en laboratoire ne sont plus en mesure de remplir les besoins grandissants de l'industrie agroalimentaire. En conséquence, la production alimentaire a connu un changement radical en passant de ce mode de contrôle inférentiel vers un mode de mesure directe des paramètres critiques du produit (van den Berg et al., 2013).

Ce changement de paradigme a été facilité par l'introduction de nouveaux concepts et systèmes qui ont premièrement fait leurs preuves dans le domaine pharmaceutique : les approches *Quality by Design* (QbD) et *Process Analytical Technology* (PAT). En effet, ayant observé le même enjeu entourant la qualité des produits pharmaceutiques, l'agence européenne des médicaments (*European Medicines Agency* [EMA], 2003) et la *Food and Drug Administration* (FDA, 2004) ont chacun mis sur pied des guides dédiés à ces nouvelles approches. Jusqu'à ce moment, les réglementations en force étaient surtout dédiées à assurer un haut niveau de qualité des produits sans encourager l'innovation manufacturière (Bakeev, 2010).

Les approches QbD et PAT ont chacun leur historique, lesquelles ne font pas l'objet du présent mémoire, mais peuvent être consultés dans plusieurs articles et livres de référence (Bakeev, 2010 ; Chew et Sharratt, 2010 ; Cullen et al., 2014; Hitzmann et al., 2015). La section actuelle se penche surtout sur leur définition, leurs caractéristiques et leur utilisation dédiée à améliorer la qualité des produits du secteur agroalimentaire.

2.1.1. Quality by Design

Ce terme désigne une approche globale qui dicte que la qualité d'un produit ne devrait pas être testée, mais garantie par la conception (Bakeev, 2010). L'application du QbD est guidée par l'*International Conference on Harmonisation* (ICH) dans différents documents : Q8 - *Pharmaceutical Development* (*International Conference on Harmonisation*, 2006), Q9 - *Quality Risk Management* (*International Conference on Harmonisation*, 2005) et Q10 - *Pharmaceutical Quality System* (*International Conference on Harmonisation*, 2008).

L'élément central du QbD est l'espace de design. Il est défini par l'ICH comme étant un espace multivarié contenant la relation entre les variables d'entrée et les paramètres du procédé responsables d'assurer la qualité du produit. Travailler à l'intérieur de cet espace n'est pas considéré comme un changement à l'égard de la réglementation et ne requiert pas d'approbation additionnelle. Ceci accorde alors une meilleure flexibilité dans les ajustements du procédé afin d'atteindre la qualité ciblée (Hitzmann et al., 2015).

Un espace de design multivarié réel ainsi que ses stratégies de contrôle associées peuvent être mis en œuvre grâce à l'utilisation de modèles multivariés appliqués au suivi et au contrôle d'un procédé manufacturier. Il existe alors un lien évident entre cet espace et la science de l'analytique du procédé dans toutes les phases du cycle de vie d'un produit (Bakeev, 2010). Parmi les outils du QbD, un plan d'expérience peut aider à développer l'espace de design et identifier les sources de variabilité dans un procédé (Zachariassen et al., 2005).

Le QbD implique également des stratégies venant du *Process Analytical Technology* (PAT), en partie pour réduire les risques associés à la qualité du produit (FDA, 2004), mais aussi parce que ce dernier permet d'obtenir une compréhension profonde des matières premières et de la dynamique du procédé à travers un processus d'apprentissage continu (Rathore et Kapoor, 2016).

2.1.2. Process Analytical Technology

Le PAT est défini comme étant :

Un système innovant pour le design, l'analyse et le contrôle des procédés manufacturiers au travers de mesures fréquentes des paramètres critiques de qualité et de performances relatifs aux produits, aux matières premières et aux procédés dans le but d'assurer un produit final de qualité (FDA, 2004).

L'EMA (2003) le décrit de façon similaire, ajoutant que ce système est souvent utilisé en tant qu'élément d'une stratégie de qualité par design.

La FDA (2004) a donc introduit le PAT en tant que nouveau cadre réglementaire dans le but d'accroître la compréhension et le contrôle des procédés manufacturiers. Cette approche est cohérente avec leur philosophie dictant que « la qualité ne peut pas être testée dans un produit,

mais doit être intégrée ou assurée par design » [traduction libre]. Bien qu'ils aient été développés initialement pour le secteur pharmaceutique, certains outils relatifs au PAT étaient utilisés depuis des décennies dans le secteur agroalimentaire à des fins de suivis de production. Avec le temps, ces techniques de mesures en direct ont permis de développer une meilleure compréhension ainsi qu'un contrôle amélioré des procédés de fabrication permettant d'assurer en permanence une qualité prédéfinie (Johnsen, 2014).

Dans son ensemble, PAT est un système qui s'inscrit dans le cadre de concepts dédiés à la gestion de la qualité tels l'approche réglementaire fondée sur les risques, la compréhension du procédé, la qualité par design (QbD), les attributs critiques de qualité, le suivi du procédé, le contrôle du procédé par statistiques multivariées, la gestion de la qualité par les risques, la libération en temps réel, l'approche systémique intégrée, l'amélioration continue et la gestion des connaissances (Chew et Sharratt, 2010). À lui seul, PAT contient quatre composants principaux : la compréhension initiale des facteurs affectants la dynamique du procédé et la qualité du produit, l'analyse du procédé, l'analyse multivariée et le contrôle du procédé (Grassi et Alamprese, 2018). Ceux-ci permettent d'atteindre une compréhension plus profonde du procédé et un apprentissage continu de sa dynamique tel que présenté à la figure 2.1.

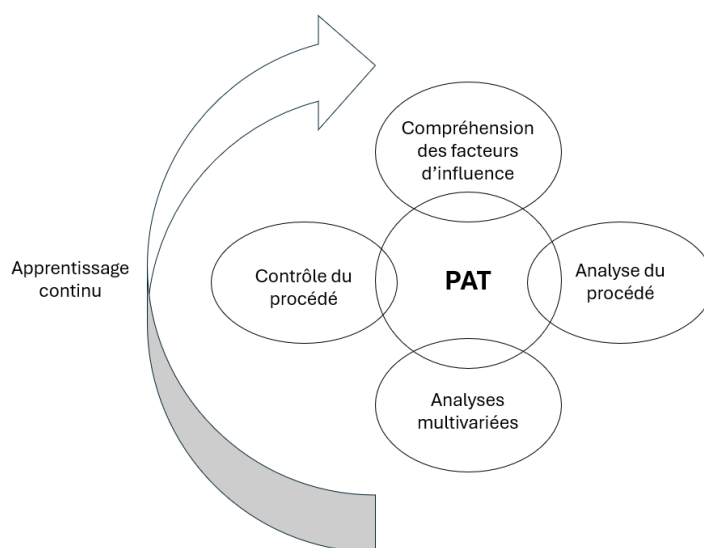


Figure 2.1 : Illustration des quatre composants principaux du PAT

Adapté de Grassi et Alamprese (2018)

Afin d'atteindre la compréhension requise du procédé, des données plus complètes sont nécessaires. Celles-ci sont idéalement obtenues en temps réel sur les matières premières, le procédé lui-même et ses sous-étapes, mais surtout sur les attributs de qualité identifiés. Bien entendu, les connaissances provenant de plusieurs domaines sont nécessaires telles la physique, la chimie, la biologie, les mathématiques, l'informatique, l'ingénierie et la science de la nutrition (Hitzmann et al., 2015). Les intervenants provenant de ces différentes spécialités doivent alors travailler ensemble en intégrant plusieurs sous-disciplines telles la chimie des procédés, la spectroscopie, les statistiques multivariées, la chimiométrie ainsi que l'instrumentation et le contrôle des procédés (Chew et Sharratt, 2010).

2.1.2.1. Avantages d'une approche PAT

Outre une augmentation de la qualité et de la sécurité des produits, l'approche PAT offre des avantages économiques et compétitifs telles l'efficacité accrue des procédés, la réduction des coûts d'opération et une meilleure validation du procédé (Cullen et al., 2014). Une implémentation réussie entraîne une utilisation optimisée des matières premières, une baisse de variation dans la qualité du produit fini, une baisse ou une élimination complète des rejets et un temps de cycle diminué. De plus, les tests de laboratoire coûteux et chronophages peuvent être remplacés par des mesures en temps réel.

En plus d'améliorer la communication avec les fournisseurs et les consommateurs, l'approche PAT a le potentiel d'aider à identifier les points de contrôle critiques, leurs limites ainsi que leurs méthodes de suivi et de contrôle (Hitzmann et al., 2015). Le processus d'apprentissage continu qui résulte alors de ce changement engendre l'innovation dans les produits et les procédés (van den Berg et al., 2013).

Selon Cullen et al. (2014), l'industrie agroalimentaire doit être encore plus alerte quant à l'utilisation du PAT en tant que cadre dédié à l'innovation des procédés et de l'assurance qualité. Les entreprises de ce domaine doivent constamment développer de nouvelles implémentations de ce système, car la demande accrue des consommateurs et des organismes réglementaires exige des produits de haute qualité, sécuritaires, à prix abordable et fabriqués de manière responsable et durable (van den Berg et al., 2013).

2.2. Les modes d'implémentation d'une mesure en industrie

Il existe différentes façons de mettre en œuvre un suivi des paramètres critiques d'un procédé industriel. L'appareil de mesure peut être installé sur la ligne de production, en proximité ou éloigné de celle-ci. Le choix de cette implémentation varie selon plusieurs facteurs relatifs aux diverses contraintes, qu'elles soient environnementales, économiques, techniques, fonctionnelles, ou relatives aux requis du capteur sélectionné. Callis et al. (1987) présentent les quatre possibilités d'implémentation, appelées aires, et fournissent leur définition respective :

- Analyse hors-ligne « *off-line* » : Caractérisé par l'extraction manuelle d'un échantillon et son transport vers un laboratoire situé à l'externe de la ligne de production.
- Analyse sur place « *at-line* » : Un instrument dédié est installé en proximité de la ligne de production. L'extraction et le transport sont faits manuellement, mais les échantillons peuvent être conditionnés de façon similaire au produit circulant sur la ligne.
- Analyse en direct « *on-line* » : Une méthode d'échantillonnage automatisée ou un système de contournement intégré dans le flux du produit rend possible l'analyse du produit dans les conditions réelles de production.
- Analyse in situ « *in-line* » : Mesure des échantillons directement sur la ligne de production en temps réel et dans les conditions exactes de production.

Selon Porep et al. (2015), « outre ces définitions, en pratique, les termes « in situ » et « en direct » sont apparemment utilisés de manière interchangeable ou les deux termes sont utilisés comme terme générique » [traduction libre]. En effet, la littérature utilise abondamment le terme « en direct » pour signifier l'utilisation d'un équipement directement au-dessus du flux de matière sur la ligne de production. Par souci de standardisation, c'est également ce terme qui est utilisé dans ce mémoire.

2.3. Les différents capteurs dans le secteur agroalimentaire

Comme cité précédemment, les produits du secteur agroalimentaire sont de nature complexe et les paramètres critiques de qualité associés peuvent être difficiles à mesurer de façon précise et fiable. Or, cette mesure essentielle constitue un pilier du déploiement de l'approche PAT, car

elle permet de fournir les données relatives à la qualité en temps réel. Heureusement, les avancées technologiques dans plusieurs domaines ont fait en sorte que différentes méthodes soient commercialisées et mises à la disposition des industries.

Parmi les différents capteurs dédiés à l'analyse et la mesure des produits alimentaires, ceux qui utilisent les méthodes optiques bénéficient d'une certaine popularité. Selon Hitzmann et al. (2015), cette popularité est expliquée par le fait que les enjeux inhérents à ces produits fragiles et complexes imposent une hygiène impeccable et une interaction physique minimale que d'autres technologies ne peuvent respecter. Les technologies relatives à l'optique utilisent toutes les interactions de l'analyte avec les radiations provenant des différentes régions du spectre afin d'identifier la nature chimique des molécules.

La région optique du spectre électromagnétique couvre la plage de longueurs d'onde allant de 100 nm à 1000 μm (Porep et al., 2015). Comme l'indique la figure 2.2, elle est divisée en trois sections principales : l'ultraviolet (100 nm à 380 nm), le visible (380 nm à 780 nm) et l'infrarouge (780 nm à 1000 μm). La zone infrarouge est séparée en trois sous-sections : proche infrarouge (NIR, de 780 nm à 2500 nm), infrarouge moyen (MIR, de 2.5 μm à 25 μm) et infrarouge lointain (FIR, de 25 μm à 1000 μm). L'illumination d'un analyte par une source optique génère un spectre d'émission ou d'absorption qui correspond à des informations précieuses sur sa structure et sa composition.

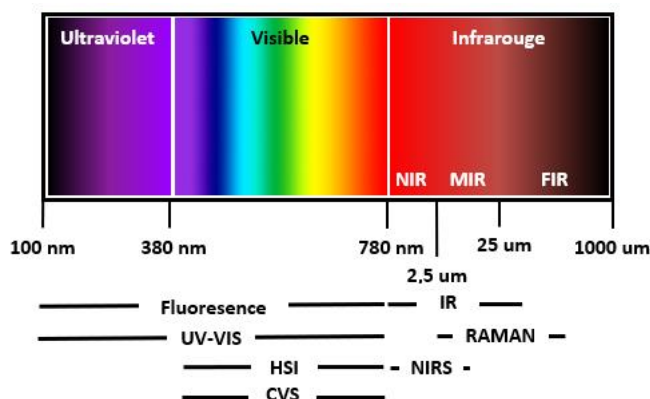


Figure 2.2 : Région optique du spectre électromagnétique

La figure 2.2 montre également les principales technologies de mesure spectroscopique couvrant la plupart des régions du spectre optique. Elles utilisent toutes le phénomène d'interaction avec les radiations provenant de différentes régions du spectre électromagnétique afin d'identifier la nature chimique des molécules (van den Berg et al., 2013). Les sections suivantes décrivent chacune d'elles.

2.3.1. La spectroscopie dans l'ultraviolet et le visible

La spectroscopie dans l'ultraviolet et le visible (UV-VIS) a été longtemps limitée à des applications de mesure des couleurs ou de concentration de composants chimiques par des photomètres à filtre (Liau et al., 2010). Cependant, les avancées technologiques concernant les filtres optiques UV, les matrices de détecteur à photodiodes et la chimimétrie ont popularisé cette technique en tant que méthode de mesure en direct. L'instrumentation relative aux appareils UV-VIS se divise en quatre catégories : les instruments à balayage numériques, les instruments à matrice de diodes, les photomètres et les spectrophotomètres contenant une matrice de photodiodes ou un dispositif à transfert de charge. Parmi les applications de ce type de spectroscopie, l'identification de certains gaz polluants est particulièrement efficace puisqu'un nombre important de ceux-ci contiennent des signatures dans la région UV (NO_x , SO_x , CO , O_3 , benzène, toluène et xylène). D'autres applications concernent l'analyse des eaux usées ou le suivi du nettoyage de réservoirs impliqués dans la fabrication de produits alimentaires ou pharmaceutiques (Ojeda et Rojas, 2009).

2.3.2. La spectroscopie par fluorescence

La spectroscopie par fluorescence utilise également les plages ultraviolet et visible du spectre optique. Elle est particulièrement attrayante en raison de sa sélectivité et de sa sensibilité accrue comparativement aux autres méthodes utilisant l'absorption telle que l'UV-VIS (Sikorska et al., 2019). Tous les instruments de spectroscopie par fluorescence contiennent une source de lumière UV, un monochromateur sur cette source et sur le signal émis par l'analyte ainsi qu'un détecteur muni d'un tube photomultiplicateur. La fluorescence d'une molécule est observée après son exposition à une source de lumière UV ou visible. Les électrons de cette molécule ayant atteint des niveaux supérieurs d'excitation retournent alors à leur état initial en libérant

des photons. L'utilisation de cette technologie reste cependant limitée à un groupe restreint de molécules, car peu de composés organiques présentent une fluorescence (Valeur et Berberan-Santos, 2012 ; van den Berg et al., 2013). Elle est cependant très utile à des fins de vérification de sécurité des produits, par exemple, pour détecter certains contaminants sur les feuilles d'épinard (Everard et al., 2014 ; Lee et al., 2014) ou la présence de résidus de charcuterie responsables de la contamination croisée sur des lames de tranches (Beck et al., 2015). Elle est également utile pour suivre le processus de fermentation au levain dans les produits de boulangerie (Grote et al., 2014).

2.3.3. La spectroscopie Raman

La spectroscopie Raman peut être appliquée dans les cas où aucune fluorescence n'est observée dans la bande d'excitation correspondante (Hitzmann et al., 2015). C'est une technologie couvrant la région spectrale située de 2.5 μm à 200 μm de longueur d'onde adaptée à la fois à l'analyse des solides, des liquides ou des gaz ou toute forme intermédiaire ou combinatoire (Jestel, 2010). L'instrument Raman est principalement constitué d'un laser monochromatique en tant que source d'illumination et d'un filtre dédié à supprimer la diffusion Rayleigh. La lumière réfléchiée par l'analyte est collectée par un détecteur puis séparée en différentes longueurs d'onde par un spectrographe. Le principe fait appel au processus de diffusion inélastique impliquant un transfert d'énergie entre la lumière incidente et les molécules cibles illuminées. Les matières idéales pour ce type de spectroscopie sont celles qui comportent des liens moléculaires doubles ou triples, différents isomères ou encore les composés soufrés et symétriques. Elle est également plus performante que la spectroscopie infrarouge en ce qui concerne l'analyse de substances composées majoritairement d'eau. Grâce à cette technologie, He et al. (2015) ont été en mesure d'identifier certains additifs alimentaires interdits (le Sudan I, la rhodamine B et le vert malachite) dans les aliments et le poisson. De leur côté, Ilaslan et al. (2015) ont réussi à mesurer la quantité de glucose, fructose et sucrose dans les boissons gazeuses et Wang et al. (2014) se sont concentrés sur la quantification de sucre, d'éthanol et de glycérol durant le processus de fermentation du vin.

2.3.4. La spectroscopie infrarouge

Alors que la spectroscopie Raman implique des mouvements vibratoires symétriques, la spectroscopie infrarouge est plutôt dédiée à l'analyse des mouvements asymétriques. Également utilisée pour analyser les solides, les liquides et les gaz, sa région fonctionnelle occupe le spectre de l'infrarouge moyen (MIR) et celui du proche infrarouge (NIR) (Coates, 2010). Le MIR (2.5 à 25 μm) couvre les vibrations fondamentales de la plupart des liaisons chimiques courantes comportant les éléments légers à moyens, en particulier la plupart des gaz et des composés organiques. Le NIR (780 nm à 2.5 μm) couvre les vibrations harmoniques et combinatoires des groupes fonctionnels reliés à l'atome d'hydrogène et produit une information extrêmement utile sur les composés organiques (Simpson, 2010). Les instruments dédiés à la spectroscopie infrarouge permettent d'effectuer des lectures autant en transmittance qu'en réflectance. Des équipements simples utilisant une longueur d'onde fixe sont non dispersifs et permettent de mesurer ou de détecter les gaz. D'autres équipements peuvent générer un spectre complet représentatif de l'absorption ou de l'émission de l'analyte et fournir une analyse par transformée de Fourier via un interféromètre. Dans tous les cas, une source illumine l'échantillon et un détecteur enregistre l'information. Certains équipements disposent d'un détecteur à matrice et nécessitent un filtre variable ou un réseau de diffraction pour générer le spectre.

2.3.5. La vision assistée par ordinateur

La vision assistée par ordinateur permet l'inspection de l'apparence d'un produit, ce qui touche directement à l'un des principaux aspects concernant les préférences des consommateurs à l'égard des produits alimentaires (Hitzmann et al., 2015). Les applications de vision ont été étudiées avant celles utilisant la spectroscopie et leur nombre a souvent été plus élevé. Ceci s'explique par la disponibilité des équipements nécessaires et la simplicité d'intégration. En effet, pour réaliser une application de vision assistée par ordinateur, il faut une caméra numérique, un ordinateur et un logiciel de traitement d'image. Souvent, une source lumineuse, soit intégrée à la caméra ou externe à celle-ci, doit être ajustée pour éviter la sous-exposition ou la surexposition des images. Plusieurs méthodes de traitement d'image sont disponibles, notamment différents réseaux de neurones, l'algorithme Viola-Jones ou une combinaison de

ceux-ci. Cette technique est particulièrement utile pour suivre le processus de cuisson des produits de boulangerie (Paquet-Durand et al., 2012) ou pour classer automatiquement les fruits en identifiant leurs défauts (Zhang et al., 2017).

2.3.6. Autres technologies

D'autres technologies sont également utilisées dans des contextes plus précis ou en combinaison avec les techniques de spectroscopie. Parmi celles-ci, les systèmes acoustiques par ultrasons sont utilisés pour analyser la qualité de différentes farines et certaines étapes dans les procédés de fabrication de produits de boulangerie (Jerome et al., 2019). La tomographie par rayons X est largement appliquée dans cette industrie pour l'analyse rapide et en ligne des niveaux de fermentation, de l'énergie libérée, de l'augmentation du volume et de l'analyse des cavités gazeuses complexes. Autrement, elle est la méthode la plus utilisée pour l'analyse in situ des processus cinétiques impliquant des solides cristallins (Pomerantsev et Rodionova, 2012).

D'une autre part, l'imagerie hyperspectrale peut être utilisée en commun avec la spectroscopie proche infrarouge. Cette technique produit une dimension spatiale supplémentaire qui permet entre autres l'identification et la localisation des composés chimiques dans les échantillons non homogènes (Manley, 2014).

2.4. La spectroscopie proche infrarouge

La spectroscopie proche infrarouge (NIRS) est la technique de mesure vibrationnelle et indirecte la plus répandue dans le secteur agroalimentaire. Elle existe depuis plus de cinquante ans et son historique est décrit de façon exhaustive par plusieurs auteurs (Niemöller et Behmer, 2008 ; Ozaki et al., 2006 ; Shenk et al., 2007). Les précurseurs de cette technologie, Norris et Ben-Gera (1968) sont à l'origine de son utilisation et de sa popularisation en agriculture. La spectroscopie NIR peut être définie comme « l'analyse des matières quant à leur tendance à absorber la lumière d'une certaine région du spectre électromagnétique » [traduction libre] (Niemöller et Behmer, 2008). La mesure est en fait une empreinte moléculaire de l'échantillon permettant une identification et une quantification très précise de l'analyte.

2.4.1. Instrumentation

Un spectromètre proche infrarouge typique se compose d'une source de rayonnement, d'un dispositif de sélection de longueurs d'onde (monochromateur), d'un détecteur photoélectrique et d'un système informatique pour l'acquisition et le traitement des données spectrales (Porep et al., 2015). Il existe plusieurs types de spectromètre, qu'ils soient à longueur d'onde unique ou à spectre complet. La figure 2.3 montre les principaux types : monochromateurs à réseau, polychromateurs à matrice de photodiodes (DA) et analyseurs par transformée de Fourier (FTNIR).

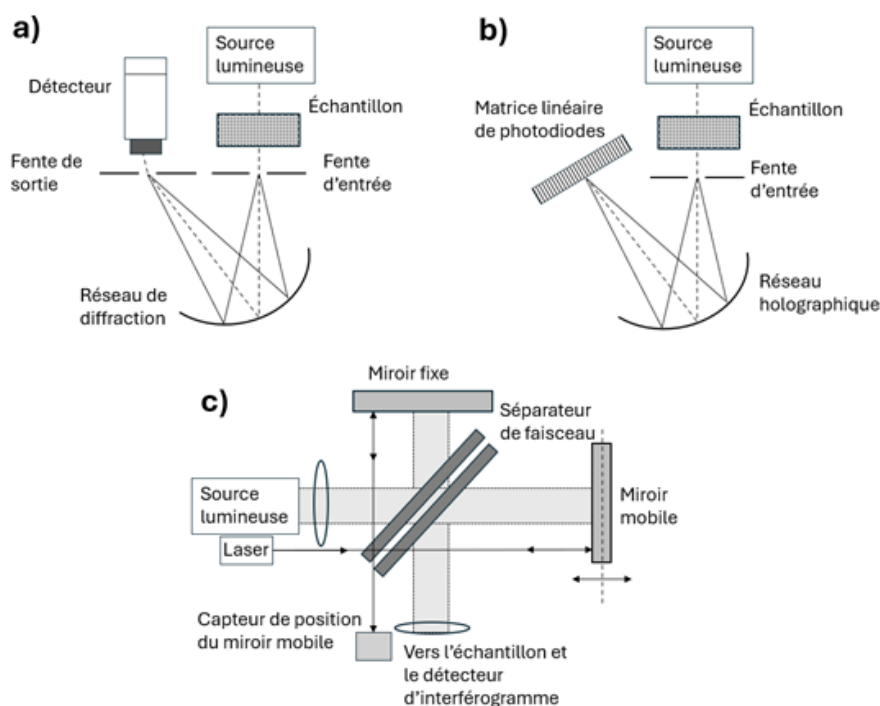


Figure 2.3 : Différents types de spectromètre proche infrarouge
a) monochromateur à réseau, b) polychromateur à matrice de photodiode, c) FTNIR.

Adapté de Skvaril et al. (2017)

Selon Skvaril et al. (2017), l'extraction du spectre NIR d'un échantillon peut se faire via différents modes de mesures dont les prédominantes sont la transmittance, la transflexion, la réflexion diffuse et la transmittance à travers un milieu diffusant. Ces modes sont représentés dans la figure 2.4.

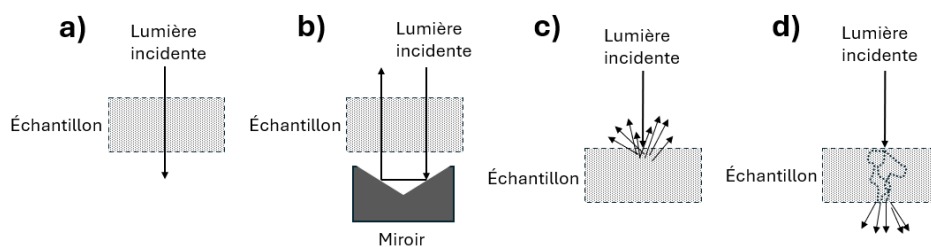


Figure 2.4 : Modes de mesure par spectroscopie proche infrarouge
 a) Transmittance, b) Transflexion, c) Réflexion diffuse, d) Transmittance à travers un milieu diffusant.

Adapté de Skvaril et al. (2017)

Shenk et al. (2007) précisent de leur côté que la lumière monochromatique d'un instrument NIR interagit avec les matières organiques en réflexion, réfraction, absorption, diffraction et transmittance. Ces modes sont présentés à la figure 2.5.

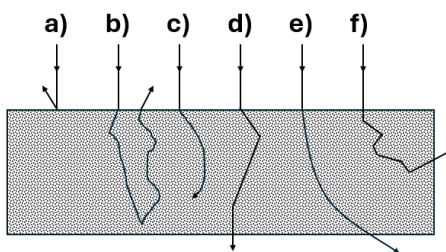


Figure 2.5 : Interactions de la radiation proche infrarouge avec une particule solide
 a) Réflexion spéculaire, b) Réflexion diffuse, c) Absorption, d) Transmittance, e) Réfraction, f) Diffusion.

Adapté de Shenk et al. (2007)

2.4.2. Avantages

Plusieurs auteurs s'entendent sur le fait que la spectroscopie NIR procure des avantages exceptionnels comparativement aux méthodes d'analyse standards en laboratoire (Cen et He, 2007 ; Niemöller et Behmer, 2008 ; Pérez-Marín, 2014 ; Porep et al., 2015 ; Qiu et al., 2005). Ces avantages sont notamment les suivants :

- Une analyse rapide des données : l'appareil génère un spectre complet presque instantanément et les composants chimiques sont déterminés rapidement au travers d'un modèle de calibration ;

- Une opération simplifiée : l'opération comprend peu d'étapes et peut être exécutée par un seul opérateur avec un minimum de formation ;
- Une bonne efficacité d'analyse : le spectre NIR contient beaucoup d'information. Ainsi, de multiples composants chimiques peuvent être testés simultanément ;
- Des tests non destructifs : l'énergie NIR est très basse et n'endommage pas les échantillons. Ceux-ci peuvent être testés tels quels sans préparation spécifique ;
- Un faible coût d'opération : les ressources humaines et matérielles requises sont minimales comparativement aux méthodes de tests standards en laboratoire ;
- Une technologie non polluante et sécuritaire : aucun agent chimique n'est utilisé et aucun déchet dangereux n'est rejeté, ce qui rend cette méthode respectueuse de l'environnement et sécuritaire pour les humains ;
- Une installation en direct ou in situ : les systèmes de mesure peuvent être installés directement sur la ligne de production pour fournir les informations en temps réel.

De façon globale, la spectroscopie NIR assiste le personnel dans différents processus de production en fournissant des solutions en ligne adaptées aux besoins d'une entreprise manufacturière. Ces solutions peuvent aider à maximiser les profits de l'entreprise en détectant les problèmes et en réagissant le plus rapidement possible en ajustant le processus (Baeten et al., 2016). De plus, son utilisation peut amener les entreprises à obtenir une meilleure compréhension des matières premières, des produits et de leur procédé de fabrication. Ainsi, elle représente un potentiel immense pour l'amélioration de la qualité, la réduction des coûts, l'optimisation des temps de production et l'augmentation du rendement (Blanco et Villarroya, 2002 ; Camino-Sánchez et al., 2015; Hoyer et al., 2005 ; Irudayaraj et Reh, 2008 ; Niemöller et Behmer, 2008).

2.4.3. Limites et obstacles

La spectroscopie offre donc beaucoup d'avantages comparativement aux méthodes traditionnelles. Par contre, alors que la plupart des auteurs mentionnent un coût d'opération faible, Cen et He (2007) ainsi que Huang et al. (2008) stipulent que le prix d'acquisition des équipements NIR est élevé et représente un frein autant aux applications pratiques qu'à celles

dédiées à la recherche. En outre, le processus de calibration des appareils NIR requiert une quantité appréciable d'échantillons dédiés à construire un modèle robuste et précis. Ces échantillons doivent tous être analysés en laboratoire de façon manuelle entraînant des coûts de matières et de main-d'œuvre considérables (Camino-Sánchez et al., 2015; Maslovaric et al., 2011).

La performance et la précision des modèles de calibration dépendent directement des mesures de références effectuées en laboratoire ainsi que de la qualité de la distribution des lots d'échantillons utilisés. Ceux-ci doivent impérativement représenter l'étendue de la production observée incluant l'écart entre les différentes variantes de produits, la variabilité des procédés et des matières premières ainsi que l'effet des facteurs externes et des conditions ambiantes changeantes. Le processus de construction du modèle doit alors s'étendre sur une longue période de temps (des mois ou des années) afin de tenir compte de tous ces facteurs (Camino-Sánchez et al., 2015). Ozaki et al. (2006) précisent même que dans bien des cas, il est nécessaire de développer une calibration spécifique à chaque commodité et constituant, ce qui augmente la charge de travail requise afin d'atteindre l'objectif.

Selon Manley (2014), l'un des désavantages majeurs de la spectroscopie NIR consiste en sa dépendance aux méthodes de références standards et à la chimométrie nécessaires à la construction d'un modèle de calibration. Le processus requiert des connaissances en statistiques et en analyses multivariées afin d'établir les relations entre les concentrations mesurées en laboratoires et les caractéristiques du spectre obtenu. Cette analyse représente une tâche complexe, longue et ardue.

L'environnement industriel reste un milieu hostile pour les spectromètres et collecter des spectres représentatifs et fiables dans de telles conditions reste difficile (Porep et al., 2015). En effet, l'inhomogénéité des échantillons intacts, leur mouvement ainsi que leur orientation variable constituent un réel défi. De plus, les lectures souffrent de fluctuations causées par la température. Ainsi beaucoup d'études ont été faites sur la spectroscopie NIR, mais pratiquement aucune n'a réussi à implanter un système de suivi in situ ou en direct dans un environnement de production réelle. En effet, ces études ont plutôt réalisé leurs mesures dans

des environnements contrôlés afin d'éliminer l'impact de ces éléments externes perturbateurs. Les applications en industrie restent donc peu nombreuses (Grassi et Alamprese, 2018).

2.4.4. Liens moléculaires des produits agroalimentaires

Les produits agroalimentaires sont constitués de liens moléculaires primaires réagissant particulièrement aux longueurs d'onde de la région proche infrarouge du spectre électromagnétique (Qiu et al., 2005). Cette particularité rend alors possible la mesure des composés organiques et biologiques (H. Huang et al., 2008). En effet, plusieurs types de vibrations moléculaires existent dans la région NIR entre autres sous la forme Carbone-Hydrogène (C–H), Azote-Hydrogène (N–H) et Oxygène-Hydrogène (O–H) (H. Huang et al., 2008 ; Manley, 2014; Ozaki et al., 2006 ; Shenk et al., 2007). Ces vibrations hydrogéniques sont causées par l'étirement, la flexion ou la déformation des liens sous l'effet de l'exposition à la lumière. Alors que les vibrations fondamentales de ces liens génèrent une absorption dans la région MIR, des bandes spectrales reliées à leur seconde et troisième harmonique peuvent être mesurées en proche infrarouge.

En plus d'être caractérisé par les vibrations harmoniques, le spectre NIR d'un analyte comporte également des combinaisons de celles-ci. En effet, dans les molécules polyatomiques, deux ou plusieurs modes vibratoires peuvent interagir de manière à provoquer des changements d'énergie simultanés et donner lieu à des bandes d'absorption combinatoires (Blanco et Villarroya, 2002 ; Kumaravelu et Gopal, 2015). Ceci fait en sorte que la spectroscopie NIR représente une solution idéale pour l'identification des paramètres majeurs des aliments tels que l'humidité, le gras, les protéines, l'amidon ou les acides aminés (Niemöller et Behmer, 2008). Outre la composition chimique de l'échantillon, les différentes interactions des molécules, par exemple les liaisons d'hydrogènes et les interactions dipolaires, sont également en mesure de fournir des informations sur les propriétés physiques de l'analyte.

2.4.5. Spectres dans le proche infrarouge

Une bande d'absorption est produite lorsque la radiation NIR à une fréquence spécifique vibre à la même fréquence que le lien moléculaire de l'échantillon (Shenk et al., 2007). Le spectre d'un

produit agroalimentaire contient typiquement entre sept à dix pics. La plupart de ces pics représentent la composition de plusieurs autres pics individuels ne pouvant être résolus séparément, comme le présente la figure 2.6. La représentation privilégiée d'un spectre NIR est la forme logarithmique de l'inverse de la réflectance ($\log [1/R]$). Dans cette représentation, l'absorption est assumée linéaire avec la concentration de l'élément dans l'analyte.

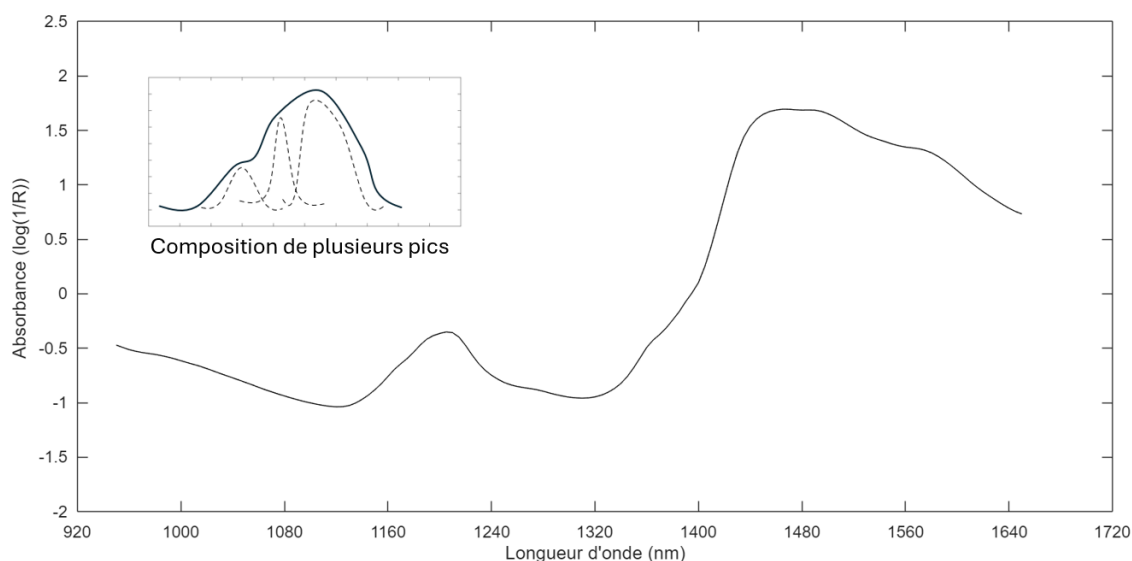


Figure 2.6 : Exemple de l'apparence d'un spectre NIR

Adapté de Shenk et al. (2007)

Les bandes d'un spectre NIR sont définies par trois paramètres : leur emplacement, leur amplitude et leur largeur. Cependant, puisque ces bandes sont le résultat de la superposition de plusieurs pics et que l'amplitude des vallées entre ceux-ci a des formes différentes causées par les bandes adjacentes et superposées, l'estimation de l'emplacement reste la mesure la plus précise des trois paramètres. Comme le montre la figure 2.7, en plus de pics et de vallées, le spectre NIR est caractérisé par une ligne de base et contient un effet multiplicatif relié à la variation de grosseur des particules.

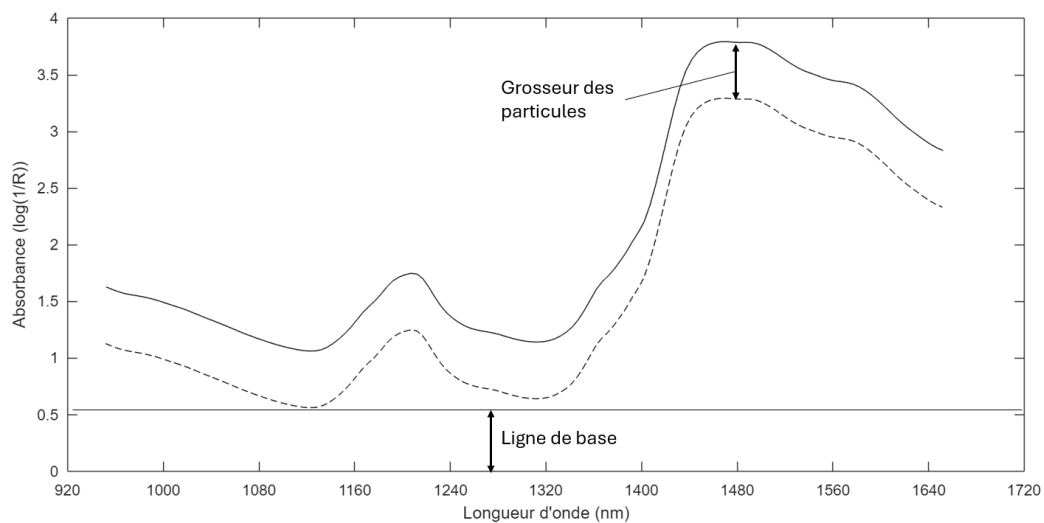


Figure 2.7 : Caractéristiques générales d'un spectre NIR

Adapté de Shenk et al. (2007)

En plus des éléments de la figure 2.7, le signal contient inévitablement du bruit provenant de l'instrument lui-même ainsi que de son environnement. En conséquence, le spectre NIR d'un échantillon reste complexe et ne permet pas d'associer directement une bande et une amplitude particulière à un élément chimique distinct. Ainsi, afin d'utiliser la spectroscopie NIR en tant qu'outil pour l'identification des propriétés des aliments, l'obtention du spectre seul n'est pas suffisante. Le processus de mesure complet doit également inclure un prétraitement des données, l'acquisition des concentrations précises par les méthodes standards et le développement du modèle de calibration (Cen et He, 2007). Ce processus est présenté à la section 2.5.

2.4.6. Distance de pénétration

Il existe quelques articles scientifiques ayant étudié la distance de pénétration de la lumière proche infrarouge dans les produits agroalimentaires. Selon Zude (2008), celle-ci peut être influencée par la quantité d'eau présente dans l'échantillon ainsi que par plusieurs facteurs provoquant de l'absorption et de la dispersion. De façon générale, la distance de pénétration de la lumière proche infrarouge ne dépasse pas 10 cm pour la région de 780 à 1100 nm, mais est

normalement plus basse que 10 mm pour les fruits. D'ailleurs, les études menées par (Wedding et al., 2024) ont permis d'évaluer une distance de pénétration de 3.3 mm à 1074 nm et 4.8 mm dans la région 860 à 834 nm pour les avocats. De leur côté, Huang et al. (2016) ont pu déterminer qu'à partir de 3 mm, les performances de mesure par NIR se détérioraient grandement dans la poudre de lait. Le tableau 2.1, adapté de Tyagi et al. (2020), contient différents produits ainsi que leur distance de pénétration respective. Les données de ce tableau permettent de constater que la distance est plus grande dans les produits à faible teneur en eau.

Tableau 2.1 : Distances de pénétration de la lumière proche infrarouge

Adapté de Tyagi et al. (2020)

Produit	Pic spectral (μm)	Distance de pénétration (mm)
Pâte, blé	1.0	4 à 6
Pain, blé	1.0	11 à 12
Pain, biscuit, séché	0.88	12
Grain, blé	1.0	2
Carottes	1.0	1.5
Pâte de tomate (70% à 85% d'eau)	1.0	1
Pommes de terre crues	1.0	6
Pommes de terre séchées	0.88	15 à 18
Pommes crues	1.65	5.9

2.5. Chimiométrie

La chimiométrie représente l'élément central de la spectroscopie NIR, car elle permet de faire le lien entre les spectres mesurés et les propriétés physiques ou les concentrations réelles des constituants d'un analyte. Selon Cen et He (2007), « La chimiométrie est la science qui relie les mesures effectuées sur un système ou un processus chimique à l'état du système via l'application de méthodes mathématiques ou statistiques » [traduction libre]. Elle est largement utilisée en tant que méthode d'analyse multivariée pour la spectroscopie NIR et comporte deux éléments essentiels : le prétraitement des données spectrales et le développement d'un modèle de calibration.

2.5.1. Prétraitement des données

En plus des informations physiques et chimiques d'un analyte, les données obtenues d'un spectromètre NIR contiennent une quantité non négligeable d'interférences. Ces informations inutiles peuvent être une ligne de base, un bruit de fond ou d'autres effets causés par la température, la surface et la texture des échantillons, la grosseur des particules, l'effet de diffusion de la lumière ou encore provenir d'erreurs aléatoires. Boysworth et Booksh (2007) mentionnent effectivement que les spectres NIR sont caractérisés par de faibles pointes d'informations qui surviennent dans un fond de bruit plus intense. Il est alors important de prétraiter les données spectrales avant de développer un modèle de calibration qui sera fiable, précis et stable (Cen et He, 2007). L'objectif du prétraitement est donc de retirer les phénomènes physiques présents dans les spectres en conservant l'information pertinente afin d'améliorer la régression multivariable du modèle de calibration (Rinnan et al., 2009).

Dans les systèmes à transmittance pure sans diffusion de lumière, la loi de Lambert-Beer décrit la relation linéaire entre le spectre d'absorbance et la concentration des constituants. Cette relation est représentée par l'équation (2.1) dans laquelle A représente l'absorbance, T la transmittance, ε l'absorptivité molaire, l la longueur de trajet optique, c la concentration du constituant et λ la longueur d'onde.

$$A_{\lambda} = -\log_{10}(T) = \varepsilon_{\lambda} * l * c \quad (2.1)$$

Dans les cas où un ensemble de données ne suit pas cette loi, il faut ajouter des facteurs ou des composants dans la régression linéaire pour compenser les phénomènes non idéaux, ce qui complexifie le modèle et rend la calibration plus difficile. Les méthodes de prétraitement permettent alors de simplifier le modèle en éliminant le plus possible les phénomènes non linéaires rendant ainsi le modèle plus robuste.

Le tableau 2.2 présente les différentes techniques mentionnées dans la littérature ainsi que leur utilité (Barnes et al., 1989 ; Boysworth et Booksh, 2007 ; Cen et He, 2007 ; Ozaki et al., 2006 ; Porep et al., 2015). Parmi celles-ci, le *Multiplicative Scatter Correction* (MSC), le *Standard Normal Variate* (SNV) utilisé conjointement avec le *Detrending* (Dt), le *Mean centering* et la méthode

Savitzky-Golay sont les plus répandues. Celles-ci sont expliquées plus en détail dans les sous-sections qui suivent.

Tableau 2.2 : Méthodes de prétraitement des spectres NIR

Méthode	Utilité
<i>Moving-average</i>	<ul style="list-style-type: none"> • Lissage et réduction du bruit.
<i>Savitzky-Golay (SG) smoothing</i>	<ul style="list-style-type: none"> • Lissage et réduction du bruit.
<i>Savitzky-Golay (SG) derivative</i>	<ul style="list-style-type: none"> • Réduction du bruit ; • Réduction des signaux de fond ; • Correction de la ligne de base ; • Augmentation de la résolution spectrale.
<i>Discrete Wavelet Transform (DWT)</i>	<ul style="list-style-type: none"> • Réduction du bruit ; • Réduction des signaux de fond.
<i>Multiplicative Scatter Correction (MSC)</i>	<ul style="list-style-type: none"> • Correction de la ligne de base ; • Réduire l'effet de dispersion de la lumière.
<i>Orthogonal Signal Correction (OSC)</i>	<ul style="list-style-type: none"> • Correction de la ligne de base ; • Réduire l'effet de dispersion de la lumière ; • Correction de la dérive de l'instrument.
<i>Standard Normal Variate (SNV)</i>	<ul style="list-style-type: none"> • Correction de la ligne de base ; • Suppression de l'effet d'interférence multiplicative de dispersion, de la taille des particules et du changement de distance ; • Correction des effets de dispersion multiplicatifs et additifs.
<i>Detrending (Dt)</i>	<ul style="list-style-type: none"> • Correction de la ligne de base ; • Compensation de la curvilinearité.
<i>Difference Spectra</i>	<ul style="list-style-type: none"> • Comparaison de deux spectres ; • Exposer l'effet d'un paramètre externe.
<i>Mean Centering (MC)</i>	<ul style="list-style-type: none"> • Amélioration de la résolution.
<i>Normalization</i>	<ul style="list-style-type: none"> • Égalisation de l'amplitude.

2.5.1.1. Multiplicative Scatter Correction (MSC)

Cette méthode initialement développée par Martens et al. (1983), qui vise à corriger les effets de diffusion de la lumière, est probablement la plus utilisée en tant que technique de prétraitement des spectres NIR. Le principe vise à retirer les artéfacts et imperfections des données originales avant de procéder au développement du modèle de calibration. Cette méthode comprend deux étapes :

1. Estimer les coefficients de correction responsables des effets additifs et multiplicatifs.

$$x_{org} = b_0 + b_{ref,1} * x_{ref} + e \quad (2.2)$$

2. Corriger le spectre.

$$x_{corr_MSC} = \frac{x_{org} - b_0}{b_{ref,1}} = x_{ref} + \frac{e}{b_{ref,1}} \quad (2.3)$$

Dans ces équations, x_{org} représente le spectre original mesuré par l'instrument, x_{ref} est un spectre de référence utilisé pour prétraiter l'ensemble de données complet, e est la partie non modélisée de x_{org} , x_{cor_MSC} est le spectre corrigé et b_0 et $b_{ref,1}$ sont des paramètres scalaires qui diffèrent pour chaque échantillon. La méthode pour déterminer ces coefficients est présentée graphiquement par Rinnan et al. (2009). Le spectre de l'échantillon est tracé par rapport à un spectre de référence sélectionné et les termes de correction scalaire représentent l'origine et la pente de la ligne noire obtenue à partir de la régression des moindres carrés passant par tous les points. Dans la plupart des applications, un spectre moyen de la calibration est utilisé comme référence, mais il est aussi possible d'utiliser un spectre générique.

2.5.1.2. Standard Normal Variate (SNV)

Considéré comme la deuxième méthode la plus appliquée pour réduire les effets de diffusion de la lumière, le SNV suit le même concept que le MSC, mais sans utiliser de référence commune. Dans la méthode SNV, chaque échantillon est corrigé de façon seule et isolée des autres échantillons du jeu de données. Selon Barnes et al. (1989), la grosseur des particules, la diffusion, la multicollinéarité ainsi que leur effet multiplicatif représentent une variance beaucoup plus

élevée que la composition chimique. Cette variance peut être grandement diminuée par l'application de la méthode SNV. Une fois la transformation appliquée, la forme du spectre ne change pas, mais l'écart dû à la grosseur des particules est minimisé. Il est donc plus facile d'observer les réelles différences entre les courbes.

Le SNV centre chaque spectre, puis le met à l'échelle selon son propre écart-type, corrigeant les décalages sur l'axe $\log(1/R)$. Les spectres résultants ont alors une moyenne nulle et une variance unitaire. L'opération est représentée dans l'équation (2.4) où x_{corr_SNV} est le spectre corrigé, x_{org} est le spectre original mesuré, a_0 est la moyenne du spectre et a_1 son écart-type.

$$x_{corr_SNV} = \frac{x_{org} - a_0}{a_1} \quad (2.4)$$

Barnes et al. (1989) démontrent bien l'effet du SNV appliqué à des spectres de sucrose sous forme cristalline et en poudre. Lorsqu'ils sont représentés sur le même graphique, les spectres originaux semblent indiquer la présence de deux groupes d'échantillons distincts alors qu'en appliquant le SNV, tous les échantillons se regroupent ensemble.

2.5.1.3. Detrending

Le *detrending*, souvent appliqué en complément du SNV, est une méthode de correction visant à éliminer les tendances présentes dans les spectres d'absorption. Les études de Barnes et al. (1989) ont relevé que les traits dominants des spectres d'absorption étaient l'augmentation du niveau d'absorption dans la plage de 1100 à 2500 nm. Cette tendance est généralement linéaire, mais devient curvilinéaire pour les échantillons plus compactés, comme dans le cas des poudres. De manière générale, les solides présentent une absorption plus intense à des longueurs d'onde plus élevées, car les combinaisons de bandes deviennent plus probables à l'approche des vibrations fondamentales.

Le *detrending* est effectué à l'aide d'un polynôme du second ordre dans une analyse de régression, où la valeur spectrale d'un échantillon est la réponse (ou variable dépendante) et la variable indépendante est donnée par les longueurs d'onde, corrigeant les effets dus à la courbure de la ligne de base (Fernández-Cabanás et al., 2006). Cette technique est donc utile

dans les cas de présence d'une ligne de base complexe autre qu'une valeur constante ou une fonction linéaire. Elle est également suggérée dans les cas d'acquisition de longue durée dans un environnement où la ligne de base peut changer et évoluer avec le temps.

L'opération, représentée par l'équation (2.5), consiste premièrement à calculer une estimation des valeurs du spectre original $\hat{X}_{\log(\frac{1}{R}),i}$ à l'aide d'une fonction quadratique dont les coefficients A, B, C sont déterminés par la méthode des moindres carrés, e_i représentant les résidus spectraux de l'ajustement des moindres carrés et W la valeur de longueur d'onde. Par la suite, le spectre est corrigé en retirant l'estimation du spectre original en utilisant l'équation (2.6).

$$\hat{X}_{\log(1/R),i} = A + B * W + C * W^2 + e_i \quad (2.5)$$

$$X_{DT,i} = X_{\log(1/R),i} - \hat{X}_{\log(1/R),i} \quad (2.6)$$

Barnes et al. (1989) démontrent bien l'effet du *detrending* alors que les spectres d'échantillons de sucrose corrigés par SNV révèlent la présence d'une tendance. Après l'application du *detrending*, cette tendance est retirée du signal. Cette opération produit un spectre de moyenne nulle et de variance non unitaire.

2.5.1.4. Mean centering

L'opération de *mean centering* consiste à retirer la moyenne du spectre en repositionnant le centroïde des données à l'origine du système de coordonnées. Cette opération est essentielle pour le développement de modèle de calibration par *Partial Least Squares* (PLS), car elle permet de réduire le nombre de variables à utiliser simplifiant grandement le modèle (Manley, 2014). L'opération est décrite par l'équation (2.7) où x_{corr_MC} est le spectre corrigé, x_{org} est le spectre original mesuré et a_0 est la moyenne du spectre. Cette opération produit un spectre dont la moyenne est nulle et la variance distribuée autour de zéro. Il est important de noter que d'autres techniques de prétraitement comme le SNV effectuent un centrage de la moyenne dans leurs opérations.

$$x_{corr_MC} = x_{org} - a_0 \quad (2.7)$$

2.5.1.5. Savitzky-Golay

La méthode Savitzky-Golay (Savitzky et Golay, 1964) est une technique de dérivation numérique qui inclut une étape de lissage. La méthode consiste à appliquer un polynôme dans une fenêtre symétrique sur les valeurs du spectre original. Par la suite, les fonctions dérivées peuvent être calculées de façon analytique.

L'opération de lissage est particulièrement nécessaire afin d'augmenter le ratio signal sur bruit. Bien que la technique populaire de moyenne mobile puisse être appliquée, celle-ci ne permet pas de conserver les pics dans les signaux, lesquels contiennent l'information importante d'un spectre d'absorption. La technique Savitzky-Golay se base plutôt sur la méthode des moindres carrés qui permet d'évaluer la courbe de façon beaucoup plus précise. Un polynôme est évalué pour estimer le point central d'une fenêtre glissante appliquée à tout le spectre de façon séquentielle. Elle est notamment très efficace pour retirer le bruit et conserver l'information d'un spectre NIR.

De son côté, la différentiation permet d'améliorer les caractéristiques spectrales apparentes. Les dérivées ont notamment la capacité de supprimer les effets additifs et multiplicatifs dans les spectres et sont utilisées en spectroscopie analytique depuis des décennies (Rinnan et al., 2009). De plus, elles permettent de supprimer les principales tendances du signal, comme les lignes de base et les signaux basse fréquence, tout en conservant les détails importants. Selon Manley (2014), elles peuvent résoudre le principal problème de la spectroscopie NIR, soit le chevauchement des pics. Une fois que la dérivée est appliquée, il est plus facile d'identifier les bandes spectrales correspondantes à l'information recherchée. Les opérations de première et de deuxième dérivée sont les plus rencontrées dans la spectroscopie proche infrarouge. La figure 2.8 montre un exemple du signal dérivé d'un spectre obtenu à partir d'échantillon de nourriture animale.

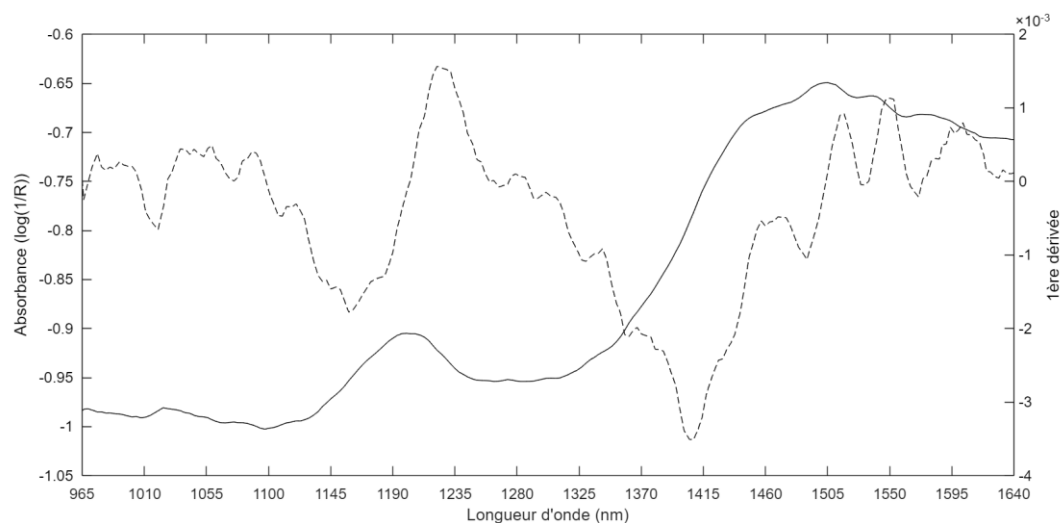


Figure 2.8 : Spectre d'absorption de nourriture animale et sa première dérivée
Ligne pleine : spectre original, ligne pointillée : première dérivée.

L'application de la méthode Savitzky-Golay requiert l'ajustement de trois paramètres, soit la dimension de la fenêtre, l'ordre du polynôme et l'ordre de la dérivée. Ces paramètres sont généralement sélectionnés de façon empirique en vérifiant les performances des modèles de calibration.

2.5.2. Calibration multivariée

La spectroscopie proche infrarouge est une méthode de mesure indirecte. Les données spectrales mesurées doivent donc être transformées afin d'obtenir les concentrations des constituants ou les propriétés physiques réelles. Or, les spectres NIR se composent de plusieurs bandes sévèrement superposées résultant en une forte multicolinéarité (Ozaki et al., 2006). On y observe également la présence de non-linéarités reliées à la nature complexe des composants et leur hétérogénéité (Pérez-Marín, 2014). Par conséquent, les méthodes populaires utilisées pour le prétraitement des spectres en général (transformations linéaires, filtres de Fourier, dérivées et régressions polynomiales) donnent des résultats peu satisfaisants, car elles sont adaptées à des analyses univariées et ne permettent pas de gérer ces particularités. La transformation des spectres en caractéristiques utiles n'est donc possible que par le

développement d'un modèle de calibration utilisant des méthodes d'analyses multivariées (Blanco et Villarroya, 2002 ; Maslovaric et al., 2011).

Selon Boysworth et Booksh (2007), l'analyse multivariée détient trois capacités intrinsèques qui font d'elle un outil puissant pour les applications de calibration :

- Une sélectivité complète du signal n'est pas nécessaire pour produire une calibration réussie. Dans une analyse univariée, il faut trouver le signal correspondant à la caractéristique recherchée (par exemple la protéine). Or, à l'intérieur d'un échantillon alimentaire, il est impossible de déterminer une longueur d'onde unique associée à cette caractéristique ;
- L'analyse multivariée permet la calibration d'un échantillon à plusieurs constituants. Le modèle peut être développé pour plus d'une caractéristique à la fois et il peut être construit pour une caractéristique entourée d'interférences multiples ;
- L'analyse multivariée permet le développement d'outils de diagnostic puissants pour aider à évaluer la qualité et la fiabilité des futurs spectres, ce qui est impossible avec les modèles univariés.

La calibration d'un système NIR s'effectue en utilisant les données provenant de mesures standards de référence effectuées en laboratoire, lesquelles fournissent la composition et les propriétés réelles des échantillons. Une fois le spectre NIR obtenu sur ces mêmes échantillons, des techniques mathématiques et statistiques permettent d'obtenir une équation empirique qui associe les données chimiques aux éléments principaux de ce spectre (Maslovaric et al., 2011). L'objectif de cette méthode est donc de construire un modèle capable d'estimer précisément les caractéristiques et les propriétés des échantillons inconnus. Blanco et Villarroya (2002) énumèrent les étapes impliquées dans ce processus, lesquelles sont détaillées dans les sous-sections suivantes.

2.5.2.1. Sélection du lot de calibration

Le développement d'une méthode analytique par spectroscopie NIR utilise à priori un lot de calibration fiable. Ce lot permet de collecter les spectres provenant d'échantillons dont la

composition ainsi que les conditions font l'objet d'une variabilité représentative de la production. La variance contenue dans ce lot détermine directement la robustesse et la précision du modèle. Le tableau 2.3 donne quelques exemples de lot de calibration utilisés pour différents produits. Bien qu'aucune des études citées n'implémente de mesure en direct sur une ligne de production, les lots présentés dans le tableau comportent tout de même une variabilité ainsi qu'une quantité satisfaisante selon l'application.

Tableau 2.3 : Exemples de lots de calibration

Produit	Caractéristique	Lot de calibration	Auteurs
Blé	Humidité	Échantillons de blé provenant de quinze régions géographiques de l'ouest du Canada.	Mohan et al. (2005)
Aliments composés	Blé Tournesol	7523 échantillons commerciaux.	Pérez-Marín et al. (2008)
Fromage	Humidité Gras Sels inorganiques	64 échantillons de fromage fondu.	Blazquez et al. (2004)
Aliments composés	Protéine Fibres Soja Tournesol	400 échantillons d'aliments composés finis.	Fernández-Ahumada et al. (2008)
Arachides	Huile	64 échantillons de 47 géotypes cultivés en saison des pluies ou en été.	Misra et al. (2000)
Nourriture extrudée pour chien	Humidité Gras Protéine Fibres Peroxyde	649 échantillons de différents producteurs sélectionnés au hasard entre 2013 et 2017.	De Marchi et al. (2018)
Nourriture extrudée pour chien	Minéraux	119 échantillons d'un producteur italien entre 2017 et 2019.	Goi et al. (2019)

Tableau 2.3 : Exemples de lots de calibration (suite)

Produit	Caractéristique	Lot de calibration	Auteurs
Nourriture extrudée pour chien	Amidon Amidon gélatinisé Fibres	119 échantillons d'un producteur italien entre 2018 et 2019.	Goi et al. (2020)

2.5.2.2. Mesures par méthodes standards

Outre la qualité du lot de calibration et la quantité de ses échantillons, les mesures effectuées par méthodes standards en laboratoire ont un impact direct sur les performances du modèle de calibration. Les informations collectées durant le processus de mesure représentent les données de référence pour bâtir un modèle dédié à estimer la composition des échantillons inconnus. En effet, la régression des données spectrales sur les mesures de référence permet d'estimer les coefficients nécessaires à la calibration du modèle. Ceux-ci peuvent alors être appliqués à un nouvel échantillon collecté afin de produire une estimation de la caractéristique ciblée (Shenk et al., 2007).

2.5.2.3. Mesure de l'humidité

L'humidité dans les produits agroalimentaires est souvent déterminée par des méthodes de mesure de la perte de poids par séchage (Irudayaraj et Reh, 2008 ; Niemöller et Behmer, 2008 ; Shenk et al., 2007). Ces méthodes de séchage, qu'elles soient classiques, au four, sous vide, par infrarouge ou au micro-ondes, mesurent essentiellement la masse nette perdue et ne permettent pas de faire la distinction entre l'eau et d'autres substances volatiles. Bien que très reproductible, cette méthode peut conduire à des résultats différents lorsque les paramètres sont modifiés. Ainsi, il est important de tenir compte de cette particularité si l'application de mesure peut en être affectée.

Les méthodes standards de détermination de l'humidité qui traitent de nourriture animale sont résumées dans le tableau 2.4 accompagnées de leurs paramètres principaux.

Tableau 2.4 : Méthodes de détermination de l'humidité dans les aliments pour animaux

Méthode	Poids	Température (°C)	Durée (h)
NFTA 2.1.4	10 à 20 g	105	3
AOAC 935.29	2 à 10 g (typ. 5 g)	104	3
AOAC 945.15	2 à 10 g	104	3

2.5.2.4. Acquisition et prétraitement des spectres NIR

Dans le but de construire des équations fiables, les mêmes échantillons analysés par méthodes standards doivent être mesurés par le spectromètre NIR. Les spectres sont ensuite traités afin de conserver l'information importante comme mentionné dans la section 2.5.1. Dans la plupart des études, il est nécessaire d'obtenir plusieurs spectres pour le même échantillon et de calculer la moyenne.

2.5.2.5. Développement du modèle et validation

Le développement d'un modèle de calibration NIR est basé sur l'utilisation de méthodes d'analyse multivariées connues et établies depuis plusieurs années. La sélection de la méthode dépend notamment du type d'analyse (quantitatif ou qualitatif), de la forte présence d'effets non linéaires et de la puissance de calcul disponible. Pérez-Marín et al. (2012) et Skvaril et al. (2017) recommandent de toujours utiliser la méthode la plus simple, par exemple, utiliser une méthode non linéaire seulement si les méthodes linéaires ne sont pas satisfaisantes. Ils ajoutent que la présence de larges plages de paramètres et l'hétérogénéité des échantillons d'étalonnage sont des indicateurs typiques d'un besoin d'algorithmes plus sophistiqués.

La figure 2.9 présente la classification des méthodes d'analyse multivariées les plus rencontrées dans le domaine de la spectroscopie NIR. Les méthodes présentées à gauche de la figure sont associées aux analyses quantitatives, dont l'objectif est d'estimer la concentration d'un paramètre donné dans un échantillon. En revanche, les méthodes situées à droite sont destinées aux analyses qualitatives, visant à détecter la présence ou l'absence d'un composant spécifique. Selon Skvaril et al. (2017), les méthodes de régression linéaire représentées principalement par PLS donnent des résultats acceptables dans la plupart des cas d'études quantitatives. Cette

approche est d'ailleurs la plus répandue dans la littérature en raison de sa simplicité et de son efficacité. Dans les cas où les performances du modèle PLS sont insuffisantes, des méthodes non linéaires plus avancées, telles que les réseaux de neurones artificiels, peuvent être envisagées. Bien qu'elles puissent améliorer la qualité des prédictions, ces méthodes exigent une puissance de calcul plus importante.

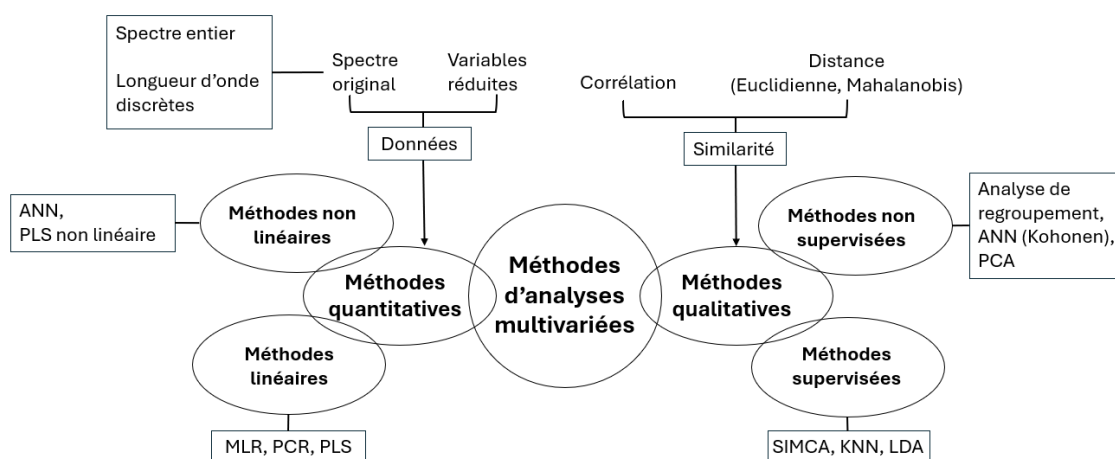


Figure 2.9 : Classification des méthodes d'analyse multivariées

Adapté de Blanco et Villarroya (2002)

En somme, bien que les méthodes d'analyse multivariée appliquées au développement de modèles de calibration NIR soient bien établies, leur sélection doit s'appuyer sur une évaluation rigoureuse des besoins spécifiques de l'application. Une utilisation judicieuse privilégie la simplicité méthodologique, tout en reconnaissant que la complexité des données, notamment en présence de non-linéarités ou d'une forte hétérogénéité, peut justifier le recours à des approches plus avancées.

2.5.2.6. Valeurs aberrantes

Les modèles développés doivent faire l'objet d'un processus de détection des valeurs aberrantes avant d'être soumis à la vérification de leurs performances. La distance de Mahalanobis est souvent utilisée pour effectuer cette tâche (Bakeev, 2010 ; Irudayaraj et Reh, 2008 ; Ozaki et al., 2006). Celle-ci est « simplement une distance euclidienne pondérée, où chacune des dimensions est inversement pondérée par sa variance globale dans les données d'étalonnage » [traduction

libre] (Bakeev, 2010). D'autres auteurs choisissent d'utiliser le test de Student (Castrillo et al., 2005, 2005 ; De Marchi et al., 2018 ; Granato et al., 2014 ; Hervera et al., 2012 ; Pérez-Marín et al., 2001) ou sa généralisation, le Hotelling T^2 (Bakeev, 2010 ; Irudayaraj et Reh, 2008 ; Porep et al., 2015 ; Xu et al., 2020).

2.5.2.7. Sélection et validation du modèle

La sélection du modèle et sa validation représentent les dernières étapes du processus de développement avant son déploiement. L'objectif de ce processus est de choisir le modèle de calibration le plus précis et d'estimer ses performances avec les échantillons futurs. Un lot indépendant de validation est normalement nécessaire, mais plusieurs études utilisent la validation croisée lorsque la quantité d'échantillons est limitée. La sélection du modèle le plus performant est faite en évaluant différents indicateurs, lesquels sont expliqués par Porep et al. (2015) et résumés dans les sous-sections suivantes.

RMSEC - Root Mean Square Error of Calibration

Le RMSEC permet de déterminer dans quelle mesure le modèle correspond aux données de calibration. Cette validation interne représente une estimation empirique et optimiste des performances du modèle. L'équation (2.8) permet de calculer cet indicateur. L'erreur est calculée entre les données NIR (\hat{y}_i) et les données de référence du lot de calibration (y_i), puis divisée par le nombre d'échantillons (n) ou le degré de liberté ($n - A - 1$) où A correspond au nombre de facteurs utilisés dans le modèle. La valeur de cet indicateur va toujours diminuer en ajoutant des facteurs.

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n - A - 1}} \quad (2.8)$$

RMSECV - Root Mean Square Error of Cross-Validation

Le RMSECV est uniquement basé sur les données de calibration en utilisant la méthode de validation croisée sur les données. La plupart des études utilisent une validation croisée de type *leave-one-out* qui teste le modèle sur chaque échantillon en utilisant tous les autres comme lot

de calibration. L'équation (2.9) permet de calculer cet indicateur. L'erreur est calculée entre les données NIR (\hat{y}_i) et les données de référence du lot de calibration (y_i), puis divisée par le nombre d'échantillons (n). Cet indicateur peut être considéré comme une estimation du RMSEC puisque la seule différence réside dans la manière dont les échantillons sont choisis pour la validation.

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (2.9)$$

RMSEP - Root Mean Square Error of Prediction

Le RMSEP est une estimation de la variation des valeurs de référence et des valeurs prédites d'un lot d'échantillons différents de ceux utilisés pour la calibration. Il représente ainsi le meilleur estimé des performances futures en quantifiant la variabilité des différences entre les valeurs mesurées par spectroscopie NIR et les références obtenues en laboratoire (Boysworth et Booksh, 2007). Une valeur près de zéro indique que le modèle est fiable (Maslovaric et al., 2011). L'équation (2.10) permet de calculer cet indicateur. L'erreur est calculée entre les données prédites par NIR (\hat{y}_i) et les données de référence du lot d'échantillons (y_i), puis divisée par le nombre d'échantillons total (n_{val}). Cet indicateur représente une validation indépendante du modèle et dépend grandement de la qualité du lot d'échantillon de test utilisé. Une représentation graphique du RMSEP en relation avec le nombre de facteurs du modèle représente une méthode très utile pour déterminer le nombre de composants à utiliser (Porep et al., 2015).

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_{val}} (\hat{y}_i - y_i)^2}{n_{val}}} \quad (2.10)$$

Biais

Le biais indique l'erreur systématique du modèle et représente « la moyenne des différences entre les valeurs prédites (\hat{y}_i) et les valeurs de référence (y_i) du nombre respectif d'échantillons (n) » [traduction libre] (Porep et al., 2015). L'équation (2.11) permet de calculer cet indicateur.

$$Biais = \sum_{i=1}^n \frac{(\hat{y}_i - y_i)}{n} \quad (2.11)$$

SEC, SECV, SEP

Les indicateurs « Standard Error of Calibration » (SEC), « Standard Error of Cross-Validation » (SECV) et « Standard Error of Prediction » (SEP) sont des mesures des erreurs corrigées par le biais et représentent d'autres indicateurs communs rencontrés dans l'analyse des performances des modèles NIR. Ils indiquent l'écart-type des résiduels prédits et sont interprétés par la moyenne des différences entre les données prédites par NIR (\hat{y}_i) et les données de référence (y_i) et décrivant la variance attribuable aux erreurs aléatoires inexplicables dans le lot de prédiction. L'équation (2.12) permet de calculer cet indicateur.

$$SEP = \sqrt{\frac{\sum_{i=1}^{n_{val}} (\hat{y}_i - y_i - Bias)^2}{n_{val} - 1}} \quad (2.12)$$

R^2

Le coefficient de détermination R^2 fournit l'information sur la qualité de régression du modèle. Par conséquent, cette évaluation statistique permet de déterminer le degré de variation des données correctement modélisées. La formule de R^2 est fournie dans l'équation (2.13) où les valeurs prédites (\hat{y}_i), les valeurs de références (y_i) et leurs moyennes (\bar{y}) sont représentées. Dans le secteur agroalimentaire, une valeur de R^2 plus grande que 0.90 est considérée comme excellente, une valeur entre 0.70 et 0.89 est considérée comme bonne alors qu'une valeur entre 0.50 et 0.69 permet une bonne séparation en trois groupes (haut, moyen et bas) (Perez et al., 2002).

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.13)$$

RPD - Ratio of Prediction to Deviation

Ce rapport est un autre indicateur utile et utilisé couramment pour l'interprétation des calibrations. Il représente le rapport entre la déviation standard des valeurs de référence (SD) et

le RMSEP ou le RMSECV. Des valeurs plus élevées de RPD indiquent une meilleure capacité de prédiction normalisée à l'écart-type. L'équation (2.14) permet de calculer cet indicateur.

$$RPD = \frac{SD}{RMSEP \text{ ou } RMSECV} \quad (2.14)$$

2.5.2.8. Utilisation et suivi du modèle

Le modèle sélectionné offrant les meilleures performances est finalement soumis à de nouveaux échantillons en production. La calibration doit être suivie périodiquement en effectuant d'autres tests selon les méthodes standards et en comparant avec les résultats estimés par le modèle. Ce dernier devra faire l'objet d'ajustements, car il est impossible de couvrir concrètement la totalité de la variabilité avec le lot initial dédié à la calibration. De plus, lors de l'introduction de nouveaux produits ou de variantes de ceux-ci, le processus doit reprendre du début en intégrant ces échantillons en tant que lot de calibration.

2.6. Facteurs d'influence d'une mesure par proche infrarouge en direct

Lorsque les calibrations des spectromètres proche infrarouge en laboratoire ont commencé à fournir des résultats satisfaisants, les scientifiques ont débuté les essais pour élaborer des calibrations sur des capteurs en direct dans des environnements de production. Bien que des solutions commerciales de capteurs proche infrarouge soient disponibles, des efforts considérables ont dû être déployés pour élaborer des techniques de calibration afin de contrer les effets de plusieurs facteurs pouvant influencer les mesures. Comme mentionné à la section 2.4.3, il fut constaté que, bien qu'une calibration puisse donner de très bons résultats dans un environnement contrôlé sur des échantillons broyés et conditionnés, celle-ci ne pouvait pas être directement transférée sur un appareil de mesure en direct ni adaptée facilement à des conditions de mesures en environnement industriel.

L'analyse des différences entre les deux méthodes de mesure met en lumière la nécessité de prendre en compte une multitude de facteurs d'origines diverses dans la calibration d'un spectromètre proche infrarouge en direct. Afin de bien connaître les enjeux auxquels le développement d'une calibration doit faire face, cette section présente les facteurs identifiés

dans la littérature en les regroupant selon leur origine. Par la suite, un diagramme de Pareto est proposé afin de classer ces facteurs selon leur occurrence et fournir une représentation visuelle plus claire.

2.6.1. Description des sources consultées

Au total, 33 articles de différents types ont été consultés, ceux-ci couvrant différents produits dans le secteur agroalimentaire. La figure 2.10 montre la répartition des types d'articles et celle des produits couverts. L'ensemble des articles analysés sont majoritairement des études effectuées pour développer des calibrations NIR dans lesquelles les auteurs mentionnent les obstacles rencontrés ainsi que les méthodes mises en place pour contrer leurs effets. Neuf résumés de la technologie et de l'état de l'art ainsi que deux chapitres de livre sur la spectroscopie proche infrarouge permettent de compléter la revue. La liste comprend également une norme internationale. En ce qui a trait aux produits ciblés par les articles, la majeure partie de ceux-ci cible les produits agroalimentaires en général. Cette catégorie rassemble les articles de type résumé, les chapitres et la norme internationale. Pour leur part, les études ciblent des produits de différentes natures dont les viandes sont les plus représentées suivies des produits laitiers, des produits céréaliers, des grains, des aliments composés et des fruits et légumes. Il n'existe qu'un seul article portant sur la nourriture animale ainsi que sur les vins.

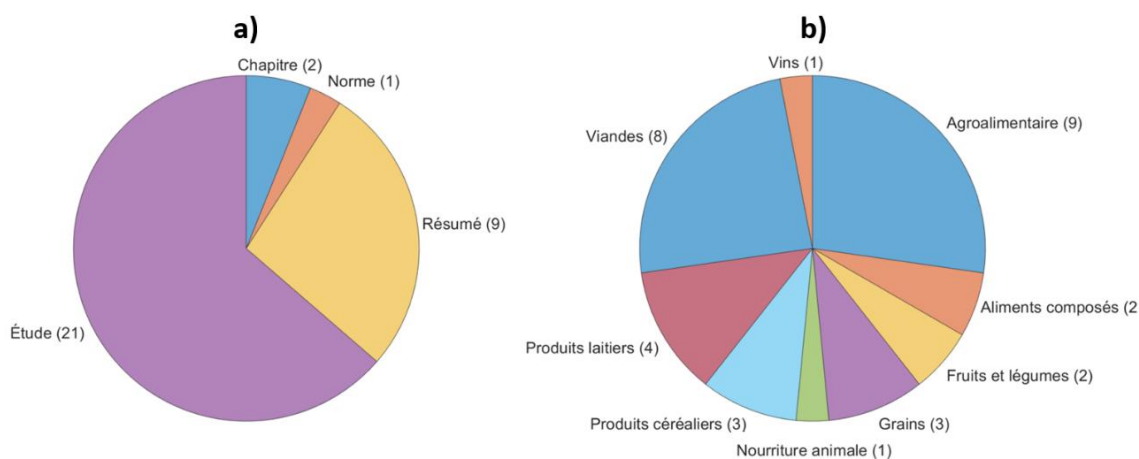


Figure 2.10 : Répartition des articles consultés
a) types de document, b) produits étudiés.

2.6.2. Environnement

L'environnement correspond au milieu physique dans lequel un capteur NIR en direct effectue ses mesures. Il inclut l'ensemble des éléments externes et situés autour du capteur susceptibles d'influencer les mesures de quelque manière que ce soit. Selon Wang et Paliwal (2007), il est essentiel de tenir compte de ces conditions environnementales dès l'achat ou la conception d'un système NIRS. Melfsen et al. (2013) soulignent également l'importance de ces effets en recommandant d'intégrer les variations environnementales aux modèles de calibration afin d'améliorer la précision des prédictions.

Parmi les facteurs environnementaux influençant les mesures, la température ambiante joue un rôle majeur. Lors de leurs essais, Holroyd et al. (2013) ont observé que les conditions d'opération du capteur provoquaient fréquemment une augmentation interne de la température de l'instrument au-delà des seuils de tolérance, allant jusqu'à causer des défaillances de celui-ci. Ces auteurs mentionnent également que la présence de poussière peut affecter l'instrument et compromettre son fonctionnement. De leur côté, Bock et Connelly (2008) soulignent que, pour garantir leur fiabilité, les capteurs doivent être conçus pour résister non seulement aux températures élevées et à la poussière, mais aussi aux vibrations inhérentes aux environnements industriels.

Au-delà des risques de défaillance matérielle, les fluctuations de températures peuvent non seulement affecter la précision des composantes internes d'un capteur (Niemöller et Behmer, 2008), mais aussi les propriétés de l'analyte lui-même en affectant les liaisons O-H et en provoquant un déplacement du pic d'absorption composite de l'eau (Paul et Pfitzner, 2004 ; Shenk et al., 2007).

L'installation d'un capteur en direct doit également tenir compte du caractère discontinu du flux de production. En effet, plusieurs situations peuvent mener à un flux de produit plus faible, intermittent ou parfois même interrompu pendant un certain temps. Dans ces situations, un capteur peut produire des signaux hors norme. Holroyd et al. (2013) ont mis en évidence ce phénomène dans le cadre de la mesure de produits laitiers, où la présence de bulles d'air dans le flux génère des valeurs aberrantes. De semblables anomalies peuvent être dues à

l'obstruction du capteur, situation observée par Bock et Connelly (2008) qui, en plus de détecter des manques de matières, obtenaient de mauvaises lectures lorsque l'équipement de brassage de la pâte à pain bloquait la zone de mesure du capteur.

2.6.3. Spectromètre

Les spectromètres sont des appareils complexes constitués de plusieurs composantes électroniques, électromécaniques, pneumatiques et optiques. Cette complexité introduit des défis quant à l'opération dans un milieu industriel. En règle générale, un spectre récolté par un capteur NIR contient non seulement les informations fournies par l'interaction de la lumière avec l'analyte, mais aussi des artefacts instrumentaux et les erreurs de collecte de données et de calcul (Shenk et al., 2007). L'instrument peut également être sujet à une dérive qui introduit un biais graduel dans les mesures (Tonolini et al., 2023) ou produire des données avec une pente et un biais différents à la suite du remplacement d'une composante.

Selon Niemöller et Behmer (2008), il est essentiel de sélectionner le meilleur instrument pour la tâche à réaliser en tenant compte des caractéristiques telles les spécifications de l'instrument, l'interface utilisateur, le logiciel d'opération, le coût ainsi que le support fourni par le fabricant. Toutefois, les auteurs soulignent que bien que chaque appareil puisse mesurer les mêmes paramètres, il existe une différence à long terme dans la stabilité et la précision des résultats entre les différentes technologies. Par ailleurs, les variations provenant directement de l'instrument doivent être incluses dans les modèles de calibration (Melfsen et al., 2013).

L'installation et la mise en service d'un capteur doivent respecter certaines exigences techniques, car plusieurs paramètres influencent directement les performances. Des éléments tels que la fréquence d'acquisition des données et le rapport signal sur bruit sont cruciaux pour assurer la qualité des mesures. La distance entre le capteur et l'échantillon constitue également un paramètre déterminant, puisqu'une distance réduite tend à augmenter l'amplitude des signaux d'absorption (O'Farrell et al., 2010; Pérez-Marín, 2014; Wold et al., 2011). Par ailleurs, l'utilisation d'une fibre optique dans certaines installations peut entraîner une dégradation notable du signal (Tonolini et al., 2023).

2.6.4. Équipements

Les différents équipements présents autour du capteur peuvent influencer les mesures optiques. Bien que la spectroscopie proche infrarouge n'utilise qu'une portion du spectre optique, la présence de lumière environnante en arrière-plan peut affecter les performances en fonction du type de source lumineuse. Une variation de cette lumière peut introduire un biais dans les prédictions ou une dérive de la lecture (Tonolini et al., 2023).

Dans d'autres situations, les équipements transportant le produit analysé peuvent réfléchir la lumière émise par le capteur et produire des effets parasites ou limiter la réflexion (Niemöller et Behmer, 2008 ; O'Farrell et al., 2010). Lorsque le produit circule sur un convoyeur, le tapis peut également induire une absorption plus marquée lorsque la quantité de matière est insuffisante, perturbant ainsi les mesures (Montes et al., 2006 ; Welle et al., 2005). Il est donc essentiel que le système prenne en compte ces phénomènes afin d'éviter toute altération des résultats.

2.6.5. Échantillons

La spectroscopie proche infrarouge permet de mesurer non seulement les propriétés chimiques, mais également les propriétés physiques d'un échantillon. Ainsi, toutes les caractéristiques pouvant affecter l'une ou l'autre de ces propriétés représentent un facteur pouvant influencer les mesures. Cet effet peut être recherché dans le cas de la mesure de quantité des composants chimiques, mais il peut également introduire des biais lorsqu'il ne fait pas directement l'objet de la mesure.

Parmi les différents facteurs d'influence identifiés dans la littérature, les propriétés physiques et chimiques des échantillons sont les plus souvent mentionnées. L'impact de la taille des échantillons, de l'état de ceux-ci (intact, en poudre, en granule, moulu) et de la grosseur des particules une fois moulu est particulièrement bien documenté dans la littérature, et ce, pour différents produits. Fernández de Ahumada et al. (2005) ont notamment associé des variations importantes causées par l'état des échantillons. De leur côté, O'Farrell et al. (2010) ont fait face à un défi lors de l'analyse de pièces de viande pour lesquelles la grosseur provoquait une variabilité. De même, Paul et Pfitzner (2004) ont constaté des différences d'écart-type en

fonction de la longueur de coupe de mélanges fourragers, tandis que Long et al. (2008) ont observé des erreurs associées à une taille plus grande des échantillons de grain. La grosseur des particules est un enjeu majeur, car elle influence directement la diffusion de la lumière. Celle-ci est considérée comme étant une importante source de variance qui complique l'interprétation d'un spectre (Montes et al., 2006).

D'autres propriétés physiques peuvent également affecter les mesures, notamment l'état de surface ainsi que l'épaisseur de l'échantillon. L'effet principal de ces facteurs est de diminuer la distance de pénétration de la lumière dans le spectre proche infrarouge à l'intérieur de l'échantillon empêchant ainsi sa caractérisation complète. Certaines études ont même mis en évidence des variations liées à la couleur des échantillons (O'Farrell et al., 2010; Wold et al., 2011).

Outre les propriétés physiques des échantillons, leur disposition, leur orientation ainsi que la densité de ceux-ci sous le capteur peuvent également contribuer à provoquer un phénomène de diffusion ou un déplacement de la ligne de base tandis que le mouvement des échantillons représente un défi supplémentaire identifié par plusieurs auteurs.

Du côté de la composition chimique, l'hétérogénéité des échantillons intacts influence également les mesures (Porep et al., 2015 ; Prieto et al., 2009 ; Wold et al., 2011). Une lecture effectuée sur une section d'un échantillon hétérogène risque de ne pas refléter fidèlement la concentration des constituants, entraînant ainsi des écarts significatifs (Gou et al., 2013). Par ailleurs, la présence de composés chimiques indésirables peut également affecter les mesures. L'effet préjudiciable de tels composés a été observé par Tonolini et al. (2023) durant l'analyse des matrices de lactosérum, nécessitant le développement d'une technique de calibration spécifique pour en atténuer l'impact.

Lorsqu'il est question de traiter les facteurs externes à l'aide de méthodes de prétraitement, la température du produit est le paramètre le plus étudié (Roger et al., 2003). Ceci s'explique par le fait que la température influence significativement les mesures spectroscopiques proche infrarouge en modifiant les propriétés chimiques et physiques des échantillons. L'effet le plus documenté concerne le déplacement des bandes d'absorption de l'eau provoqué par les

vibrations des liens O-H (Shenk et al., 2007). Pour minimiser ces influences, il est essentiel de standardiser la température des échantillons avant l'analyse, d'adopter des méthodes de prétraitement adaptées et d'intégrer cette variabilité dans les modèles de calibration afin d'assurer la robustesse et la précision des mesures.

2.6.6. Méthodes

Les méthodes associées à la spectroscopie proche infrarouge englobent à la fois les procédures suivies par le personnel, incluant toutes les étapes nécessaires à la calibration, ainsi que les techniques utilisées pour le prétraitement des spectres et la construction des modèles d'analyse.

La série d'opérations liées à l'échantillonnage pour les mesures de référence joue un rôle clé dans la précision des résultats. Shenk et al. (2007) suggèrent fortement de standardiser les procédures de traitement des échantillons ainsi que la façon de les présenter à l'instrument tout en mentionnant que des différences dans ces procédures sont inévitables. Ces opérations comprennent les techniques d'échantillonnage, les manipulations faites sur les échantillons, les méthodes d'emballage et de stockage ainsi que toute autre étape reliée à la préparation des échantillons. Les auteurs avertissent qu'une attention particulière doit être accordée à l'humidité et à la température de l'échantillon lors de la mesure puisque ceux-ci ont tendance à s'équilibrer avec les conditions de l'environnement à moins qu'ils ne soient conservés dans des sacs hermétiques. Ce phénomène d'équilibrage est régi par le processus de diffusion de la vapeur d'eau de l'intérieur de l'échantillon vers sa surface. La vitesse et l'efficacité de cette diffusion dépendent de la quantité de matière, de la taille des particules et du type d'emballage, ce qui peut entraîner des variations en fonction du temps et de la préparation des échantillons (Niemöller et Behmer, 2008).

En outre, la méthode de mesure de référence elle-même comporte une incertitude qui doit être évaluée afin d'éviter l'apparition de biais ou de variations de la pente de régression du modèle. En effet, puisque la spectroscopie NIR est une méthode indirecte, celle-ci absorbe l'erreur de la méthode de référence en plus de l'erreur intrinsèque (Bock et Connelly, 2008). D'ailleurs, Prieto et al. (2009) énoncent que des erreurs importantes dans les données de référence peuvent

réduire la prévisibilité d'un appareil NIR. La précision et la répétabilité de cette méthode sont donc très importantes pour assurer une calibration précise.

La calibration multivariée repose sur l'application de différentes techniques de prétraitement et d'analyse qui doivent être sélectionnées avec rigueur. Comme présenté dans la section 2.5.1, les méthodes de prétraitement permettent de corriger certains effets indésirables influençant le spectre mesuré. Le modèle de calibration utilisé a également un effet direct sur les performances d'estimation puisque c'est à l'intérieur de ce modèle que la relation entre le spectre et les mesures de référence est établie.

2.6.7. Diagramme de Pareto

Le diagramme de Pareto présenté à la figure 2.11 illustre les facteurs classés par fréquence d'occurrence, en ne comptabilisant qu'une seule mention par article pour chaque facteur. Deux facteurs liés aux propriétés des échantillons se démarquent par leur fréquence élevée, soit l'hétérogénéité et la température du produit. Les facteurs suivants, également fréquemment mentionnés, incluent la grosseur des particules, la température ambiante, le mouvement et l'orientation des échantillons, l'état de surface, la distance de lecture ainsi que les méthodes de préparation des échantillons. Un second groupe de facteurs, apparaissant avec une occurrence un peu plus basse, mais toujours significative, comprend le type d'échantillon, la technologie du capteur, la taille et la densité de l'échantillon, la longueur du trajet optique, la température interne du spectromètre, le spectre du convoyeur, ainsi que les mesures de référence. Enfin, les facteurs complétant les 80 % cumulés des occurrences comprennent la lumière environnante, le modèle de calibration, les techniques d'échantillonnage, puis la disposition et l'épaisseur des échantillons.

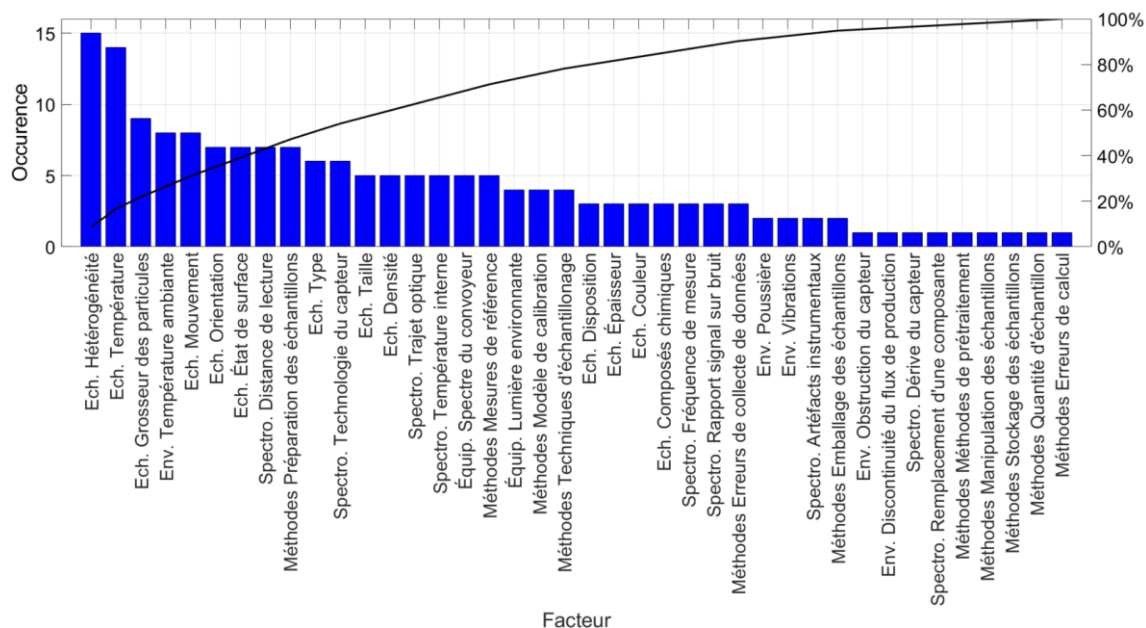


Figure 2.11 : Diagramme de Pareto des facteurs d'influence des mesures NIR
Ech. = Échantillon, Equip. = Équipements, Env. = Environnement, Spectro. = Spectromètre.

Bien que l'occurrence de certains facteurs soit plus élevée, il est trop tôt pour tirer des conclusions sur leurs effets réels dans l'application, car cela ne garantit pas nécessairement sa prédominance. En effet, les résultats obtenus à cette étape sont influencés par le nombre d'articles analysés et l'échantillonnage ne peut pas être considéré représentatif de toutes les études effectuées sur la spectroscopie en direct. Ces résultats permettent essentiellement de représenter de façon plus claire l'espace des éléments perturbateurs potentiels.

2.7. Conclusion

L'état de l'art en gestion de la qualité dans le secteur agroalimentaire est de plus en plus orienté vers un nouveau paradigme innovant guidant les industries vers l'implémentation d'une gestion de la qualité par design au lieu de contrôles inférentiels basés sur des tests de qualité tardifs et discontinus. Le pilier le plus important de cette nouvelle méthode consiste à installer une mesure en continu des paramètres critiques de qualité du produit, dont la plus adaptée aux produits alimentaires est la spectroscopie proche infrarouge. Bien qu'offrant de réels avantages comparativement aux méthodes traditionnelles, la spectroscopie NIR demeure complexe à

implémenter dans un environnement industriel. En effet, les multiples méthodes de traitement de données et d'analyses multivariées rendent possible le développement de calibrations performantes et précises en laboratoire, mais très peu de preuves existent sur le succès d'une implémentation in situ ou en direct dans un environnement industriel. La liste des facteurs d'influence est longue et contient des éléments pouvant provenir de différentes sources. De plus, les auteurs ayant identifié ces facteurs ne semblent pas parvenir à quantifier leurs effets réels. Ainsi, le développement d'une calibration robuste pour la lecture de paramètres critiques en direct représente un défi de taille.

CHAPITRE 3 - MÉTHODOLOGIE

Les objectifs de ce projet de recherche visent à identifier et à quantifier les facteurs qui influencent la calibration d'un spectromètre proche infrarouge et la précision de mesure de l'humidité. Pour y parvenir, une méthodologie structurée doit être développée afin d'encadrer rigoureusement les différentes étapes de l'analyse. Le présent chapitre est constitué de deux sections principales orientées vers l'atteinte de ces objectifs.

La première section décrit la méthodologie associée à l'identification des facteurs pertinents et mesurables pour l'étude. Basée sur la liste des facteurs d'influence potentiels rapportés lors de la revue de la littérature, une analyse du contexte industriel est effectuée en tenant compte des particularités du procédé de fabrication et des caractéristiques du produit étudié. À partir de ces deux sources d'information, les facteurs sont ensuite filtrés pour ne retenir que ceux jugés pertinents et mesurables dans le cadre spécifique de cette étude.

La deuxième section vise à quantifier l'influence de ces facteurs à l'aide d'une méthodologie adaptée pour chacun d'eux. La stratégie d'analyse adoptée repose sur le choix de méthodes spécifiques, déterminées en fonction de la nature du facteur étudié et du type de donnée disponible. Ainsi, selon les caractéristiques propres à chaque facteur, l'analyse pourra prendre la forme d'une analyse statistique descriptive ou inférentielle, d'une étude de la causalité ou encore d'une expérimentation contrôlée.

3.1. Identification des facteurs d'influence

La première étape du projet de recherche consiste à identifier les facteurs d'influence pouvant affecter la mesure d'humidité en direct du capteur NIR pour l'application spécifique de l'entreprise. Les activités relatives à cette étape contribuent à produire la base des analyses et des essais dédiés à la quantification des facteurs considérés les plus importants. Ainsi, le contenu de cette section comprend à la fois la méthodologie, les informations recueillies ainsi que les facteurs sélectionnés.

Pour compléter la liste des facteurs recensés dans la revue de la littérature, un relevé des éléments spécifiques à l'application de l'entreprise est réalisé. Cette analyse porte sur

l'environnement du capteur, les équipements avoisinants, les aspects liés au capteur NIR, les propriétés physiques et chimiques des produits fabriqués, ainsi que les techniques et méthodes utilisées pour la calibration et la vérification des performances.

Les spécifications du capteur sont examinées afin d'identifier ses limites et ses contraintes, tout en vérifiant dans quelle mesure elles sont respectées. Par ailleurs, plusieurs rencontres avec le fabricant sont prévues pour bénéficier de leur expertise et approfondir la compréhension de la technologie. La première rencontre est consacrée à la revue technique du capteur, abordant les éléments critiques ainsi que les fonctions intégrées. Une seconde séance est orientée sur l'analyse des facteurs d'influence potentiels et les méthodes de correction appliquées sur les spectres. Finalement, une discussion détaillée sur le modèle de calibration et les performances attendues vient clore ces échanges.

Un diagramme d'Ishikawa (Ishikawa, 1985) est ensuite construit en identifiant les catégories principales pouvant regrouper les éléments qui affectent les mesures d'humidité par proche infrarouge. Utilisé abondamment dans les processus de résolution de problème, ce format visuel permet de représenter les causes potentielles d'un problème de manière efficace. Celui-ci regroupe d'une part les constats issus de la revue de la littérature, et d'autre part, les observations tirées de l'analyse de l'application industrielle. Ainsi, les variables indépendantes représentant les facteurs potentiels d'influence sont rassemblées par catégorie sans égard à leur importance ou prédominance.

Le diagramme d'Ishikawa doit être filtré en ne conservant que les éléments qui sont pertinents et utiles pour l'application précise de l'entreprise. Ainsi, les éléments qui ne sont pas applicables dans le contexte opérationnel doivent être retirés de cette étude. Il en va de même pour les facteurs considérés comme constants, c'est-à-dire ceux qui ne sont pas susceptibles d'introduire de variabilité dans les mesures. Enfin, les facteurs pour lesquels aucune méthode de mesure n'est envisageable doivent également être exclus de l'analyse.

3.1.1. Étude de l'application

Des séances d'observations et de relevés d'information ont été organisées sur place en milieu industriel. Ces activités ont notamment permis d'étudier l'environnement réel du capteur, les caractéristiques des produits fabriqués ainsi que les opérations traitant de l'échantillonnage des lots de calibration et de vérification. Les sous-sections suivantes contiennent les observations effectuées durant l'étude.

3.1.1.1. Environnement du capteur

Le capteur est installé à une distance fixe de 257 mm au-dessus du flux de produits semi-finis à la fin de la ligne de production. Il est positionné au-dessus d'un convoyeur qui achemine la matière dans des bacs de collecte. Des guides latéraux sont installés de part et d'autre du convoyeur afin de maintenir le produit centré dans la zone de mesure du capteur. Sous celui-ci, les croquettes circulent à une vitesse constante de 0,38 mètre par seconde. Toutefois, leur disposition et leur orientation demeurent aléatoires, en raison des rebonds provoqués par les transferts entre convoyeurs.

Le flux de produit sous le capteur est normalement continu, mais peut devenir intermittent dépendamment des événements qui surviennent en production. Bien que le débit soit généralement stable et constant pendant une séquence de fabrication, des perturbations surviennent fréquemment au démarrage, à l'arrêt ou lors d'un changement de produit. De plus, certains bris d'équipement peuvent entraîner une interruption temporaire du flux. Dans ces situations, il est possible que le capteur effectue des mesures erronées ou mesure directement un spectre associé au tapis du convoyeur.

Le capteur est bien protégé des contacts possibles avec d'autres équipements ou des impacts provoqués par les activités humaines de la production. De plus, aucun autre équipement ou pièce mobile ne peuvent obstruer sa zone de mesure. Le spectromètre est fixé solidement à une barre horizontale, laquelle est vissée sur un poteau ancré au mur et découplé du convoyeur, ce qui empêche les vibrations des autres équipements de se propager jusqu'au capteur. En effet,

aucune vibration détectable, que ce soit par observation visuelle ou par perception tactile, n'a été observée.

L'endroit où le spectromètre est installé ne reçoit aucune lumière du soleil. La seule source lumineuse de l'environnement est constituée de trois unités de ballast comprenant chacun quatre tubes Philips InstantFit à DEL de puissance lumineuse fixe. Chaque ballast peut s'éteindre et s'allumer selon un détecteur de mouvement provoquant ainsi un changement d'intensité du fond lumineux dans l'environnement du capteur. Le centre du point illuminé du capteur se trouve à une distance diagonale de 1550 mm du ballast le plus proche.

Les chariots élévateurs circulent régulièrement à proximité de la ligne où le capteur NIR est installé, notamment pour remplacer les bacs de produits semi-finis. Ils sont munis d'une lumière bleue dirigée vers le sol qui s'allume lorsque le chariot recule. En raison de l'angle de la lumière et des obstacles, il est physiquement impossible pour celle-ci d'illuminer le convoyeur où circule le produit et la zone de mesure du capteur NIR.

L'environnement est caractérisé par la présence de poussière causée à la fois par la proximité d'un tamis et par le transfert de la matière entre les sections distinctes des convoyeurs. Aucun système de ventilation ni d'extraction n'est présent, ce qui provoque l'accumulation considérable de poussière sur les équipements, notamment sur le capteur. La température ambiante de cette section de l'usine peut varier selon les saisons, bien qu'elle bénéficie d'une certaine régulation par un système de chauffage et d'air climatisé.

3.1.1.2. Caractéristiques des produits

Plus d'une centaine de produits sont fabriqués sur la ligne de production de l'entreprise. Ces produits peuvent varier selon leur forme, leur taille, leur densité, les ingrédients utilisés pour l'ajout de saveur appliqués en surface ainsi que leur recette. Il existe 147 ingrédients différents pouvant être utilisés dans la conception des recettes. Les différentes formes possibles sont la forme standard représentant un cône tronqué, l'étoile, le lotus ainsi que le cœur. Le diamètre approximatif peut varier de 8 à 33 mm alors que leur épaisseur couvre la plage de 5 à 9 mm. La couleur du produit fini est principalement brune, mais peut aller du brun foncé au brun clair ou

même au beige. L'hétérogénéité des croquettes peut varier, car il n'y a pas de dispositif ni de méthode pour assurer une pâte homogène autre qu'un temps de brassage fixe au mélangeur. Ce temps de brassage peut parfois être augmenté selon le jugement de l'opérateur. De façon visuelle, l'hétérogénéité d'une croquette semble principalement affectée par la présence de morceaux de gras provenant de la viande utilisée.

Les principales caractéristiques des produits sont évaluées périodiquement en laboratoire par le personnel du département de la qualité. Les analyses portent principalement sur trois composantes majeures : le taux d'humidité, le taux de protéine et le taux de gras. Ces paramètres sont exprimés en pourcentage de la masse totale du produit, ce qui implique une interdépendance entre eux. En effet, toute variation dans la proportion d'un composant influence les proportions des autres. Par exemple, une augmentation du taux de protéine entraîne, à masse constante, une diminution relative du taux d'humidité et du taux de gras. Ce phénomène est souvent utilisé comme stratégie de contrôle rapide pour l'atteinte des cibles d'opération.

La température du produit qui circule sous le capteur est régulée par un refroidisseur. La logique de contrôle n'est cependant pas précise, car elle ne fait que sélectionner l'entrée d'air extérieur ou celle de l'usine dépendamment de la température extérieure. En effet, afin d'économiser de l'énergie, si la température extérieure est dans la plage entre 0 et -15 °C, l'air utilisé pour refroidir le produit provient de l'extérieur. Autrement, l'air de l'usine, parfois climatisée, est acheminé au refroidisseur. L'objectif de cette logique n'est pas de réguler la température du produit en fin de ligne. Elle permet simplement d'empêcher la condensation dans les bacs de produits semi-finis en s'assurant que la température du produit ne soit pas trop élevée par rapport à la température ambiante de la salle.

3.1.1.3. Procédure d'échantillonnage

L'entreprise s'est dotée d'une procédure d'échantillonnage dédiée à la construction et la vérification des modèles de calibration NIR. Cette procédure débute par la planification des produits à échantillonner à partir de la cédule de production. Les produits d'intérêt sont ensuite sélectionnés en fonction de plusieurs critères afin d'assurer une construction adéquate du

modèle de calibration. Un produit est sélectionné pour l'échantillonnage si celui-ci n'a pas été mesuré auparavant ou si les écarts de mesure NIR observés lors des échantillonnages antérieurs sont jugés trop élevés. Autrement, si le produit a déjà été sélectionné et que les performances de mesure sont satisfaisantes, il est généralement exclu de la liste des nouveaux échantillons. Pour vérifier le modèle, la sélection des produits repose plutôt sur leurs spécifications et leur cible d'humidité. L'objectif principal est de couvrir la plus large plage d'humidité possible afin de garantir la robustesse du modèle.

Une fois les produits identifiés, un tableau d'échantillonnage est fourni aux superviseurs de production qui vont coordonner la prise d'échantillon sur la ligne. Lors de l'échantillonnage, l'opérateur doit appuyer sur un bouton dans l'interface usager du capteur NIR pour démarrer la prise de mesure. Il doit ensuite nommer l'échantillon puis aller récolter une quantité de produit sous la chute à l'aide d'un bol. La mesure s'effectue pendant une durée de 10 secondes sur le flux de produit. Durant ces 10 secondes, environ 450 spectres sont récoltés et le capteur fournit une mesure moyenne contenant notamment les valeurs d'absorbance à chaque longueur d'onde, l'humidité estimée ainsi que des informations additionnelles. Ces données sont horodatées et enregistrées dans la base de données interne du capteur. Pour que l'échantillon et la mesure soient valides, l'opérateur doit donc s'assurer d'obtenir le produit sous la chute pendant la prise de mesure du capteur. Autrement, ceci peut contribuer à augmenter l'erreur associée à l'échantillonnage.

L'opérateur transfère ensuite le contenu du bol dans un sac, il le scelle, puis applique une étiquette sur laquelle il inscrit le nom de l'échantillon, sa description, le lot de production, le numéro de ligne, le quart de travail et l'heure de l'échantillonnage. Le sac est entreposé dans un bac sur une étagère dans le local de l'assurance qualité. Ce local est à proximité du capteur, donc partage les mêmes conditions environnementales que l'usine à l'exception d'une unité d'air climatisé qui peut être en fonction. La porte du local reste cependant ouverte en tout temps pour faciliter les allées et venues du personnel. Un délai d'entreposage est requis, car la livraison au laboratoire externe se fait selon une cédule établie qui optimise les transports d'échantillon tout en respectant une quantité minimum afin de réduire les coûts. Ainsi, compte tenu du

moment de l'échantillonnage, la durée de stockage des échantillons peut varier considérablement.

La dernière étape de la chaîne d'échantillonnage est le transport des échantillons au laboratoire externe pour effectuer les mesures de référence. Le laboratoire accrédité ISO/IEC 17025:2017 se situe à moins de 5 km des installations de l'entreprise. Le trajet ne dure normalement que quelques minutes. Arrivés au laboratoire, les échantillons sont stockés avant d'être testés selon les standards établis. Les données officielles relatives à la précision de ces mesures sont présentées dans le tableau 3.1. Ce tableau présente les indicateurs relatifs à la validation des méthodes utilisées, soit la limite de quantification (LOQ), la limite de détection (LOD), l'incertitude de la mesure ainsi que le coefficient de variation (CV) représentant l'écart-type divisé par la moyenne.

Tableau 3.1 : Précision des méthodes de mesure de référence de l'humidité

Test	Méthodes de références	LOQ	LOD	Incertainde	CV
Humidité 104 °C	Méthode 2.1.4, NFTA				
	Méthode 935.29, AOAC	1.97 %	0.59 %	0.44 %	≤ 3.35 %
	Méthode 945.15, AOAC				

Le laboratoire envoie un certificat d'analyse pour chaque échantillon mesuré contenant l'identification de l'échantillon, sa date d'analyse et les résultats obtenus. Ces résultats sont ensuite entrés dans l'interface utilisateur du capteur en tant que mesure de référence pour chaque échantillon. Une table contenant toutes les informations peut être exportée et envoyée au fournisseur du spectromètre pour construire, vérifier ou mettre à jour le modèle de calibration.

3.1.1.4. Spécifications techniques du capteur NIR DA 7440™

Perten, membre du groupe Perkin Elmer, propose une solution complète de mesure en direct par spectroscopie incluant l'équipement, l'installation, la formation, la calibration et le service après-vente pour les entreprises qui fabriquent des produits de toutes sortes. Le capteur DA 7440™ représente, selon la brochure du produit, « un capteur avancé, moderne et polyvalent

pour les mesures en direct ». Celui-ci permet notamment de procéder à la mesure de multiples constituants en temps réel. Adapté pour les environnements industriels, l'indice de protection de cet équipement est IP67 et la plage de température d'opération est de 0 à 40 °C. Cet appareil doit être relié à un système d'air comprimé afin de maintenir une pression interne positive et assurer que la lentille reste propre et exempte de contaminants.

Muni d'une matrice de 256 diodes InGaAs, le détecteur optique couvre une plage de longueurs d'onde de 950 à 1650 nm séparée en 141 longueurs d'onde, soit une valeur à tous les 5 nm. Selon les documents fournis par le fabricant, cette technologie moderne de haute performance est particulièrement adaptée aux applications de mesures en direct. L'appareil offre un haut rapport signal sur bruit grâce à l'utilisation du matériau InGaAs et à un système de refroidissement thermoélectrique. L'absence de pièce mobile et la capacité à acquérir un spectre simultané sur toutes les longueurs d'onde représentent également des avantages.

Le spectromètre contient également une source de référence interne ainsi qu'un réseau de diffraction stationnaire recouvert d'or. Il est muni de sa propre source d'illumination consistant en une lampe tungstène-halogène activée en permanence. L'annexe III contient un tableau des spécifications techniques principales du capteur et son principe d'opération est disponible en ligne (PerkinElmer AB, 2014).

Le capteur est conçu pour être installé sur une barre positionnée à 426 mm (-50 + 100 mm) du produit à analyser (PerkinElmer AB, 2023). Ceci donne une distance de lecture de 250 mm (200-350 mm). Le capteur illumine une section circulaire directement en dessous et la mesure est faite au milieu de celui-ci tel que présenté à la figure 3.1.

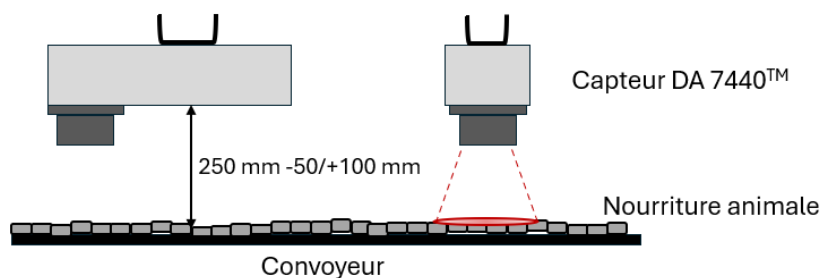


Figure 3.1 : Montage du capteur NIR DA 7440™

Le manuel d'installation et de maintenance du produit mentionne que la fonction de stabilisation continue de la longueur d'onde et les vérifications automatiques avec une référence interne garantissent des mesures précises, une stabilité dans le temps et une transférabilité d'étalonnage entre différents instruments. La brochure du produit indique que le capteur est en mesure de détecter automatiquement s'il y a un produit ou non sur le convoyeur et peut gérer les variations de hauteur, de couleur et de température du produit sans aucun effet indu sur les résultats de mesure.

3.1.1.5. Entretien avec les experts

Tout au long du projet de recherche, les différents intervenants de Perten ont été consultés dans le but de mieux connaître la technologie du capteur DA 7440TM et d'identifier les méthodes implémentées pour la correction ou à la compensation des facteurs d'influence de la mesure NIR. Le groupe d'intervenants fut constitué d'un scientifique d'application, d'un spécialiste des instruments spectroscopiques ainsi que d'un spécialiste d'application. Les premiers entretiens ont permis de clarifier certaines spécifications et d'identifier les techniques utilisées visant à assurer une précision et une répétabilité adéquate de l'instrument.

Le capteur procède à la lecture d'un spectre complet à une fréquence de 30 spectres par seconde. Dans une application réelle, l'appareil est normalement configuré pour produire une valeur moyenne toutes les 10 secondes. Il est possible de modifier la fréquence de lecture selon le besoin. Il est rare que ce paramètre change une fois qu'il est ajusté à l'application. Dans le cas actuel, la valeur par défaut de 10 secondes est utilisée.

Une fonction de vérification périodique de la référence interne est implémentée dans le capteur. Cette fonction permet d'effectuer la mesure d'une référence stable toutes les 30 minutes. Lors de la procédure de mesure de référence, une lampe au xénon agissant comme source lumineuse secondaire s'active, tandis qu'un obturateur se ferme pour bloquer la source principale. La matrice de diode du détecteur est alors illuminée et produit un spectre brut qui peut être comparé avec une référence mémorisée. Le pourcentage de différence entre le signal obtenu et le signal mémorisé est calculé par l'appareil et comparé à une valeur seuil. La figure 3.2 montre

un graphique de comparaison des références. Celui-ci est accessible à partir de l'interface utilisateur du capteur.

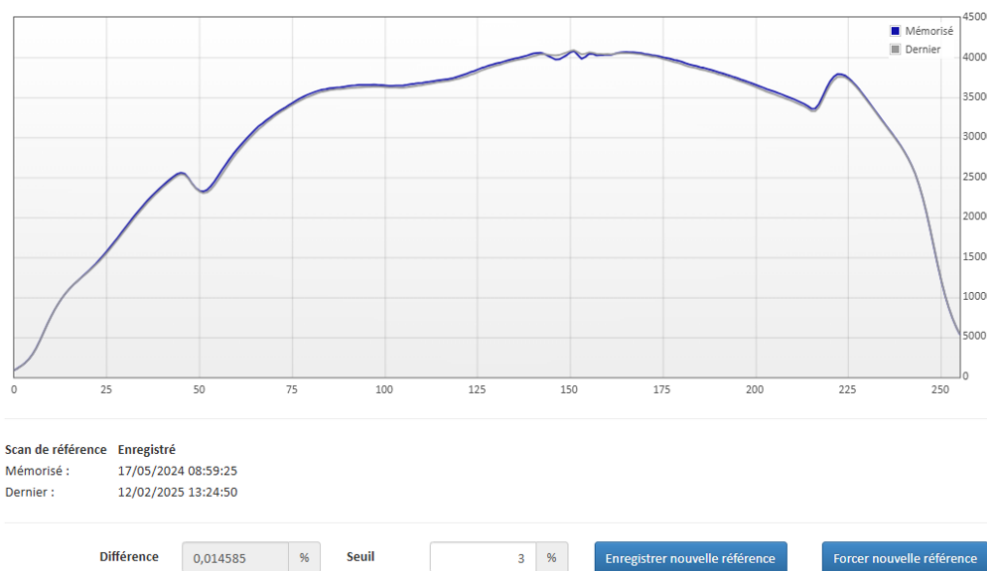


Figure 3.2 : Comparaison des références internes du capteur NIR DA 7440™

Outre les capacités de mesure de référence interne, une référence externe est également disponible afin d'évaluer la dérive dans les mesures. La mesure de cette référence doit être effectuée de façon régulière comme mentionné dans le manuel d'installation et de maintenance de l'appareil. La cellule utilisée pour la mesure comprend une rondelle en plastique polystyrène poly (1-phényléthylène) caractérisée par une signature spectrale de réflectance définie. Toute déviation de plus de 1 % doit être corrigée.

Une fonction de normalisation permet le transfert de calibration d'un capteur à un autre sans avoir à reconstruire le modèle de calibration. La normalisation doit être effectuée lors du remplacement d'une pièce ou lors d'une maintenance préventive.

Lorsqu'interrogée sur les facteurs potentiels pouvant mener à un écart de lecture, l'équipe mentionne l'importance de bien contrôler l'étape d'échantillonnage et de s'assurer que les méthodes de référence en laboratoire sont précises. En effet, un modèle de calibration NIR ne peut pas être plus précis que les mesures de référence. La précision de ces mesures inclut non

seulement la précision de la méthode utilisée, mais également toutes les sources d'erreurs relatives à la chaîne d'échantillonnage.

Dans les environnements où la poussière est présente, celle-ci peut avoir un impact sur la lecture seulement si son origine est organique. Cependant, considérant la conception du capteur et son branchement à un système d'air comprimé, les chances d'observer un impact sont très faibles. Celles-ci sont encore plus faibles si les procédures d'entretiens et de nettoyage spécifiées dans le manuel d'utilisation sont effectuées.

Les produits fabriqués par l'entreprise peuvent faire l'objet d'ajout de saveurs externes. Ces saveurs sont constituées de poudre, d'huile ou de gras appliqués à l'externe du produit affectant ainsi leurs propriétés de surface. Selon l'équipe d'experts, le gras pourrait notamment bloquer une certaine partie de la lumière et affecter la profondeur de pénétration. Cependant, selon leur expérience, cet effet reste négligeable.

Pour gérer les effets de la plupart des facteurs générant des spectres erronés, créant une dispersion de la lumière ou provoquant une ligne de base dans le signal, différentes techniques de prétraitement sont implémentées dans le logiciel du capteur. Premièrement, l'instrument effectue le calcul de la distance de Mahalanobis (Mahalanobis, 1936) pour chaque mesure et peut rejeter des mesures aberrantes, souvent associées à l'absence de produit ou à la lecture du spectre du convoyeur. Deuxièmement, les méthodes SNV et *detrend* (Barnes et al., 1989) sont appliquées aux spectres afin d'éliminer les effets relatifs à la grosseur des particules, leur orientation ainsi que les différences dans les propriétés de surface. La méthode de lissage et de première dérivée de Savitzky-Golay (Savitzky et Golay, 1964) est également appliquée. Les écarts reliés à l'hétérogénéité du produit sont réduits par l'implémentation d'une moyenne constituée de milliers de lectures rapides. Finalement, pour compenser les effets de la température du produit, la technique d'orthogonalisation des paramètres externes (EPO) (Roger et Boulet, 2018)) est appliquée lors de la construction du modèle de calibration. Cette technique est largement utilisée en raison de sa simplicité à compenser les effets des facteurs préjudiciables.

Au niveau du modèle de calibration, le fournisseur utilise la méthode Honigs Regression (Honigs et Åberg, 2012). Cette technique développée à l'interne est constituée d'une régression PLS

combinée à un apprentissage par proche voisin (KNN). Elle permet notamment de gérer les variations entre les produits. La philosophie derrière cette régression d'apprentissage énonce qu'il est plus facile d'améliorer une calibration existante en l'adaptant à de nouveaux échantillons que de tenter de développer une calibration robuste à tous les effets possibles et futurs. Ainsi, une première calibration standard est développée par méthode PLS traditionnelle basée sur une bibliothèque de base d'échantillons. Celle-ci est ensuite appliquée aux échantillons futurs en obtenant les distances voisines entre ces nouveaux échantillons et ceux présents dans la bibliothèque. De nouveaux groupes d'échantillons peuvent alors être détectés et une mise à jour de la bibliothèque peut ensuite être effectuée pour améliorer les performances de prédiction.

Puisque Perten déploie des instruments de mesure NIR dans plusieurs secteurs de l'agroalimentaire, ceux-ci peuvent bénéficier d'une bibliothèque contenant près de 16 000 échantillons d'origine différente. Normalement, lors de l'installation de leur capteur, ils utilisent une calibration de base utilisant cette bibliothèque, puis mettent à jour le modèle avec les échantillons du client. Au début du projet de recherche, le dernier modèle de calibration mis à jour présentait un SECV de 0.499 tel que le montre la figure 3.3.

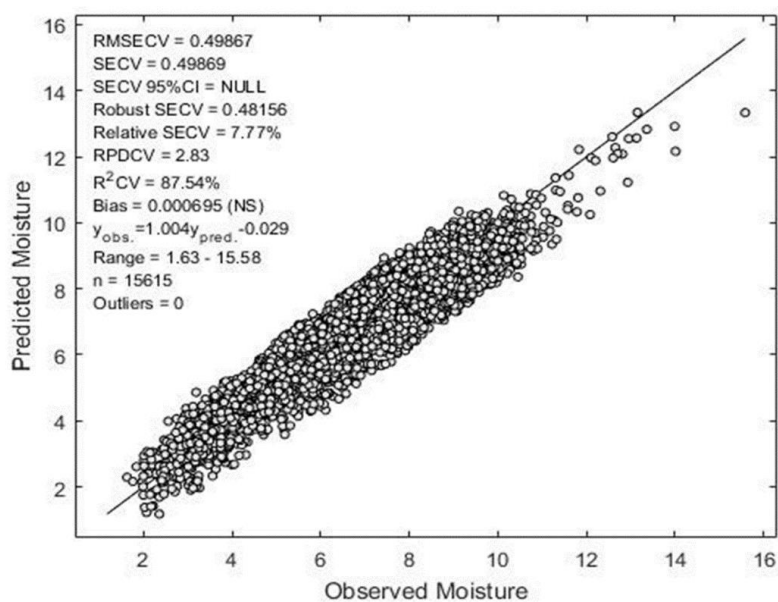


Figure 3.3 : Performances du modèle de calibration initial

3.1.2. Diagramme d'Ishikawa initial

En fusionnant les facteurs identifiés dans la revue de la littérature avec les éléments observés durant l'étude de l'application ainsi que les informations provenant du fournisseur, un diagramme Ishikawa global peut être élaboré. Dans ce diagramme présenté à la figure 3.4, cinq catégories principales sont identifiées. Celles-ci permettent de regrouper les facteurs selon leur origine : l'environnement, les échantillons, le spectromètre, les équipements auxiliaires ainsi que les méthodes.

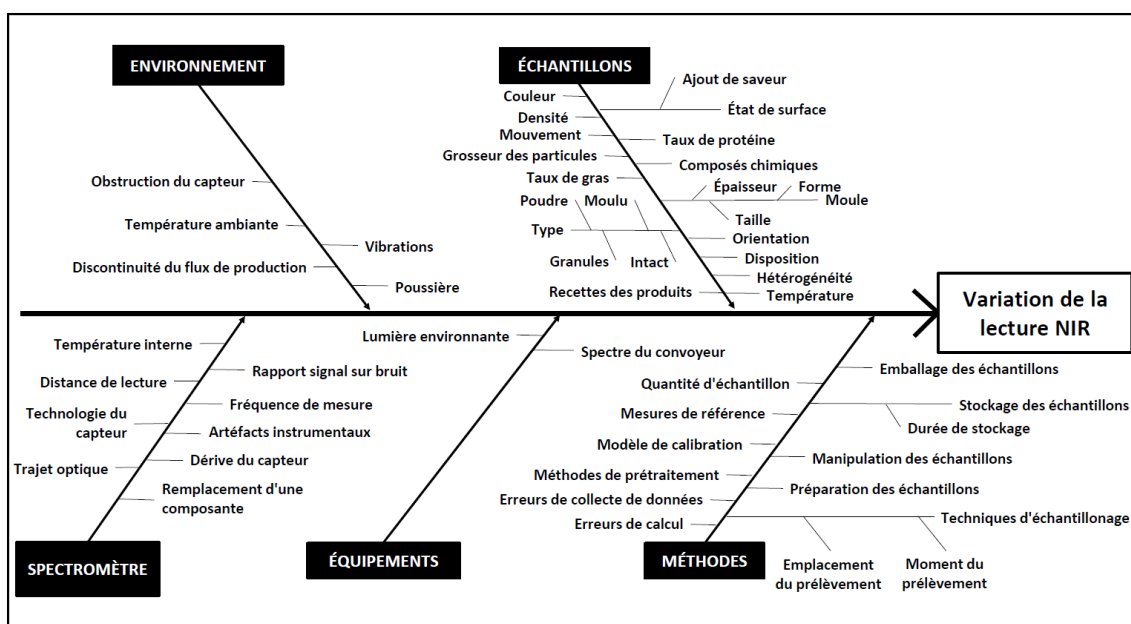


Figure 3.4 : Diagramme d'Ishikawa initial des facteurs d'influence

3.1.3. Sélection des facteurs

Le diagramme d'Ishikawa de la figure 3.4 contient plus de 45 facteurs potentiels. Parmi ces facteurs, certains ne sont pas applicables au cas faisant l'objet de l'étude. D'autres sont constants et n'engendrent aucune variabilité dans la mesure d'humidité. Enfin, une autre partie de ceux-ci ne sont pas contrôlables ni mesurables. Tous ces facteurs doivent être retirés pour ne révéler que ceux qui sont pertinents pour l'étude.

3.1.3.1. Facteurs non applicables

Les facteurs potentiels identifiés dans l'étude ne sont pas tous applicables pour le cas de cette entreprise. Cette applicabilité peut dépendre autant du type de produit mesuré, du processus de fabrication ou des procédures en vigueur. Par exemple, aucune pièce n'a été remplacée durant l'étude, ce qui implique qu'aucune variabilité ne peut être attribuée au facteur « remplacement d'une composante ». Le trajet optique n'est pas applicable, car il implique l'utilisation d'une fibre optique, ce qui n'est pas le cas pour cette installation. De plus, l'environnement de ce capteur ne comporte pas de vibrations et aucune obstruction n'est possible. En ce qui concerne la présence de poussière, une procédure de nettoyage et de vérification de la pression d'air a été mise en place à l'entreprise suivant les recommandations du fournisseur et assurant ainsi que le capteur reste propre et exempt de poussière.

Au niveau des caractéristiques des produits, aucun composé chimique ne peut affecter les mesures. La grosseur des particules n'est pas applicable, car celle-ci désigne principalement la dimension des particules une fois le produit moulu. Puisque l'étude ne couvre que des produits intacts, ce facteur, ainsi que le type d'échantillon peuvent être retirés.

3.1.3.2. Facteurs constants

Tous les facteurs qui ne varient pas dans l'application peuvent être retirés de l'analyse. Ainsi, la technologie du capteur et les facteurs relatifs à ses spécifications comme la fréquence de mesure et le rapport signal sur bruit ne sont pas applicables. Au niveau de la dérive du capteur, celle-ci est vérifiée périodiquement et sa stabilité peut être vérifiée au début et à la fin de l'étude. La distance de lecture ne change pas puisque le capteur est installé à une hauteur fixe conformément au requis du fournisseur. Il est possible que la hauteur du produit circulant sous le capteur change, mais cette variation est considérée négligeable et son effet est compensé par les différentes techniques de prétraitement.

En ce qui concerne les caractéristiques du produit, certaines variables ont été exclues de l'analyse en raison de leur faible variabilité ou de leur absence d'effet. La couleur, dont la plage de variation se limite du brun au beige, a été jugée négligeable. Le mouvement des échantillons

est également écarté de l'étude, car la vitesse du convoyeur reste constante. Puisque l'étude ne couvre que des échantillons intacts, la préparation de ceux-ci peut être retirée de l'analyse. De plus, l'emplacement du prélèvement d'échantillon reste toujours le même et le moment du prélèvement est considéré comme n'ayant pas d'effet tant que celui-ci se trouve à l'intérieur de la plage de 10 secondes d'échantillonnage du spectromètre.

En ce qui concerne les méthodes de prétraitement, le fournisseur du capteur utilise les techniques traditionnelles éprouvées depuis plus de 50 ans pour le traitement des spectres proche infrarouge. Ces méthodes n'ont pas changé depuis l'installation du capteur sur la ligne de production. Ainsi, tous les spectres analysés dans cette étude ont été traités par les mêmes techniques.

Finalement, les méthodes utilisées pour les mesures de référence font l'objet d'un contrôle qualité strict dont les limites et la précision sont bien définies. Il est donc acceptable de considérer les résultats de ces mesures comme ayant une précision fixe.

3.1.3.3. Facteurs non mesurables

Les facteurs non mesurables de cette étude sont constitués des variables qualitatives ou dont les efforts de mesure sont élevés ou impossibles à déployer. Premièrement, les artefacts instrumentaux représentent des effets indésirables provoqués par un instrument pouvant survenir de façon imprévue et ponctuelle. Ceux-ci dépendent généralement de la technologie utilisée ainsi que de l'effet de l'environnement dans lequel l'instrument évolue. Généralement, les fabricants implémentent des mesures pour contrer leurs effets ou les réduire au maximum ce qui les rend négligeables et difficiles à distinguer du bruit. En ce qui concerne les erreurs de calcul, les erreurs de collecte de données, ainsi que la manipulation des échantillons, celles-ci sont d'origine humaine et ne peuvent pas être contrôlées ni mesurées. Leurs effets se constatent généralement trop tard et il est difficile de prouver leur présence et de quantifier leur effet.

Au niveau des produits, la disposition des croquettes et leur orientation demeurent des variables aléatoires ne pouvant être mesurées ni contrôlées. L'hétérogénéité et la densité sont également difficilement mesurables, surtout dans un contexte industriel où le flux de production est rapide.

Par ailleurs, le contrôle de celles-ci reste difficile, car elles dépendent à la fois de la préparation des ingrédients et du mélange de la pâte, deux étapes influencées par de multiples phénomènes et contraintes opérationnelles.

3.1.4. Diagramme d'Ishikawa final

Une fois les facteurs non pertinents pour l'analyse retirés, le diagramme d'Ishikawa final, tel que présenté à la figure 3.5, peut être utilisé pour poursuivre l'étude.

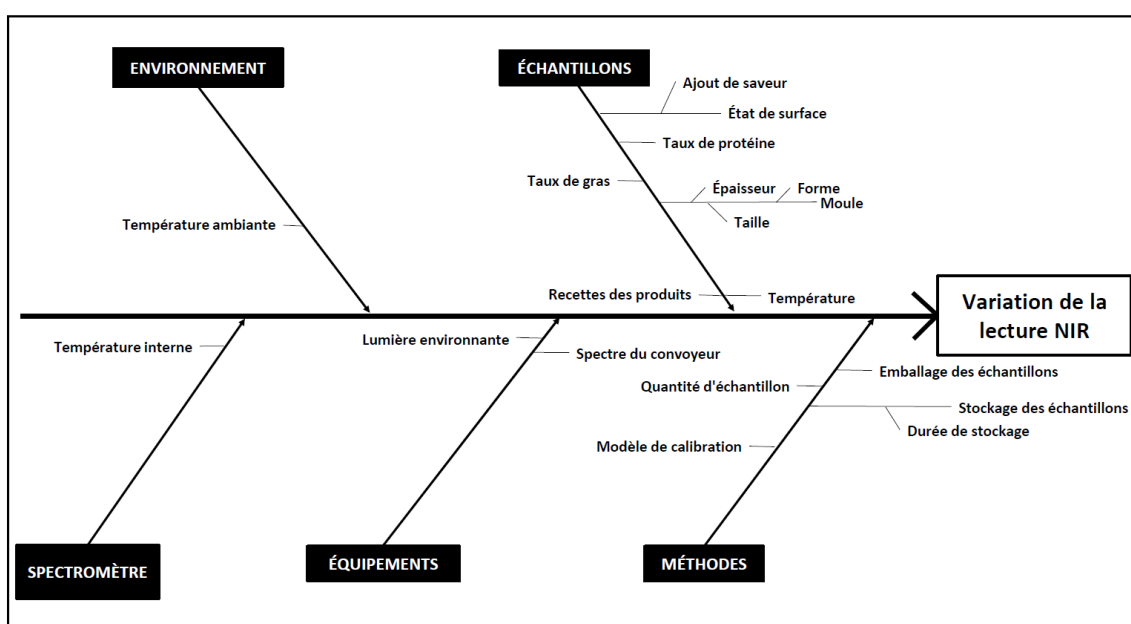


Figure 3.5 : Diagramme d'Ishikawa final des facteurs d'influence

3.2. Quantification des facteurs d'influence

Chacun des facteurs présents dans le diagramme de la figure 3.5 doit faire l'objet d'une méthodologie spécifique établie dans le but de quantifier leur impact potentiel sur la variation de mesure de l'humidité.

Cette section vise à détailler cette approche méthodologique. Elle débute par la description de la variable dépendante étudiée, en précisant son calcul et sa pertinence dans le cadre de l'étude. Ensuite, le jeu de données utilisé pour les analyses est présenté, incluant sa provenance, sa composition et les étapes de traitement préalable. Les méthodes d'analyse statistique sélectionnées sont ensuite exposées, en expliquant les raisons justifiant leur choix en fonction des types de données et des objectifs de l'analyse, puis chaque facteur inclus dans le jeu de données est présenté, accompagné de la méthode spécifique employée pour évaluer son influence. Enfin, la méthodologie associée aux facteurs faisant l'objet d'essais expérimentaux spécifiques est présentée.

3.2.1. Variable dépendante

Dans cette étude, la manière de caractériser la variabilité de la lecture est d'utiliser l'écart de mesure d'humidité entre la prédiction faite par NIR (y_{NIR}) et la valeur de référence obtenue en laboratoire (y_{ref}). La variable dépendante correspond donc à l'écart de prédiction NIR (e) et est calculée par l'équation (3.1). En utilisant ce calcul, une valeur positive d'écart correspond à une prédiction plus humide alors qu'une valeur négative correspond à une prédiction plus sèche.

$$e = y_{NIR} - y_{ref} \quad (3.1)$$

L'hypothèse de base énonce que si deux échantillons sont identiques, une variabilité sur la lecture NIR de l'humidité de ceux-ci se traduit par la même variabilité sur l'écart puisque la valeur de référence ne change pas. De plus, l'utilisation de l'écart permet d'analyser l'effet des facteurs externes à la lecture spectroscopique qui affectent la mesure de référence tels que les erreurs liées aux méthodes d'échantillonnage et les effets du stockage et de l'emballage.

Dans l'étude, il existe deux versions de la variable dépendante. La première nommée *Ecart_H* représente l'écart de mesure NIR généré par le capteur en utilisant les différents modèles de calibration qui ont été développés durant la construction du jeu de données. La deuxième *Ecart_Pred_H*, représente la valeur prédite par un seul modèle de façon hors ligne. Celle-ci est dédiée à retirer l'effet des performances des modèles de calibration pouvant masquer l'influence des autres facteurs.

3.2.2. Jeu de données de calibration

Les activités relatives à la calibration du capteur NIR DA 7440™ ont débuté avant le projet de recherche actuel. Ces activités furent orientées vers le développement d'un modèle global de calibration pouvant estimer l'humidité de tous les produits semi-finis fabriqués sur la ligne de production. Ainsi, un jeu de données comprenant 422 échantillons prélevés sur une durée de 1 an et 4 mois est disponible. Ce jeu de données contient notamment les valeurs d'absorbance du spectre obtenues sur les 141 longueurs d'onde, la valeur de l'humidité estimée par le modèle, la valeur de l'humidité mesurée par la méthode de référence en laboratoire ainsi que la variable dépendante étudiée, soit l'écart de mesure. D'autres informations sont présentes, telles que le nom de l'échantillon, la date de la prise de mesure, le nombre de mesures effectuées ainsi que le nom du modèle de calibration. Afin de permettre les analyses visant à quantifier l'effet de certains facteurs, ce jeu de données doit être fusionné avec d'autres informations associées aux variables indépendantes. Le tableau 3.2 présente ces variables, leur origine ainsi que les opérations effectuées sur celles-ci avant leur fusionnement avec le jeu de données de calibration.

Tableau 3.2 : Informations fusionnées avec le jeu de données de calibration

Variable	Origine	Opérations sur la donnée
Numéro de produit	Nom de l'échantillon	<ul style="list-style-type: none"> • Extraire le nom du produit du nom de l'échantillon ; • Valider le nom du produit ; • Uniformiser les noms de produit.
Numéro de moule	Cédule de production	<ul style="list-style-type: none"> • Extraire le numéro de moule associé au produit dans la cédule de production.
Forme du produit	Dessin technique du moule	<ul style="list-style-type: none"> • Extraire la forme du produit à partir des informations du moule.
Diamètre du produit	Dessin technique du moule	<ul style="list-style-type: none"> • Extraire le diamètre à partir des informations du moule.
Hauteur du produit	Dessin technique du moule	<ul style="list-style-type: none"> • Extraire la hauteur du produit à partir des informations du moule.
Volume du produit	Dessin technique du moule	<ul style="list-style-type: none"> • Calculer le volume du produit à partir de sa hauteur et son diamètre.
Délai de mesure en laboratoire	Base de données des échantillons, Certificat d'analyse	<ul style="list-style-type: none"> • Extraire la date de la mesure inscrite dans le certificat d'analyse ; • Calculer le délai en jour entre la prise d'échantillon NIR et la mesure en laboratoire.
Informations sur l'ajout de saveurs	Spécifications des produits	<ul style="list-style-type: none"> • Extraire les informations de saveur en poudre, gras et saveur liquide associés au produit.
Groupe de produit	Liste des ingrédients des produits	<ul style="list-style-type: none"> • Regrouper les produits selon leurs ingrédients en 5, 10 et 15 groupes par <i>clustering</i>.
Taux de protéine	Certificat d'analyse	<ul style="list-style-type: none"> • Extraire les résultats de mesure du certificat d'analyse du laboratoire externe.
Taux de gras	Certificat d'analyse	<ul style="list-style-type: none"> • Extraire les résultats de mesure du certificat d'analyse du laboratoire externe.

Tableau 3.2 : Informations fusionnées avec le jeu de données de calibration (suite)

Variable	Origine	Opérations sur la donnée
Température du produit et température ambiante	Base de données de l'assurance qualité	<ul style="list-style-type: none"> • Procéder à une requête sur la base de données du serveur de l'assurance qualité pour récupérer la température du produit et la température ambiante mesurée en fin de ligne par les opérateurs ; • Joindre les données de température au jeu de donnée de calibration selon le champ d'horodatage en utilisant la méthode d'interpolation des données.

Le jeu de données ainsi formé contient à la fois des données de type quantitatives et qualitatives. Des valeurs discrètes, continues et catégorielles sont présentes, ce qui empêche l'utilisation d'une méthode unique pour analyser les effets des variables indépendantes. De plus, la cueillette des données n'a pas été faite à partir d'un plan d'expérience équilibré. Ceci complexifie l'analyse, car les méthodes traditionnelles d'analyse de la variance comme l'ANOVA ou l'ANCOVA requièrent la mise en place d'un plan équilibré afin d'obtenir des résultats fiables. Bien que les analyses non équilibrées soient possibles, les résultats obtenus doivent généralement être interprétés avec prudence, surtout si l'ampleur du déséquilibre est important. Par ailleurs, il faut souvent procéder à une étude plus détaillée des groupes ou requérir à d'autres méthodes avant de conclure. Ainsi, afin de travailler avec cette contrainte, une stratégie spécifique doit être élaborée. Cette étude utilisera donc une combinaison de différentes méthodes statistiques afin de quantifier l'effet des facteurs.

3.2.3. Analyse des facteurs du jeu de données

La démarche méthodologique liée à l'analyse des facteurs compris dans le jeu de données débute par l'application de la forêt aléatoire, qui permet d'examiner l'influence des variables de manière globale sur l'ensemble des données. Ensuite, la matrice de covariance est calculée pour évaluer les relations entre les variables continues et la variable dépendante. Les variables

catégorielles, ainsi que les variables continues pouvant être regroupées en classes, sont ensuite analysées de façon individuelle à l'aide de la méthode ANOVA.

3.2.3.1. Forêt aléatoire

Afin de classer les facteurs par ordre d'importance dans un jeu de données déséquilibré, il est justifié de faire appel à des méthodes d'apprentissage statistiques évoluées telles les forêts aléatoires (Breiman, 2001). Le concept derrière ce type d'apprentissage consiste à faire évoluer un grand nombre d'arbres en fonction d'un vecteur aléatoire, puis d'utiliser la moyenne des sorties de chaque arbre comme valeur de prédiction. Chaque arbre est échantillonné indépendamment, mais en utilisant la même distribution pour chacun d'eux. Bien que les forêts aléatoires soient utilisées amplement pour la prédiction ou la classification, c'est plutôt leur propriété d'établir l'importance des variables qui sera utilisée dans cette étude. L'analyse par développement de forêt aléatoire est implémentée par la fonction MATLAB `TreeBagger()` en utilisant la méthode de régression et un total de 300 arbres. Bien que cette technique puisse classer les facteurs selon l'importance de leur effet sur une variable dépendante, elle ne permet pas de visualiser en quelle mesure celle-ci varie ou si la variance est provoquée par certains groupes particuliers. Afin d'obtenir cette information, d'autres méthodes d'analyses comme la matrice de covariance et l'ANOVA doivent être effectuées.

3.2.3.2. Matrice de covariance

En ce qui concerne les variables continues, la matrice de covariance est un outil utile pour quantifier la causalité entre deux variables. Cependant, avant de l'utiliser, les données doivent être normalisées pour éviter que l'échelle des variables de grande amplitude génère des covariances élevées par rapport aux variables de faible amplitude. Pour la normalisation, la méthode de la cote Z est utilisée. Celle-ci consiste à normaliser les données afin d'obtenir une moyenne nulle et un écart-type unitaire. L'équation (3.2) permet de normaliser les données par cette méthode. Dans cette équation, z représente la valeur normalisée, x la valeur originale, u la moyenne et σ l'écart-type des données.

$$z = \frac{x - u}{\sigma} \quad (3.2)$$

3.2.3.3. ANOVA

Des analyses ANOVA à un facteur sont effectuées sur les variables catégorielles ainsi que les variables continues pouvant être catégorisées afin d'étudier plus en détail les groupes pouvant être problématiques. Pour les variables continues, les critères utilisés pour déterminer la plage et les limites des groupes sont les suivants :

1. Orienter les limites des groupes selon les spécifications des produits ;
2. Utiliser des intervalles fixes pour assurer une comparabilité ;
3. Définir de trois à six groupes pour assurer une puissance statistique et éviter de masquer certains effets ;
4. Adapter les groupes selon la distribution des données à analyser.

Pour toutes les analyses ANOVA de cette étude, on considère l'hypothèse nulle H_0 énonçant que le facteur n'a pas d'effet significatif sur l'écart de mesure NIR de l'humidité. L'hypothèse alternative H_1 énonce que le facteur a un effet significatif sur l'écart de mesure NIR de l'humidité. Le seuil de signification α est fixé à 0.05.

Considérant l'ampleur du déséquilibre dans les données, la valeur p sera consultée, mais l'analyse reposera plus précisément sur l'étude des données statistiques produites par l'ANOVA. Ce déséquilibre complexifie ou empêche également l'analyse des interactions entre les facteurs. Pour cette raison, l'étude actuelle ne couvrira pas les interactions. Ainsi, la visualisation des résultats sous forme de diagramme de boîtes ainsi que l'affichage des réponses moyennes et des intervalles de confiance pourront apporter des informations plus pertinentes sur l'effet respectif des différents niveaux du facteur.

L'utilisation de l'ANOVA traditionnelle requiert que les données soient distribuées selon la loi normale et que les variances entre les différents groupes soient considérées comme égales. Pour vérifier ceci, un test de normalité d'Anderson-Darling sera appliqué à la variable dépendante. Pour ce test, on considère l'hypothèse nulle H_0 énonçant que les données suivent une loi normale. L'hypothèse alternative H_1 énonce que les données ne suivent pas une loi normale. Le seuil de signification α est fixé à 0.05. Dans le cas où les données ne suivent pas une loi normale,

une analyse devra être faite pour identifier la source de la non-normalité, puis vérifier si les données peuvent être utilisées de façon conditionnelle ou si une méthode doit être implémentée pour rendre leur distribution normale.

Afin de vérifier l'égalité des variances entre les groupes, le test de Levene sera utilisé. Pour ce test, on considère l'hypothèse nulle H_0 énonçant que les variances des groupes sont égales. L'hypothèse alternative H_1 énonce qu'au moins un groupe a une variance significativement différente. Le seuil de signification α est fixé à 0.05. Dans le cas où les variances ne sont pas égales, l'ANOVA de Welch sera utilisée comme technique d'analyse.

Finalement, les catégories composées de moins de 10 échantillons seront considérées comme statistiquement non représentatives et seront retirées du lot de données avant de réaliser chaque ANOVA.

Le schéma de la figure 3.6 présente le processus d'analyse des facteurs du jeu de données.

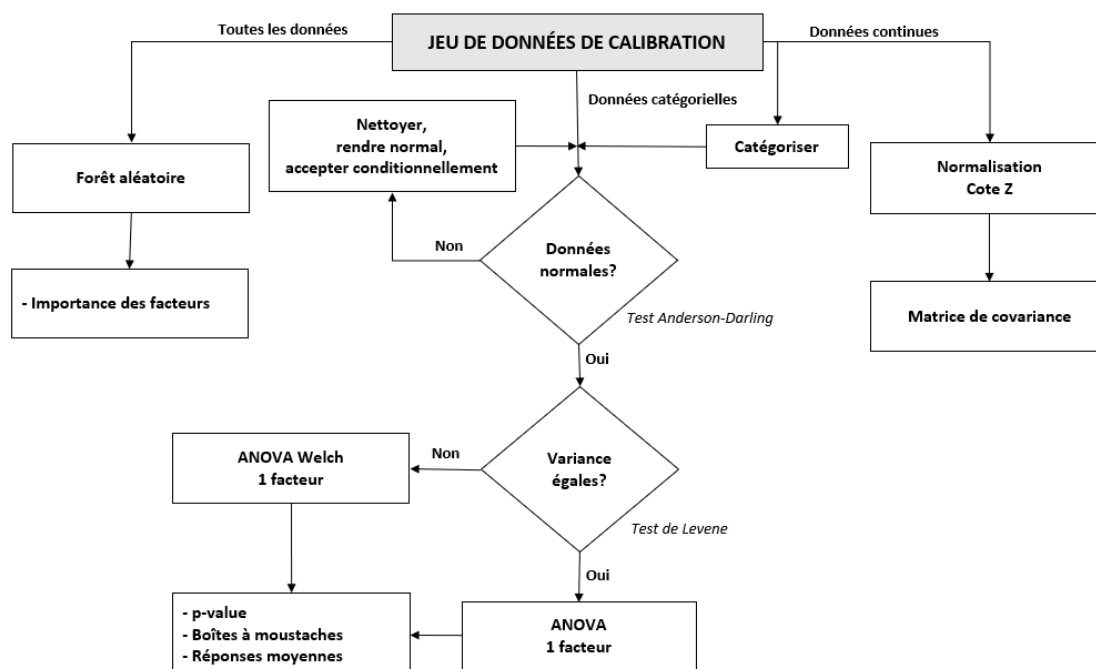


Figure 3.6 : Schéma des méthodes d'analyse du jeu de données

3.2.3.4. Modèle de calibration

Le modèle de calibration est une variable catégorielle correspondant au nom du modèle utilisé pour estimer l'humidité du produit. Le jeu de données est constitué de plusieurs modèles différents, car le processus de calibration itératif a nécessité plusieurs développements et ajustements. Ainsi, un total de sept différents modèles de calibration sont présents et chacun d'eux a une performance d'estimation différente. L'effet d'un modèle sur l'estimation de l'humidité peut affecter de façon considérable les écarts observés et masquer les effets des autres facteurs identifiés. Il est donc important de quantifier préalablement cet effet et de prendre certaines mesures pour l'éliminer avant d'analyser les autres facteurs.

La méthode d'analyse retenue pour ce facteur est l'ANOVA. Celle-ci sera en mesure de déterminer si le modèle a un impact sur l'écart d'humidité mesuré. La variable dépendante ciblée pour cette analyse représente l'écart de mesure NIR généré par le capteur, nommée *Ecart_H* dans le jeu de données. L'interprétation du résultat devra cependant être faite avec précaution, car certains modèles ont été appliqués à une faible quantité de données comme le montre la figure 3.7.

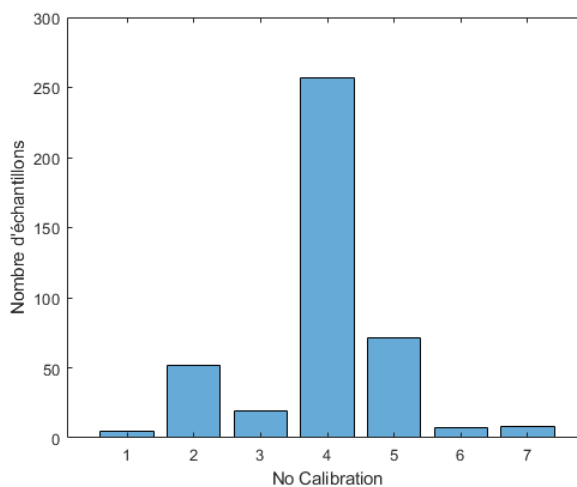


Figure 3.7 : Nombre d'échantillons estimés par modèle de calibration

Dans le cas où le modèle de calibration exerce un impact significatif sur l'écart de mesure, il sera possible de procéder à l'estimation en mode hors-ligne du jeu de données entier avec le dernier

modèle valide développé, soit le modèle 7. Ceci générera une nouvelle variable dépendante, nommée *Ecart_Pred_H*, représentant l'écart entre l'estimation d'humidité hors-ligne du modèle et les mesures de laboratoire. Dépourvue ainsi de la variance induite par l'influence des modèles, elle sera utilisée pour analyser les effets des autres variables indépendantes de l'étude.

3.2.3.5. Taux de protéine

Le taux de protéine est une variable continue dont la mesure est fournie par la méthode de référence en laboratoire externe. La méthode utilisée pour analyser l'effet de ce facteur consiste à calculer sa covariance avec l'écart de mesure d'humidité. De plus, afin d'obtenir plus d'information sur son effet potentiel, une analyse ANOVA est réalisée en catégorisant la variable en six groupes de 15 à 20, 20 à 25, 25 à 30, 30 à 35, 35 à 40 et 45 et plus. Ces groupes ont été définis à partir des cibles de taux de protéine inscrites dans les spécifications des produits et en consultant l'histogramme des mesures réelles de la figure 3.8. Celui-ci expose la présence d'un groupe distinct de données caractérisées par un taux de protéine plus élevé qui doit être étudié adéquatement.

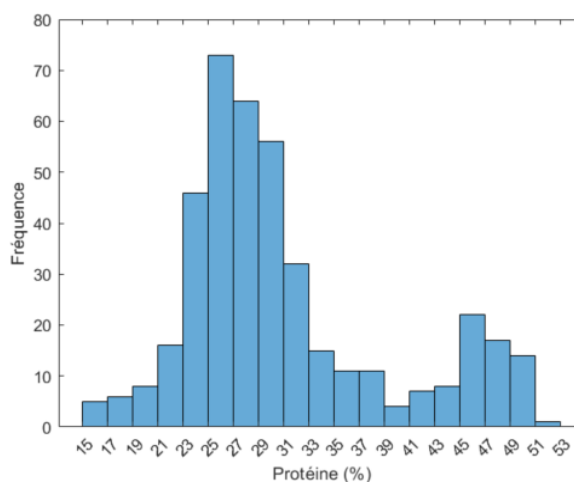


Figure 3.8 : Histogramme du taux de protéine

3.2.3.6. Taux de gras

Le taux de gras est une variable continue dont la mesure est fournie par la méthode de référence en laboratoire externe. La méthode utilisée pour analyser l'effet de ce facteur consiste à calculer

sa covariance avec l'écart de mesure d'humidité. Une analyse ANOVA est également réalisée en catégorisant la variable en cinq groupes de 10 à 13, 13 à 16, 16 à 19, 19 à 22 et 22 et plus. Ces groupes sont déterminés à partir des spécifications de taux de gras des produits et en visant la séparation des données de l'histogramme de la figure 3.9 en groupes de plage identique, sauf pour le groupe de 22 et plus dans laquelle on retrouve moins d'échantillons.

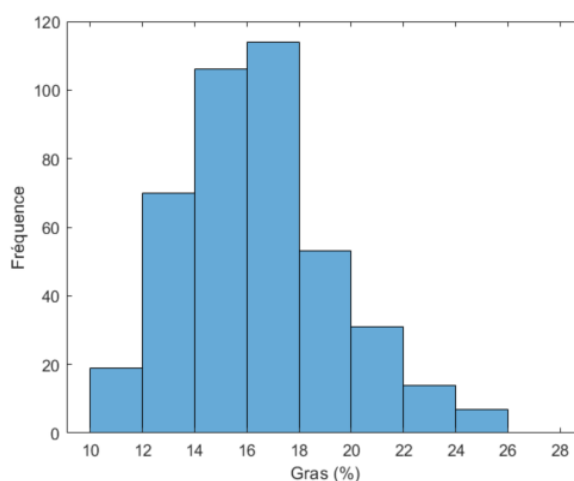


Figure 3.9 : Histogramme du taux de gras

3.2.3.7. Température du produit

Les températures du produit sont obtenues à partir de relevés manuels réalisés toutes les heures par le personnel d'assurance qualité à l'aide d'un thermomètre infrarouge sans contact. Ces données sont enregistrées dans le système de gestion de la qualité, puis extraites par requête sur la base de données avant d'être intégrées au jeu de données de calibration.

Étant donné que l'heure de mesure ne coïncide pas toujours avec celle de l'échantillonnage, une interpolation est effectuée afin d'associer une température estimée à chaque échantillon. Cette approche demeure fiable, dans la mesure où le produit est préalablement refroidi avant l'échantillonnage par le refroidisseur, rendant peu probable une variation significative de température entre deux mesures successives.

Cette mesure peut être affectée par la précision de l'instrument, les erreurs de collecte de données ou les erreurs d'entrée manuelle. Toutefois, cette donnée étant essentielle au contrôle

de la qualité du produit, la présence de valeurs aberrantes est peu probable. En effet, toute mesure atypique doit être justifiée et validée par le département de qualité avant la libération des lots.

La température du produit représente une variable continue qui peut faire l'objet d'une analyse de covariance. En plus de celle-ci, une analyse ANOVA est utilisée afin de fournir de plus amples informations. L'ANOVA est réalisée en utilisant six groupes de 10 à 17, 17 à 20, 20 à 23, 23 à 26, 26 à 29, 29 °C et plus. Les groupes sont déterminés en visant la répartition de trois groupes de chaque côté de la température d'opération normale de 23 °C. Chaque groupe couvre un intervalle de 3 °C, sauf les groupes localisés aux extrêmes. Ces groupes ont été ajustés de façon à obtenir un nombre de données statistiquement représentatif. L'histogramme de la figure 3.10 aide à créer ces groupes.

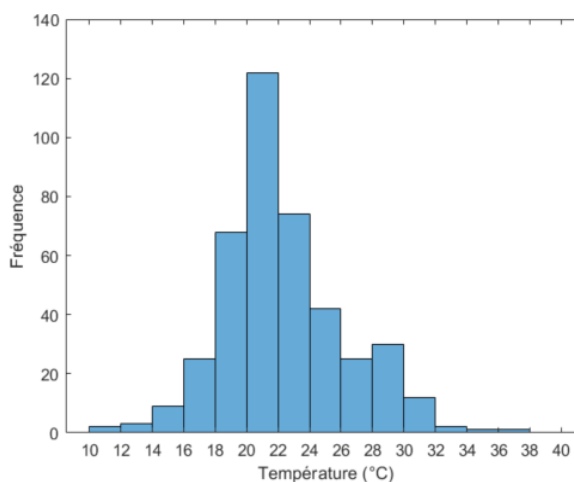


Figure 3.10 : Histogramme de la température du produit

3.2.3.8. Température ambiante

Tout comme les données de température du produit, les données de température ambiante de la section de l'usine où le capteur est installé sont obtenues à partir de relevés manuels effectués par les opérateurs d'assurance qualité chaque heure. Ces données font ainsi l'objet du même traitement et sont soumises aux mêmes hypothèses et sources d'imprécision que pour la température du produit.

La température ambiante représente également une variable continue pouvant faire l'objet d'une analyse de covariance combinée avec une ANOVA dont l'historgramme de la figure 3.11 aide à créer les groupes. Étant donné l'étendue plus petite de la plage de température, l'ANOVA est réalisée en utilisant cinq groupes de 18 à 20, 20 à 22, 22 à 24, 24 à 26, 26 à 28 et 28 °C et plus. Les groupes sont déterminés en considérant la plage de température d'opération normale de 22 à 24 °C, puis les autres intervalles sont établis de manière symétrique autour de celle-ci, avec des amplitudes équivalentes, sauf pour la plage 28 °C et plus dans laquelle on retrouve moins d'échantillons.

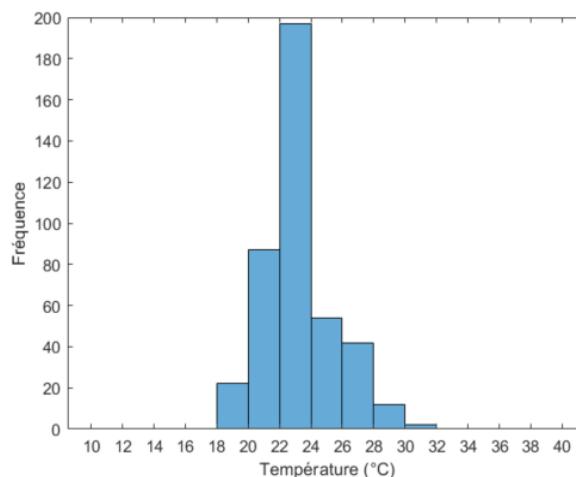


Figure 3.11 : Histogramme de la température ambiante

3.2.3.9. Dimensions du produit

Les dimensions des produits sont constituées de plusieurs variables quantitatives qui dépendent toutes du moule utilisé. Dans cette application précise, chacun des moules est unique et ceux-ci ne partagent aucune caractéristique dimensionnelle. Ainsi, le diamètre, la hauteur, le volume ainsi que le numéro du moule représentent des variables totalement colinéaires. Cependant, le volume représente la variable de choix à utiliser, car sa valeur est reliée à la dimension totale du produit.

Le calcul du volume du produit utilise le diamètre et la hauteur obtenus à partir du dessin du moule. Le volume est estimé à partir de la formule mathématique du volume d'un cône tronqué.

Pour les formes différentes, la même formule est utilisée afin de simplifier les calculs. Bien que la précision soit plus faible pour ces formes, le calcul permet tout de même d'obtenir une donnée spécifique et représentative de la variable à étudier.

Le volume est donc une variable quantitative dont l'effet peut être analysé en utilisant la matrice de covariance combinée à une ANOVA. La méthode ANOVA permettra d'analyser l'effet de cette variable sur l'écart de mesure en considérant chaque volume comme étant une catégorie spécifique. Cependant, comme le montre la figure 3.12, il existe un grand déséquilibre dans le nombre d'échantillons entre certains groupes, ce qui pourrait affecter la puissance statistique du test.

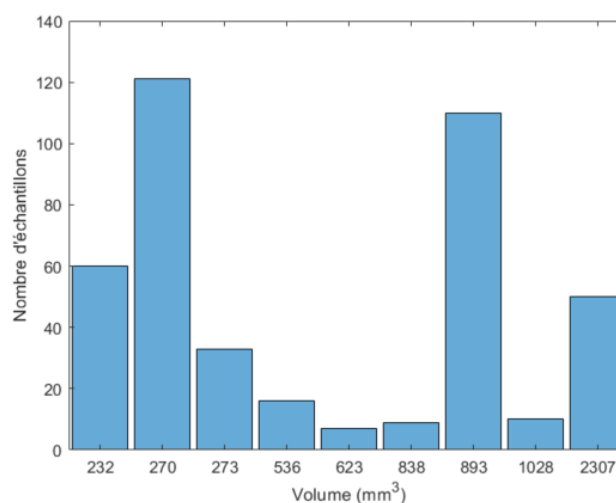


Figure 3.12 : Nombre d'échantillons par volume des produits

L'origine du déséquilibre s'explique par le fait que certains volumes sont associés à un produit ou à une famille particulière de produits. En effet, le volume 232 est strictement associé aux produits de la famille de nourriture pour chat de forme standard. Le volume 273 correspond à de la nourriture pour chien fabriquée dans un moule en forme de lotus. Le volume 536 contient exclusivement des produits de forme étoile alors que le volume 623 est formé d'un seul produit spécifique de forme standard. Le volume 838 est associé à quelques produits dont le moule est en forme de lotus. Finalement, le volume 1028 est constitué d'un seul produit en forme de cœur. Ainsi, afin d'obtenir une analyse de variance valide qui cible exclusivement l'effet du volume, il

est nécessaire de ne conserver que les groupes qui contiennent des produits fabriqués dans des dimensions différentes, soit les groupes 270, 893 et 2307.

3.2.3.10. *Forme du produit*

La forme du produit constitue une variable catégorielle qui dépend du moule utilisé. Une même forme peut être partagée par plus d'un moule, car les dimensions peuvent changer. Il existe quatre formes distinctes possibles, soit la forme standard représentant un cône tronqué, le lotus, l'étoile et le cœur. La figure 3.13 montre le nombre d'échantillons associés à chaque forme. La majorité des produits sont de forme standard alors que la forme lotus est moins fréquente. Les formes de cœur et d'étoile sont associées à des produits rarement fabriqués sur cette ligne de production. La méthode ANOVA permet d'analyser l'effet de cette variable sur l'écart de mesure, mais sera limitée par le déséquilibre du nombre d'échantillons entre les catégories.

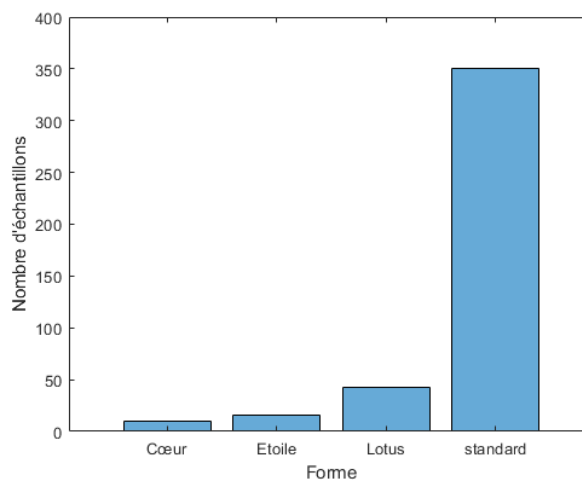


Figure 3.13 : Nombre d'échantillons par forme des produits

3.2.3.11. *Recette des produits*

Bien que certains produits se ressemblent ou soient constitués de la même recette, mais avec des dimensions différentes, les combinaisons d'ingrédients possibles peuvent avoir un impact sur les mesures d'humidité. La méthode suggérée pour analyser cette relation est de procéder à l'ANOVA sur le numéro de produit. Cependant, le graphique de la figure 3.14 montre un déséquilibre majeur dans la représentativité des produits, car plusieurs de ceux-ci ont moins de

dix échantillons. Ces produits risquent de ne pas être statistiquement représentatifs et affecter la puissance statistique de l'analyse. Cependant, en retirant tous ces produits de l'analyse, il ne reste que 14 produits sur les 67 du jeu de données original, ce qui risque de ne pas générer assez de variance.

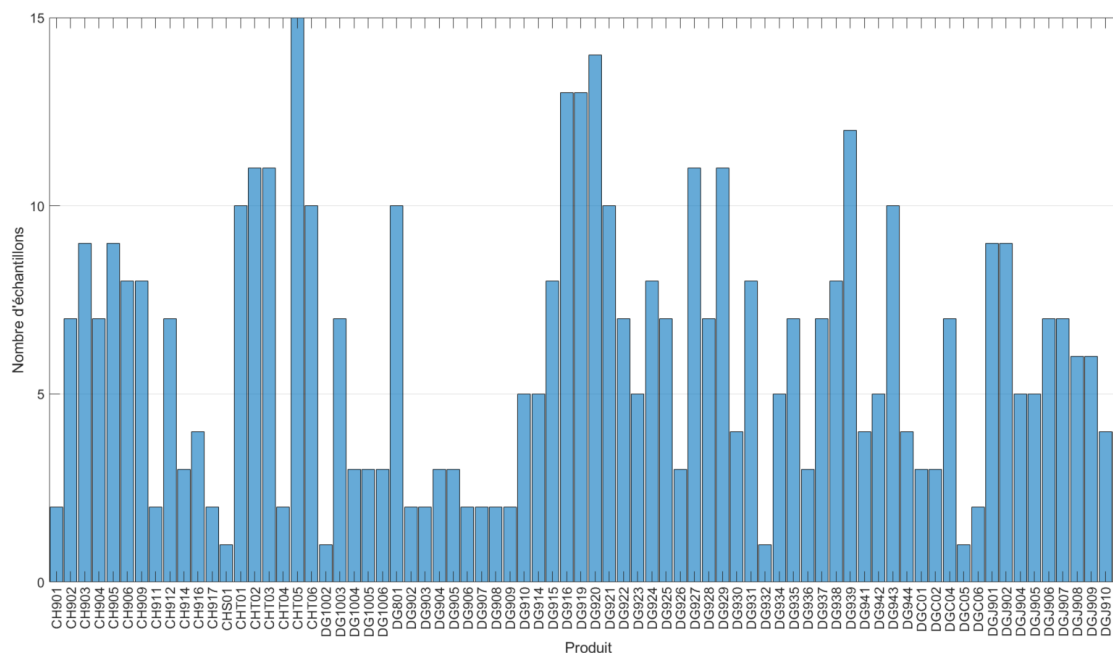


Figure 3.14 : Nombre d'échantillons par produits

Pour régler ce problème, il est possible de regrouper les produits similaires selon leur recette et ainsi créer des groupes contenant plus d'échantillons qui partagent des ingrédients similaires. En plus de réduire le nombre de catégories à analyser, cela permet d'établir si une calibration est plus performante sur un groupe de produit plutôt que sur un autre ou d'observer si certains ingrédients exercent une influence sur la mesure NIR.

Puisqu'il existe une incertitude sur le nombre idéal de groupes de produits à utiliser, ceux-ci seront partitionnés en 5, 10 et 15 groupes. Le partitionnement est effectué à partir de la liste d'ingrédients en utilisant les fonctions `linkage()` et `cluster()` de MATLAB. Ces fonctions génèrent un dendrogramme des produits regroupés selon les ingrédients comme montré aux figure 3.15 à Figure 3.17.

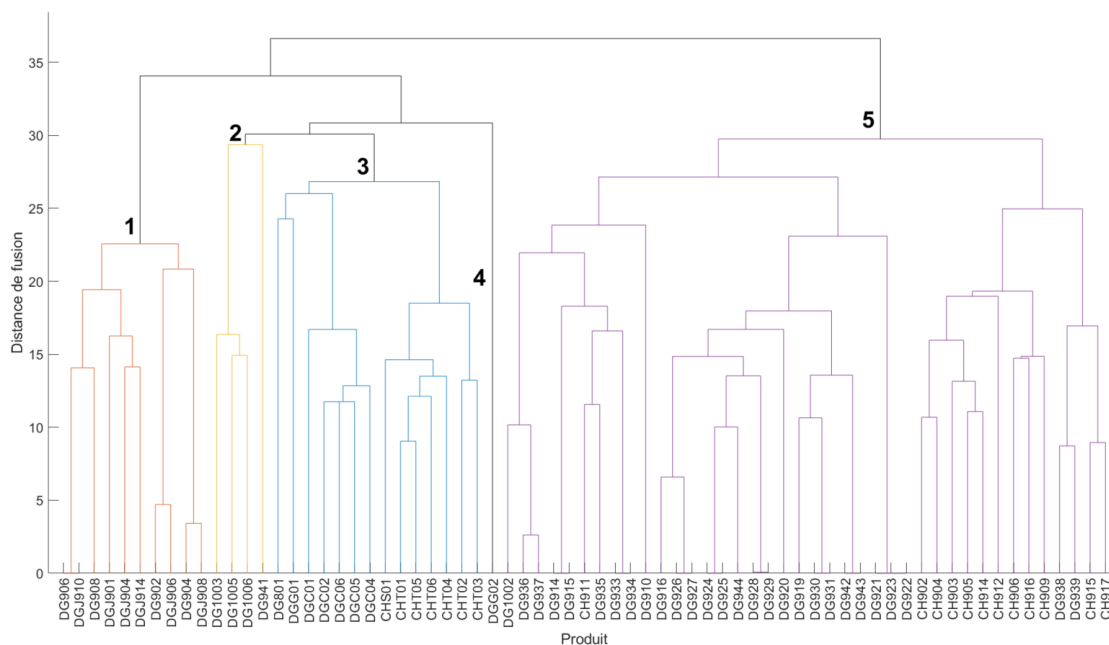


Figure 3.15 : Dendrogramme pour 5 groupes de produits

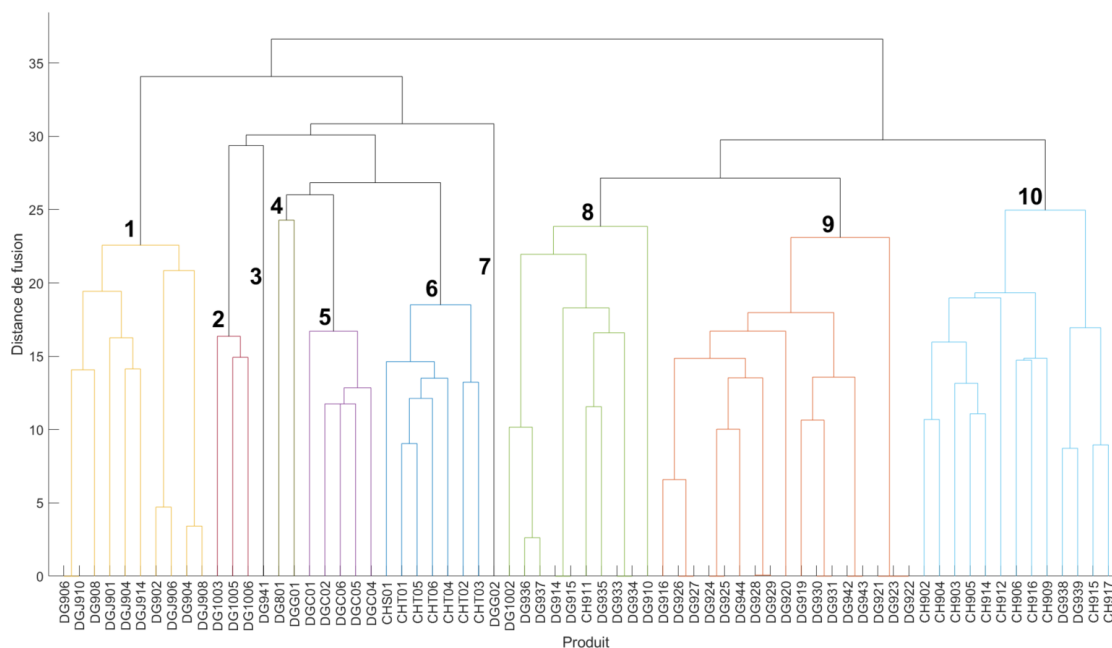


Figure 3.16 : Dendrogramme pour 10 groupes de produits

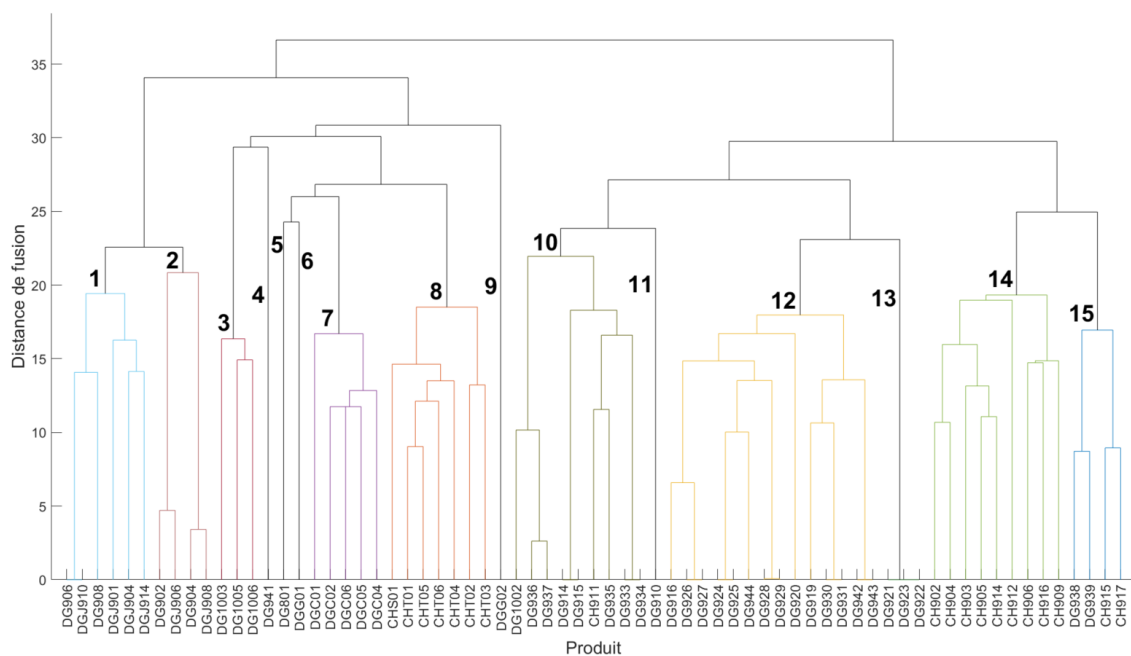


Figure 3.17 : Dendrogramme pour 15 groupes de produits

Les graphiques du nombre d'échantillons par groupe de produit de la figure 3.18 montrent que, bien que les produits soient regroupés, un déséquilibre majeur persiste dans la représentativité des catégories qui seront analysées par l'ANOVA et cela risque d'affecter la puissance statistique du test. De plus, l'augmentation du nombre de groupes dévoile la présence de certains produits isolés dont le nombre d'échantillons est inférieur à dix. Ces groupes non statistiquement valides devront donc être retirés des analyses.

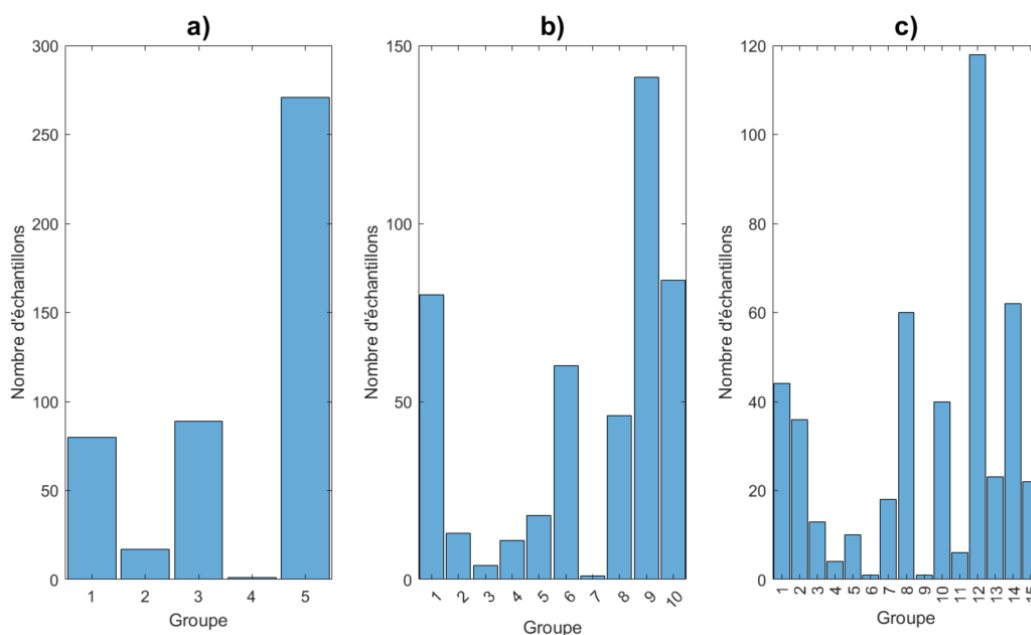


Figure 3.18 : Nombre d'échantillons par regroupement de produits
a) regroupement en 5 groupes, b) regroupement en 10 groupes, c) regroupement en 15 groupes.

3.2.3.12. Ajout de saveur externe

Pour certains produits, l'ajout de saveur externe vise notamment à améliorer l'appétence. Cette étape est effectuée entre les opérations de séchage et de refroidissement. Les ingrédients utilisés peuvent être une combinaison de poudre protéinée, de saveur liquide ou de gras. Dans tous les cas, la matière grasse est appliquée avant la poudre afin que celle-ci adhère bien au produit. Le fait d'ajouter de tels ingrédients peut affecter le taux de pénétration de la lumière infrarouge par absorption ou diffusion. Ainsi, le type d'ingrédient ajouté ou le simple fait d'appliquer des ingrédients additionnels en surface peuvent avoir un impact sur la lecture NIR.

L'effet des trois variables catégorielles d'ajout de saveur (poudre, liquide et gras) sur l'écart de lecture NIR sera étudié par la méthode ANOVA. La figure 3.19 montre le nombre d'échantillons par ingrédient de saveur ajoutée. Bien que chaque catégorie compte au moins dix échantillons, il existe un déséquilibre pouvant affecter les résultats de l'analyse.

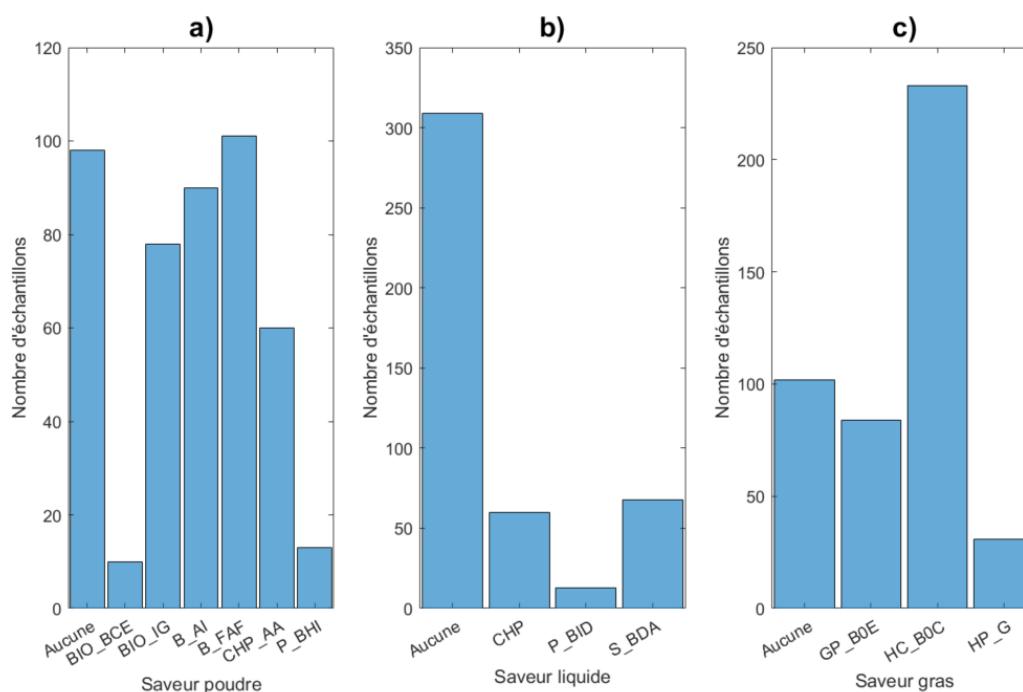


Figure 3.19 : Nombre d'échantillons par saveur externe
 a) distribution de la saveur en poudre, b) distribution de la saveur liquide, c) distribution de la saveur en gras.

3.2.3.13. Délai de mesure de référence

Le protocole d'échantillonnage prévoit une période de stockage de tous les échantillons après le prélèvement sur la ligne de production. Pendant cette période, un phénomène d'équilibrage se produit. En effet, un produit semi-fini qui sort de la ligne ne peut pas être considéré à l'équilibre, car il vient de subir diverses transformations chimiques. De plus, sa température interne n'est pas stabilisée, ce qui fait que chaque croquette peut être dans un état différent. La matière ainsi conservée dans un sac va tendre à s'équilibrer avec le temps et la durée requise sera différente d'un échantillon à l'autre. Ainsi, un échantillon mesuré en laboratoire après quelques heures pourrait ne pas fournir les mêmes résultats que lorsque le test est fait après 7 jours de stockage.

La méthode utilisée pour analyser l'effet de ce facteur consiste à calculer sa covariance avec l'écart de mesure d'humidité. De plus, afin d'obtenir plus d'information sur son effet potentiel, une analyse ANOVA sera réalisée en considérant chaque jour comme une catégorie. Selon la figure 3.20, les échantillons sont restés un minimum de 2 jours en stockage, mais certains ont

reposé plus de 2 semaines avant d'être mesurés. Ces données, relatives à des échantillons laissés trop longtemps en stockage selon l'entreprise représentent des valeurs aberrantes qui doivent être retirées de l'analyse.

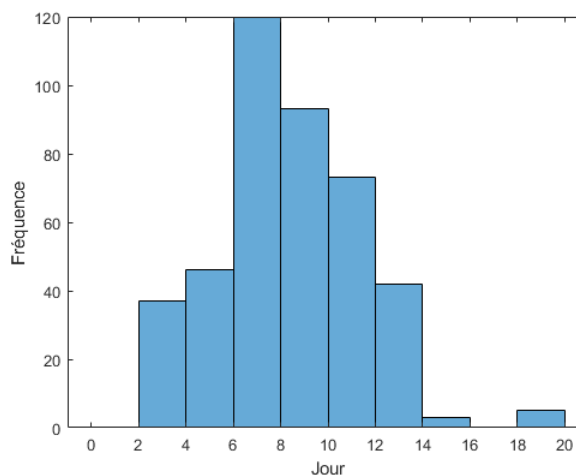


Figure 3.20 : Histogramme du délai de mesure

3.2.4. Essais expérimentaux

La méthodologie visant à quantifier l'effet des facteurs faisant l'objet d'essais expérimentaux est détaillée dans cette section. Contrairement aux variables continues et catégorielles du jeu de données, les facteurs ciblés par des essais font l'objet d'une méthodologie unique visant à obtenir des données quantifiables sur leur effet potentiel ainsi qu'à analyser leur influence sur la lecture d'humidité NIR.

3.2.4.1. Quantité d'échantillons

Lors des mesures de référence en laboratoire, la quantité de matière utilisée peut influencer les résultats si l'échantillon est considéré peu représentatif du lot mesuré par le capteur NIR. En ce qui concerne l'application ciblée dans l'étude, le laboratoire précise qu'une quantité de 5 g est prélevée de l'échantillon principal pour effectuer le test de séchage au four.

De son côté, lors de l'échantillonnage, le capteur NIR fournit une moyenne du taux d'humidité du produit mesuré sur une durée de 10 secondes. Ainsi, pendant la prise de mesure, l'appareil

effectue environ 450 lectures et produit une moyenne associée à la quantité de matière qui est passée sous celui-ci. Cette quantité de matière peut être considérable compte tenu de la grosseur et de la densité du produit. Elle est évaluée approximativement à 4.5 kg pour des croquettes de petite taille. L'écart entre une lecture moyenne d'humidité sur cette quantité de matière et une seule mesure sur un sous-échantillon de 5 g peut être significatif si le lot est caractérisé par une grande variabilité.

Dans le but d'évaluer l'erreur associée à la quantité de sous-échantillons utilisés comme référence, un essai expérimental est entrepris. Celui-ci vise à quantifier la variabilité des mesures d'humidité obtenues sur des sous-échantillons prélevés à l'intérieur d'un même sac d'échantillons. Trois produits distincts sont sélectionnés selon la procédure d'échantillonnage de l'entreprise, puis soumis à l'analyse par séchage au four conformément à la méthode de référence utilisée en laboratoire. Pour chacun de ces produits, trois sous-échantillons de 230 g et huit sous-échantillons de 5 g sont analysés et une simulation monte carlo avec $N = 10\,000$ est effectuée afin d'obtenir une distribution simulée la plus représentative possible de la réalité. L'analyse repose sur la comparaison des moyennes et des écarts-types obtenus pour chaque masse.

L'essai expérimental est également prolongé afin d'évaluer l'effet du délai de mesure sur les sous-échantillons de 5 g. Cette étape vise à quantifier plus précisément l'influence du temps de stockage sur la stabilité des mesures de référence. Pour chacun des trois produits, huit sous-échantillons de 5 g sont analysés par séchage au four à intervalles réguliers de 2 jours ou plus selon la disponibilité du laboratoire. La durée totale de cet essai est établie à un minimum de 20 jours afin de représenter la même étendue qu'observée dans le jeu de données de calibration. L'évolution de la moyenne et de l'écart-type des mesures en fonction du temps permettra d'estimer la variabilité potentielle induite par la durée de stockage avant la mesure de référence.

Finalement, une analyse ANOVA est effectuée afin de vérifier si la durée de stockage a un impact sur la mesure de référence. Pour cette analyse, on considère l'hypothèse nulle H_0 énonçant que le délai n'a pas d'effet significatif sur la mesure d'humidité. L'hypothèse alternative H_1 énonce

que le délai a un effet significatif sur la mesure d'humidité. Le seuil de signification α est fixé à 0.05. Un test de Levene est effectué avant l'ANOVA afin de vérifier l'homogénéité des variances entre les groupes. Pour ce test, on considère l'hypothèse nulle H_0 énonçant que les variances des groupes sont égales. L'hypothèse alternative H_1 énonce qu'au moins un groupe a une variance significativement différente. Le seuil de signification α est fixé à 0.05. Dans le cas où les variances ne sont pas égales, l'ANOVA de Welch sera utilisée comme technique d'analyse.

3.2.4.2. Emballage des échantillons

Tel que mentionné par Shenk et al. (2007), une attention particulière doit être accordée à l'humidité et à la température de l'échantillon pour éviter que celui-ci s'équilibre avec les conditions ambiantes. Cet effet peut générer une erreur considérable, car l'humidité mesurée par le capteur NIR et celle mesurée par la méthode de référence seront faites sur un échantillon dont la valeur du paramètre aura changé. Il est d'ailleurs impossible d'effectuer ces deux mesures simultanément dans les conditions actuelles. Ainsi, l'échantillon doit être conservé dans un sac scellé qui fournit une barrière efficace contre la diffusion de la vapeur d'eau.

Les enjeux communs d'une entreprise manufacturière imposent constamment de faire des choix qui représentent un compromis entre les performances et les coûts. Étant donné le haut volume de consommation de sacs scellés par l'entreprise, il est financièrement inconcevable d'utiliser des sacs de haute qualité à un coût élevé si ce choix n'est pas justifié par des arguments basés sur des données fiables et objectives. La simple revue des spécifications techniques du type de sac peut fournir une appréciation générale de l'efficacité, mais cette appréciation doit être vérifiée de façon quantitative par rapport à l'application réelle afin de fournir les données nécessaires dans ce contexte de prise de décision. L'objectif de cet essai vise donc à obtenir une mesure quantitative de la capacité du sac à agir comme barrière contre la diffusion de la vapeur d'eau et ainsi préserver l'état des échantillons entre la prise de mesure NIR et le test en laboratoire.

Pour atteindre cet objectif, deux essais expérimentaux doivent être réalisés. Le premier essai doit être effectué dans l'environnement réel de production afin de répéter les conditions dans lesquelles les échantillons sont conservés. Le deuxième essai, exécuté dans des conditions

extrêmes, permet d'évaluer le pire cas pouvant provoquer la plus grande erreur possible associée à ce facteur.

Dans les deux cas, un sac EPL-4575 de la compagnie LABPLAS (Sainte-Julie, Québec) constitué d'un mélange de polyéthylène est sélectionné pour conserver les échantillons. Ce modèle est utilisé par l'entreprise pour le transport et le stockage de tous les échantillons prélevés sur la ligne de production. Les spécifications générales du sac sont présentes à l'annexe IV. Le type de produit utilisé pour ces essais n'est pas considéré comme ayant un impact dans les résultats. Cependant, il est requis d'utiliser un produit fréquemment fabriqué afin que les résultats représentent les cas les plus rencontrés dans la réalité. De plus, le même échantillon sera utilisé pour les deux essais.

La méthode utilisée pour évaluer l'efficacité des sacs consiste à mesurer l'évolution de la masse de l'échantillon une fois ensaché. Cette méthode gravimétrique basée sur la norme ASTM D 5229/D 5229M – 92 (*American Society for Testing and Materials, 2004*) consiste à mesurer l'évolution de l'humidité d'un échantillon en se basant sur la masse de celui-ci, lorsqu'exposé à un environnement spécifique. Cette méthode requiert une balance de précision de 1 mg pour des échantillons de plus de 50 g.

Le modèle de balance PS 1000.R1 du fabricant RADWAG (Radom, Pologne) est configuré afin de fournir une mesure à toutes les 15 minutes sur son port de communication série. Afin de proposer une solution d'acquisition de données utilisant peu d'espace physique, un appareil IoT2050 de Siemens (Munich, Allemagne) est relié à la balance par un câble série et programmé pour que la lecture soit horodatée et enregistrée dans un fichier.

Afin de fournir plus d'informations durant l'essai, les conditions environnementales à l'intérieur et à l'extérieur du sac sont enregistrées toutes les 10 minutes à l'aide d'appareils RHT10. Le premier est inséré dans le sac avec l'échantillon de produit et le deuxième est déposé à proximité de la balance.

L'échantillon doit être laissé sur la balance pendant 8 jours, représentant le délai moyen observé dans le jeu de données de calibration. Les données sont analysées dans le logiciel MATLAB afin

de fournir la courbe de l'évolution de la masse, de la température et de l'humidité relative de l'échantillon en fonction du temps. La différence de masse calculée entre le début et la fin de l'essai représente la masse d'eau absorbée ou désorbée. Celle-ci peut être représentée en fonction de la masse initiale de l'échantillon pour fournir une mesure de pourcentage en fonction de la masse humide.

Pour le premier essai en conditions réelles, l'échantillon est installé dans le local dédié aux opérations du département de l'assurance qualité, lieux où les échantillons sont laissés avant leur acheminement au laboratoire externe. Le deuxième essai utilise le même échantillon que l'essai précédent. Cependant, il est requis d'assécher le produit par la méthode de séchage au four à une température de 104 °C pendant 5 heures. Par la suite, l'échantillon est ensaché et laissé au repos pendant 20 minutes pour le refroidir. Afin de garantir un environnement stable et contrôlé, une étuve TOV-5 de Cascade Tek (Plano, États-Unis) est utilisée. La température est fixée à 27 °C, légèrement au-dessus de la température ambiante de la salle afin de garantir un contrôle stable. Afin d'assurer un taux d'humidité relative ambiant maximal, un seau d'eau est laissé à l'intérieur de l'étuve.

Pour ces essais, il est entendu que l'augmentation ou la perte de masse est attribuable uniquement au gain ou à la perte d'eau. Les autres éléments volatils sont considérés comme négligeables dans l'analyse.

Il est également entendu que, puisque le processus de diffusion de la vapeur d'eau dépend à la fois du taux d'humidité du produit et de l'air ambiant, l'essai dans les conditions réelles ne représentera qu'un seul cas spécifique de la population. C'est pourquoi le deuxième essai, représentant le pire cas, est prévu.

3.2.4.3. Lumière environnante

Les différentes sources lumineuses disposées autour d'un capteur NIR peuvent affecter les mesures en influençant le signal d'arrière-plan dans les spectres. Selon le manuel d'installation et de maintenance de l'appareil, celui-ci ne doit pas être installé à proximité d'une source lumineuse intense dont l'intensité varie. Bien que les effets générés par ces sources puissent

être corrigés par les techniques de prétraitement, il est important de quantifier l'amplitude de ce signal et ainsi juger de son importance par rapport au signal de l'échantillon.

Le capteur NIR DA 7440™ n'est cependant pas en mesure d'effectuer ce type de mesure. Premièrement, celui-ci est fixé au-dessus du convoyeur et ne peut pas effectuer des mesures dans d'autres orientations. De plus, les spectres qu'il fournit sont déjà prétraités, ce qui empêche de mesurer un signal d'arrière-plan. Ainsi, les mesures expérimentales des spectres provenant de l'arrière-plan sont effectuées en utilisant le spectromètre de laboratoire Qneo AFBR-S20N1N256 du fabricant Broadcom (Palo Alto, États-Unis) ainsi que le logiciel propriétaire Waves. Cet appareil est muni de la même technologie de matrice de diode InGaAs composée de 256 détecteurs et couvre la plage de longueurs d'onde de 950 à 1700 nm. Le plan de mesure est résumé dans le tableau 3.3. Celui-ci permet de réaliser des mesures dans l'environnement réel tout en caractérisant le pire cas possible. Pour chaque cas, 10 spectres sont mesurés.

Tableau 3.3 : Plan des mesures d'arrière-plan

Localisation	Orientation	Détails	Justification
À proximité du capteur NIR.	Vers le convoyeur	Convoyeur à l'arrêt	Obtenir une lecture représentative de l'orientation du capteur dans l'application réelle.
À proximité du capteur NIR.	Vers le convoyeur	Convoyeur en marche	Mesurer l'effet du mouvement du convoyeur en situation réelle.
À proximité du capteur NIR.	Vers la source lumineuse la plus proche	Convoyeur à l'arrêt	Obtenir une lecture de l'amplitude de la source lumineuse à la distance réelle.
À 200 mm de la source lumineuse.	Vers la source lumineuse	Convoyeur à l'arrêt	Obtenir une lecture de l'amplitude dans le pire cas à proximité de la source lumineuse.

3.2.4.4. Spectre du convoyeur

Comme mentionné dans certaines études sur le développement de calibration NIR, le spectre du convoyeur peut générer des erreurs de lecture ou des variations si le flux de produit devient intermittent ou si la quantité de produit diminue et provoque des zones libres exposant le tapis du convoyeur. L'amplitude et la forme du spectre mesuré dépendent dans ce cas du matériau de la surface réfléchissante. Ainsi, compte tenu de l'application, l'effet du spectre du convoyeur peut représenter une source d'erreur importante si cette situation n'est pas gérée de façon adéquate.

Afin d'obtenir le spectre du tapis du convoyeur avec le capteur NIR DA 7440™, les mesures doivent être effectuées lors d'un arrêt de production. Deux mesures sont prévues soit une mesure lorsque le convoyeur est à l'arrêt et une mesure lorsque le convoyeur est en marche. Ceci permettra de voir s'il y a une différence notable entre les deux états. Avant de procéder à la prise de mesure, la configuration du capteur doit être modifiée pour désactiver la fonction de détection de produit. Ceci permet au capteur de mesurer un spectre sur un convoyeur vide, autrement la mesure n'est pas enregistrée. Les lectures sont comparées avec un spectre typique d'un produit afin d'évaluer la différence et d'établir une confiance satisfaisante en la fonction de détection de produit.

3.2.4.5. Température interne du spectromètre

La température ambiante autour du capteur influence directement ses composantes internes, notamment celles qui constituent le système optique. Celles-ci sont particulièrement impactées par des variations de température. C'est pourquoi le capteur NIR DA 7440™ est équipé d'un système à refroidissement thermoélectrique qui s'assure que la température des éléments optiques reste à l'intérieur d'une plage adéquate pour leur opération. Cependant, il est possible que certaines conditions extrêmes surviennent ayant comme conséquence d'atteindre les limites d'opération sans que cela soit connu. L'objectif de cet essai est donc de vérifier le bon fonctionnement du système de refroidissement en évaluant la température interne des composantes du capteur en fonction de la température ambiante.

L'essai doit couvrir une durée assez longue pour permettre de mesurer les conditions extrêmes de l'usine. Ainsi, un système d'acquisition de données basé sur la solution Ignition Historian de la compagnie Inductive Automation (Folsom, États-Unis) est utilisé et un script python est déployé afin de transférer les données de température interne du capteur NIR DA 7440™ dans une base de données de type séries chronologiques. Les données de température ambiante proviennent d'un capteur installé en production à proximité du capteur NIR. Les données disponibles couvrent une plage d'une année à partir du 7 mai 2024. Cette plage de données permet d'observer les effets des mois les plus chauds et les plus froids de l'année au Québec.

Dans cette analyse, la variable indépendante est représentée par la température ambiante de la salle et les quatre variables dépendantes sont les températures de la matrice de diodes, des composantes optiques, des composantes de contrôle et de la carte électronique. L'effet de la température ambiante sur la température interne des différentes composantes est étudié par une analyse des statistiques descriptives, ainsi que le calcul de la covariance en utilisant le logiciel MATLAB.

Premièrement, les données sont affichées dans un graphique temporel afin de déceler les points où la qualité des données est problématique. Ceci peut être causé par des moments durant lesquels les systèmes sont en panne ou que les capteurs fournissent de mauvaises données. Toutes les données aberrantes sont retirées du jeu de données. La relation entre les variables est ensuite vérifiée visuellement à l'aide d'histogrammes. Finalement, une matrice de covariance est calculée afin de quantifier l'effet de la variable indépendante sur les quatre variables dépendantes.

CHAPITRE 4 - RÉSULTATS

Ce chapitre présente les résultats obtenus à la suite des différentes analyses statistiques et essais expérimentaux réalisés sur chacun des facteurs potentiels identifiés dans le cadre de l'étude. Conformément à la structure définie dans la section de méthodologie, les résultats sont d'abord présentés pour les facteurs inclus dans le jeu de données de calibration. L'analyse débute par l'évaluation de l'importance relative des facteurs à l'aide de la méthode de forêt aléatoire, suivie de l'examen de la matrice de covariance pour les variables continues. La normalité de la variable dépendante est ensuite vérifiée, avant de présenter les résultats d'analyse pour chacun des facteurs du jeu de données. Enfin, les résultats des essais expérimentaux menés sur certains facteurs spécifiques sont exposés et un tableau récapitulatif synthétise l'ensemble des analyses effectuées.

4.1. Forêt aléatoire

L'analyse de l'ensemble du jeu de données de calibration débute par l'application de la méthode de forêt aléatoire. Cette analyse se concentre sur la variable dépendante correspondant à l'écart prédit hors ligne à l'aide du dernier modèle de calibration valide, soit la variable nommée *Ecart_Pred_H*. Ceci est requis pour retirer l'effet du modèle de calibration et exposer la variance des autres facteurs. L'effet du modèle de calibration fait l'objet d'une analyse spécifique dans la section plus bas réservée à ce facteur. La figure 4.1 présente les résultats de l'importance normalisée des facteurs pour une forêt aléatoire de 300 arbres. Selon ces résultats, le volume des croquettes et le taux de protéine représentent les variables ayant une importance élevée. Le délai de mesure en laboratoire ainsi que la température du produit ont une importance moyenne alors que la température ambiante, l'ajout de saveur en poudre ainsi que le taux de gras, exercent une influence faible, mais non négligeable. Tous les facteurs ont une importance inférieure à 0.3 et sont considérés comme exerçant une influence non nulle, mais négligeable.

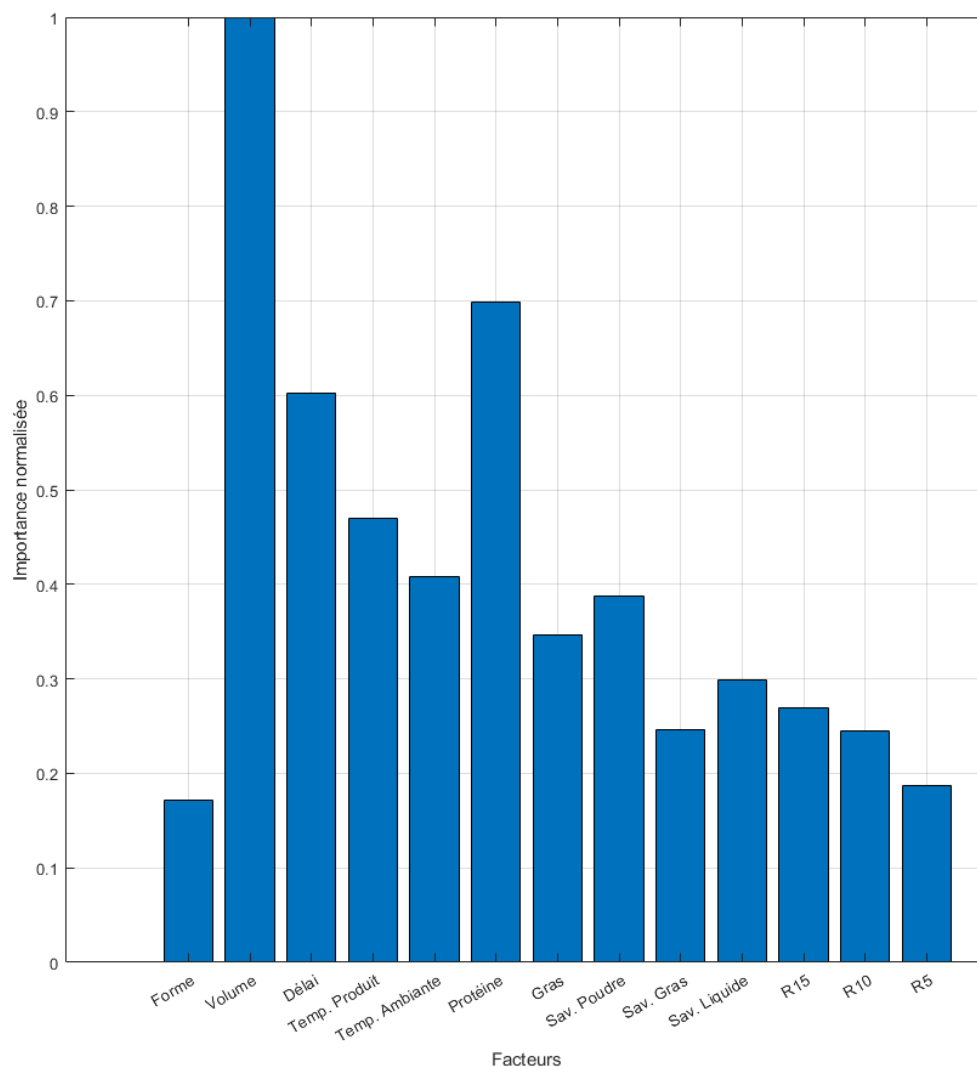


Figure 4.1 : Importance des facteurs selon une forêt aléatoire

4.2. Matrice de covariance

Les facteurs représentant des variables continues ont été analysés en calculant la matrice de covariance afin de quantifier leur lien causal avec l'écart de mesure prédit par mode hors ligne ($Ecart_Pred_H$). Toutes ces variables ont été normalisées en utilisant la technique de cote z.

Selon la figure 4.2, les deux variables continues ayant le lien causal le plus élevé dans la matrice sont le taux de protéine ainsi que le volume du produit. La température du produit ainsi que le

taux de gras ont une covariance un peu plus faible, mais significative alors que la température ambiante et le délai de mesure ont des covariances négligeables. Tous les facteurs ont des covariances négatives, ce qui implique que leur augmentation tend à causer une mesure NIR plus sèche que la réalité.

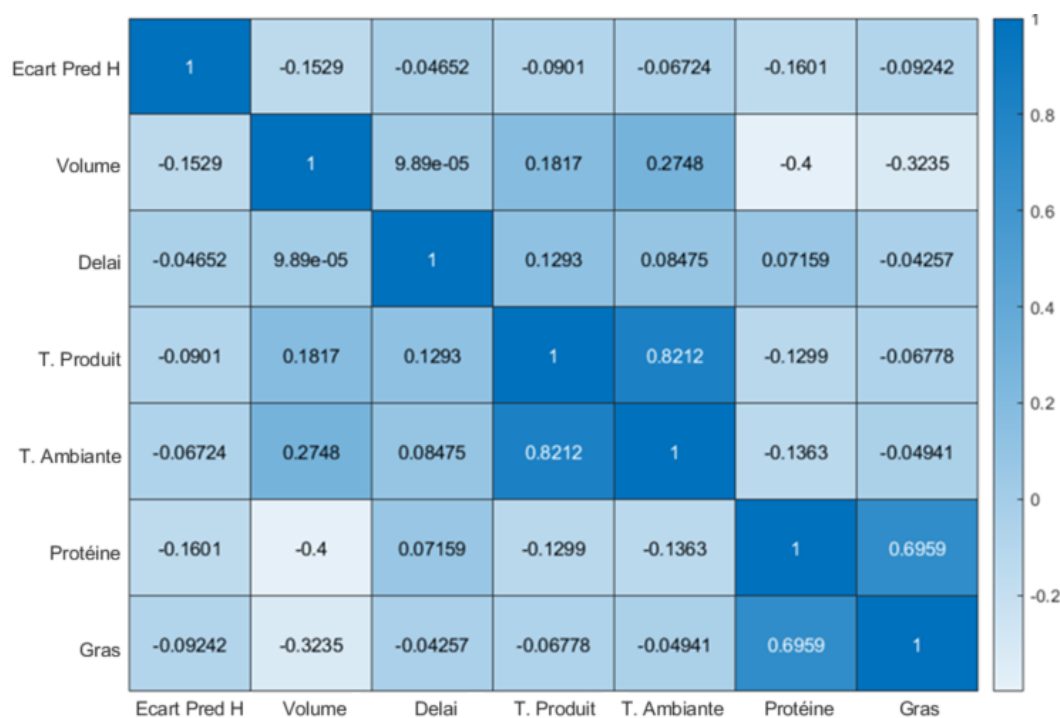


Figure 4.2 : Matrice de covariance des variables continues

4.3. Test de normalité de la variable dépendante

Avant de procéder à l'analyse des facteurs par la méthode ANOVA, le test d'Anderson-Darling a été effectué afin de vérifier si les deux variables dépendantes suivent une loi normale. Tel que présenté dans le graphique de gauche de la figure 4.3, la valeur p du test d'Anderson-Darling est de 0.007, ce qui confirme que les données de la variable *Ecart_H* ne suivent pas une distribution normale. Cependant, le résultat du test semble affecté par la présence de trois valeurs aberrantes principalement présentes à des écarts négatifs de -5 % et moins. En retirant ces valeurs du jeu de données, la valeur p devient 0.544 et la distribution des écarts peut être considérée comme normale comme le montre le graphique de droite. Ces valeurs aberrantes,

produites durant les premiers mois suivant l'installation du capteur, ne contiennent pas d'information pertinente pour l'étude, car la raison de l'écart n'a pas été documentée. Leur origine pourrait être reliée à la mesure de produits dont les caractéristiques ne sont pas couvertes par le premier modèle général de calibration appliqué ou par le manque d'expérience du personnel face à une nouvelle procédure d'échantillonnage.

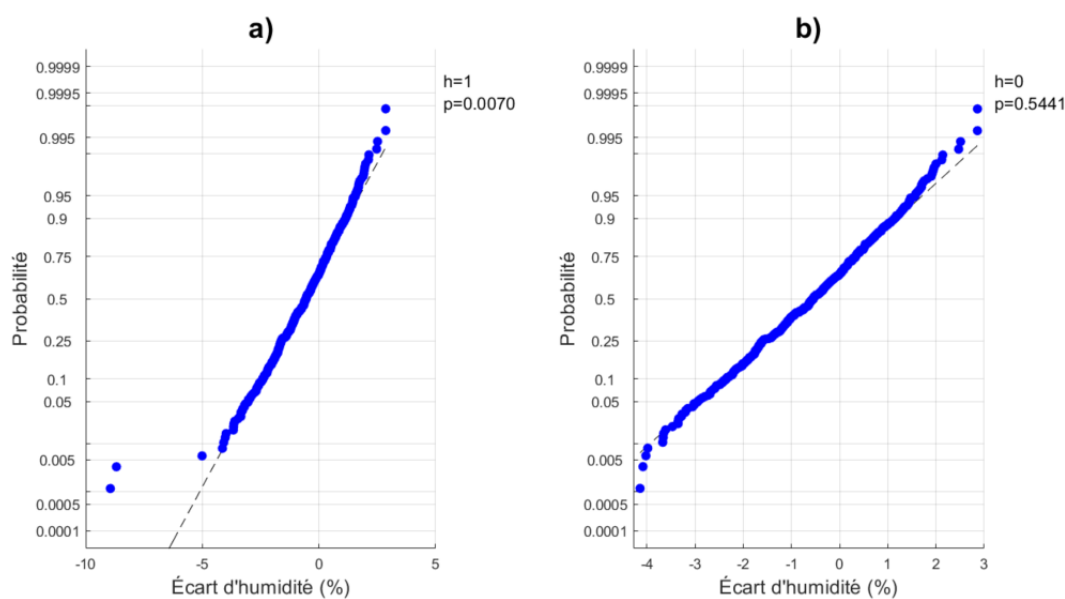


Figure 4.3 : Graphique de probabilité normale — écart d'humidité
a) jeu de données complet, b) jeu de données sans valeurs aberrantes.

La figure 4.4 présente les résultats d'analyse de la normalité des données pour la variable *Ecart_Pred_H*. Le graphique de gauche montre une valeur p de 0.0005, ce qui confirme que le test d'Anderson-Darling échoue et que les données ne suivent pas une loi normale. Une fois de plus, la présence de données aberrantes semble affecter le résultat. Lorsque ces données retirées, le test échoue toujours, car la valeur p est de 0.0006. Ce résultat est expliqué par le comportement des données dont l'écart est supérieur à 1 % et inférieur à -1 %. Ceci pourrait s'expliquer par la présence de certains phénomènes modifiant la distribution des données ou la présence de certaines familles de produits dont les propriétés sont différentes. Puisque ces données sont utiles à l'analyse, elles seront conservées et les analyses effectuées par ANOVA devront tenir compte de cette particularité.

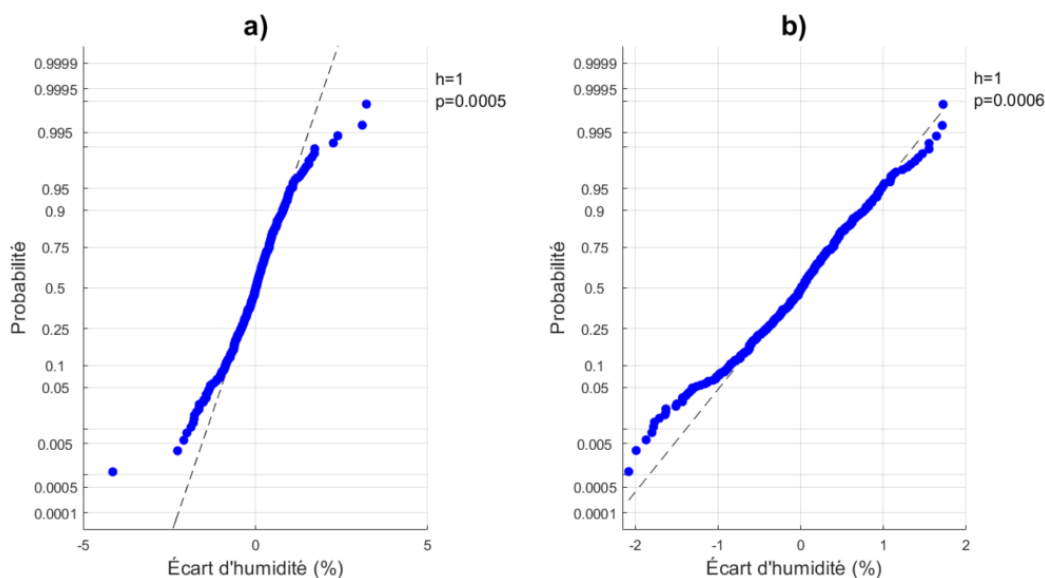


Figure 4.4 : Graphique de probabilité normale — écart d'humidité prédit
a) jeu de données complet, b) jeu de données sans valeurs aberrantes.

4.4. Analyse des facteurs du jeu de données

Les effets de tous les facteurs présents dans le jeu de données de calibration ont été analysés selon leur méthodologie respective présentée au chapitre précédent. Afin d'assurer la clarté et la concision des sections suivantes, seuls les résultats finaux des tests de Levene sont présentés dans le corps du texte. Les tableaux détaillés résultants de ces tests sont disponibles en annexe I. De la même manière, les valeurs p issues des analyses ANOVA sont rapportées dans le texte et les tableaux complets présentant les résultats détaillés sont fournis en annexe II.

Les résultats des analyses ANOVA sont présentées sous forme de deux graphiques complémentaires. L'un affiche les résultats sous forme de boîtes à moustaches et permet de comparer les distributions des différents groupes en fournissant une représentation visuelle des statistiques de base. L'autre graphique affiche les réponses moyennes de chaque groupe ainsi que leur intervalle de confiance, ce qui permet de constater si des groupes ont des moyennes significativement différentes.

4.4.1. Modèle de calibration

L'analyse de variance à un facteur (ANOVA) a été effectuée sur la variable représentant le modèle de calibration dans le but d'évaluer si celle-ci a un effet significatif sur l'écart de mesure de l'humidité, soit la variable *Ecart_H*.

Alors que les données suivent une loi normale, le test de Levene produit une valeur p de 0.008 et conclut à une inégalité des variances entre les groupes. Dans ce cas, une ANOVA de Welch est plus appropriée comme méthode d'analyse. Lorsqu'appliquée sur les données, cette analyse produit une valeur p de 0.010, indiquant que le modèle de calibration a un effet significatif sur l'écart de mesure d'humidité. Le graphique des réponses moyennes de la figure 4.5 permet d'observer les différences entre les modèles. La majorité des modèles produisent des écarts moyens négatifs alors que le modèle 3 produit un écart moyen positif. Puisque le modèle de calibration a une influence significative sur l'écart de mesure, la variable dépendante *Ecart_Pred_H* sera utilisée pour analyser l'effet de tous les autres facteurs.

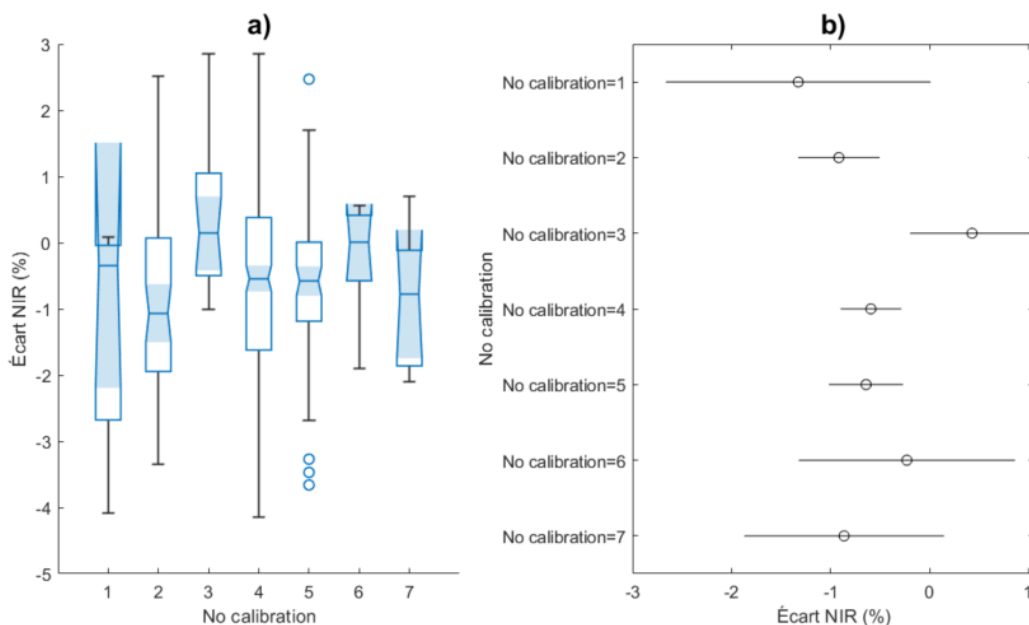


Figure 4.5 : Résultats graphiques ANOVA — modèle de calibration
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

4.4.2. Taux de protéine

Les analyses globales, réalisées à l'aide d'une forêt aléatoire et d'une analyse de la covariance, ont mis en évidence un impact important du taux de protéine sur l'écart de mesure de l'humidité par NIR. Ce facteur exerce un effet significatif, se traduisant par une sous-estimation du taux d'humidité du produit. Afin de mieux caractériser cette relation, une ANOVA a été menée en utilisant les groupes définis dans la section méthodologie.

Le test de Levene produit une valeur p de 0.170, ce qui indique que les variances peuvent être considérées comme homogènes entre tous les groupes. L'ANOVA, dont les résultats sont présentés à la figure 4.6, a produit une valeur p de 0.009, indiquant que le taux de protéine a un effet significatif sur l'écart de mesure. Les graphiques montrent que pour un faible taux de protéine, l'écart est positif alors qu'il tend à devenir significativement négatif à partir de 40 %.

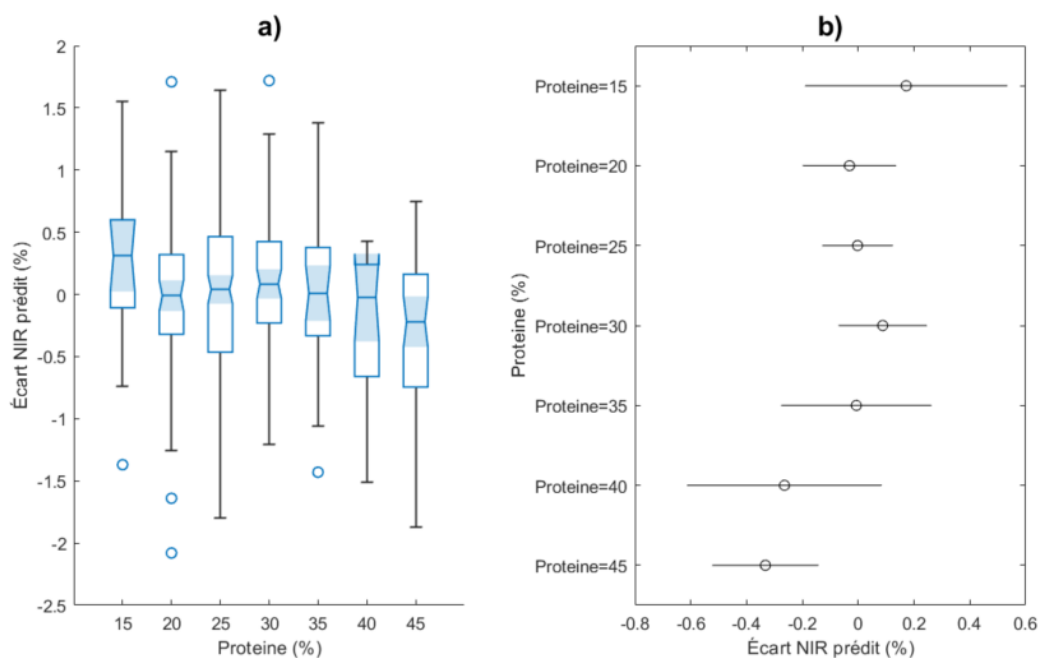


Figure 4.6 : Résultats graphiques ANOVA — taux de protéine
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

4.4.3. Taux de gras

D'après le classement de l'importance des variables fourni par la forêt aléatoire, ainsi que la valeur de covariance plus faible observée dans la matrice de covariance, le taux de gras semble exercer un effet modéré sur l'écart de mesure NIR. Pour valider cette observation, une ANOVA a été réalisée en utilisant les groupes définis dans la section méthodologie.

Selon les résultats du test de Levene, l'hypothèse d'homogénéité des variances entre les groupes est respectée, car la valeur p obtenue est de 0.100. L'ANOVA a produit une valeur p de 0.009, suggérant que le taux de gras exerce un effet significatif sur l'écart de mesure. Le graphique des réponses moyennes de la figure 4.7 illustre que le groupe 19 tend à générer un écart significativement négatif.

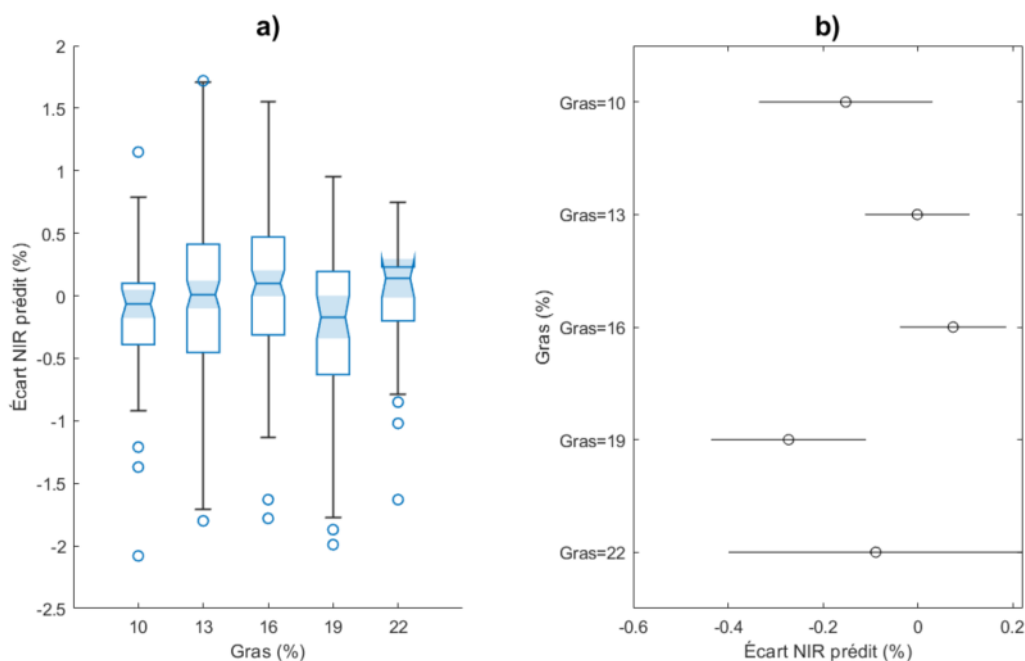


Figure 4.7 : Résultats graphiques ANOVA — taux de gras
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

En étudiant le groupe 19, il est observé que la majorité des produits qui le compose est constituée de produits de nourriture pour chat ayant un taux de protéine de 40 % et plus. La distribution de ces échantillons à l'intérieur des groupes de taux de gras est présentée à la

figure 4.8. Les résultats sont donc influencés par cette variable confondante. En retirant les 60 échantillons ciblés, le test de Levene conclut toujours à une égalité des variances entre les groupes en produisant une valeur p de 0.083. Cependant, le groupe représentant un taux de gras de 22 doit être retiré de l'analyse, car il ne contient plus que trois échantillons. L'ANOVA, dont les résultats sont présentés à la figure 4.9, produit alors une valeur p de 0.202 prouvant qu'en l'absence de l'effet confondant du taux de protéine, le taux de gras n'exerce pas d'effet significatif sur l'écart de mesure NIR.

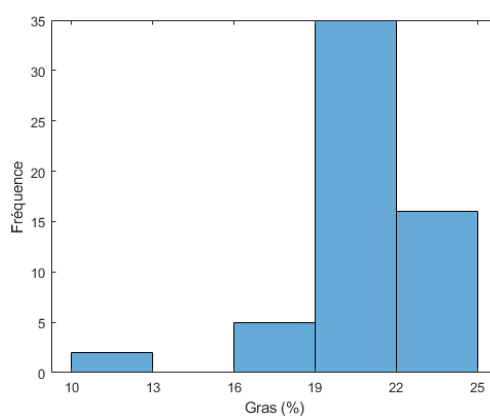


Figure 4.8 : Distribution des échantillons à haut taux de protéine selon le taux de gras

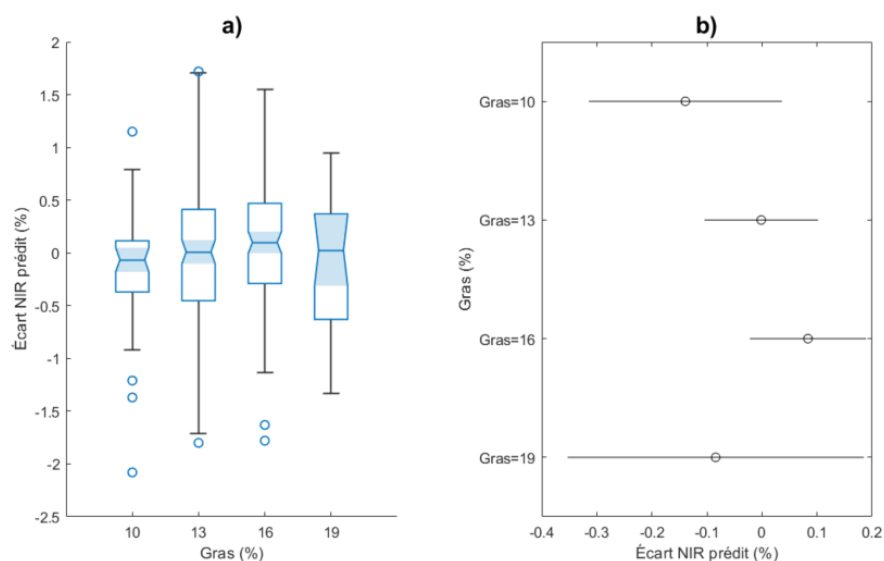


Figure 4.9 : Résultats graphiques ANOVA — taux de gras sans variable confondante
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

4.4.4. Température du produit

D'après le classement de l'importance des variables fourni par la forêt aléatoire, ainsi que la covariance observée dans la matrice de covariance, la température du produit aurait un effet modéré sur l'écart de mesure NIR. Pour approfondir cette relation, une ANOVA a été effectuée en considérant les groupes définis dans la méthodologie.

Selon les résultats du test de Levene, les variances peuvent être considérées comme homogènes entre les six groupes puisque la valeur p est de 0.695. De son côté, l'ANOVA a généré une valeur p de 0.080, suggérant que la température du produit n'a pas d'effet statistiquement significatif sur l'écart de mesure NIR. Toutefois, les graphiques de la figure 4.10 montrent que le groupe constitué des échantillons dont la température est de 29 °C et plus provoquerait un écart généralement négatif.

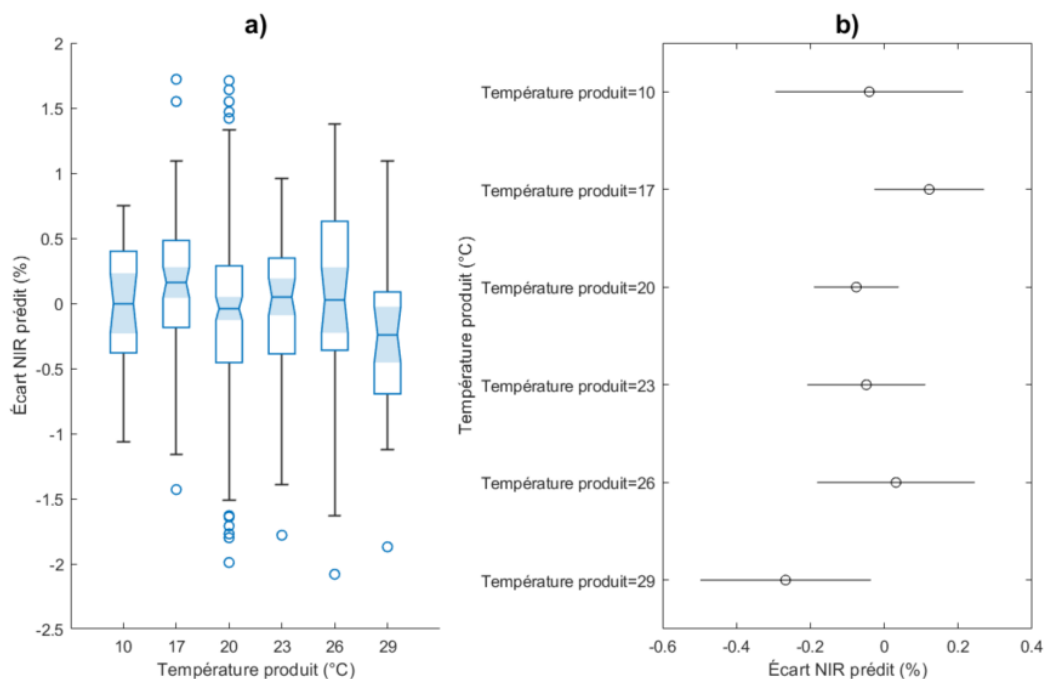


Figure 4.10 : Résultats graphiques ANOVA — température du produit
 a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

La figure 4.11 montre que l'effet produit par le groupe de température élevé ne peut pas être attribuable à l'effet de la variable confondante du taux de protéine, car ces échantillons sont

distribués dans tous les groupes. Afin de confirmer le tout, le groupe à haute teneur en protéine a été retiré du jeu de données et une analyse ANOVA supplémentaire a été effectuée. Selon les résultats du test de Levene, les variances peuvent être considérées comme homogènes entre les cinq groupes, car la valeur p est de 0.610. L'ANOVA a généré une valeur p de 0.184, démontrant que la température du produit n'a pas d'effet statistiquement significatif sur l'écart de mesure NIR lorsque les produits à haute teneur en protéine sont retirés de l'analyse. Cependant, le diagramme de boîtes de la figure 4.12 présente le même phénomène observé précédemment. Bien que les réponses moyennes soient statistiquement similaires, le groupe de température plus haute produit nettement une moyenne généralement négative. Ce phénomène est également présenté dans le graphique de la figure 4.13 dans lequel l'effet des températures situées aux extrêmes sur l'écart de mesure NIR de l'humidité est mis en évidence. Basé sur ces résultats, la température du produit aurait un effet sur la mesure d'humidité NIR, mais seulement pour les valeurs extrêmes.

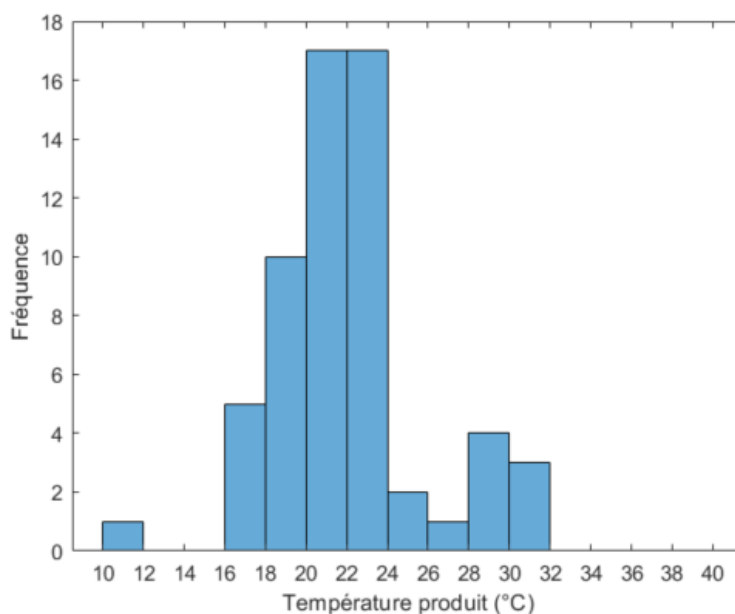


Figure 4.11 : Distribution des échantillons à haut taux de protéine selon la température du produit

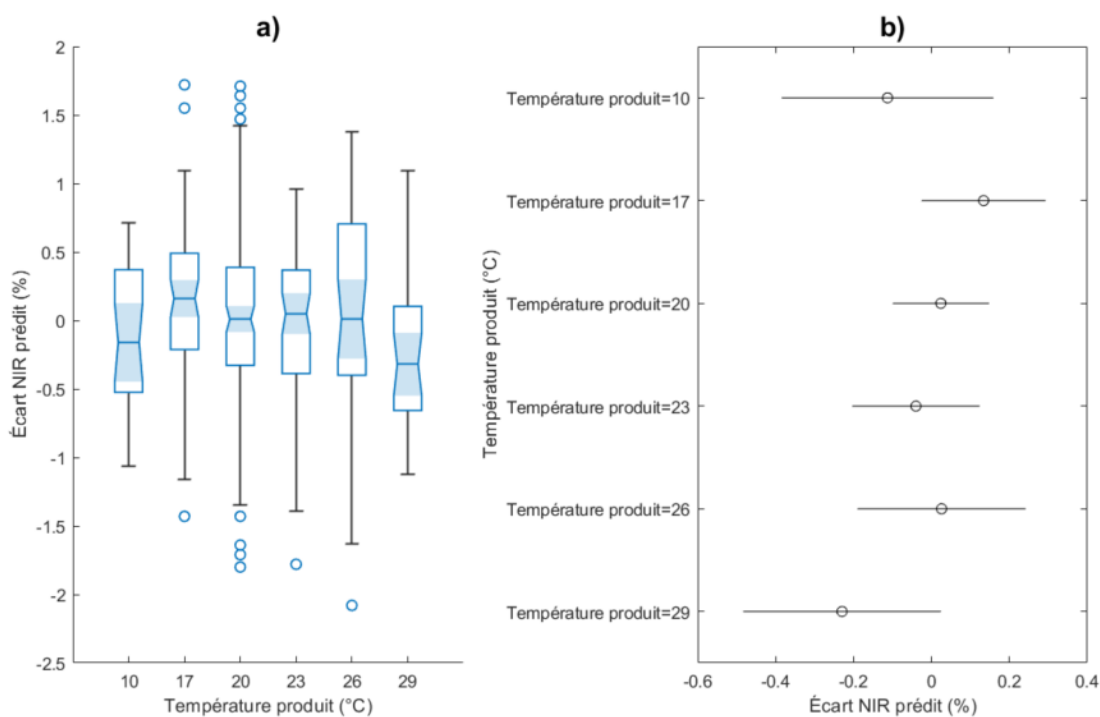


Figure 4.12 : Résultats graphiques ANOVA — température du produit sans variable confondante
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

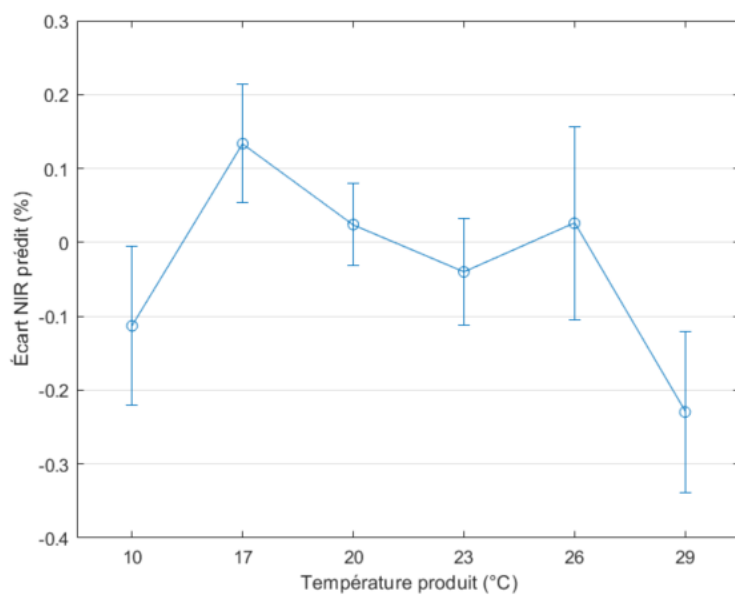


Figure 4.13 : Tendence moyenne de l'écart par température du produit

4.4.5. Température ambiante

Les résultats d'analyse par forêt aléatoire et par covariance obtenus pour la température ambiante indiquent que celle-ci a un effet limité, mais non négligeable sur l'écart de mesure NIR. Afin de mieux caractériser cet effet, une ANOVA a été effectuée en considérant les groupes présentés dans la méthodologie.

Le test de Levene produit une valeur p de 0.060, ce qui indique que les variances peuvent être considérées comme homogènes entre les six groupes. L'ANOVA a produit une valeur p de 0.049 signifiant que la température ambiante a un effet significatif sur l'écart de mesure NIR. En analysant les graphiques de la figure 4.14, il est observé que le groupe associé à la température ambiante la plus élevée semble responsable de mesures généralement négatives. Ce phénomène est également présenté dans le graphique des moyennes et des erreurs de la figure 4.15 dans lequel la température semble avoir un effet notable sur l'écart de mesure NIR de l'humidité qu'à partir de 28 °C.

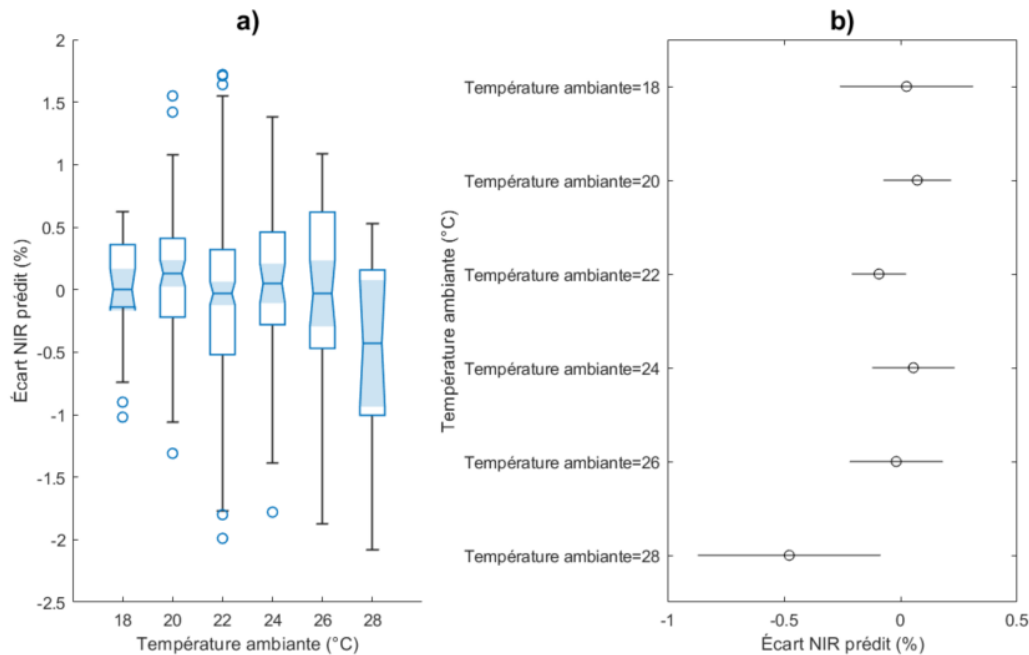


Figure 4.14 : Résultats graphiques ANOVA — température ambiante
 a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

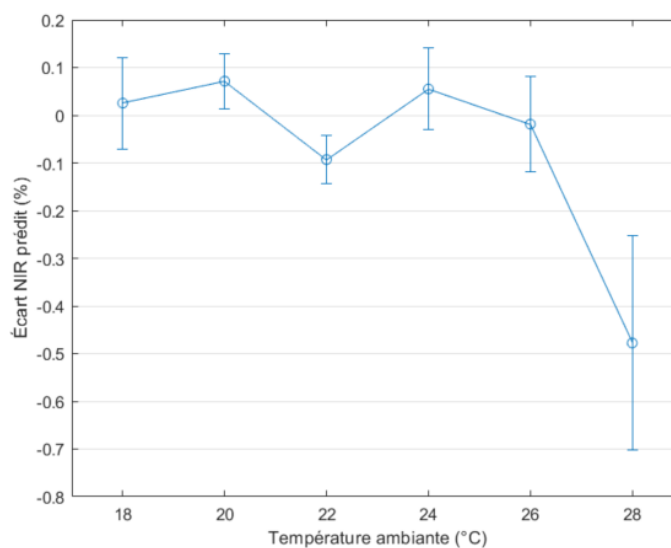


Figure 4.15 : Tendence moyenne de l'écart par température ambiante

Comme pour la température du produit, l'effet observé pour les températures plus élevées ne peut pas être attribué à la variable confondante du taux de protéine. En effet, comme le montre la figure 4.16, les échantillons à haut taux de protéine sont répartis sur toute la plage de température ambiante et peu d'entre eux occupent la plage de 28 °C et plus.

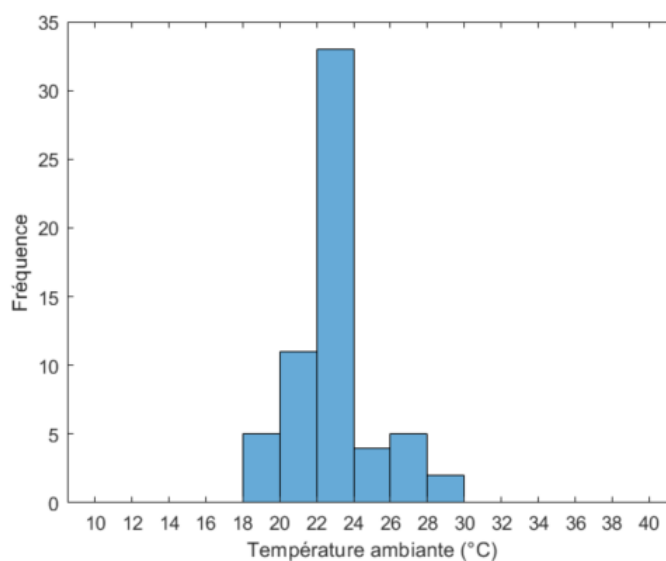


Figure 4.16 : Distribution des échantillons à haut taux de protéine selon la température ambiante

4.4.6. Dimensions du produit

Selon les résultats des analyses par forêt aléatoire et par covariance, le volume, principal indicateur des dimensions du produit, s'avère être le facteur le plus influent sur l'écart de mesure NIR. Une ANOVA a été réalisée afin d'approfondir la compréhension de l'effet associé aux groupes définis dans la section de méthodologie.

Selon les résultats du test de Levene, l'hypothèse d'homogénéité des variances entre les groupes n'est pas respectée, car la valeur p est de 0.031. L'ANOVA de Welch, plus adaptée dans ce contexte, a produit une valeur p de 0, indiquant un effet significatif de ce facteur sur l'écart de mesure. Les graphiques de la figure 4.17 révèlent une tendance selon laquelle l'écart devient de plus en plus négatif à mesure que le volume des croquettes augmente, ce qui suggère que la mesure d'humidité par NIR tend à sous-estimer la teneur réelle en eau des produits de plus grande taille.

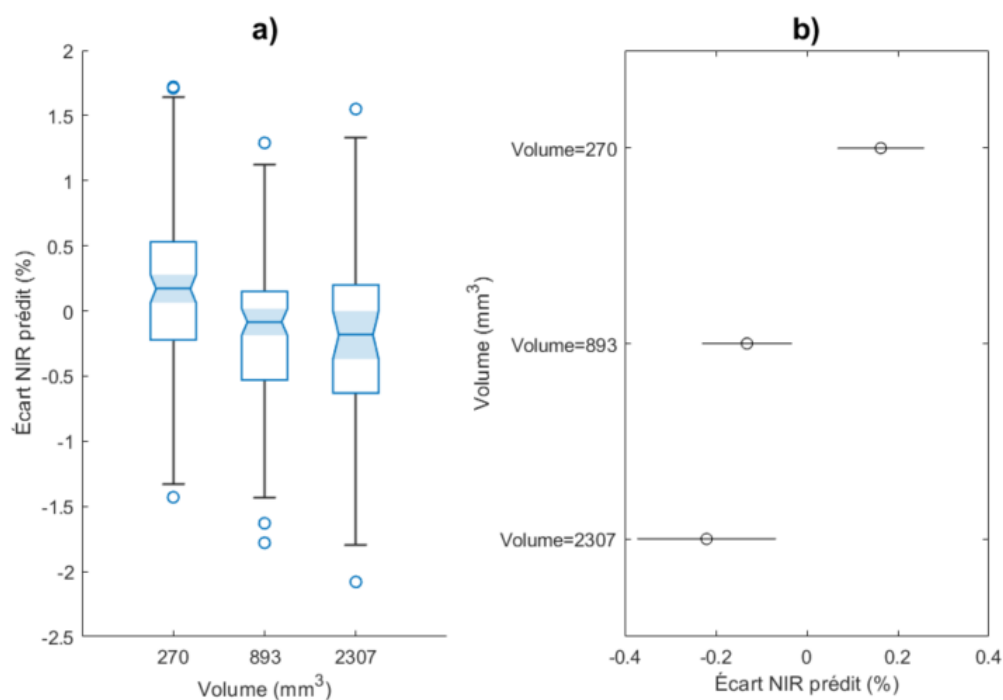


Figure 4.17 : Résultats graphiques ANOVA — volume du produit
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

4.4.7. Forme du produit

Selon le classement de l'importance des variables fourni par la forêt aléatoire, la forme du produit n'apparaît pas comme un facteur influençant l'écart de mesure NIR. Pour valider cette observation, une ANOVA a été réalisée en considérant les quatre formes possibles en tant que groupes distincts.

Selon les résultats du test de Levene, l'hypothèse d'homogénéité des variances entre les groupes est respectée, car la valeur p résultante est de 0.912. L'ANOVA a produit une valeur p de 0.068, indiquant que la forme du produit n'a pas d'effet significatif sur l'écart de mesure. Cette conclusion est appuyée par les graphiques de la figure 4.18, qui montrent une réponse moyenne similaire entre les groupes, sans qu'aucun ne se distingue nettement des autres. Il est important de noter que dans cette analyse, les produits à haute teneur en protéine sont présents dans le groupe de forme standard. Ceci expliquerait pourquoi la valeur moyenne associée à ce groupe est un peu plus faible.

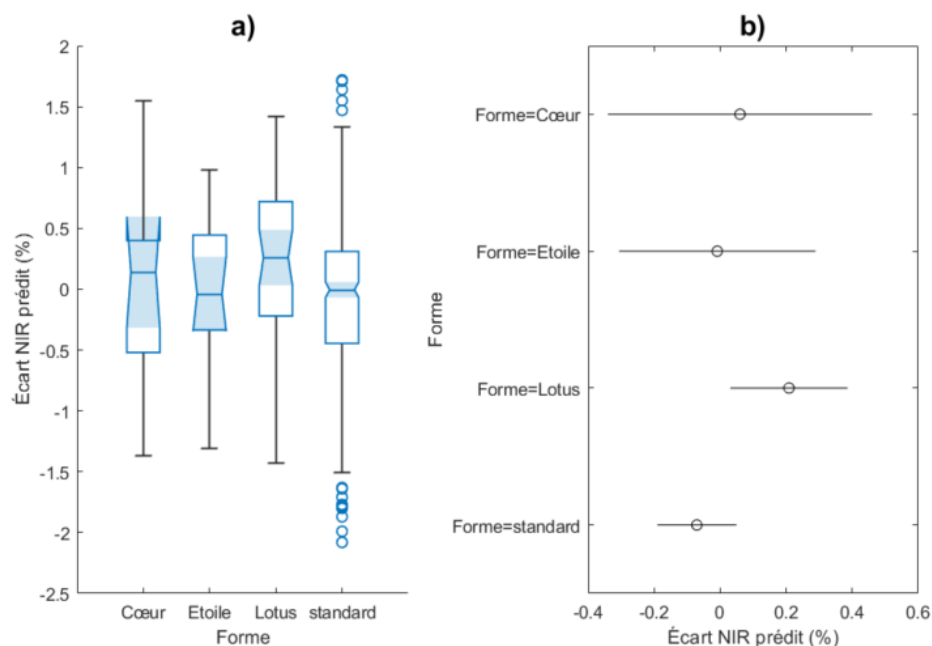


Figure 4.18 : Résultats graphiques ANOVA — forme du produit
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

4.4.8. Recette des produits

Selon les analyses effectuées par forêt aléatoire, les recettes n'exerceraient pas d'influence sur l'écart de mesure NIR. Toutefois, une ANOVA a été menée afin d'examiner plus en détail l'effet potentiel des groupes pour ces trois variables indépendantes.

Avant de procéder à l'analyse, les familles de produit contenant moins de dix échantillons ont été exclus pour ne conserver que celles dont le nombre est statistiquement représentatifs. Le test de Levene appliqué aux trois catégories indique que l'hypothèse d'homogénéité des variances est respectée dans tous les cas. En effet, les valeurs p des regroupements en 5, 10 et 15 groupes sont respectivement de 0.664, 0.643 et 0.342. De son côté, le résultat de l'ANOVA classique indique une valeur p de 0.012 pour la catégorie de 5 groupes, 0.017 pour la catégorie de 10 groupes et 0.049 pour la catégorie de 15 groupes. Bien que la valeur p ait une tendance croissante avec le nombre de groupe utilisé, les trois tests concluent que la recette du produit exerce une influence sur l'écart de mesure NIR.

Les graphiques des réponses moyennes des figures suivantes montrent d'ailleurs une variabilité marquée entre certaines recettes. Pour le partitionnement en 5 groupes de produit, le groupe 1 produit une réponse moyenne positive alors que le groupe 3 produit une réponse moyenne négative. Les partitionnements en 10 et 15 groupes permettent d'obtenir plus de détails sur les groupes de produits plus précis étant à l'origine de cette différence. Les graphiques de la figure 4.20 et de la figure 4.21 montrent que l'écart positif est principalement expliqué par les premiers sous-groupes de produits distincts alors que le groupe représentant la nourriture pour chat à haute teneur en protéine est le seul responsable de l'écart négatif.

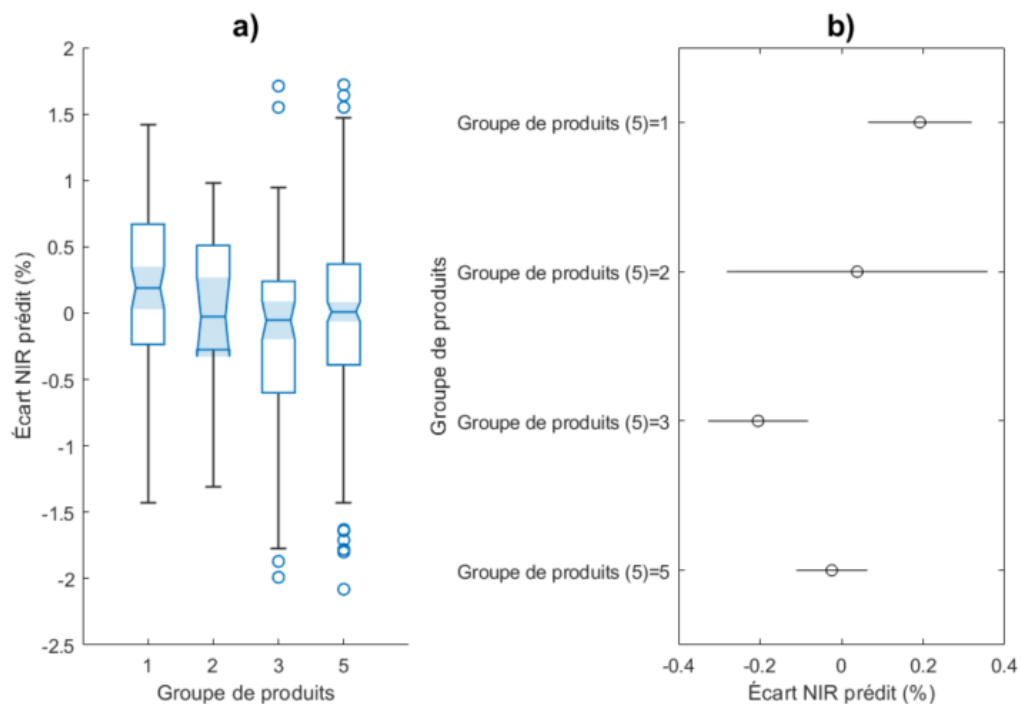


Figure 4.19 : Résultats graphiques ANOVA — groupe de produit (5)
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

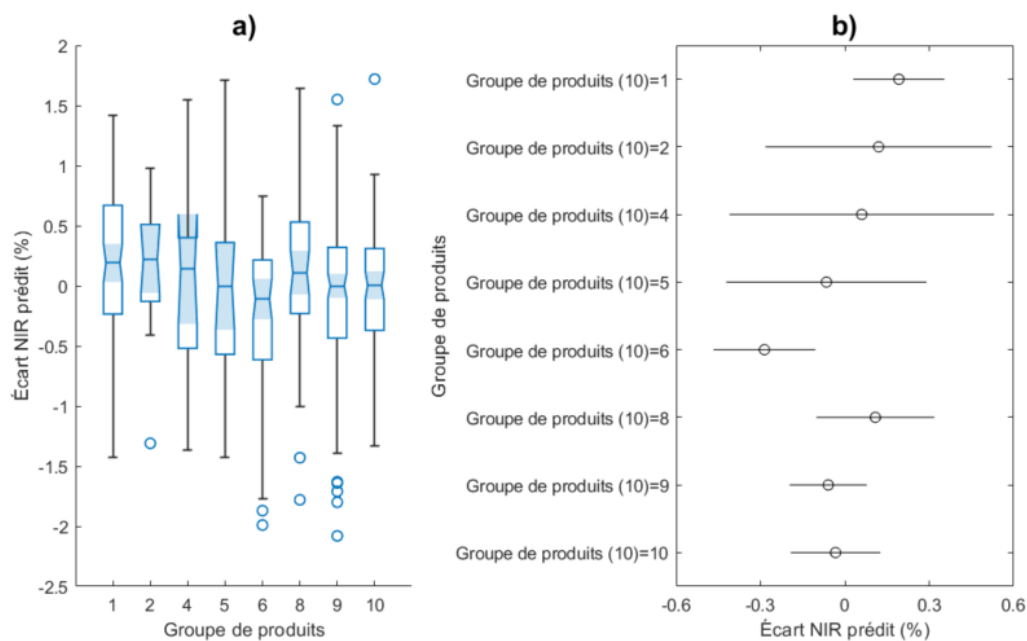


Figure 4.20 : Résultats graphiques ANOVA — groupe de produit (10)
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

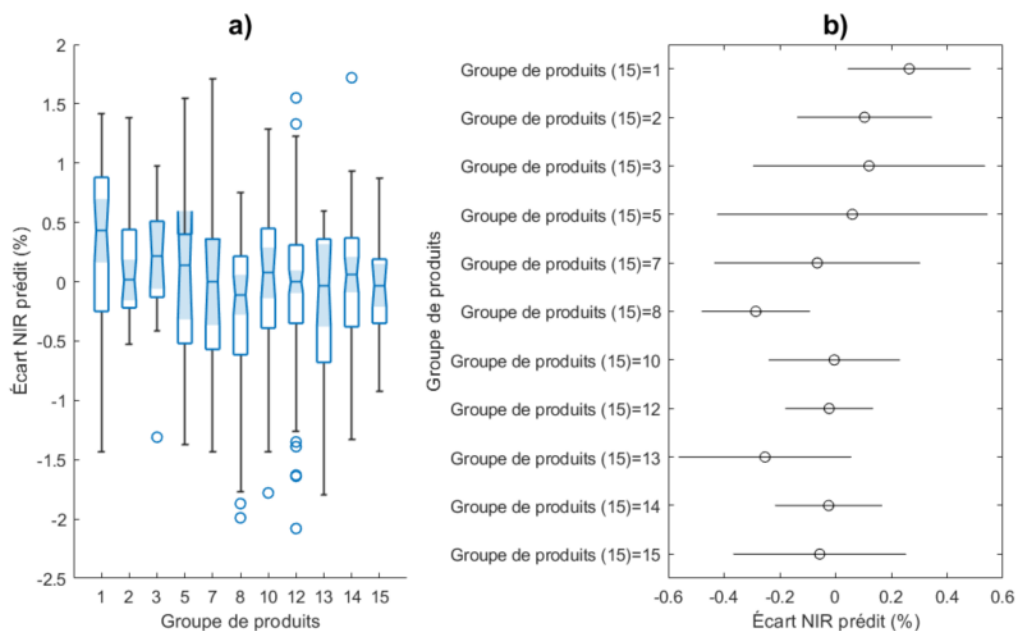


Figure 4.21 : Résultats graphiques ANOVA — groupe de produit (15)
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

Le résultat semble donc affecté par la teneur en protéine plutôt que par un groupe d'ingrédients en particulier. Pour vérifier cette hypothèse, les 60 échantillons concernés par ce groupe ont été retirés de l'analyse. Les résultats de l'ANOVA obtenus indiquent alors que la recette du produit n'exerce pas d'influence significative sur l'écart de mesure NIR. En effet, les valeurs p ont toutes augmenté à 0.151 pour la catégorie de 5 groupes, 0.449 pour la catégorie de 10 groupes et 0.449 pour la catégorie de 15 groupes.

4.4.9. Ajout de saveur externe

Selon les analyses effectuées par forêt aléatoire, l'ajout de saveur en poudre exercerait une faible influence sur l'écart de mesure NIR alors que les effets des ajouts de saveurs liquides et de gras seraient négligeables. Selon les résultats du test de Levene, les variances peuvent être considérées comme homogènes entre tous les groupes. En effet, les valeurs p associées à la saveur en poudre, la saveur liquide et l'ajout de gras sont respectivement de 0.481, 0.593 et 0.856.

Toutefois, les analyses ANOVA révèlent un effet significatif pour chacun des trois types d'ajouts. En effet, les valeurs p obtenues sont de 0.025 pour la saveur en poudre, 0.014 pour la saveur liquide, et 0.016 pour l'ajout de gras. Ce résultat peut être expliqué par la présence d'une famille de produit de nourriture pour chat caractérisée par un haut taux de protéine. Celle-ci est également la seule à avoir la saveur en poudre CHP_AA, la saveur liquide CHP et occupe une grande partie des échantillons du groupe de saveur de gras GP_B0E. Ces groupes démontrent en effet une réponse moyenne plus basse dans les figures Figure 4.22, Figure 4.23 et Figure 4.24, ce qui a pour effet de produire des valeurs p plus basses dans les analyses ANOVA.

En retirant les 60 échantillons de cette famille de produit, les résultats de l'ANOVA ne démontrent pas d'effet significatif des trois saveurs sur l'écart de mesure NIR. En effet, les valeurs p passent à 0.448 pour la saveur en poudre, 0.660 pour la saveur liquide et 0.431 pour le gras.

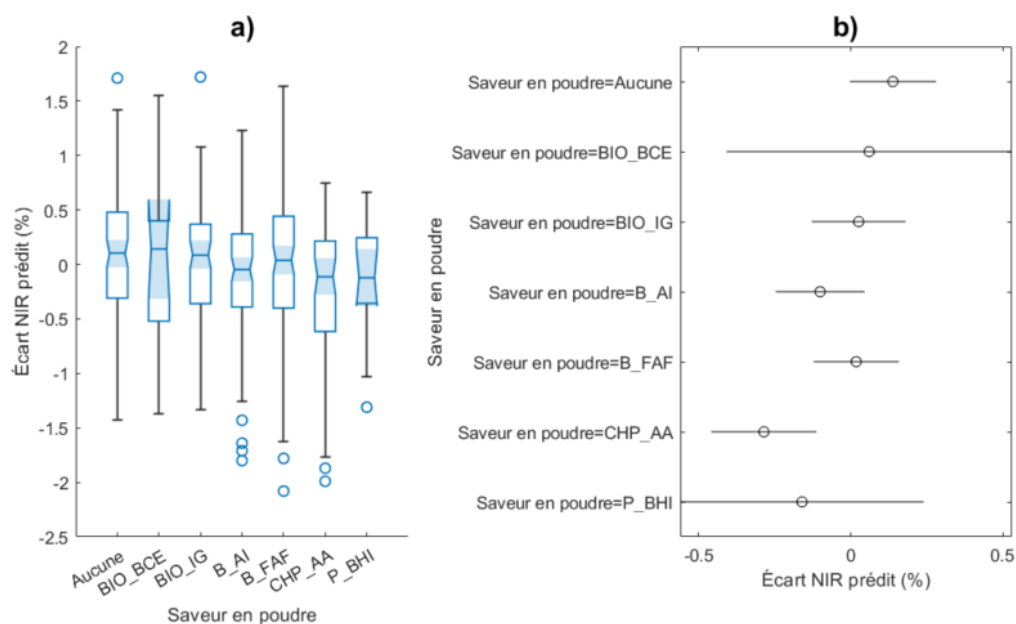


Figure 4.22 : Résultats graphiques ANOVA — saveur poudre
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

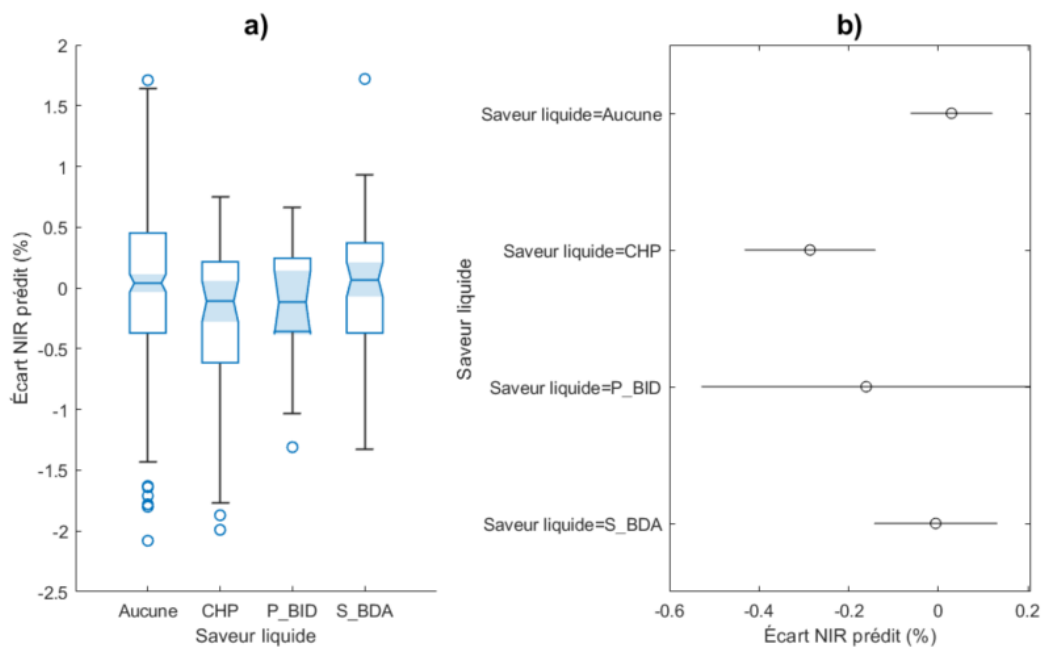


Figure 4.23 : Résultats graphiques ANOVA — saveur liquide
 a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

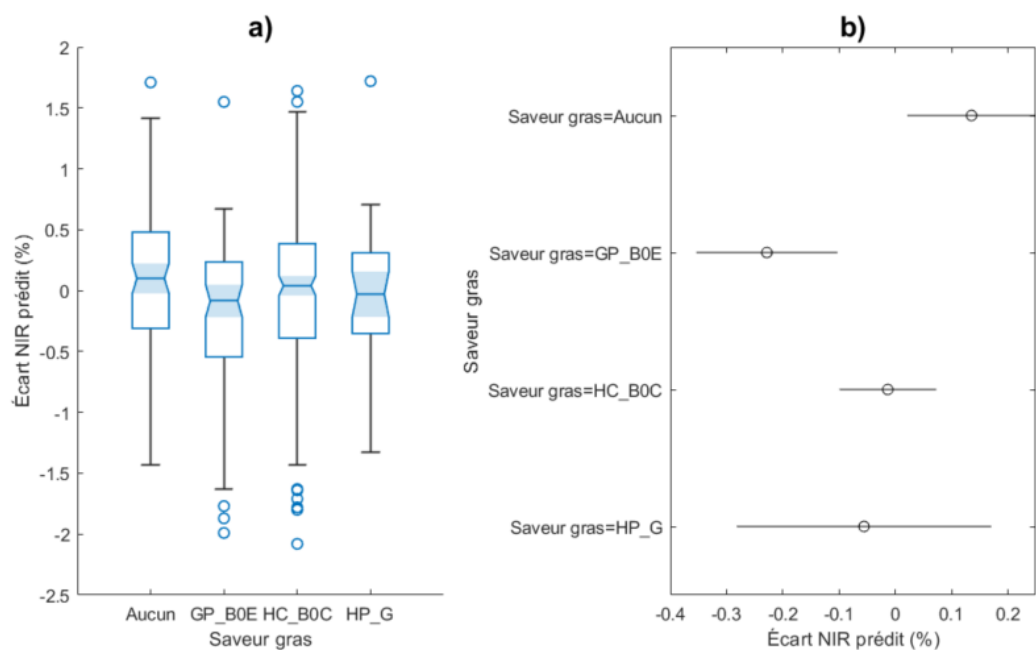


Figure 4.24 : Résultats graphiques ANOVA — saveur gras
 a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

4.4.10. Délai de mesure de référence

La variable représentant le délai de mesure des échantillons en laboratoire a été identifiée comme exerçant un effet moyen sur l'écart de mesure NIR lors de l'analyse par forêt aléatoire. De son côté, l'analyse de la covariance démontre que sa covariance est faible. Afin de confirmer le tout, une ANOVA a été effectuée en utilisant chaque jour de délai comme un groupe spécifique. Selon les résultats du test de Levene, les variances sont égales pour tous les groupes, car la valeur p résultante est de 0.280. L'analyse ANOVA produit une valeur p de 0.152 prouvant que ce facteur n'a pas d'effet significatif sur l'écart de mesure. Selon les graphiques de la figure 4.25, aucune tendance précise ne peut être identifiée.

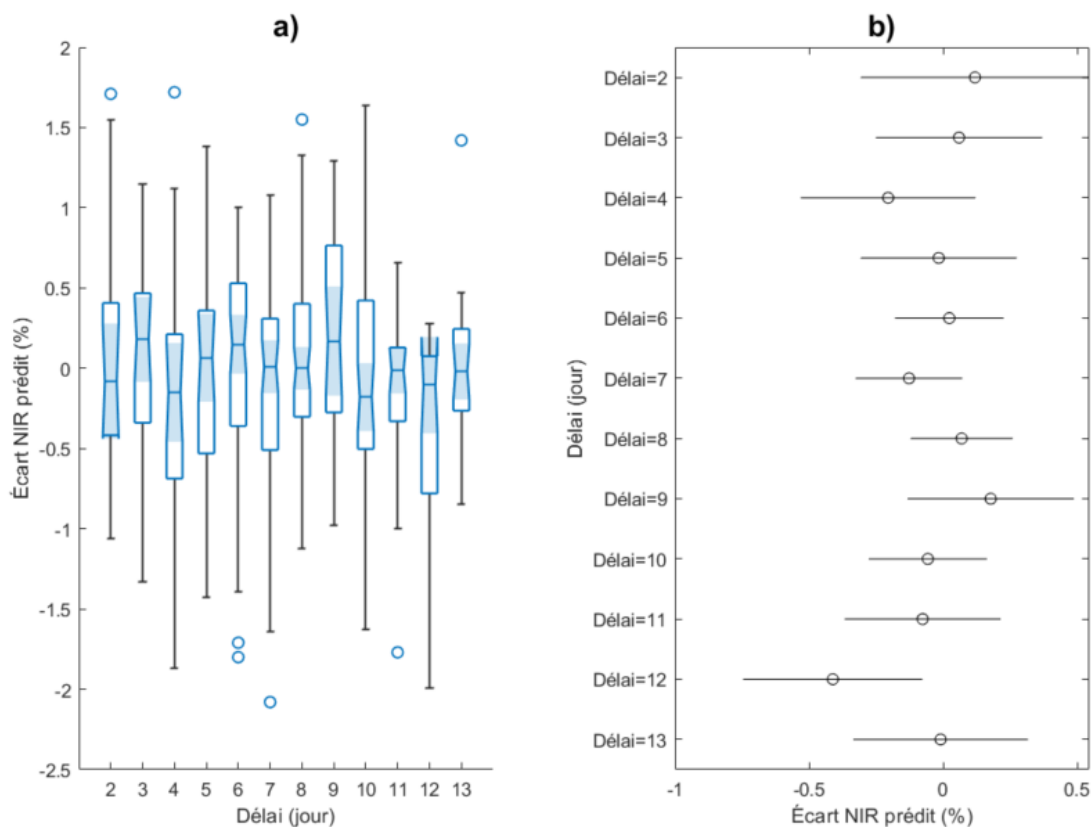


Figure 4.25 : Résultats graphiques ANOVA — délai de mesure
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

4.5. Essais expérimentaux

Les effets de tous les facteurs faisant l'objet d'essais expérimentaux ont été analysés selon leur méthodologie respective présentée au chapitre précédent. L'analyse débute par l'étude de l'impact de la quantité d'échantillon, suivie de l'influence de l'emballage. Ensuite, les résultats relatifs à l'effet de la lumière ambiante et du spectre émis par le convoyeur sont présentés. Enfin, l'influence de la température interne du spectromètre est examinée en détail.

4.5.1. Quantité d'échantillons

Un échantillon des produits suivants a été prélevé en fin de ligne selon la procédure interne d'échantillonnage de l'entreprise :

- DG920, nourriture pour chien grande race, forme standard, volume 2300 mm³ ;
- DGJ901, nourriture pour chien adulte, forme lotus, volume 1150 mm³ ;
- CHT05, nourriture pour chat, forme standard, volume 250 mm³.

Suite au prélèvement, trois sous-échantillons de 230 g et huit sous-échantillons de 5 g ont été mis au four à une température de 104 °C pendant 3 heures. La quantité restante pour chaque produit a été conservée dans un sac EPL-4575 de format moyen pour la suite des essais. Les résultats de mesure d'humidité par séchage au four ont été enregistrés et associés au jour 0. Par la suite, huit sous-échantillons de 5 g ont été soumis à la même procédure à un intervalle d'environ 2 jours pendant un minimum de 20 jours.

Le tableau 4.1 présente les statistiques descriptives des deux masses testées pour les différents produits suite à la simulation monte carlo avec $N = 10\ 000$. Les résultats démontrent que les écarts-types des sous-échantillons de 5 g sont plus élevés que ceux des sous-échantillons de 230 g, et ce, pour les trois produits. De plus, l'écart-type diminue avec le volume du produit, ce qui semble indiquer que les produits constitués de petites croquettes sont caractérisés par une variabilité plus faible. Les graphiques de la figure 4.26 permettent de visualiser ces différences sous forme d'histogramme. L'échelle de chacun de ces graphiques a été ajustée sur l'axe des ordonnées afin de représenter adéquatement les distributions. Basé sur ces résultats, la quantité

de produit utilisée pour les mesures par séchage au four est un facteur pouvant influencer grandement la mesure de référence d'humidité, donc l'écart mesuré par la méthode NIR.

Tableau 4.1 : Statistiques descriptives — échantillons de 230 g et 5 g

Produit	Masse	Moyenne	Écart-type	Intervalle de confiance 95 %
DG920	5 g	8.88	0.86	7.18 à 10.57
	230 g	8.91	0.31	8.30 à 9.53
DGJ901	5 g	9.22	0.77	7.72 à 10.71
	230 g	8.92	0.11	8.72 à 9.13
CHT05	5 g	7.80	0.44	6.93 à 8.65
	230 g	7.70	0.01	7.67 à 7.72

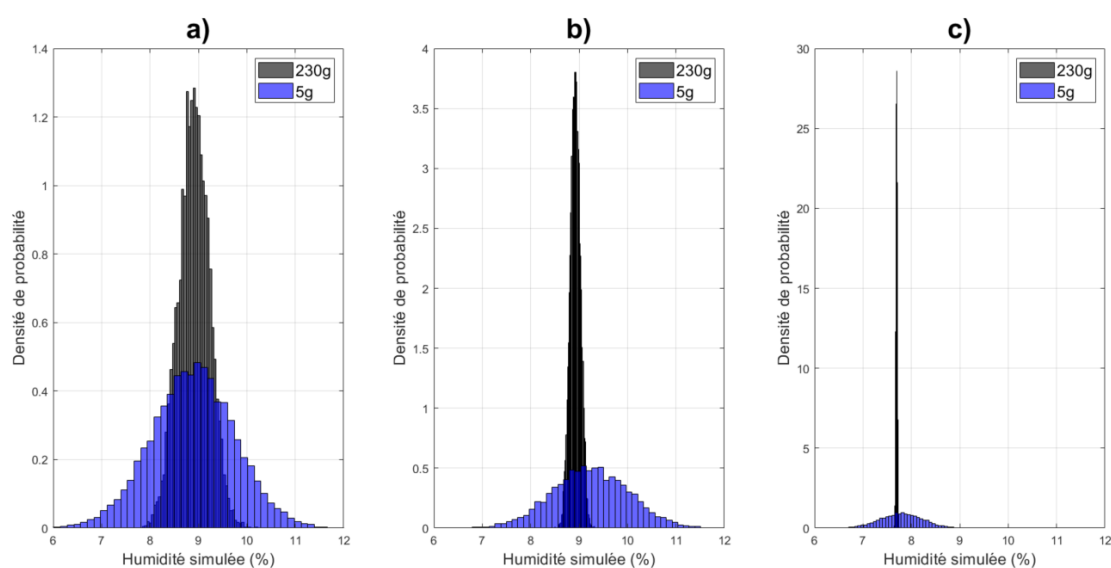


Figure 4.26 : Distributions estimées — échantillons de 230 g et 5 g
a) DG920, forme standard, volume 2300 mm³, b) DGJ901, forme lotus, volume 1150 mm³, c) CHT05, forme standard, volume 250 mm³.

Les graphiques de la figure 4.27 montrent l'évolution des mesures d'humidité par séchage au four sur les échantillons de 5 g pour la durée totale prévue dans la méthodologie. Ces graphiques permettent de constater visuellement les différents écart-types du tableau 4.1. Le produit DG920 formé de plus grosses croquettes est caractérisé par une instabilité accrue alors que le plus petit produit CHT05 présente une plus grande stabilité entre les mesures. Le produit

de format moyen DGJ901 se situe entre les deux. Tous les produits ont perdu de l'humidité comparativement à la mesure des sous-échantillons effectuée dès leur prélèvement sur la ligne de production, ce qui indique qu'une mesure prise plus tard a peu de chance de représenter la réalité.

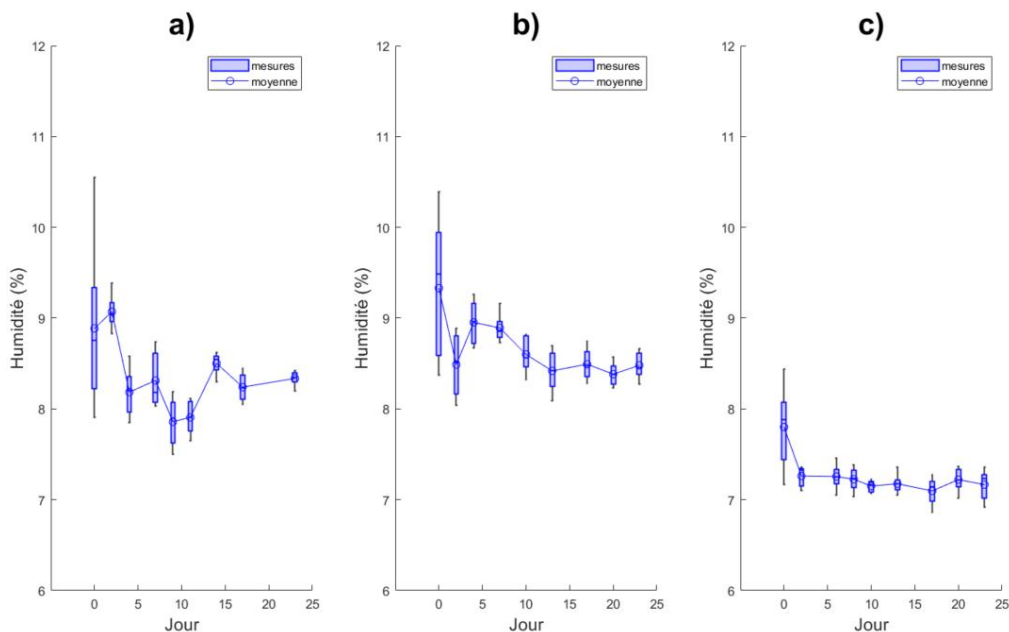


Figure 4.27 : Évolution des mesures de référence d'humidité (5 g) pour trois produits
a) DG920, forme standard, volume 2300 mm³, b) DGJ901, forme lotus, volume 1150 mm³, c) CHT05, forme standard, volume 250 mm³.

Afin de vérifier statistiquement si la durée de stockage a un impact sur la mesure de référence, une ANOVA a été effectuée sur chaque produit en utilisant le jour comme catégorie. Les tests de Levene démontrent que l'hypothèse d'homogénéité des variances entre les groupes n'est pas respectée pour les trois produits. En effet, les valeurs p obtenues sont de 0 pour les trois produits. De son côté, L'ANOVA de Welch a produit une valeur p de 0 pour le produit DG920, 0 pour le produit DGJ901 et 0.014 pour le produit CHT05, indiquant un effet significatif de ce facteur sur la mesure d'humidité dans tous les cas. Cependant, en retirant les valeurs du jour 0, le résultat de l'ANOVA de Welch devient différent pour le produit CHT05 seulement avec une valeur p de 0.183. En effet, pour ce produit, à partir de 2 jours de stockage, le délai n'a plus

d'impact significatif sur la mesure de l'humidité. Les résultats graphiques des ANOVA pour les trois produits sont présentés dans les figures figure 4.28 à figure 4.30.

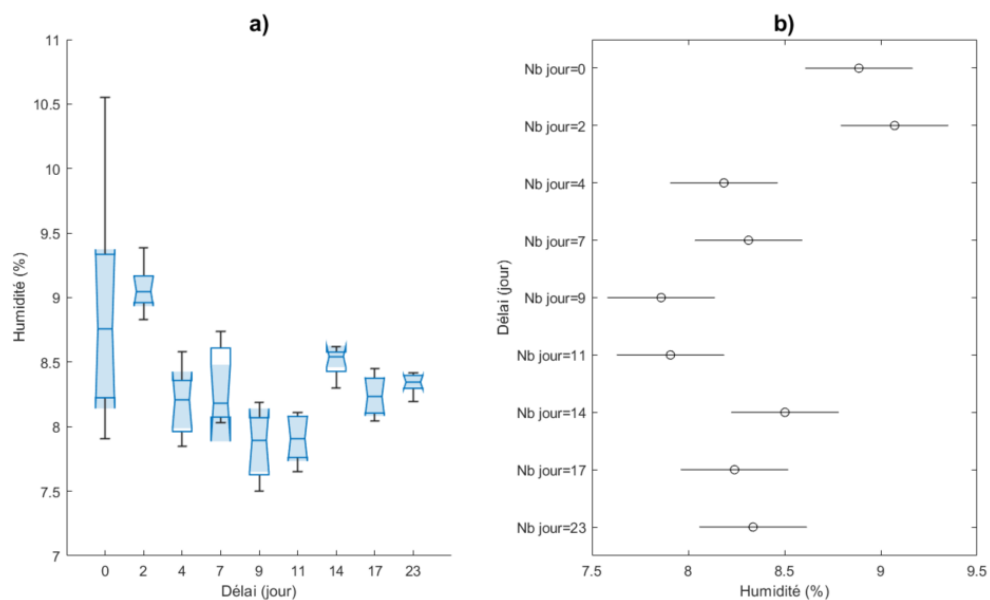


Figure 4.28 : Résultats graphiques ANOVA — humidité DG920
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

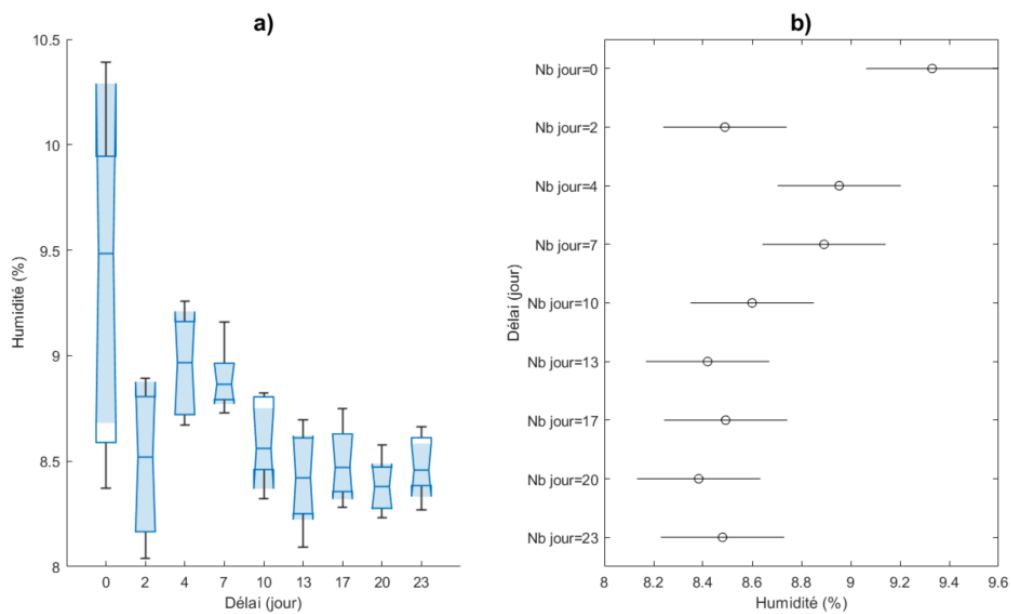


Figure 4.29 : Résultats graphiques ANOVA — humidité DGJ901
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

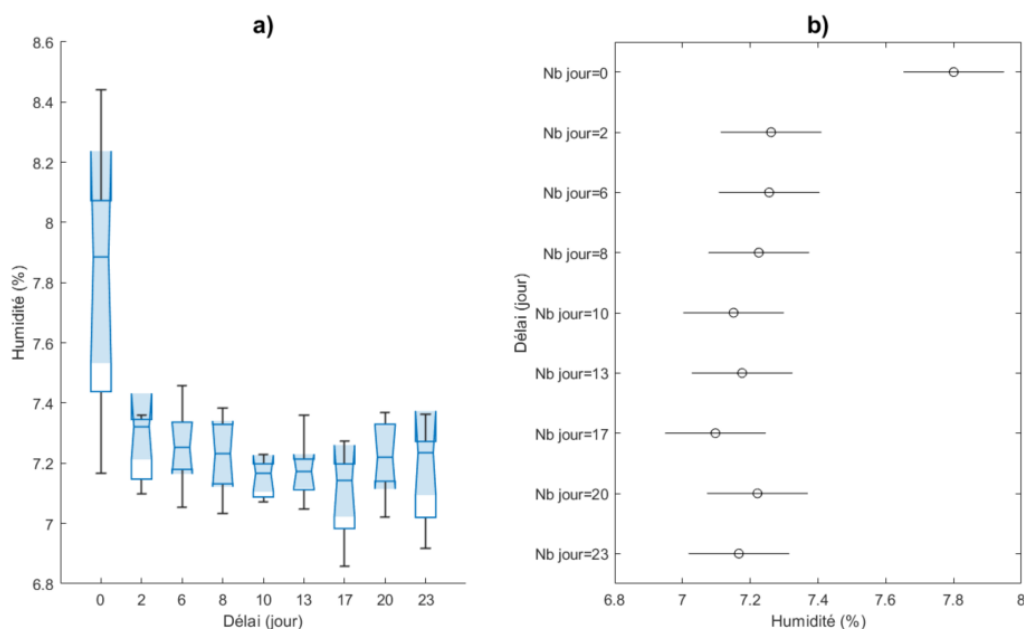


Figure 4.30 : Résultats graphiques ANOVA — humidité CHT05
a) comparaison des distributions entre groupes, b) réponses moyennes et intervalle de confiance des groupes.

4.5.2. Emballage des échantillons

Un échantillon d'un produit populaire de nourriture pour chat a été prélevé en fin de ligne en s'assurant de suivre la procédure interne de l'entreprise pour l'enregistrement des données NIR. Le produit ainsi qu'un capteur de température et d'humidité RHT10 ont été ensachés dans un sac EPL-4575 de format moyen, puis le sac a été déposé sur la balance de précision. Au début de l'essai, le poids de l'échantillon était de 491.41 g. Le sac a été laissé sur la balance pour une durée de 8 jours en enregistrant périodiquement la masse et les conditions environnantes à l'intérieur et à l'extérieur du sac scellé. Les graphiques de la figure 4.31 montrent l'évolution de la masse ainsi que celle des températures et d'humidité relatives.

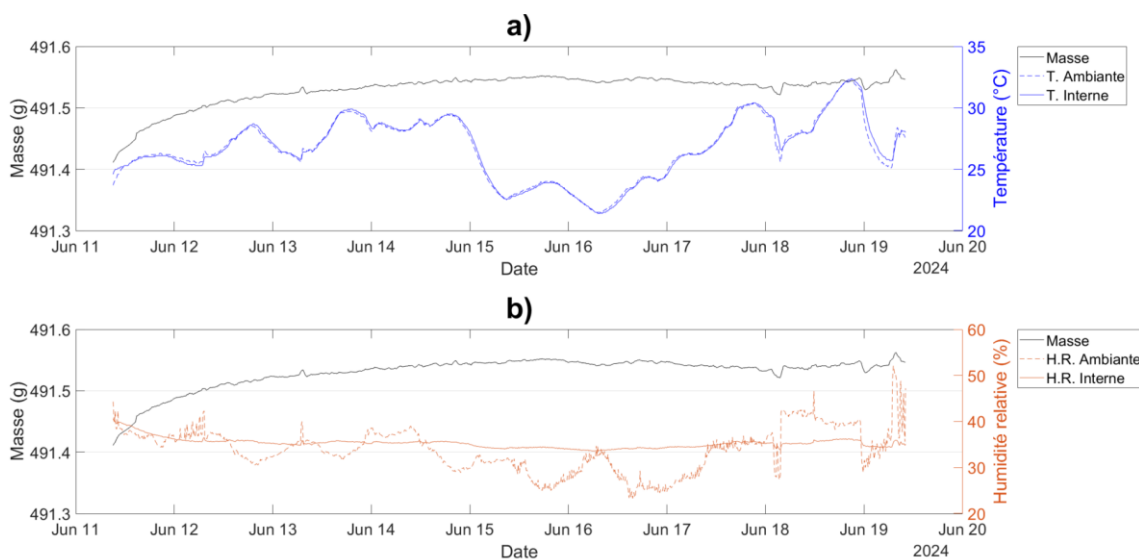


Figure 4.31 : Évolution de la masse — échantillon ensaché en conditions réelles
 a) évolution de la masse et de la température ambiante, b) évolution de la masse et de l'humidité relative.

La courbe de la masse en fonction du temps indique la présence d'un régime transitoire d'une durée approximative de 3 jours durant lequel l'échantillon a gagné de la masse de façon exponentielle. Le régime permanent qui suit cette phase montre que la masse s'est stabilisée et a continué d'évoluer faiblement avec les conditions environnantes. Sur l'ensemble de la période d'essai, bien que la température ait fluctué, aucune corrélation notable avec la masse de l'échantillon n'a été observée. Comme l'illustre la figure 4.31a, la température interne du sac suit celle de l'environnement avec un certain décalage temporel. En revanche, l'humidité relative externe semble avoir exercé une influence plus marquée sur la masse de l'échantillon. En effet, à quatre reprises, des variations brusques et significatives de l'humidité relative externe ont entraîné des modifications mesurables de la masse. Il est à noter que l'amplitude des variations d'humidité relative mesurées à l'intérieur du sac est nettement inférieure à celle observée à l'extérieur. Cette différence s'explique par le fait que le capteur interne mesure la pression de vapeur d'eau émise par l'échantillon lui-même, et non celle de l'air ambiant.

Le gain de masse durant le régime transitoire et les variations en régime permanent ne peuvent être expliqués que par le transport d'une quantité d'eau au travers de la paroi du sac. Il est alors prouvé que le sac EPL-4575 ne peut pas créer une barrière parfaite contre les conditions

ambiantes de la salle dans laquelle l'échantillon est conservé. Cependant, la quantité d'eau totale absorbée par l'échantillon correspond à 0.135 g en 8 jours, ce qui représente une augmentation de 0.016 g par jour en moyenne. La teneur en eau selon la base humide correspondante à cette masse est de 0.026 %, soit un gain moyen de 0.0032 % par jour. À partir de ce résultat, il est peu probable que l'utilisation de ce type de sac soit responsable d'un écart de mesure NIR. Cependant, puisque l'échantillon s'équilibre selon les conditions ambiantes de la salle, il serait possible que des conditions extrêmes provoquent des changements de taux d'humidité considérables.

Pour vérifier cette hypothèse, le même échantillon a été séché au four à 104 °C pendant 5 heures afin d'assurer d'évaporer la plus grande quantité d'eau possible. Une étuve TOV-5 a été conditionnée à une température de 27 °C et un seau d'eau a été placé à l'intérieur afin d'assurer une humidité relative ambiante maximale. Suite au séchage, l'échantillon de produit ainsi qu'un capteur de température et d'humidité RHT10 ont été ensachés dans un nouveau sac EPL-4575 de format moyen, puis le sac a été déposé sur la balance de précision. Au début de l'essai, le poids de l'échantillon était de 175.03 g. Le sac a été laissé sur la balance pour une durée de 10.8 jours en enregistrant périodiquement la masse et les conditions environnantes à l'intérieur et à l'extérieur du sac scellé. Les graphiques de la figure 4.32 montrent l'évolution de la masse ainsi que celle des températures et d'humidité relatives.

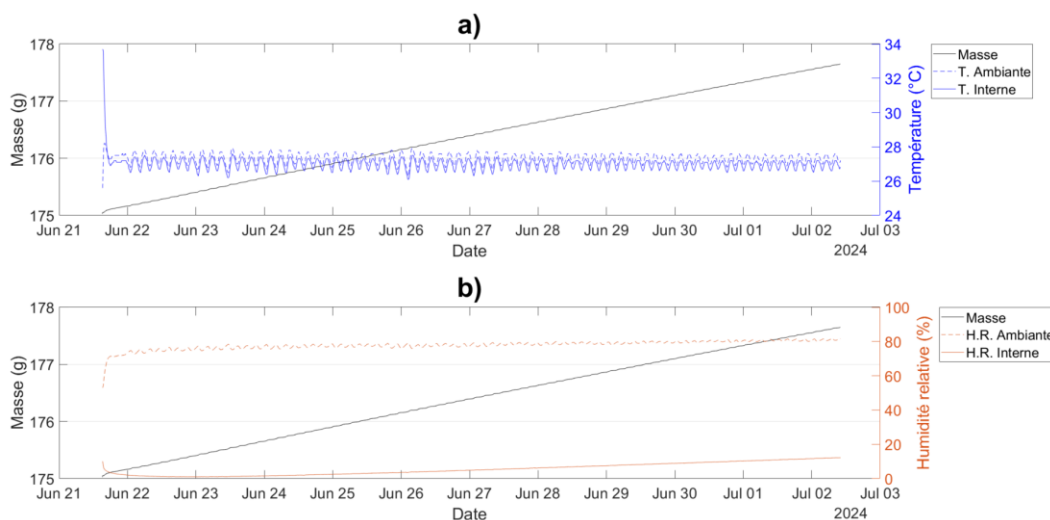


Figure 4.32 : Évolution de la masse — échantillon ensaché en pires conditions
 a) évolution de la masse et de la température ambiante, b) évolution de la masse et de l'humidité relative.

Pour ce cas extrême, la courbe de masse de l'échantillon montre également le même phénomène transitoire au début du test, mais évolue de façon plus rapide. L'échantillon gagne de la masse alors que l'humidité relative à l'intérieur du sac diminue de façon exponentielle. Par la suite, la courbe prend une allure linéaire et n'atteint pas de régime permanent. Le gain de masse durant l'essai s'explique par le transport d'une quantité d'eau au travers de la paroi du sac provenant de l'environnement très humide de l'étuve. Cependant, ce phénomène se fait lentement, car la masse totale d'eau absorbée sur les 10.8 jours est de 2.62 g, soit une augmentation de 0.24 g par jour. Cette masse correspond à une teneur en eau selon une base humide de 1.48 % et un taux d'augmentation de 0.14 % par jour. Cette valeur pourrait affecter l'écart de mesure NIR pour des échantillons laissés longtemps en stockage sous des conditions extrêmes.

Les deux essais ont démontré que les sacs utilisés pour entreposer les échantillons ne sont pas complètement imperméables à la vapeur d'eau. La quantité d'eau pouvant être absorbée par un échantillon plus sec dépend à la fois de sa teneur en eau et de l'humidité relative ambiante. Le cas en situation réelle a démontré que la masse d'eau absorbée est minime alors que le test en condition extrême mène à un gain de masse non négligeable.

4.5.3. Lumière environnante

Les mesures des spectres bruts ont été effectuées aux emplacements et orientations spécifiés dans le tableau 3.3. Étant donné les faibles valeurs d'amplitude générées, un exemple de spectre du produit CHT01 est fourni dans les graphiques afin de mieux évaluer l'ordre de grandeur. La figure 4.33 présente les différents spectres dont l'amplitude correspond au compte normalisé du détecteur, soit la valeur d'intensité brute lue par celui-ci.

Dans tous les cas, les mesures d'amplitude sont négligeables par rapport à la mesure sur un produit. Le pire cas, représenté par une mesure à 200 mm de la source lumineuse, montre une amplitude un peu plus élevée pour les longueurs d'onde de 1030 à 1330 nm, mais les valeurs restent tout de même très faibles, car elles représentent moins de 1 % de l'amplitude du produit. Tous les spectres d'arrière-plan obtenus constituent une ligne de base de basse amplitude, laquelle peut être facilement retirée par la plupart des techniques de prétraitement telles que le SNV, *detrend* et Savitzky-Golay sans aucun risque d'affecter l'information importante du signal.

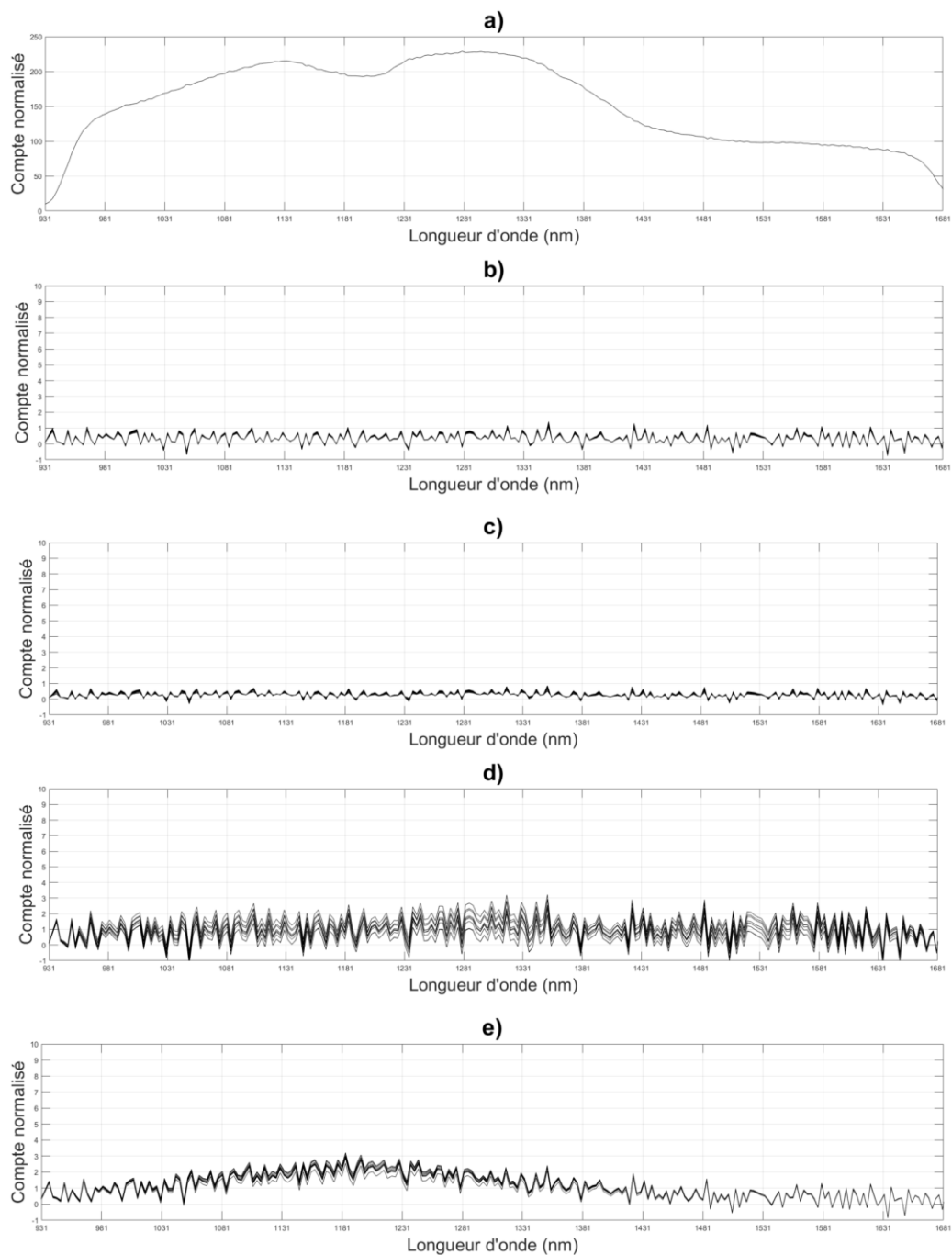


Figure 4.33 : Spectres de l'arrière-plan à différents endroits
a) spectre d'un produit, b) arrière-plan, proximité du capteur, vers le convoyeur, convoyeur en arrêt, c) arrière-plan, proximité du capteur, vers le convoyeur, convoyeur en marche, d) arrière-plan, proximité du capteur, vers la source lumineuse, e) arrière-plan, 200 mm de la source lumineuse.

4.5.4. Spectre du convoyeur

La mesure du spectre du tapis du convoyeur a été effectuée avec le capteur NIR DA 7440™ en désactivant la fonction de détection du produit. Une mesure a été prise avec le convoyeur à l'arrêt et une deuxième lorsque le convoyeur était en marche. Afin de juger de l'ordre de grandeur de l'amplitude des mesures, un spectre du produit CHT01 est fourni dans les graphiques. Les résultats sont présentés à la figure 4.34 dans laquelle l'amplitude représente l'absorbance mesurée par le capteur.

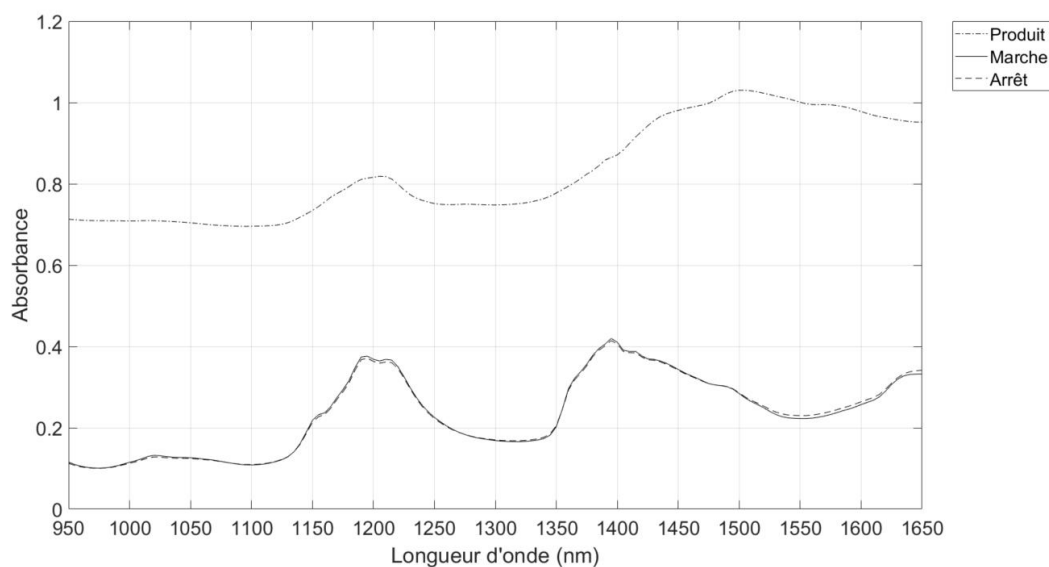


Figure 4.34 : Spectres du convoyeur

Les deux spectres obtenus avec le convoyeur en arrêt et en marche ne présentent pas de différence notable. Ceux-ci ont une moyenne relativement basse comparée à celle du signal d'un produit. Ceci s'explique par le fait que le produit est plus près du capteur, car il est empilé sur une certaine hauteur. Ainsi, la quantité de lumière réfléchie qui revient au détecteur est plus grande. C'est en partie ce qui permet de distinguer entre la présence d'un produit et son absence. L'autre distinction est possible par la forme du signal. Les pics et les vallées des deux signaux ne sont pas situés aux mêmes longueurs d'onde, ce qui peut être détecté par le capteur NIR pour détecter la présence ou l'absence du produit.

4.5.5. Température interne du spectromètre

Les données de température ambiante et interne du capteur NIR DA 7440™ ont été enregistrées sur la période allant du 7 mai 2024 au 7 mai 2025. Une analyse visuelle des séries chronologiques a permis d'identifier et d'exclure certaines plages de données, notamment celles correspondant à des arrêts prolongés de production ou à des périodes de défaillance des systèmes de transfert et d'enregistrement. Par ailleurs, les valeurs aberrantes relevées par le capteur de température ambiante ont également été retirées de l'analyse.

Le graphique temporel résultant de ces opérations est présenté à la figure 4.35. L'analyse de l'évolution des différentes températures met en évidence une sensibilité des températures de la carte électronique et de la section de contrôle aux variations de la température ambiante, contrairement à celles de la matrice et des composants optiques, qui demeurent stables. Cette observation est appuyée par les histogrammes illustrés à la figure 4.36 ainsi que le diagramme à boîtes de la figure 4.37. En effet, l'étendue des valeurs mesurées pour la matrice et les composants optiques est très faible, tandis que celle des températures de la carte électronique et de la section de contrôle présente une variabilité plus marquée.

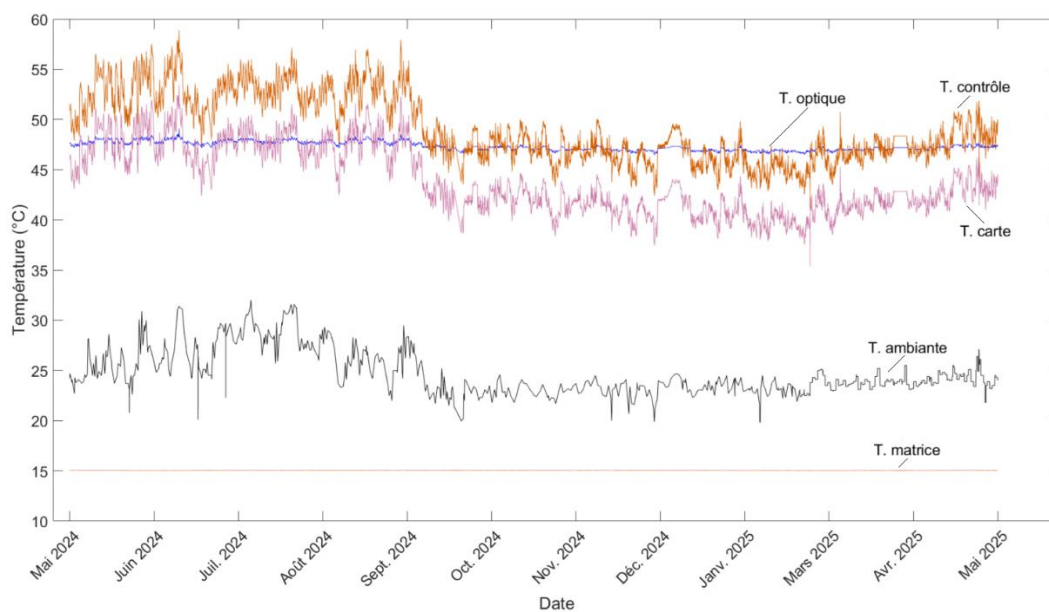


Figure 4.35 : Graphique temporel des températures du capteur NIR DA 7440™

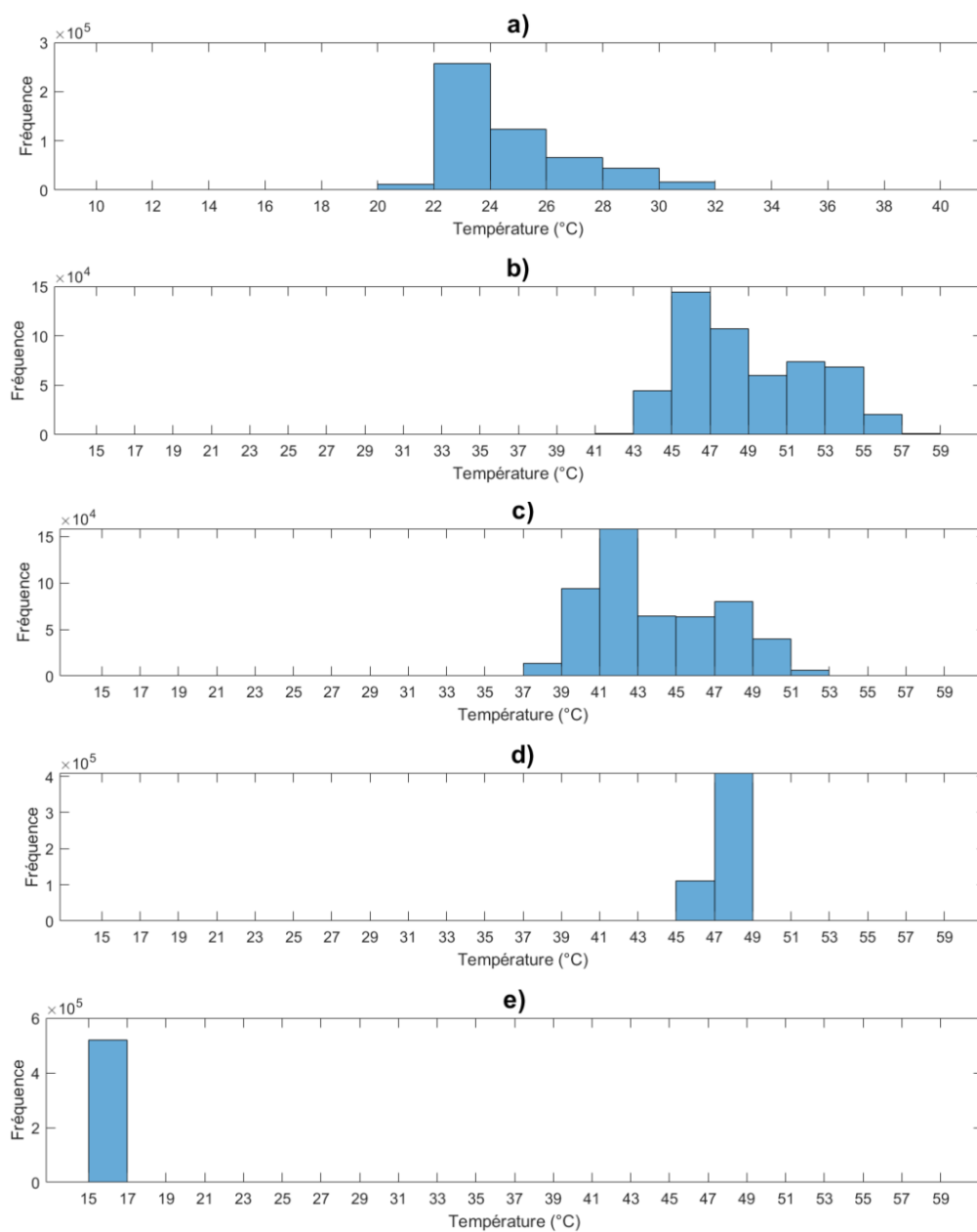


Figure 4.36 : Histogrammes des températures du capteur NIR DA 7440™
 a) température ambiante, b) température de la section contrôle, c) température de la carte,
 d) température de la section optique, e) température de la matrice.

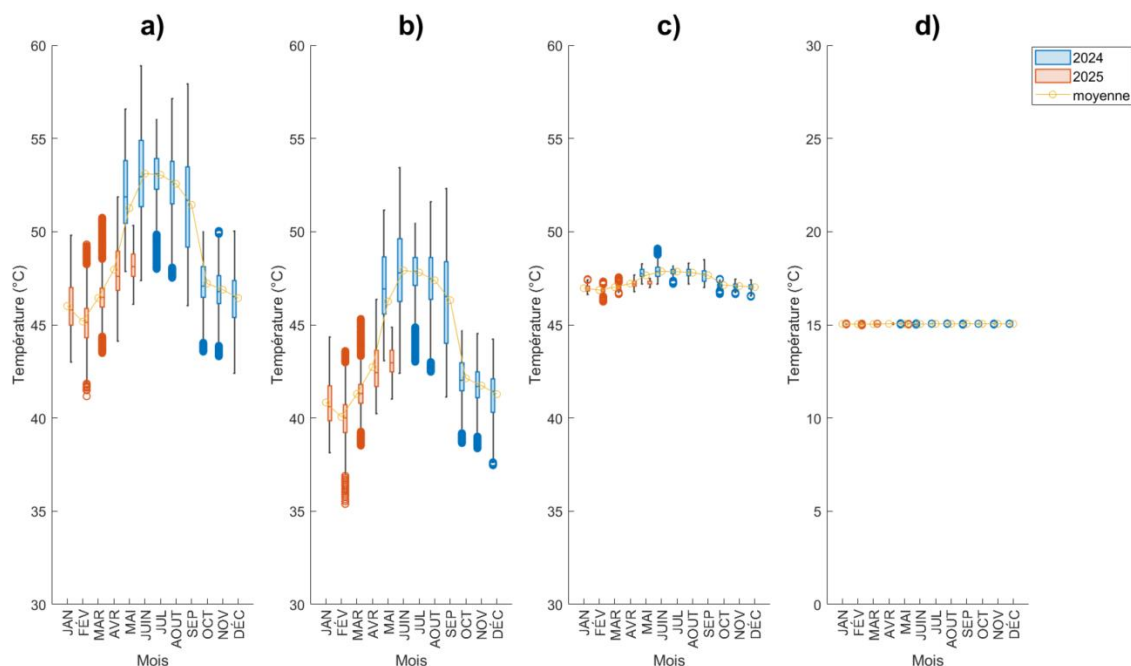


Figure 4.37 : Diagramme de boîtes des températures internes du capteur NIR DA 7440™
a) température de la section contrôle, b) température de la section carte, c) température de la section optique,
d) température de la section matrice.

La figure 4.37 montre également que les températures sont plus élevées pendant la période estivale. Le même phénomène est présent à la figure 4.38 pour la température ambiante de la salle où se trouve le capteur. Cette différence s'explique par une meilleure capacité du système de chauffage et de climatisation à réguler une température en chauffant plutôt qu'en climatisant. En effet, le système de climatisation ne semble pas être en mesure de garantir une température fixe lors des chaudes journées d'été. Les valeurs aberrantes basses signalées en février, juillet, octobre et décembre sont principalement reliées à des arrêts de production, notamment causées par l'arrêt du séchoir situé à proximité.

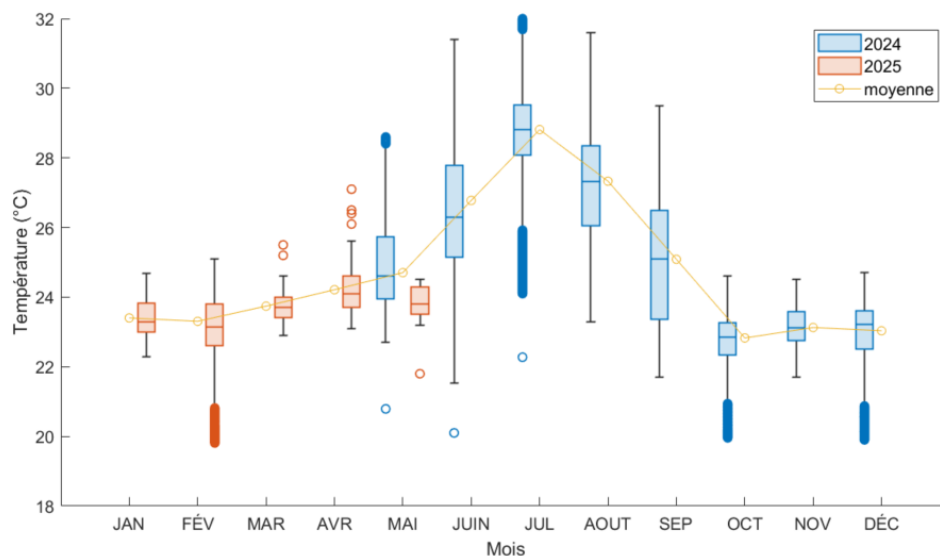


Figure 4.38 : Diagramme de boîtes de la température ambiante

La matrice de covariance de la figure 4.39 montre que la température ambiante a un effet sur la température du circuit de contrôle et de la carte électronique. Cependant, son effet est très faible sur les composantes optiques et nul sur la matrice de diode. Le système de refroidissement thermoélectrique est alors très performant et assure une température stable des éléments critiques du capteur indépendamment des conditions extérieures.

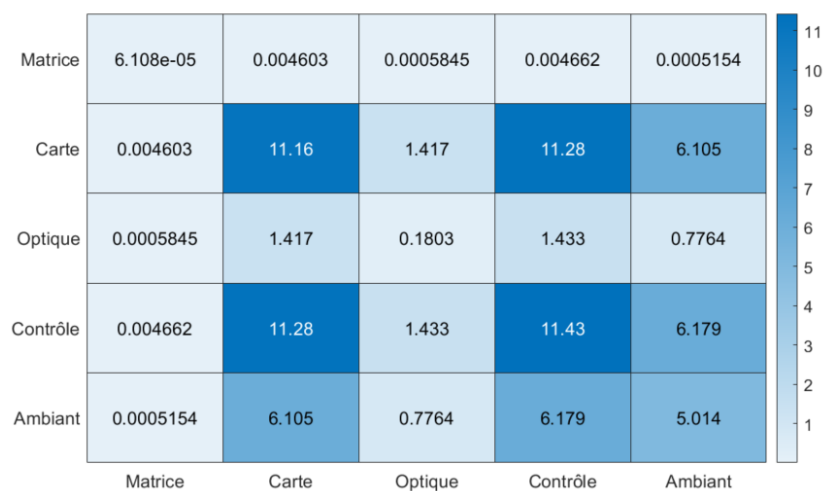


Figure 4.39 : Matrice de covariance des températures du capteur

4.6. Résumé des analyses

Cette section présente une synthèse structurée des principaux résultats obtenus au cours de l'étude. Premièrement, les résultats issus des analyses menées à partir du jeu de données de calibration sont présentés, accompagnés de la quantification des facteurs identifiés comme exerçant une influence statistiquement significative. Par la suite, les résultats des essais expérimentaux sont résumés et leur influence respective est quantifiée.

4.6.1. Facteurs du jeu de données de calibration

Le tableau 4.2 contient les résultats des analyses pour chaque facteur identifié dans le jeu de donnée de calibration. Ce tableau résume les résultats de chacune des trois analyses effectuées sur le facteur lorsqu'applicable, soit la forêt aléatoire, la matrice de covariance et l'ANOVA.

Le tableau 4.3 regroupe les facteurs exerçant une influence haute ou moyenne sur l'écart de mesure NIR de l'humidité accompagné des données quantifiables sur leur effet.

Tableau 4.2 : Résumé des analyses du jeu de calibration

Facteur	Forêt aléatoire	Covariance	ANOVA	Influence
	Importance		Valeur p	
Modèle de calibration	Non évalué	Non applicable	0.010	Haute
Taux de protéine	Haut	-0.160	0.009	Haute
Taux de gras	Faible	-0.092	0.202	Négligeable
Température du produit	Moyen	-0.090	0.080	Moyenne
Température ambiante	Faible	-0.067	0.049	Moyenne
Dimension du produit	Haut	-0.153	0.000	Haute
Forme du produit	Négligeable	Non applicable	0.068	Négligeable
Recette des produits				
Partition de 15 groupes	Négligeable	Non applicable	0.449	Négligeable
Partition de 10 groupes	Négligeable	Non applicable	0.449	Négligeable
Partition de 5 groupes	Négligeable	Non applicable	0.151	Négligeable
Ajout de saveur externe				
Saveur en poudre	Faible	Non applicable	0.448	Négligeable
Saveur liquide	Négligeable	Non applicable	0.660	Négligeable
Gras	Négligeable	Non applicable	0.431	Négligeable
Délai de mesure de référence	Moyen	-0.047	0.152	Négligeable

Tableau 4.3 : Quantification des facteurs influents du jeu de calibration

Facteur	Seuil	Écart moyen	Écart-type
Modèle de calibration	Dépendant du modèle	-1.326 à 0.430	0.853 à 1.794
Taux de protéine	40 à 45 %	-0.264	0.664
	45 % et plus	-0.333	0.616
Dimension du produit (volume)	270 mm ³	0.163	0.617
	893 mm ³	-0.132	0.558
	2307 mm ³	-0.220	0.785
Température du produit	10 °C et moins	-0.041	0.527
	29 °C et plus	-0.267	0.604
Température ambiante de la salle	28 °C et plus	-0.478	0.812

4.6.2. Facteurs vérifiés par essais expérimentaux

Le tableau 4.4 contient les résultats des essais expérimentaux réalisés sur certains facteurs potentiels. Pour chacun de ceux-ci, la condition de test est décrite, accompagnée du résultat et de l'évaluation de l'influence du facteur sur l'écart de mesure NIR d'humidité.

Tableau 4.4 : Résumé des essais expérimentaux

Facteur	Condition	Résultat	Influence
Quantité d'échantillons	Produit DG920	$\sigma_{230\text{ g}} = 0.32, \sigma_{5\text{ g}} = 0.88$	Haute
	Produit DGJ901	$\sigma_{230\text{ g}} = 0.11, \sigma_{5\text{ g}} = 0.77$	Haute
	Produit CHT05	$\sigma_{230\text{ g}} = 0.01, \sigma_{5\text{ g}} = 0.44$	Haute
Délai de mesure	Produit DG920	ANOVA p = 0.00	Haute
	Produit DGJ901	ANOVA p = 0.00	Haute
	Produit CHT05	ANOVA p = 0.17	Nulle
Emballage des échantillons	Environnement réel	+0.0032 %/jr	Négligeable
	Environnement extrême	+0.14 %/jr	Haute
Lumière environnante	200 mm de la source	Rapport de signal < 1 %	Nulle ¹
Spectre du convoyeur	Sans produit	Rapport de signal < 49 %	Nulle ²
Température interne du spectromètre	Matrice de diode	Cov. = 0.00	Nulle
	Carte électronique	Cov. = 6.62	Moyenne
	Composants optiques	Cov. = 0.84	Négligeable
	Composants de contrôle	Cov. = 6.70	Moyenne

¹ L'effet de ce facteur est retiré par les techniques de prétraitement.

² L'effet de ce facteur est retiré par une technique de détection du spectre du convoyeur.

CHAPITRE 5 - ANALYSE DES RÉSULTATS ET DISCUSSIONS

Ce chapitre propose une interprétation critique des résultats présentés précédemment. Son contenu vise premièrement à identifier les limites méthodologiques susceptibles d'avoir influencé les résultats. Par la suite, les facteurs ayant démontré un effet significatif sur l'écart de mesure NIR sont examinés plus en détail, tout comme ceux dont l'influence s'est révélée négligeable. Enfin, des discussions sont entamées sur les résultats en les replaçant dans le contexte des objectifs de l'étude ainsi que des connaissances disponibles dans la littérature, puis des pistes d'amélioration et de recherche future sont discutées en lien avec les conclusions tirées.

5.1. Analyse critique des résultats

Les résultats obtenus dans le cadre de cette étude dépendent fortement de la méthodologie mise en œuvre ainsi que des contraintes propres à l'environnement industriel ciblé. Il est donc essentiel de reconnaître les limitations et d'en proposer une analyse critique. Les sections qui suivent présentent d'abord les principales limites méthodologiques relevées, l'influence potentielle de la colinéarité et des variables confondantes, puis exposent l'analyse des facteurs jugés influents ainsi que les justifications ayant mené à l'exclusion de certains facteurs non influents.

5.1.1. Limites méthodologiques

La stratégie utilisée dans cette étude repose principalement sur l'utilisation d'un jeu de données existant n'ayant pas été construit à partir d'un plan d'expérience structuré. Bien que cette option soit peu coûteuse et n'impacte pas les opérations de production de l'entreprise, la fiabilité des résultats dépend directement de la qualité des données recueillies. Ainsi, les efforts habituellement investis dans la planification et l'exécution d'un plan d'expérience rigoureux sont plutôt concentrés dans les étapes de traitement et d'analyse des données, sur la vérification des conditions d'application des différentes méthodes statistiques ainsi que sur l'interprétation complexifiée des résultats.

L'utilisation d'un tel jeu de données ne garantit pas que la plage optimale des variables d'intérêt soit couverte. Par conséquent, les effets détectés par les analyses peuvent parfois être faibles ou difficiles à distinguer, ce qui peut nuire à la solidité des conclusions. Par ailleurs, l'absence de structure expérimentale équilibrée empêche l'évaluation rigoureuse des interactions entre les facteurs au moyen des méthodes standards. Ainsi, pour assurer la fiabilité des résultats, les analyses doivent être menées avec une grande rigueur méthodologique et une prudence accrue dans l'interprétation des observations.

La majorité des valeurs aberrantes présentes dans les données de cette étude ont été retirées des analyses sans que leur origine puisse être déterminée. Bien que cette approche présente l'avantage d'être simple et rapide, elle comporte un risque de biais si ces valeurs résultent de l'influence réelle d'un facteur pertinent. Idéalement, chaque valeur aberrante devrait faire l'objet d'une investigation afin d'en identifier son origine et d'évaluer si son retrait est justifié. C'est d'ailleurs la méthode à suivre selon la norme internationale ISO 12099 (ISO, 2017). Toutefois, dans cette étude, puisque le jeu de données a été produit par différentes personnes sur plusieurs mois, il a été impossible de retracer les causes précises de ces anomalies. Celles-ci ont donc été retirées par précaution sans en avoir identifié la cause. Cependant, considérant leur faible proportion, l'impact de la suppression de ces données est jugé sans impact significatif sur les résultats globaux.

L'utilisation d'une méthode d'estimation hors ligne du taux d'humidité, calculée en utilisant le dernier modèle de calibration valide, tend à produire des valeurs optimistes dans le jeu de données. En effet, bien que chacune des valeurs aient été estimées de la même manière qu'en temps réel en utilisant le spectre original, leur prédiction est effectuée à l'aide d'un modèle calibré sur ces mêmes données. Ainsi, la variance des prédictions est artificiellement réduite par rapport à une utilisation en conditions réelles, ce qui peut contribuer à masquer les effets de certains facteurs. L'étude aura toutefois réussi à identifier les causes principales de variabilité en utilisant cette technique.

L'utilisation de forêts aléatoires comme méthode d'analyse s'accompagne d'une certaine variabilité dans le classement des indices d'importance des facteurs. Cette variabilité s'explique

par le caractère aléatoire de la méthode, qui repose sur la génération répétée d'arbres construits à partir de sous-échantillons de données et de sous-ensembles de variables sélectionnés de manière aléatoire. En conséquence, d'une exécution à l'autre, des fluctuations peuvent être observées dans l'ordre d'importance des facteurs. Toutefois, cette instabilité reste généralement limitée aux facteurs de contribution moyenne. Les variables les plus influentes conservent systématiquement leur position dominante, tandis que les variables les moins influentes conservent leur faible importance. Cette stabilité partielle dans les extrêmes de l'échelle d'importance confère une certaine robustesse à l'interprétation des résultats, malgré la variabilité inhérente à la méthode. C'est pourquoi d'autres analyses plus poussées sont nécessaires afin d'en confirmer les résultats. Malgré cela, la forêt aléatoire demeure un outil pertinent pour effectuer des analyses exploratoires en raison de sa capacité à gérer des relations non linéaires, à prendre en compte des interactions complexes entre les variables et à fournir une mesure directe de l'importance des facteurs sans nécessiter de fortes hypothèses statistiques.

Lors des analyses ANOVA, certains groupes ont dû être retirés de l'analyse par manque de représentativité statistique. En effet, des groupes comportant peu d'échantillons peuvent fausser les résultats en augmentant l'erreur de variance ou en produisant des conclusions non généralisables. Leur retrait vise donc à préserver la robustesse et la validité des analyses. Cependant, cette démarche risque de masquer ou de sous-estimer l'impact de certains groupes et peut limiter la portée des conclusions. Ce compromis méthodologique souligne l'importance d'un échantillonnage équilibré et suffisamment dense pour les études futures.

Les valeurs de température du produit et de température ambiante proviennent de relevés manuels. Pour ces données, il existe un risque d'erreur et d'imprécision provoquée par l'être humain. De plus, l'utilisation de l'interpolation pour attribuer une valeur à un échantillon reste une méthode moins précise qu'une mesure réelle. Cependant, bien que l'analyse de l'effet de ces facteurs à l'aide de trois approches distinctes ait mené à des résultats divergents, la source de cette ambiguïté semble plutôt reliée à une relation non linéaire plutôt qu'à l'imprécision des données collectées.

En ce qui concerne les essais expérimentaux réalisés sur l’emballage des échantillons, ceux-ci ne couvrent qu’un seul produit, lequel semble le moins affecté par les délais de stockage comme l’ont dévoilé les résultats de l’évolution des mesures de références sur plusieurs jours. Les résultats pourraient être différents en utilisant un produit de plus grande dimension. De plus, les essais n’ont pas couvert le phénomène de désorption d’un échantillon humide laissé dans un environnement sec par manque d’équipement dédié au contrôle environnemental. Ainsi, bien que les essais aient pu mesurer une certaine influence de l’emballage, il existe une incertitude sur la plage réelle de l’effet potentiel de ce facteur.

5.1.2. Variables colinéaires et confondantes

Dans cette étude, il existe plusieurs liens de dépendance entre les facteurs étudiés. Ceci tend à complexifier l’analyse, car ce phénomène crée à la fois de la multicollinéarité et provoque un effet de confusion. Certains liens sont plus faciles à identifier que d’autres. Par exemple, il existe une hiérarchie entre le numéro de produit, la recette, la forme, la dimension, les teneurs en protéine et en gras et l’ajout des saveurs externes. Le potentiel de colinéarité entre ces variables est important et peut influencer les résultats. De plus, puisque les produits à haute teneur en protéine sont uniquement associés à une famille de produit de nourriture pour chat, l’influence de cette variable génère un effet de confusion. Pour certains facteurs, ce phénomène a été identifié et minimisé, mais la relation confondante peut affecter les résultats de façon plus subtile dans d’autres cas.

La figure 5.1 contient les résultats d’analyse de la corrélation entre les variables indépendantes numériques du jeu de données. Plusieurs corrélations sont présentes dans le jeu de données, notamment entre le groupe de produit « Recette5 » et la plupart des variables relatives aux caractéristiques des produits. De plus, le taux de protéine, facteur le plus influent identifié dans les résultats, est caractérisé par de hauts coefficients de corrélation avec le taux de gras et l’ajout de saveurs externes. Cette relation a notamment été constatée lors de la présentation des résultats au chapitre précédent.

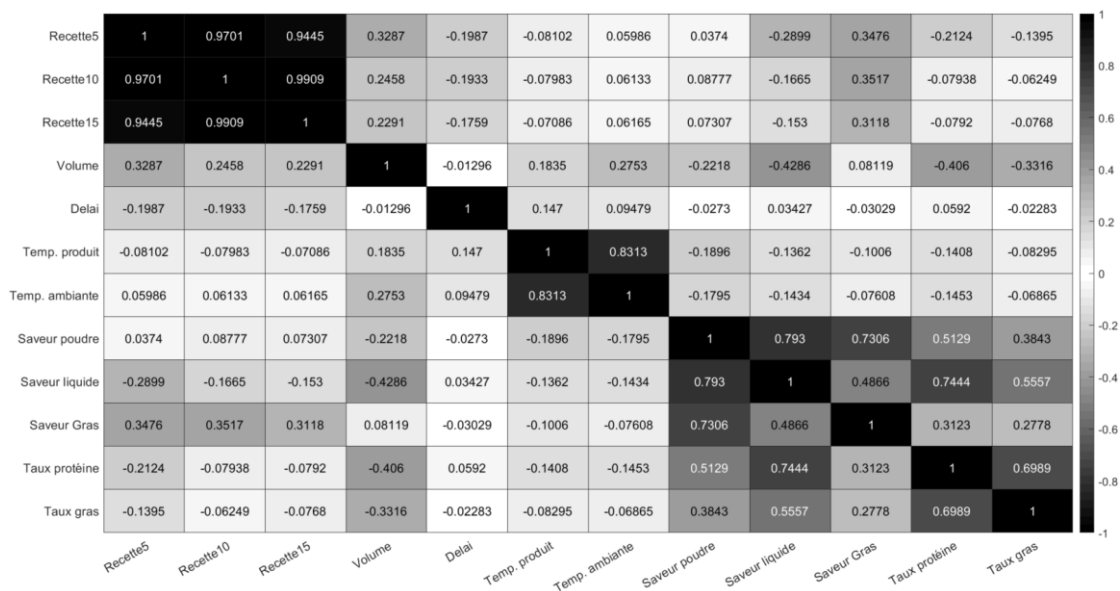


Figure 5.1 : Matrice de corrélation du jeu de données

5.1.3. Facteurs influents

Dès les premières analyses effectuées dans cette étude, il a été observé que la variable dépendante étudiée pouvait varier en fonction du modèle de calibration en place dans le capteur NIR. Le jeu de donnée utilisé était donc compartimenté en blocs représentant les différentes itérations de développement du modèle de calibration. Proposant des performances différentes, chacun de ces modèles pouvait provoquer des écarts majeurs et ainsi masquer l'effet des autres variables indépendantes de l'étude. De plus, ce facteur n'était pas nécessairement utile à l'étude, car il apporte peu de bénéfice scientifique. En effet, c'est plutôt une évidence mathématique que si un modèle n'est pas adapté, il produira un écart considérable. Il est d'ailleurs le premier facteur suspecté dès qu'une calibration produit de faibles performances. Heureusement, l'étude a pu se poursuivre en utilisant l'estimation hors ligne à partir d'un unique modèle. Ceci a été possible grâce à la présence du spectre complet d'absorbance des échantillons dans le jeu de données.

Le développement d'une forêt aléatoire a permis d'identifier deux facteurs d'influence se distinguant clairement des autres. En effet, en utilisant la fonction d'identification de

l'importance des variables, la dimension du produit ainsi que le taux de protéine ont été identifiés comme facteurs exerçant une haute influence sur l'écart de mesure NIR de l'humidité. Ce constat a été renforcé par l'étude des covariances des variables continues du jeu de données. Cette analyse a contribué à identifier plus précisément que ces deux facteurs étaient négativement corrélés avec l'écart de mesure, c'est-à-dire que l'augmentation de ces facteurs provoquait une sous-estimation de l'humidité du produit.

Les analyses par méthode ANOVA ont pu apporter plus de détails en identifiant les groupes responsables de la variance observée. Il a été établi que les échantillons ayant un taux de protéine de 40 % et plus étaient principalement responsables des écarts négatifs de mesure d'humidité. Ce groupe, représentant les produits de nourriture pour chat, a influencé les résultats de plusieurs autres analyses de variance effectuées dans l'étude. En effet, ces produits représentaient souvent un groupe précis lors de l'analyse de l'effet de la dimension du produit, de la recette ainsi que de l'ajout de saveurs. Il a fallu retirer ce groupe de ces analyses afin d'éviter un biais dans les résultats causés par l'effet de cette variable confondante.

L'ANOVA effectuée sur les dimensions du produit a démontré que l'écart devenait de plus en plus négatif à mesure que le volume des croquettes augmentait. Ce phénomène peut être expliqué par la difficulté de la lumière à pénétrer suffisamment les échantillons plus épais, produisant ainsi une lecture en surface peu représentative du cœur de la croquette. Cependant, en comparant les dimensions des produits fournis à la section 3.1.1.2 avec les distances de pénétration énoncées dans la section 2.4.6, il devient peu probable que ce soit le cas. L'explication pourrait plutôt provenir des résultats de l'expérimentation effectuée sur la quantité d'échantillons. Celle-ci démontre que les produits de plus grandes dimensions souffrent d'une variabilité accrue des mesures de référence. Cette variabilité est également amplifiée par l'utilisation de sous-échantillons de 5 g. Ainsi, la plus grande partie de l'écart relié au volume du produit serait attribuable à la variabilité induite par les techniques de mesure de référence plutôt que par l'effet optique.

Cette même expérimentation a apporté plus de détails sur la quantification de l'erreur possible due à l'échantillonnage tardif sur un produit conservé dans un sac scellé. Il a été observé que les

mesures d'humidité effectuées sur des échantillons de plus grosse taille souffraient d'une instabilité majeure dans le temps affectant ainsi la mesure de référence. Cette instabilité semble diminuer avec la dimension des croquettes. De plus, il a été prouvé que l'état d'équilibre de ces échantillons pouvait être affecté par les conditions ambiantes, car la diffusion de la vapeur d'eau au travers du sac utilisé pouvait influencer l'état d'équilibre lors d'entreposage à long terme. Ces phénomènes non linéaires expliquent pourquoi l'ANOVA effectuée sur le délai de mesure n'a pas réussi à démontrer l'existence d'un effet significatif. En effet, l'influence de délai de mesure est répartie sur une longue période et ne cible pas une journée de stockage en particulier. Les analyses par forêt aléatoire et par covariance, moins sensibles à ces effets, ont pu faire ressortir un certain lien causal.

En ce qui concerne l'effet de la température du produit et de la température ambiante, les résultats obtenus par le développement d'une forêt aléatoire et par l'analyse de covariance suggèrent un effet non négligeable de ces facteurs sur la mesure NIR d'humidité. Les analyses par ANOVA ont démontré que les températures élevées pouvaient particulièrement causer une estimation biaisée de l'humidité en sous-estimant sa valeur. Considérant que la calibration du capteur utilise la méthode EPO dédiée à limiter l'effet de la température du produit sur le spectre NIR, il est attendu que ces facteurs ne soient pas parmi les plus influents, car les effets de ceux-ci sont normalement compensés lors de l'étape de prétraitement des spectres. Cependant, les limites de cette méthode restent inconnues à ce jour et peu d'information est partagée par le fournisseur.

Le phénomène d'équilibrage thermique pourrait alors expliquer l'écart observé. Lors de la collecte des échantillons, certains ont été mesurés par NIR alors qu'ils étaient à des températures extrêmes. Toutefois, avant d'être analysés par la méthode de référence en laboratoire, ces échantillons ont eu le temps de se stabiliser à température ambiante normale. Cette différence dans les conditions de mesure entre la méthode NIR et la méthode de référence crée un déséquilibre important, puisque la température a un effet bien documenté sur la teneur en eau des matériaux organiques. En effet, à température plus élevée, une partie de l'eau libre peut s'évaporer ou se redistribuer, modifiant ainsi l'humidité mesurable par la méthode de

référence. Par conséquent, un écart entre les mesures obtenues par NIR et celles produites par méthode de référence peut survenir, non pas à cause d'une étape de prétraitement défailante, mais en raison de la différence d'état physique du produit au moment de chaque mesure.

5.1.4. Facteurs non influents

La lumière environnante, le spectre du convoyeur et la température interne du spectromètre sont tous des facteurs n'exerçant aucune influence dans l'application. Ces trois facteurs font l'objet de techniques efficaces pour en contrôler les effets potentiels. En effet, l'utilisation de méthodes de prétraitement éprouvées ainsi que la mise en place de techniques de détection de matière sur le convoyeur produisent des lectures de qualité. De plus, la présence d'un système de refroidissement thermoélectrique assure une stabilité des mesures et un haut rapport signal sur bruit.

Les résultats d'analyses effectuées sur la forme du produit ont conclu que ce facteur n'a pas d'influence sur la mesure NIR d'humidité. Cependant, la représentativité limitée de certaines formes peut avoir influencé les résultats. Afin d'obtenir des résultats plus précis, un même produit aurait dû être mesuré sous différentes formes tout en contrôlant la dimension, car celle-ci est considérée comme influente. Ce type d'étude n'a pas été effectué en raison des contraintes opérationnelles du partenaire industriel. En effet, ceci aurait monopolisé la ligne de production et engendré des pertes de matière.

En ce qui concerne le taux de gras, les recettes des produits et les ajouts de saveurs externes, les analyses ont permis de conclure qu'ils n'exercent pas d'effet significatif sur l'écart de mesure NIR d'humidité. En effet, après avoir retiré l'effet confondant du taux de protéine du jeu de données, les analyses ANOVA ont confirmé l'absence d'effet significatif de ces facteurs.

5.2. Discussions

5.2.1. Retour sur les objectifs de recherche

L'objectif principal de cette étude était de contribuer à l'optimisation d'une ligne de production de nourriture pour animaux en réduisant la variabilité de la qualité du produit fini. S'inscrivant dans le cadre plus large d'un programme de recherche, les activités ont été concentrées autour d'un élément critique à la gestion de la qualité, soit l'implantation d'une mesure en temps réel des paramètres critiques de qualité par spectroscopie NIR. Cependant, la calibration insatisfaisante du capteur utilisée pour l'estimation de l'humidité du produit limitait la portée des interventions envisagées dans le programme de recherche. En effet, face à un problème de fiabilité des données relatives à l'humidité du produit, le contrôle et l'optimisation du procédé étaient impossibles. Il devenait donc nécessaire d'identifier les facteurs influençant ce type de mesure spectroscopique et de quantifier leur effet sur l'écart de mesure.

Bénéficiant de plusieurs décennies de recherche, la spectroscopie proche infrarouge fait l'objet de plusieurs articles scientifiques, lesquels ont étudié le développement de calibrations dans le secteur agroalimentaire. Au travers de ces études, les auteurs ont observé un grand nombre de facteurs pouvant influencer la précision de la calibration et des mesures NIR. Ainsi, en fonction des données d'entrée identifiées dans un nombre représentatif d'articles, il a été possible de dresser une liste préliminaire de facteurs potentiels. Cette liste a pu être complétée par les observations sur le site ainsi que la consultation d'experts en fabrication et en calibration de capteurs NIR. Elle a ensuite été filtrée pour conserver les facteurs pertinents et prioritaires pour l'application spécifique.

Une fois les facteurs identifiés, des analyses approfondies ont été entreprises afin de quantifier leur effet sur l'écart de mesure NIR de l'humidité. Dépendamment du type de donnée relative au facteur, les analyses ont consisté en une expérimentation, une analyse par méthodes statistiques, une analyse de causalité ou une combinaison de celles-ci.

Les résultats de ces analyses ont mené à la conclusion que les facteurs principaux responsables des plus grands écarts de mesure NIR de l'humidité étaient le modèle de calibration, la

dimension du produit, la teneur en protéine, la température du produit, la température ambiante, la quantité de matière utilisée pour les mesures de référence, le délai de mesure en laboratoire ainsi que la qualité des sacs scellés utilisés pour l'emballage des échantillons.

5.2.2. Comparaison avec la littérature

La revue de la littérature menée dans le cadre de cette étude visait à identifier les facteurs d'influence potentiels d'une mesure spectroscopique de façon générale dans le but de répertorier le plus grand nombre d'éléments possibles. Bien qu'un nombre significatif de publications ait été consulté, peu d'entre elles portaient spécifiquement sur la nourriture pour animaux ou sur des produits comparables. Il est donc plausible que les travaux réalisés dans cette étude aient permis de mettre en lumière de nouveaux facteurs ou encore des effets particuliers, jusque-là peu documentés, influençant la mesure de l'humidité par spectroscopie proche infrarouge.

De manière générale, les articles consultés se limitent à énumérer les facteurs d'influence sans toutefois quantifier leur impact réel sur la précision des mesures. À ce titre, la présente étude se distingue en cherchant à mesurer explicitement l'ampleur de ces effets ou à décrire leurs seuils de manifestation. Il a ainsi été constaté que, malgré l'utilisation d'une méthode de prétraitement visant à atténuer l'effet de la température du produit, la mesure de l'humidité présente une variabilité accrue lorsque la température du produit s'écarte de la plage normale d'opération. Un phénomène similaire a été observé pour la température ambiante : les échantillons mesurés durant les périodes les plus chaudes de l'année présentent également une variabilité plus élevée. Ces informations plus précises concernant l'effet de certains facteurs semblent peu accessibles dans la littérature.

Par ailleurs, certains facteurs, tels que le taux de protéine, qui n'avaient pas été identifiés initialement dans la revue de la littérature, ont été intégrés à l'analyse et ont démontré une influence significative. En effet, les résultats obtenus ont révélé qu'un taux de protéine élevé pouvait engendrer des écarts significatifs dans les mesures NIR, un phénomène non abordé dans les articles étudiés, possiblement en raison des plages de concentration plus faibles dans les produits ciblés par ces recherches.

L'étude a également mis en évidence l'importance cruciale d'un contrôle rigoureux de la chaîne d'échantillonnage, en particulier lorsqu'il s'agit de mesurer un produit qui n'a pas atteint son équilibre hygroscopique. Cet enjeu, bien qu'occasionnellement mentionné dans la littérature, est rarement discuté en profondeur et les recommandations restent généralement superficielles. Les essais expérimentaux menés ici ont démontré que des aspects tels que les conditions de stockage des échantillons et la masse utilisée pour les mesures de référence ont une influence déterminante sur la fiabilité des résultats. De plus, la dimension physique du produit s'est révélée être un facteur significatif de variabilité, un aspect peu ou pas considéré dans les travaux antérieurs. Ainsi, les observations issues de cette étude contribuent à enrichir la compréhension des facteurs influents dans un contexte industriel réel et complexe.

Finalement, plusieurs des facteurs identifiés dans la littérature sont déjà largement pris en charge par des technologies éprouvées ou par des méthodes mathématiques robustes. C'est notamment le cas de la lumière ambiante, du spectre émis par le convoyeur ou encore de la température interne du spectromètre, tous étudiés dans le cadre de cette recherche. Les analyses menées sur ces éléments visaient à valider la fiabilité des solutions existantes, ce qui a été confirmé par les résultats obtenus.

5.2.3. Pistes d'amélioration futures

Bien que les résultats obtenus dans cette étude aient permis de mieux comprendre les facteurs influençant la mesure NIR de l'humidité, certaines limites méthodologiques et scientifiques subsistent. Ainsi, plusieurs pistes de recherche sont envisageables afin d'approfondir les connaissances acquises et d'optimiser davantage les outils et méthodes développés.

5.2.3.1. Améliorations méthodologiques

Les travaux réalisés dans le cadre de cette étude reposent majoritairement sur l'analyse de données issues d'un environnement de production réel, sans qu'un plan d'expérience contrôlé n'ait été mis en place initialement. Bien que cette approche présente l'avantage de respecter les contraintes industrielles, elle limite la capacité à isoler l'effet individuel de certains facteurs et génère plusieurs incertitudes quant à la validité des conclusions. À cet égard, une piste

d'amélioration consiste à développer un plan d'expérience structuré permettant d'étudier les interactions possibles entre les facteurs, de mieux contrôler les conditions expérimentales et d'assurer une représentativité statistique et une étendue suffisante pour tous les groupes analysés. Un tel plan d'expérience peut être développé de façon plus efficace en utilisant les observations issues de l'étude actuelle.

De plus, la gestion des valeurs aberrantes pourrait être optimisée par l'intégration d'un protocole d'identification rigoureux visant à documenter leur origine avant leur exclusion. L'interprétation des résultats gagnerait ainsi en robustesse, en évitant de retirer des données potentiellement révélatrices d'un effet réel. La méthode de gestion des valeurs aberrantes est clairement décrite dans la norme ISO 12099 (ISO, 2017).

Enfin, une attention particulière pourrait être portée à la réduction de la colinéarité entre les variables. L'utilisation de techniques de sélection de variables, ou la conception d'un jeu de données minimisant les effets confondants, permettrait de mieux évaluer l'influence indépendante de chaque facteur. Une analyse rigoureuse de la colinéarité aurait également comme avantage potentiel de réduire la quantité de données à recueillir et à analyser.

5.2.3.2. Améliorations techniques et technologiques

Sur le plan technologique, plusieurs améliorations peuvent être envisagées pour accroître la stabilité et la fiabilité des mesures spectroscopiques. Par exemple, l'intégration de capteurs environnementaux pourrait enrichir le modèle de calibration en permettant une compensation active de certaines fluctuations externes. Un capteur de température du produit en fin de ligne aurait également comme avantage d'informer le personnel si la température du produit est en dehors de la plage normale d'opération couverte par le modèle. D'ailleurs, un meilleur contrôle de la température du produit par le refroidisseur contribuerait à diminuer la variance observée sur ce facteur.

Concernant le modèle de calibration, la méthode populaire consiste à utiliser des méthodes de régressions linéaires afin de produire une estimation. Or, plusieurs effets de seuil et non linéarité ont été découverts dans cette étude confirmant certaines limites de ce genre de modèle pour

l'application. Le développement de modèle basé sur les réseaux de neurones artificiels ou la mise en œuvre de méthodes de correction en ligne plus sophistiquées, telles que les modèles adaptatifs ou les algorithmes d'apprentissage automatique, pourrait améliorer la précision des prédictions en temps réel dans ces cas. En plus de gérer les effets non linéaires de certaines variables, de telles approches permettraient de tenir compte des variations lentes dans les conditions d'opération et les propriétés des produits, sans nécessiter une recalibration manuelle fréquente. D'ailleurs, le modèle utilisé actuellement basé sur une régression d'apprentissage nécessite beaucoup de temps et d'énergie avant d'obtenir des performances acceptables.

Une amélioration significative pourrait être apportée à la chaîne d'échantillonnage en standardisant les pratiques liées à la préparation et au stockage des échantillons. Premièrement, le délai de stockage entre la mesure NIR et la mesure en laboratoire devrait être mieux contrôlé afin de réduire les biais introduits par l'équilibrage d'humidité de l'échantillon avec l'environnement. L'utilisation de sacs scellés à faible perméabilité à la vapeur d'eau devraient être utilisés dans le but de figer le plus possible l'état de l'échantillon au moment de l'échantillonnage. D'une autre part, afin d'éliminer le plus possible l'effet du délai de stockage, une mesure de référence devrait être effectuée directement après l'échantillonnage en utilisant une masse représentative. Dans cette étude, il a été prouvé qu'une masse de 230 g produisant une variance beaucoup plus faible qu'une masse de 5 g. Puisque la dimension du produit provoque une variance, il serait également utile d'étudier plus en détails l'effet de la masse sur la variabilité des mesures et de proposer une valeur optimale.

5.2.3.3. Généralisation des résultats

Les résultats obtenus dans cette étude reposent sur un contexte industriel, un type de produit, ainsi qu'un environnement d'acquisition de données spécifiques. Afin d'évaluer la robustesse des conclusions tirées et leur applicabilité à d'autres contextes, il serait pertinent de reproduire les analyses dans d'autres usines, avec d'autres types de produits (formules, formats, taux d'humidité, etc.), ou en utilisant d'autres capteurs NIR. Cette démarche permettrait de valider si les conclusions peuvent être généralisées et de distinguer les phénomènes universels des effets

liés aux conditions particulières de l'étude actuelle. Une telle approche contribuerait également à bâtir des recommandations normalisées pour l'intégration de la spectroscopie NIR en industrie.

5.2.3.4. Approfondissement des connaissances scientifiques

Certains phénomènes identifiés dans cette étude, bien qu'observés de façon claire, mériteraient d'être approfondis afin d'en comprendre les mécanismes fondamentaux. Par exemple, l'effet significatif de taux élevés de protéines sur l'erreur de prédiction NIR a été constatée, mais sans en expliquer le phénomène. Une étude plus ciblée, intégrant des essais contrôlés et des analyses détaillées du spectre, pourrait permettre de mieux comprendre les interactions spectrales en jeu. La même stratégie pourrait être appliquée à l'analyse de l'effet de la température sur le spectre NIR des échantillons. Ces pistes permettraient non seulement de bonifier les connaissances fondamentales sur le phénomène, mais également d'enrichir les modèles de calibration et les stratégies de correction.

En ce qui concerne l'effet de la dimension du produit, de plus amples connaissances du phénomène de transport des molécules d'eau à l'intérieur d'une matrice organique complexe pourrait contribuer à améliorer à la fois la calibration et les procédures de gestion des échantillons. Ainsi, afin de mieux comprendre ce phénomène, des simulations et des essais expérimentaux contrôlés couvrant l'étendue des types de produits fabriqués devraient être entrepris. De plus, une mesure plus précise de la distance de pénétration aurait comme avantage de préciser les limites de la technologie face à l'application spécifique.

CHAPITRE 6 - CONCLUSION

Ce mémoire s'inscrit dans le cadre d'une démarche visant à mieux comprendre les facteurs pouvant influencer la précision des mesures d'humidité effectuées par spectroscopie proche infrarouge en direct afin d'expliquer les performances insatisfaisantes du modèle de calibration en place. Les activités effectuées dans cette étude font partie des premières étapes d'un programme de recherche complet dédié à optimiser une ligne de production de nourriture animale. L'objectif principal de cette étude était d'identifier les variables responsables de l'écart observé entre les mesures NIR et les valeurs de référence obtenues en laboratoire ainsi que d'en quantifier leur effet.

Basé sur un jeu de données obtenu expérimentalement, une méthodologie spécifique impliquant différentes techniques d'analyses a été mise en place afin d'éviter le déploiement d'un plan d'expérience complexe. Ainsi, l'utilisation de forêts aléatoires, l'analyse de covariance ainsi que l'analyse de variance (ANOVA) ont été utilisées de manière combinatoire dans le but de produire des résultats cohérents et adaptés au jeu de données.

Ces différentes analyses ont permis d'identifier plusieurs facteurs influents, tels que la dimension du produit, le taux de protéine, la température du produit, la température ambiante ou encore certains aspects liés à la chaîne d'échantillonnage et de mesure de référence. L'étude a d'ailleurs mis en évidence certaines zones de vulnérabilité précises provoquées par ces facteurs. À l'inverse, d'autres variables, bien que citées dans la littérature, se sont avérées négligeables dans le contexte spécifique de cette application. C'est le cas pour la lumière environnante, le spectre du convoyeur et la température interne du spectromètre qui peuvent être compensés par les méthodes de prétraitement traditionnelles. De leur côté, les effets de la forme du produit, du taux de gras, des recettes des produits et des ajouts de saveurs externes n'ont pas révélé d'effet significatifs.

En plus de contribuer à une meilleure compréhension des enjeux liés à la mesure NIR en contexte industriel, ce travail propose également une base méthodologique adaptable à d'autres environnements de production. Il souligne l'importance d'un contrôle rigoureux de la chaîne d'échantillonnage et offre des pistes concrètes d'amélioration pour les entreprises souhaitant

fiabiliser de telles mesures en ligne. Les conclusions tirées sur les effets non linéaire de certains facteurs mettent en lumière les limitations des modèles de calibration à base de régression linéaire et ouvrent la porte au développement de modèles mieux adaptés aux environnements industriels. En effet, avec le développement rapide du matériel et du logiciel dans le secteur de l'intelligence artificielle, de nouvelles options émergent et rendent possible le développement de modèles de calibration plus performants.

La stratégie utilisée dans cette étude a été orientée vers l'utilisation d'un jeu de données existant dans lequel la représentativité statistique des données et l'étendue des variables n'étaient pas contrôlée. Bien que plusieurs facteurs aient été étudiés, certains effets potentiels n'ont pu être explorés en profondeur en raison de contraintes de temps, de disponibilité la ligne de fabrication, de matériel ou de qualité des données. En effet, certaines analyses ont dû être réalisées en retirant des groupes peu représentés, ce qui pourrait avoir influencé la sensibilité statistique de certaines conclusions. Cependant, en utilisant une méthodologie rigoureuse et orientée vers l'utilisation de plusieurs méthodes d'analyses, les résultats obtenus ont contribué à l'identification des facteurs d'influence principaux. Ces résultats posent les bases de futurs travaux pouvant approfondir la compréhension de certains phénomènes ou généraliser les conclusions à d'autres types de produits et de procédés.

Les résultats obtenus dans le cadre de cette étude ouvrent la voie à plusieurs pistes d'amélioration et de recherche. Sur le plan méthodologique, une attention particulière devrait être portée à la chaîne d'échantillonnage, notamment en réduisant ou en éliminant le délai entre les mesures NIR et les mesures de référence, en augmentant la masse utilisée pour l'analyse en laboratoire, et en assurant des conditions de stockage optimales, par exemple en utilisant des sacs à faible perméabilité à la vapeur d'eau.

En ce qui concerne les approfondissements des connaissances scientifiques, des travaux de recherche plus poussés sur l'influence de certains facteurs contribueraient à une meilleure compréhension des mécanismes à l'origine des écarts de mesure en spectroscopie NIR. Des études ciblées pourraient notamment permettre de quantifier plus précisément l'effet de la température du produit et de sa teneur en protéines sur les spectres mesurés. Par ailleurs, la

réalisation de simulations et d'essais expérimentaux visant à étudier le phénomène de diffusion de l'eau dans ce type de produit offrirait un éclairage précieux sur la dynamique de la source de variabilité observée lors de l'équilibrage de produits de grande dimension.

En somme, les travaux présentés dans ce mémoire ont permis de mieux comprendre les facteurs influençant la précision des mesures spectroscopiques NIR dans un contexte industriel appliqué à la nourriture pour animaux. Bien que certaines limitations aient été rencontrées, les méthodologies développées, les résultats obtenus et les pistes d'amélioration dégagées constituent une base solide pour la poursuite des recherches dans ce domaine. Ce projet a également mis en lumière l'importance d'une approche rigoureuse, tant sur le plan expérimental qu'analytique, pour faire progresser les connaissances scientifiques et soutenir l'innovation technologique en milieu industriel. Il est souhaité que ce travail contribue à inspirer de nouvelles recherches, à favoriser l'implantation de solutions de mesure plus robustes, précises et adaptées aux réalités d'un environnement industriel, et ultimement, à accélérer le développement de calibrations en direct plus performantes et fiables.

LISTE DES RÉFÉRENCES

- Alomar, D., Hodgkinson, S., Abarzúa, D., Fuchslocher, R., Alvarado, C., et Rosales, E. (2006). Nutritional evaluation of commercial dry dog foods by near infrared reflectance spectroscopy. *Journal of Animal Physiology and Animal Nutrition*, 90(5-6), 223-229.
- American Society for Testing and Materials. (2004). *Standard Test Method for Moisture Absorption Properties and Equilibrium Conditioning of Polymer Matrix Composite Materials*. ASTM, D 5229/D 5229M – 92.
- Baeten, V., Pierna, J. A. F., Lecler, B., Abbas, O., Vincke, D., Minet, O., Vermeulen, P., et Dardenne, P. (2016). Near Infrared Spectroscopy for Food and Feed : A Mature Technique. *NIR News*, 27(1), 4-6.
- Bakeev, K. A. (2010). *Process Analytical Technology : Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries* (2^e éd.). Chichester : John Wiley and Sons.
- Barnes, R. J., Dhanoa, M. S., et Lister, S. J. (1989). Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43(5), 772-777.
- Beck, E. A., Lefcourt, A. M., Lo, Y. M., et Kim, M. S. (2015). Use of a portable fluorescence imaging device to facilitate cleaning of deli slicers. *Food Control*, 51, 256-262.
- Blanco, M., et Villarroya, I. (2002). NIR spectroscopy : A rapid-response analytical tool. *TrAC - Trends in Analytical Chemistry*, 21(4), 240-250.
- Blazquez, C., Downey, G., O'Donnell, C., O'Callaghan, D., et Howard, V. (2004). Prediction of moisture, fat and inorganic salts in processed cheese by near infrared reflectance spectroscopy and multivariate data analysis. *Journal of Near Infrared Spectroscopy*, 12(3), 149-157.
- Bock, J. E., et Connelly, R. K. (2008). Innovative uses of near-infrared spectroscopy in food processing. *Journal of Food Science*, 73(7), 91-98.
- Boysworth, M. K., et Booksh, K. S. (2007). Aspects of multivariate calibration applied to near-infrared spectroscopy. In D. A. Burns et E. W. Ciurczak (Éd.), *Handbook of Near-Infrared Analysis* (3^e éd., pp. 207-229). Boca Raton : CRC Press.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
- Callis, J. B., Illman, D. L., et Kowalski, B. R. (1987). Process Analytical. *Analytical Chemistry*, 59(9), 624-637.

- Camino-Sánchez, F. J., Delgado-Moreno, B., et Navarro-Fuentes, P. (2015). Quick Development of a NIRS Method for the Analysis of Moisture, Protein, and Fat in Nutritional Clinical Products. *Food Analytical Methods*, 8(7), 1758-1770.
- Castrillo, C., Baucells, M., Vicente, F., Muñoz, F., et Andueza, D. (2005). Energy evaluation of extruded compound foods for dogs by near-infrared spectroscopy. *Journal of Animal Physiology and Animal Nutrition*, 89(3-6), 194-198.
- Cen, H., et He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science and Technology*, 18(2), 72-83.
- Chew, W., et Sharratt, P. (2010). Trends in process analytical technology. *Analytical Methods*, 2(10), 1412-1438.
- Coates, J. P. (2010). Infrared Spectroscopy for Process Analytical Applications. In K. A. Bakeev (Éd.), *Process Analytical Technology : Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries* (2^e éd., pp. 157-194). Chichester : John Wiley and Sons.
- Cullen, P. J., O'Donnell, C. P., et Fagan, C. C. (2014). Benefits and challenges of adopting pat for the food industry. In P. J. Cullen, C. P. O'Donnell, et C. C. Fagan (Éd.), *Process Analytical Technology for the Food Industry. Food Engineering Series* (pp. 1-5). New York : Springer.
- De Marchi, M., Righi, F., Meneghesso, M., Manfrin, D., et Ricci, R. (2018). Prediction of chemical composition and peroxide value in unground pet foods by near-infrared spectroscopy. *Journal of Animal Physiology and Animal Nutrition*, 102(1), 337-342.
- European Medicines Agency (2003). *Quality by design*. European Medicines Agency. Consulté le 29 novembre 2022, tiré de <https://www.ema.europa.eu/en/human-regulatory/research-development/quality-design>
- Everard, C. D., Kim, M. S., et Lee, H. (2014). A comparison of hyperspectral reflectance and fluorescence imaging techniques for detection of contaminants on spinach leaves. *Journal of Food Engineering*, 143, 139-145.
- Fernández de Ahumada, E., Garrido-Varo, A., Guerrero, J. E., Pérez Marín, D., et Gómez Cabrera, A. (2005). Comparación de sistemas discontinuos versus continuos para el análisis nirs de piensos compuestos. *Xi jornadas sobre produccion animal*, 2, 635-637.
- Fernández-Ahumada, E., Guerrero-Ginel, J. E., Pérez-Marín, D., et Garrido-Varo, A. (2008). Near Infrared Spectroscopy for Control of the Compound-Feed Manufacturing Process : Mixing Stage. *Journal of Near Infrared Spectroscopy*, 16(3), 285-290.

- Fernández-Cabanás, V. M., Garrido-Varo, A., Pérez-Marín, D., et Dardenne, P. (2006). Evaluation of Pretreatment Strategies for Near-Infrared Spectroscopy Calibration Development of Unground and Ground Compound Feedingstuffs. *Applied Spectroscopy*, 60(1), 17-23.
- Figura, L. O., et Teixeira, A. A. (2023). *Food Physics : Physical Properties - Measurement and Applications* (2^e éd.). Gewerbestrasse : Springer International Publishing.
- Food and Drug Administration (2004). *Guidance for industry : PAT - A framework for innovative pharmaceutical development, manufacturing and quality assurance*. U.S. Department of Health and Human Services.
- Garrido, A., Sánchez, M. T., Cano, G., Pérez, D., et López, C. (2001). Prediction of neutral and acid detergent fiber content of green asparagus stored under refrigeration and modified atmosphere conditions by near-infrared reflectance spectroscopy. *Journal of Food Quality*, 24(6), 539-550.
- Goi, A., Manuelian, C. L., Currò, S., et De Marchi, M. (2019). Prediction of mineral composition in commercial extruded dry dog food by near-infrared reflectance spectroscopy. *Animals*, 9(9), 640-651.
- Goi, A., Manuelian, C. L., Righi, F., et De Marchi, M. (2020). At-line prediction of gelatinized starch and fiber fractions in extruded dry dog food using different near-infrared spectroscopy technologies. *Animals*, 10(5), 862-873.
- Gou, P., Santos-Garcés, E., Høy, M., Wold, J. P., Liland, K. H., et Fulladosa, E. (2013). Feasibility of NIR interactance hyperspectral imaging for on-line measurement of crude composition in vacuum packed dry-cured ham slices. *Meat Science*, 95(2), 250-255.
- Granato, D., de Araújo Calado, V. M., et Jarvis, B. (2014). Observations on the use of statistical methods in Food Science and Technology. *Food Research International*, 55, 137-149.
- Grassi, S., et Alamprese, C. (2018). Advances in NIR spectroscopy applied to process analytical technology in food industries. *Current Opinion in Food Science*, 22, 17-21.
- Grote, B., Zense, T., et Hitzmann, B. (2014). 2D-fluorescence and multivariate data analysis for monitoring of sourdough fermentation process. *Food Control*, 38(1), 8-18.
- He, S., Xie, W., Zhang, W., Zhang, L., Wang, Y., Liu, X., Liu, Y., et Du, C. (2015). Multivariate qualitative analysis of banned additives in food safety using surface enhanced Raman scattering spectroscopy. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 137, 1092-1099.

- Hervera, M., Castrillo, C., Albanell, E., et Baucells, M. D. (2012). Use of near-infrared spectroscopy to predict energy content of commercial dog food. *Journal of Animal Science*, 90(12), 4401-4407.
- Hitzmann, B., Hauselmann, R., Niemoeller, A., Sangi, D., Traenkle, J., et Glassey, J. (2015). Process analytical technologies in food industry - challenges and benefits : A status report and recommendations. *Biotechnology Journal*, 10(8), 1095-1100.
- Holroyd, S. E., Prescott, B., et McLean, A. (2013). The use of in-and on-line near infrared spectroscopy for milk powder measurement. *Journal of Near Infrared Spectroscopy*, 21(5), 441-443.
- Honigs, D., et Åberg, P. (2012). Honigs regression, LOCAL and PLS: near infrared of dry and fresh forage. *Proceedings of the 15th International Conference on Near Infrared Spectroscopy* (pp. 313-316). Cape Town : M. Manley.
- Hoyer, M., Schumann, R., Wüst, E., et Premier, G. C. (2005). Feedback control system design for a fresh cheese separator. *Proceedings of 16th IFAC World Congress* (Vol. 16, pp. 79-84). Prague : IFAC.
- Huang, H., Yu, H., Xu, H., et Ying, Y. (2008). Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages : A review. *Journal of Food Engineering*, 87(3), 303-313.
- Huang, M., Kim, M. S., Chao, K., Qin, J., Mo, C., Esquerre, C., Delwiche, S., et Zhu, Q. (2016). Penetration Depth Measurement of Near-Infrared Hyperspectral Imaging Light for Milk Powder. *Sensors*, 16(4), 441-552.
- Ilaslan, K., Boyaci, I. H., et Topcu, A. (2015). Rapid analysis of glucose, fructose and sucrose contents of commercial soft drinks using Raman spectroscopy. *Food Control*, 48, 56-61.
- International Conference on Harmonisation. (2005). *Q9 Quality risk management*. International Conference on Harmonisation, Q9.
- International Conference on Harmonisation. (2006). *Q8 Pharmaceutical development—Scientific guideline*. International Conference on Harmonisation, Q8.
- International Conference on Harmonisation. (2008). *Q10 Pharmaceutical quality system*. International Conference on Harmonisation, Q10.
- Irudayaraj, J., et Reh, C. (2008). *Nondestructive Testing of Food Quality*. Ames : Blackwell Publishing Professional.
- Ishikawa, K. (1985). *What is total quality control? : The Japanese way* (6^e éd.). Englewood Cliffs : Prentice-Hall.

- Jerome, R. E., Singh, S. K., et Dwivedi, M. (2019). Process analytical technology for bakery industry : A review. *Journal of Food Process Engineering*, 42(5).
- Jestel, N. L. (2010). Raman spectroscopy. In K. A. Bakeev (Éd.), *Process Analytical Technology : Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries* (2^e éd., pp. 195-244). Chichester : John Wiley and Sons.
- Johnsen, J. L. (2014). Food industry perspectives on the implementation of a pat strategy. In C. P. O'Donnell, C. Fagan, et P. J. Cullen (Éd.), *Process Analytical Technology for the Food Industry. Food Engineering Series* (pp. 269-291). New York : Springer.
- Kumaravelu, C., et Gopal, A. (2015). A review on the applications of Near-Infrared spectrometer and Chemometrics for the agro-food processing industries. *Proceedings - 2015 IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development, TIAR 2015* (Vol. 10, pp. 8-12). Chennai : IEEE.
- LABPLAS (s. d.). *EPL-4575—Fiche technique*. LABPLAS. Consulté le 5 février 2025, tiré de www.labplas.com
- Lee, H., Everard, C. D., Kang, S., Cho, B.-K., Chao, K., Chan, D. E., et Kim, M. S. (2014). Multispectral fluorescence imaging for detection of bovine faeces on Romaine lettuce and baby spinach leaves. *Biosystems Engineering*, 127, 125-134.
- Liau, M. A., Baylor, L. C., et O' Rourke, P. E. (2010). UV-visible Spectroscopy for On-line Analysis. In K. A. Bakeev (Éd.), *Process Analytical Technology : Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries* (2^e éd., pp. 81-106). Chichester : John Wiley and Sons.
- Long, D. S., Engel, R. E., et Siemens, M. C. (2008). Measuring grain protein concentration with in-line near infrared reflectance spectroscopy. *Agronomy Journal*, 100(2), 247-252.
- Maduro Dias, C. S. A. M., Nunes, H. P., Melo, T. M. M. V., Rosa, H. J. D., Silva, C. C. G., et Borba, A. E. S. (2021). Application of Near Infrared Reflectance (NIR) spectroscopy to predict the moisture, protein, and fat content of beef for gourmet hamburger preparation. *Livestock Science*, 254.
- Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science* (Vol. 2, pp. 49-55).
- Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging : Non-destructive analysis of biological materials. *Chemical Society Reviews*, 43(24), 8200-8214.
- Martens, H., Jensen, S. A., et Geladi, P. (1983). Multivariate linearity transformations for near infrared reflectance spectroscopy. *Proceedings - Nordic Symp. Applied Statistics* (pp. 205-234). Stokkland Forlag : Christie O.H.J.

- Maslovaric, M., Jovanović, R., Jankovic, S., et Tolimir, N. (2011). Application of NIR technology in the animal food industry. *Biotechnology in Animal Husbandry*, 27(4), 1811-1817.
- Melfsen, A., Hartung, E., et Haeussermann, A. (2013). Robustness of near-infrared calibration models for the prediction of milk constituents during the milking process. *Journal of Dairy Research*, 80(1), 103-112.
- Misra, J. B., Mathur, R. S., et Bhatt, D. M. (2000). Near-infrared transmittance spectroscopy : A potential tool for non-destructive determination of oil content in groundnuts. *Journal of the Science of Food and Agriculture*, 80(2), 237-240.
- Mohan, L. A., Karunakaran, C., Jayas, D. S., et White, N. D. G. (2005). Classification of bulk cereals using visible and NIR reflectance characteristics. *Canadian Biosystems Engineering*, 47, 7.07-7.14.
- Montes, J. M., Paul, C., Kusterer, B., et Melchinger, A. E. (2006). Near infrared spectroscopy to measure maize grain composition on plot combine harvesters : Evaluation of calibration techniques, mathematical transformations and scatter corrections. *Journal of Near Infrared Spectroscopy*, 14(6), 387-394.
- Niemöller, A., et Behmer, D. (2008). Use of Near Infrared Spectroscopy in the Food Industry. In *Nondestructive Testing of Food Quality* (pp. 67-118). Ames : Blackwell Publishing Professional.
- Norris, K. H., et Ben-Gera, I. (1968). Determination of moisture content in soybeans by direct spectrophotometry. *Israelic Journal of Agricultural Research*, 125-132.
- Novak, P., et Jednacak, T. (2014). The use of vibrational spectroscopy as a tool for In-Line process monitoring. In A. L. Varney (Éd.), *Process Control : Theory, Applications and Challenges* (pp. 69-99). New York : Nova Science Publishers, Inc.
- O'Farrell, M., Wold, J. P., Høy, M., Tschudi, J., et Schulerud, H. (2010). On-line fat content classification of inhomogeneous pork trimmings using multispectral near infrared interactance imaging. *Journal of Near Infrared Spectroscopy*, 18(2), 135-146.
- Ojeda, C. B., et Rojas, F. S. (2009). Process analytical chemistry : Applications of ultraviolet/visible spectrometry in environmental analysis: An overview. *Applied Spectroscopy Reviews*, 44(3), 245-265.
- Organisation internationale de normalisation. (2017). *Aliments des animaux, céréales et produits de mouture des céréales—Lignes directrices pour l'application de la spectrométrie dans le proche infrarouge*. ISO, 12099.
- Ozaki, Y., McClure, W. F., et Christy, A. A. (2006). *Near-Infrared Spectroscopy in Food Science and Technology*. Hoboken : John Wiley & Sons.

- Paquet-Durand, O., Solle, D., Schirmer, M., Becker, T., et Hitzmann, B. (2012). Monitoring baking processes of bread rolls by digital image analysis. *Journal of Food Engineering*, *111*(2), 425-431.
- Paul, C., et Pfitzner, C. (2004). Analytical use of NIR diode array spectrometers on forage harvesters. *Near Infrared Spectroscopy: Proceedings of the 11th International Conference* (pp. 333-338). Chichester : NIR Publications.
- Perez, D., Sanchez, M. T., Cano, G., et Garrido, A. (2002). Prediction of texture in green asparagus by near infrared spectroscopy (NIRS). *Journal of Food Quality*, *25*(4), 277-287.
- Pérez-Marín, D. (2014). Near Infrared Spectroscopy for Increasing Agri-Food Industry Competitiveness by the Improvement of Quality and Process Control : From Research to Real Applications. *NIR News*, *25*(8), 21-29.
- Pérez-Marín, D., Fearn, T., Guerrero, J. E., et Garrido-Varo, A. (2012). Improving NIRS predictions of ingredient composition in compound feedingstuffs using Bayesian non-parametric calibrations. *Chemometrics and Intelligent Laboratory Systems*, *110*(1), 108-112.
- Pérez-Marín, D., Garrido-Varo, A., Guerrero, J. E., Fearn, T., et Davies, A. M. C. (2008). Advanced nonlinear approaches for predicting the ingredient composition in compound feedingstuffs by near-infrared reflection spectroscopy. *Applied Spectroscopy*, *62*(5), 536-541.
- Pérez-Marín, D., Sanchez, M. T., Cano, G., et Garrido, A. (2001). Authentication of green asparagus varieties by near-infrared reflectance spectroscopy. *Journal of Food Science*, *66*(2), 323-327.
- PerkinElmer AB (2014). *Palm oil analysis. Complete lab solutions form upstream to downstream*. PerkinElmer AB. Consulté le 5 février 2025, tiré de <https://www.perkinelmer.com/>
- PerkinElmer AB (2023). *DA 7440 GP/SD - Installation and maintenance manual*. PerkinElmer AB. Consulté le 5 février 2025, tiré de <https://www.perkinelmer.com/>
- Pomerantsev, A. L., et Rodionova, O. Y. (2012). Process analytical technology : A critical view of the chemometricians. *Journal of Chemometrics*, *26*(6), 299-310.
- Pop, Ș. Z., Dracea, R., et Vlădulescu, C. (2018). Comparative study of certification schemes for food safety management systems in the european union context. *Amfiteatru Economic*, *20*(47), 9-29.

- Porep, J. U., Kammerer, D. R., et Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science and Technology*, 46(2), 211-230.
- Prieto, N., Roehe, R., Lavín, P., Batten, G., et Andrés, S. (2009). Application of near infrared reflectance spectroscopy to predict meat and meat products quality : A review. *Meat Science*, 83(2), 175-186.
- Qiu, Z., Feng, L., Li, W., et He, Y. (2005). Application of near infrared spectroscopy in nondestructive quality inspection of agricultural products. *Proceedings of MIPPR 2005 SAR and Multispectral Image Processing* (Vol. 6045). Wuhan : SPIE.
- Rathore, A. S., et Kapoor, G. (2016). Implementation of Quality by Design for processing of food products and biotherapeutics. *Food and Bioproducts Processing*, 99, 231-243.
- Rinnan, Å., Berg, F. van den, et Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, 28(10), 1201-1222.
- Roger, J.-M., et Boulet, J.-C. (2018). A review of orthogonal projections for calibration. *Journal of Chemometrics*, 32(9).
- Roger, J.-M., Chauchard, F., et Bellon-Maurel, V. (2003). EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemometrics and Intelligent Laboratory Systems*, 66(2), 191-204. Scopus.
- Savitzky, Abraham., et Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8), 1627-1639.
- Shenk, J. S., Workman, J. J., Jr., et Westerhaus, M. O. (2007). Application of NIR spectroscopy to agricultural products. In D. A. Burns et E. W. Ciurczak (Éd.), *Handbook of Near-Infrared Analysis* (3^e éd., pp. 347-386). Boca Raton : CRC Press.
- Sikorska, E., Khmelinskii, I., et Sikorski, M. (2019). Fluorescence spectroscopy and imaging instruments for food quality evaluation. In J. Zhong et X. Wang (Éd.), *Evaluation Technologies for Food Quality* (pp. 491-533). Cambridge : Woodhead Publishing.
- Simpson, M. B. (2010). Near—Infrared Spectroscopy for Process Analytical Technology : Theory, Technology and Implementation. In K. A. Bakeev (Éd.), *Process Analytical Technology : Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries* (2^e éd., pp. 107-156). Chichester : John Wiley and Sons.

- Skvaril, J., Kyprianidis, K. G., et Dahlquist, E. (2017). Applications of near-infrared spectroscopy (NIRS) in biomass energy conversion processes : A review. *Applied Spectroscopy Reviews*, 52(8), 675-728.
- Tonolini, M., van den Berg, F. W. J., Skou, P. B., Sørensen, K. M., et Engelsen, S. B. (2023). Near-infrared spectroscopy as a process analytical technology tool for monitoring performance of membrane filtration in a whey protein fractionation process. *Journal of Food Engineering*, 350.
- Tyagi, L., Sharma, G. P., Verma, R. C., Jain, S. K., Murdia, L. K., et Mathur, S. M. (2020). Infrared heating in food processing : An overview. *International Journal of Chemical Studies*, 8(3), 327-336.
- Valeur, B., et Berberan-Santos, M. N. (2012). *Molecular Fluorescence : Principles and Applications*. John Wiley & Sons.
- van den Berg, F., Lyndgaard, C. B., Sørensen, K. M., et Engelsen, S. B. (2013). Process Analytical Technology in the food industry. *Trends in Food Science and Technology*, 31(1), 27-35.
- Wang, Q., Li, Z., Ma, Z., et Liang, L. (2014). Real time monitoring of multiple components in wine fermentation using an on-line auto-calibration Raman spectroscopy. *Sensors and Actuators*, 202, 426-432.
- Wang, W., et Paliwal, J. (2007). Near-infrared spectroscopy and imaging in food quality and safety. *Sensing and Instrumentation for Food Quality and Safety*, 1(4), 193-207.
- Wedding, B. B., Wright, C., Grauf, S., White, R. D., Wedding, B. B., Wright, C., Grauf, S., et White, R. D. (2024). Wavelength variation of the depth of penetration of near infrared radiation in « Hass » avocado fruit. *Technology in Horticulture*, 4(1).
- Welle, R., Greten, W., Müller, T., Weber, G., et Wehrmann, H. (2005). Application of near infrared spectroscopy on-combine in corn grain breeding. *Journal of Near Infrared Spectroscopy*, 13(2), 69-75.
- Wold, J. P., O'Farrell, M., Høy, M., et Tschudi, J. (2011). On-line determination and control of fat content in batches of beef trimmings by NIR imaging spectroscopy. *Meat Science*, 89(3), 317-324.
- Xu, R., Hu, W., Zhou, Y., Zhang, X., Xu, S., Guo, Q., Qi, P., Chen, L., Yang, X., Zhang, F., Liu, L., Qiu, L., et Wang, J. (2020). Use of near-infrared spectroscopy for the rapid evaluation of soybean [Glycine max (L.) Merri.] water soluble protein content. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 224.
- Zachariassen, C. B., Larsen, J., Van Den Berg, F., et Engelsen, S. B. (2005). Use of NIR spectroscopy and chemometrics for on-line process monitoring of ammonia in Low

Methoxylated Amidated pectin production. *Chemometrics and Intelligent Laboratory Systems*, 76(2), 149-161.

Zhang, C., Zhao, C., Huang, W., Wang, Q., Liu, S., Li, J., et Guo, Z. (2017). Automatic detection of defective apples using NIR coded structured light and fast lightness correction. *Journal of Food Engineering*, 203, 69-82.

Zude, M. (2008). *Optical Monitoring of Fresh and Processed Agricultural Crops* (1^{re} éd.). Boca Raton : CRC Press.

ANNEXE I

Tableaux des résultats des tests de Levene

Tableau A1.1 : Test de Levene — modèle de calibration

Group	Count	Mean	Std Dev
1	5	-1.326	1.79357
2	52	-0.915	1.40678
3	19	0.43	1.03988
4	257	-0.5921	1.38233
5	71	-0.6411	1.09814
6	7	-0.2286	0.85285
7	8	-0.8625	1.0336
Pooled	419	-0.602	1.32076
Levene's statistic (absolute)	2.942		
Degrees of freedom	6, 412		
p-value	0.008		

Tableau A1.2 : Test de Levene — taux de protéine

Group	Count	Mean	Std Dev
15	15	0.172	0.69579
20	66	-0.0314	0.60526
25	164	-0.0025	0.70025
30	76	0.0878	0.53567
35	25	-0.0068	0.61332
40	16	-0.2644	0.64405
45	49	-0.3327	0.61649
Pooled	411	-0.0339	0.6401
Levene's statistic (absolute)	1.5199		
Degrees of freedom	6, 404		
p-value	0.1701		

Tableau A1.3 : Test de Levene — taux de gras

Group	Count	Mean	Std Dev
10	46	-0.1522	0.56588
13	149	-0.0014	0.71382
16	141	0.074	0.57774
19	57	-0.2726	0.69697
22	19	-0.0889	0.5895
Pooled	412	-0.034	0.64631
Levene's statistic (absolute)	1.9598		
Degrees of freedom	4, 407		
p-value	0.0998		

Tableau A1.4 : Test de Levene — taux de gras sans variable confondante

Group	Count	Mean	Std Dev
10	44	-0.1391	0.56856
13	149	-0.0014	0.71382
16	136	0.0835	0.57928
19	22	-0.0841	0.67073
Pooled	351	0.009	0.64433
Levene's statistic (absolute)	2.2468		
Degrees of freedom	3, 347		
p-value	0.0826		

Tableau A1.5 : Test de Levene — température du produit

Group	Count	Mean	Std Dev
10	28	-0.0407	0.52664
17	79	0.1215	0.62238
20	171	-0.0757	0.69099
23	67	-0.0487	0.56054
26	38	0.0316	0.75962
29	33	-0.267	0.60406
Pooled	416	-0.0369	0.64889
Levene's statistic (absolute)	0.6061		
Degrees of freedom	5, 410		
p-value	0.6953		

Tableau A1.6 : Test de Levene — température du produit sans variable confondante

Group	Count	Mean	Std Dev
10	24	-0.1129	0.52495
17	67	0.1336	0.65673
20	139	0.0241	0.65827
23	63	-0.04	0.56907
26	36	0.0258	0.78021
29	27	-0.2296	0.56421
Pooled	356	0.0051	0.64201
Levene's statistic (absolute)	0.719		
Degrees of freedom	5, 350		
p-value	0.6095		

Tableau A1.7 : Test de Levene — température ambiante

Group	Count	Mean	Std Dev
18	22	0.0255	0.44944
20	87	0.0715	0.54031
22	197	-0.093	0.70228
24	54	0.0552	0.62538
26	42	-0.0186	0.65117
28	13	-0.4777	0.81168
Pooled	415	-0.0375	0.64855
Levene's statistic (absolute)	2.1616		
Degrees of freedom	5, 409		
p-value	0.0575		

Tableau A1.8 : Test de Levene — volume du produit

Group	Count	Mean	Std Dev
270	121	0.1625	0.61655
893	110	-0.1315	0.55785
2307	50	-0.2202	0.78452
Pooled	281	-0.0207	0.62816
Levene's statistic (absolute)	3.5288		
Degrees of freedom	2, 278		
p-value	0.0307		

Tableau A1.9 : Test de Levene — forme du produit

Group	Count	Mean	Std Dev
Cœur	10	0.06	0.80396
Etoile	16	-0.0088	0.60463
Lotus	42	0.2083	0.62425
standard	348	-0.0706	0.64977
Pooled	416	-0.0369	0.64946
Levene's statistic (absolute)	0.1764		
Degrees of freedom	3, 412		
p-value	0.9124		

Tableau A1.10 : Test de Levene — groupe de produit (5)

Group	Count	Mean	Std Dev
1	40	0.192	0.63292
2	17	0.0382	0.58554
3	86	-0.2048	0.69855
5	267	-0.024	0.63312
Pooled	410	-0.0383	0.64558
Levene's statistic (absolute)	0.5273		
Degrees of freedom	3, 406		
p-value	0.6638		

Tableau A1.11 : Test de Levene — groupe de produit (10)

Group	Count	Mean	Std Dev
1	40	0.192	0.63292
2	13	0.12	0.57823
4	10	0.06	0.80396
5	16	-0.0656	0.78213
6	60	-0.286	0.65195
8	43	0.1084	0.71873
9	140	-0.0591	0.65411
10	84	-0.0332	0.54354
Pooled	406	-0.0364	0.64464
Levene's statistic (absolute)	0.7338		
Degrees of freedom	7, 398		
p-value	0.6434		

Tableau A1.12 : Test de Levene — groupe de produit (15)

Group	Count	Mean	Std Dev
1	22	0.2641	0.7397
2	18	0.1039	0.47764
3	13	0.12	0.57823
5	10	0.06	0.80396
7	16	-0.0656	0.78213
8	60	-0.286	0.65195
10	38	-0.0047	0.65785
12	118	-0.023	0.63716
13	22	-0.2532	0.72343
14	62	-0.0248	0.56646
15	22	-0.0568	0.4846
Pooled	401	-0.0489	0.63742
Levene's statistic (absolute)	1.1245		
Degrees of freedom	10, 390		
p-value	0.3422		

Tableau A1.13 : Test de Levene — saveur poudre

Group	Count	Mean	Std Dev
11P-Poulet pour chat	60	-0.286	0.65195
19-Base 403	90	-0.1011	0.59381
235-Bio B22024	10	0.06	0.80396
289-Poulet NCP	13	-0.1608	0.55554
616-Base 603	101	0.0176	0.70893
97-Bio F21003	78	0.0254	0.56795
Aucune	58	0.1012	0.67644
Pooled	410	-0.0442	0.64404
Levene's statistic (absolute)	0.9187		
Degrees of freedom	6, 403		
p-value	0.4813		

Tableau A1.14 : Test de Levene — saveur liquide

Group	Count	Mean	Std Dev
241-Saumon	68	-0.0056	0.56516
294-Poulet	13	-0.1608	0.55554
Aucune	269	0.0056	0.66523
Poulet pour chat	60	-0.286	0.65195
Pooled	410	-0.0442	0.64476
Levene's statistic (absolute)	0.6353		
Degrees of freedom	3, 406		
p-value	0.5926		

Tableau A1.15 : Test de Levene — saveur gras

Group	Count	Mean	Std Dev
203-Huile Canola	233	-0.0139	0.63919
205-Gras Poulet	84	-0.2286	0.64494
7-Huile poulet	31	-0.0561	0.63479
Aucun	62	0.0977	0.67155
Pooled	410	-0.0442	0.64501
Levene's statistic (absolute)	0.2576		
Degrees of freedom	3, 406		
p-value	0.8559		

Tableau A1.16 : Test de Levene — délai de mesure

Group	Count	Mean	Std Dev
2	13	0.1192	0.79661
3	23	0.0596	0.60308
4	21	-0.2048	0.82137
5	26	-0.0162	0.65897
6	58	0.0233	0.65706
7	61	-0.127	0.6751
8	69	0.0696	0.53069
9	23	0.1778	0.61979
10	47	-0.0568	0.71223
11	26	-0.0762	0.52063
12	20	-0.411	0.68316
13	21	-0.009	0.48647
Pooled	408	-0.0303	0.64203
Levene's statistic (absolute)	1.2073		
Degrees of freedom	11, 396		
p-value	0.2798		

Tableau A1.17 : Test de Levene — humidité DG920

Group	Count	Mean	Std Dev
0	8	8.8863	0.87379
2	8	9.07141	0.17259
4	8	8.1852	0.26059
7	8	8.31301	0.29258
9	8	7.8592	0.25788
11	8	7.90757	0.17365
14	8	8.50201	0.11302
17	8	8.24026	0.15746
23	8	8.33663	0.07335
Pooled	72	8.36684	0.34744
Levene's statistic (absolute)	5.93554		
Degrees of freedom	8, 63		
p-value	0.00001		

Tableau A1.18 : Test de Levene — humidité DG1901

Group	Count	Mean	Std Dev
0	7	9.3303	0.77981
2	8	8.48927	0.33486
4	8	8.95306	0.24308
7	8	8.8916	0.13737
10	8	8.59973	0.19253
13	8	8.41898	0.21467
17	8	8.4926	0.1735
20	8	8.383	0.11925
23	8	8.47967	0.1402
Pooled	71	8.66162	0.31132
Levene's statistic (absolute)	12.5701		
Degrees of freedom	8, 62		
p-value	0		

Tableau A1.19 : Test de Levene — humidité CHT05

Group	Count	Mean	Std Dev
0	8	7.80015	0.43718
2	8	7.26138	0.11067
6	8	7.25584	0.12396
8	8	7.22513	0.12354
10	8	7.1508	0.06122
13	8	7.176	0.0947
17	8	7.09724	0.14679
20	8	7.2212	0.11858
23	8	7.16642	0.15851
Pooled	72	7.26157	0.18483
Levene's statistic (absolute)	8.93493		
Degrees of freedom	8, 63		
p-value	0		

ANNEXE II

Tableaux des résultats des analyses ANOVA

Tableau A2.1 : ANOVA de Welch — modèle de calibration

	SumOfSquares	DF	MeanSquares	F	pValue
No Calibration	29.603	6	4.9339	2.8284	0.010408
Error	718.7	412	1.7444		
Total	748.3	418			

Tableau A2.2 : ANOVA — taux de protéine

	SumOfSquares	DF	MeanSquares	F	pValue
Proteine	7.1647	6	1.1941	2.9144	0.0085662
Error	165.53	404	0.40973		
Total	172.7	410			

Tableau A2.3 : ANOVA — taux de gras

	SumOfSquares	DF	MeanSquares	F	pValue
Gras	5.7498	4	1.4375	3.4412	0.0088004
Error	170.01	407	0.41772		
Total	175.76	411			

Tableau A2.4 : ANOVA — taux de gras sans variable confondante

	SumOfSquares	DF	MeanSquares	F	pValue
Gras	1.9257	3	0.6419	1.5461	0.20232
Error	144.06	347	0.41516		
Total	145.99	350			

Tableau A2.5 : ANOVA — température du produit

	SumOfSquares	DF	MeanSquares	F	pValue
Température Produit	4.1751	5	0.83502	1.9831	0.080097
Error	172.63	410	0.42106		
Total	176.81	415			

Tableau A2.6 : ANOVA — température du produit sans variable confondante

	SumOfSquares	DF	MeanSquares	F	pValue
Température Produit	3.1217	5	0.62435	1.5148	0.18443
Error	144.26	350	0.41217		
Total	147.38	355			

Tableau A2.7 : ANOVA — température ambiante

	SumOfSquares	DF	MeanSquares	F	pValue
Température Ambiante	4.7254	5	0.94507	2.2469	0.048989
Error	172.03	409	0.42062		
Total	176.76	414			

Tableau A2.8 : ANOVA de Welch — volume du produit

	SumOfSquares	DF	MeanSquares	F	pValue
Volume	7.4017	2	3.7008	9.379	0.00011436
Error	109.69	278	0.39459		
Total	117.1	280			

Tableau A2.9 : ANOVA — forme du produit

	SumOfSquares	DF	MeanSquares	F	pValue
Forme	3.0277	3	1.0092	2.3927	0.068022
Error	173.78	412	0.4218		
Total	176.81	415			

Tableau A2.10 : ANOVA — groupe de produit (5)

	SumOfSquares	DF	MeanSquares	F	pValue
Groupe Produit (5)	4.6588	3	1.5529	3.7261	0.011507
Error	169.21	406	0.41677		
Total	173.87	409			

Tableau A2.11 : ANOVA — groupe de produit (10)

	SumOfSquares	DF	MeanSquares	F	pValue
Groupe Produit (10)	7.2238	7	1.032	2.4833	0.016636
Error	165.39	398	0.41556		
Total	172.62	405			

Tableau A2.12 : ANOVA — groupe de produit (15)

	SumOfSquares	DF	MeanSquares	F	pValue
Groupe Produit (15)	7.5511	10	0.75511	1.8585	0.04948
Error	158.46	390	0.4063		
Total	166.01	400			

Tableau A2.13 : ANOVA — groupe de produit (5) sans variable confondante

	SumOfSquares	DF	MeanSquares	F	pValue
Groupe Produit (5)	2.1349	3	0.71162	1.7786	0.15098
Error	134.83	337	0.40009		
Total	136.97	340			

Tableau A2.14 : ANOVA — groupe de produit (10) sans variable confondante

	SumOfSquares	DF	MeanSquares	F	pValue
Groupe Produit (10)	2.3321	6	0.38869	0.96426	0.44941
Error	134.63	334	0.4031		
Total	136.97	340			

Tableau A2.15 : ANOVA — groupe de produit (15) sans variable confondante

	SumOfSquares	DF	MeanSquares	F	pValue
Groupe Produit (15)	3.5856	9	0.3984	0.98868	0.44908
Error	133.38	331	0.40296		
Total	136.97	340			

Tableau A2.16 : ANOVA — saveur en poudre

	SumOfSquares	DF	MeanSquares	F	pValue
Saveur Poudre	6.0748	6	1.0125	2.4409	0.024928
Error	167.16	403	0.41478		
Total	173.23	409			

Tableau A2.17 : ANOVA — saveur liquide

	SumOfSquares	DF	MeanSquares	F	pValue
Saveur liquide	4.4536	3	1.4845	3.571	0.014181
Error	168.78	406	0.41571		
Total	173.23	409			

Tableau A2.18 : ANOVA — saveur gras

	SumOfSquares	DF	MeanSquares	F	pValue
Saveur Gras	4.3228	3	1.4409	3.4635	0.016387
Error	168.91	406	0.41603		
Total	173.23	409			

Tableau A2.19 : ANOVA — saveur en poudre sans variable confondante

	SumOfSquares	DF	MeanSquares	F	pValue
Saveur Poudre	1.9652	5	0.39304	0.95162	0.44765
Error	142.08	344	0.41302		
Total	144.05	349			

Tableau A2.20 : ANOVA — saveur liquide sans variable confondante

	SumOfSquares	DF	MeanSquares	F	pValue
Saveur liquide	0.34397	2	0.17199	0.4153	0.66047
Error	143.7	347	0.41412		
Total	144.05	349			

Tableau A2.21 : ANOVA — saveur gras sans variable confondante

	SumOfSquares	DF	MeanSquares	F	pValue
Saveur Gras	1.1408	3	0.38026	0.92069	0.43095
Error	142.9	346	0.41302		
Total	144.05	349			

Tableau A2.22 : ANOVA — délai de mesure

	SumOfSquares	DF	MeanSquares	F	pValue
Délai	6.5385	11	0.59441	1.442	0.15151
Error	163.23	396	0.4122		
Total	169.77	407			

Tableau A2.23 : ANOVA de Welch — humidité DG920

	SumOfSquares	DF	MeanSquares	F	pValue
Nb jour	10.448	8	1.306	10.819	1.9356e-09
Error	7.6049	63	0.12071		
Total	18.053	71			

Tableau A2.24 : ANOVA de Welch — humidité DG1901

	SumOfSquares	DF	MeanSquares	F	pValue
Nb jour	6.0863	8	0.76079	7.8495	3.1977e-07
Error	6.0092	62	0.096922		
Total	12.095	70			

Tableau A2.25 : ANOVA de Welch — humidité CHT05

	SumOfSquares	DF	MeanSquares	F	pValue
Nb jour	2.7897	8	0.34871	10.208	0.014253
Error	2.1521	63	0.034161		
Total	4.9418	71			

Tableau A2.26 : ANOVA de Welch — humidité DG920, délai supérieur à 2 jours

	SumOfSquares	DF	MeanSquares	F	pValue
Nb jour	10.448	8	1.306	10.819	1.9356e-09
Error	7.6049	63	0.12071		
Total	18.053	71			

Tableau A2.27 : ANOVA de Welch — humidité DGJ901, délai supérieur à 2 jours

	SumOfSquares	DF	MeanSquares	F	pValue
Nb jour	2.5241	6	0.42068	13.083	8.9921e-09
Error	1.5756	49	0.032156		
Total	4.0997	55			

Tableau A2.28 : ANOVA de Welch — humidité CHT05, délai supérieur à 2 jours

	SumOfSquares	DF	MeanSquares	F	pValue
Nb jour	0.13788	6	0.02298	1.5456	0.18335
Error	0.72853	49	0.014868		
Total	0.86642	55			

ANNEXE III

Spécifications techniques du capteur DA 7440™

Tableau A3.1 : spécifications techniques du capteur DA 7440™

Adapté de PerkinElmer AB (2014)

Environnement de fonctionnement	
Température de fonctionnement	0 à 40 °C
Humidité relative	30 à 90 %
Distance avec le produit	Nominale : 250 mm Tolérance : -50/+100 mm (200 à 350 mm)
Montage du capteur	Tuyau de diamètre de 38 mm
Indice de protection	
Capteur DA 7440™	IP65 (sans connecteurs fermés) IP67 (avec connecteurs fermés)
Boîtier d'interface	IP65
Air comprimé	
Diamètre du connecteur	6mm
Pression	100 kPa (15 psi, 1 bar), maximum 400 kPa (4 bar)
Débit d'air	30 à 60 L/min
Spectromètre et optique	
Type de spectromètre	Matrice de diode
Gamme de longueurs d'onde	950 à 1650 nm
Précision en longueur d'onde	Meilleure que 0.05 nm
Détecteur	InGaAs à 256 éléments refroidi thermoélectriquement Avec source de référence interne pour l'étalonnage de la ligne
Réseau	Réseau stationnaire à traits dorés
Taux de collecte des données	Approximativement 30 spectres / seconde
Bruit	< 3x10 ⁻⁵ sur une période de mesure de 3 secondes
Référence	Système automatique Spectralon

ANNEXE IV

Spécifications techniques du sac d'échantillon

Tableau A4.1 : Spécifications générales du sac EPL-4575

Adapté de LABPLAS

Spécification	Valeur
Matériel	Mélange de polyéthylène
Couleur	Transparent
Dimension	114 x 190 mm
Épaisseur	0.0635 mm
Volume	Total : 490 ml Fonctionnel : 290 ml
Système d'ouverture	Ligne perforée
Système de fermeture	Attache à deux fils fonds
Stérile	Oui
Fin de vie du produit	Recyclable

