

Titre: Développement d'outils d'aide à la décision visant l'ajustement des paramètres contrôlables dans la production agroalimentaire : cas d'étude des minoteries biologiques
Title:

Auteur: Loïc Parrenin
Author:

Date: 2024

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Parrenin, L. (2024). Développement d'outils d'aide à la décision visant l'ajustement des paramètres contrôlables dans la production agroalimentaire : cas d'étude des minoteries biologiques [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. <https://publications.polymtl.ca/58016/>
Citation:

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/58016/>
PolyPublie URL:

Directeurs de recherche: Christophe Danjou, & Bruno Agard
Advisors:

Programme: Doctorat en génie industriel
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Développement d'outils d'aide à la décision visant l'ajustement des paramètres
contrôlables dans la production agroalimentaire : Cas d'étude des minoteries
biologiques**

LOIC PARRENIN

Département de mathématiques et de génie Industriel

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*

Génie industriel

Avril 2024

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Cette thèse intitulée :

**Développement d'outils d'aide à la décision visant l'ajustement des paramètres
contrôlables dans la production agroalimentaire : Cas d'étude des minoteries
biologiques**

présentée par **Loic PARRENIN**

en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Jean-Marc FRAYRET, président

Christophe DANJOU, membre et directeur de recherche

Bruno AGARD, membre et codirecteur de recherche

Antoine LEGRAIN, membre

Renan BONNARD, membre externe

DÉDICACE

À tous mes amis et ma famille...

REMERCIEMENTS

Tout d’abord, je voudrais adresser mes remerciements les plus sincères à ma famille. À mes parents et à ma sœur, votre confiance en moi et vos encouragements constants ont été la pierre angulaire de ma persévérance. Votre soutien indéfectible et vos conseils m’ont permis d’aller de l’avant, et je vous en serai éternellement reconnaissant.

À mes chers amis et collègues du laboratoire LID, votre amitié et votre passion commune pour la recherche ont enrichi mon parcours universitaire. Chaque interaction, discussion et collaboration ont façonnées ma pensée en tant que chercheur et en tant que personne. Je tiens tout particulièrement à remercier Laure et Suzanne, qui ont été les premières à m’embarquer dans ce voyage de recherche, m’inspirant par leur sagesse et leur vivacité d’esprit.

Une mention spéciale est due à Ambre, dont la gentillesse, l’amour et le soutien attentif ont été mon point d’ancrage dans les moments de stress et d’incertitude. Ta présence a illuminé même les jours les plus sombres, et je te suis profondément reconnaissant pour ça.

J’éprouve une grande gratitude envers mes directeurs de recherche, Bruno et Christophe, dont la supervision et les conseils ont joué un rôle déterminant dans ma trajectoire universitaire. Leur expertise, leur dévouement et leur soutien au cours des cinq dernières années ont joué un rôle essentiel dans ma réussite, que ce soit pour la publication des articles scientifiques ou la finalisation de mes études doctorales.

Je tiens également à exprimer ma reconnaissance à Camélia, dont les interventions opportunes et les réflexions stimulantes ont été si précieuses dans les moments de doutes.

Aux membres du jury, j’adresse mes sincères remerciements pour l’évaluation que vous avez fait de mon travail.

Enfin, j’exprime mes plus sincères remerciements à M. Beauchemin et à son équipe pour leur collaboration à ce projet, ainsi qu’aux membres de l’équipe du SENAI, pour leur accueil chaleureux au Brésil et leur participation à notre projet de recherche. Votre hospitalité et votre soutien ont permis d’accroître la portée et l’impact de notre travail, et je vous en suis vraiment reconnaissant.

Vous l'aurez compris, votre soutien, vos conseils et votre bienveillance ont indéniablement contribué à mon parcours et je n'aurais pu partager cet ouvrage sans votre présence collective. En un mot, Merci !

RÉSUMÉ

Face à une demande alimentaire mondiale en constante augmentation, les conséquences des conflits, des crises économiques et des problèmes climatiques actuels engendrent une insécurité alimentaire significative dans diverses régions et populations du monde. Il devient alors impératif d'améliorer l'efficacité et la résilience du secteur agroalimentaire afin qu'ils puissent répondre à ces besoins alimentaires croissants. Parallèlement, la pénurie de main-d'œuvre dans le secteur manufacturier et la perte d'expertise due aux départs à la retraite posent des défis pour la maîtrise des processus de production complexes. Dans le cas de la meunerie biologique, la complexité du processus de meunerie se révèle à travers plusieurs étapes, chacune présentant des défis spécifiques et des complexités qui nécessitent une gestion adaptée. L'état de l'art met particulièrement en lumière l'impératif et les défis à relever pour assurer une production de qualité constante de la farine. La variabilité inhérente à la qualité des grains de blé, en tant que matière première, exerce un impact direct sur le rendement et la qualité de la production finale.

Un défi majeur auquel est confronté le processus de meunerie biologique réside dans la nécessité d'ajuster chaque étape du processus pour atteindre les rendements et la qualité de farine souhaités. Cette adaptation précise est rendue essentielle par la nature variable des caractéristiques des grains biologiques. Dans ce contexte, l'intervention d'un meunier expérimenté devient cruciale. Les ajustements nécessaires à apporter au processus de production reposent sur les connaissances approfondies et l'expérience tacite des meuniers, qui les réalisent à travers des essais et erreurs continus. La nécessité d'une expertise spécialisée dans le contrôle du processus de meunerie, associée à une pénurie de main-d'œuvre et à une perte de compétences, pose un défi à la production de farine biologique. La variabilité de la qualité des grains de blé et des facteurs de production rend les instructions d'exploitation inefficaces, en particulier en l'absence de meuniers qualifiés pour les ajuster. Un défi important est donc de contrôler le processus de meunerie pour garantir une qualité constante de la farine biologique.

Pour relever ce défi, l'objectif général de cette thèse est de développer des outils d'aide à la décision visant l'ajustement des paramètres contrôlables dans la production agroalimentaire.

Ces outils visent à accompagner et à contrôler les processus de transformation agroalimentaire, en tirant des enseignements d'expériences réelles et en intégrant les comportements et les contraintes présents sur la chaîne de production. En se basant sur un cas d'étude d'une minoterie biologique, l'objectif général se décline en trois objectifs spécifiques menant chacun à des contributions scientifiques complémentaires.

La première contribution de ce travail est une revue des techniques d'intelligence artificielle appliquées à la meunerie biologique. Cette revue permet d'identifier quatre axes de recherche, à savoir l'économie d'énergie, la production d'une farine de qualité constante, l'optimisation des ressources et l'amélioration de la performance de la minoterie. Parmi ces axes de recherches, plusieurs opportunités liées à l'intelligence artificielle sont présentées.

La deuxième contribution consiste à structurer les données dans un contexte de production alimentaire grâce à l'expertise métier et à la cartographie des processus. De cette manière, une évaluation des données disponibles est effectuée, permettant d'identifier les améliorations nécessaires dans les processus à partir de la cartographie réalisée. Ce travail conduit à la définition d'un besoin spécifique axé sur la production constante d'une farine biologique de qualité.

Dans la continuité, la troisième contribution propose un outil d'aide à la décision pour le contrôle d'un processus de transformation agroalimentaire à partir des données de production disponibles. Adapté au processus de mouillage des grains de blé, cet outil a été développé pour ajuster les paramètres contrôlables afin d'obtenir la teneur en eau souhaitée dans les grains de blé. Basé sur un modèle de régression linéaire LASSO, il montre de bons résultats dans la prédiction de l'augmentation de la teneur en eau, avec une erreur moyenne de 0,428%. Les résultats mettent en évidence l'influence significative des conditions météorologiques sur l'ajustement des paramètres, soulignant l'importance d'un ajustement méticuleux des débits d'eau et de blé. Les coefficients dérivés de l'équation de régression fournissent une analyse de sensibilité pour évaluer leur impact.

Pour faire face à la variabilité de la qualité des intrants, la quatrième contribution se matérialise à travers le développement d'un outil d'aide à la décision dédié à la mesure de la qualité des aliments en production. Utilisant la spectroscopie NIR, cet outil sélectionne les modèles

de prédiction les plus performants pour différents aliments et attributs de qualité. Il identifie les longueurs d'onde clés permettant de prédire la qualité des aliments, apportant ainsi une contribution significative dans le contexte de la calibration NIR. Des résultats prometteurs ont été observés pour divers attributs, notamment la teneur en eau et en matières grasses des fèves de cacao, la teneur en eau et en protéines du blé (A), l'acidité et la teneur en vitamine C de la mangue, et la teneur en protéines du blé (B), avec des scores R^2_{cv} respectifs de 0,853, 0,727, 0,905, 0,993, 0,924, 0,663 et 0,905. Notamment, les modèles sélectionnés pour le blé (A) se sont révélés plus précis que le modèle de référence, affichant une erreur quadratique moyenne plus faible sur l'ensemble des données testées.

Ainsi, cette thèse a contribué au développement d'outils d'aide à la décision visant l'ajustement des paramètres contrôlables dans la production agroalimentaire. Ce travail de recherche propose une approche à suivre pour le développement d'outils d'aide à la décision afin de faciliter la prise de décision lors de l'ajustement des paramètres de contrôle dans la chaîne de transformation alimentaire. Cette approche a été développée sur base d'un cas spécifique dans une minoterie biologique.

ABSTRACT

Faced with ever-increasing global food demand, the consequences of current conflicts, economic crises and climatic problems are creating significant food insecurity in various regions and populations around the world. It is therefore imperative to improve the efficiency and resilience of the agri-food sector so that they can meet these growing food needs. At the same time, labor shortages in the manufacturing sector and the loss of expertise due to retirement are posing challenges for the mastery of complex production processes. In the case of organic milling, the complexity of the milling process is revealed through several stages, each presenting specific challenges and complexities that require adapted management. The state of the art particularly highlights the imperative and challenges of ensuring consistent flour quality. The inherent variability of wheat grain quality as a raw material has a direct impact on the yield and quality of the final product.

A major challenge facing the organic milling process is the need to fine-tune each stage of the process to achieve the desired yields and flour quality. This precise adaptation is made essential by the variable nature of organic grain characteristics. In this context, the intervention of an experienced miller becomes crucial. The necessary adjustments to the production process rely on the millers' in-depth knowledge and tacit experience, which they achieve through continuous trial and error. The need for specialized expertise in milling process control, coupled with labor shortages and loss of skills, poses a challenge to organic flour production. Variability in wheat grain quality and production factors makes operating instructions ineffective, particularly in the absence of skilled millers to adjust them. A major challenge is therefore to control the milling process to guarantee consistent organic flour quality.

To meet this challenge, the general aim of this thesis is to develop decision-support tools for adjusting controllable parameters in agri-food production. These tools aim to support and control agri-food transformation processes, by learning from real-life experience and integrating the behaviors and constraints present on the production line. Based on a case study of an organic flour mill, the general objective is broken down into three specific objectives, each

leading to complementary scientific contributions.

The first contribution of this work is a review of artificial intelligence techniques applied to organic milling. This review identifies four areas of research, namely energy savings, production of consistent-quality flour, optimization of resources and improvement of milling performance. Among these areas of research, several opportunities linked to artificial intelligence are presented.

The second contribution consists in structuring the data in a food production context, using business expertise and process mapping. In this way, an evaluation of the available data is carried out, enabling the identification of necessary process improvements based on the mapping carried out. This work led to the definition of a specific requirement focused on the consistent production of quality organic flour.

Continuing in this direction, the third contribution proposes a decision-support tool for the control of an agri-food transformation process based on available production data. Adapted to the wheat grain wetting process, this tool has been developed to adjust the controllable parameters in order to obtain the desired water content in wheat grains. Based on a LASSO linear regression model, it shows good results in predicting the increase in moisture content, with an average error of 0.428%. The results highlight the significant influence of meteorological conditions on parameter adjustment, underlining the importance of meticulous adjustment of water and wheat flows. The coefficients derived from the regression equation provide a sensitivity analysis to assess their impact.

To address the variability of input quality, the fourth contribution materializes through the development of a decision-support tool dedicated to the measurement of feed quality in production. Using NIR spectroscopy, this tool selects the best-performing prediction models for different feeds and quality attributes. It identifies the key wavelengths for predicting food quality, making a significant contribution in the context of NIR calibration. Promising results were observed for various attributes, including water and fat content in cocoa beans, water and protein content in wheat (A), acidity and vitamin C content in mango, and protein content in wheat (B), with R^2_{cv} scores of 0.853, 0.727, 0.905, 0.993, 0.924, 0.663 and 0.905 respectively. Notably, the models selected for wheat (A) proved more accurate than the

reference model, displaying a lower root-mean-square error on the tested data set.

Thus, this thesis has contributed to the development of decision support tools for the adjustment of controllable parameters in food production. This research work proposes an approach to be followed for the development of decision support tools to facilitate decision-making when adjusting control parameters in the food processing chain. This approach was developed on the basis of a specific case in an organic flour mill.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	vi
ABSTRACT	ix
TABLE DES MATIÈRES	xii
LISTE DES TABLEAUX	xvii
LISTE DES FIGURES	xviii
LISTE DES SIGLES ET ABRÉVIATIONS	xix
LISTE DES ANNEXES	xx
CHAPITRE 1 INTRODUCTION	1
CHAPITRE 2 REVUE DE LITTÉRATURE - ARTICLE 1: FUTURE TRENDS IN ORGANIC FLOUR MILLING: THE ROLE OF AI	5
2.1 Introduction	6
2.2 State of the art	9
2.2.1 Selection of the wheat variety	10
2.2.2 Transport / wheat analysis / wheat storage	13
2.2.3 Cleaning	16
2.2.4 Wheat blending	17
2.2.5 Tempering	20
2.2.6 Milling	25
2.2.7 Organic wheat flour end product	30

2.2.8	Discussion	34
2.3	Research areas towards smart wheat milling	35
2.3.1	Energy savings	36
2.3.2	Consistent quality	37
2.3.3	Resource optimization	38
2.3.4	Mill performance	38
2.4	Conclusion	39
2.5	Acknowledgements	40
CHAPITRE 3 PROBLÉMATIQUE ET DÉMARCHE DU PROJET DE RECHERCHE		41
3.1	Problématique	41
3.2	Cadre de recherche et contribution	42
3.3	Contexte industriel	43
3.4	Méthodologie de recherche	44
CHAPITRE 4 ARTICLE 2 : DÉVELOPPEMENT D'UN OUTIL DE SUIVI DES MOUVEMENTS DES GRAINS DE BLÉ BIOLOGIQUE AU NIVEAU DE LA CHAÎNE D'APPROVISIONNEMENT		48
4.1	Introduction	48
4.2	État de l'art	50
4.2.1	Caractéristiques des grains de blé	50
4.2.2	Chaîne d'approvisionnement de la récolte du grain de blé aux usines de transformation	51
4.2.3	Contrôle de la qualité du grain de blé au niveau du processus d'approvisionnement	52
4.3	Méthodologie	53
4.3.1	Cartographie de l'approvisionnement	54
4.3.2	Création de bases de données et d'interfaces	55
4.3.3	Créer des indicateurs et des tableaux de bord	56
4.4	Cas d'étude	57
4.4.1	Contexte industriel	57

4.4.2	Cartographie de l’approvisionnement	57
4.4.3	Création de bases de données et d’interfaces	58
4.4.4	Indicateurs et tableaux de bord	60
4.5	Conclusion	62
4.6	Remerciements	63

CHAPITRE 5 ARTICLE 3: A DECISION SUPPORT TOOL FOR THE FIRST STAGE OF THE TEMPERING PROCESS OF ORGANIC WHEAT GRAINS IN

A MILL	64
5.1 Introduction	65
5.2 Literature review	66
5.2.1 Tempering process in a wheat mill plant	66
5.2.2 Analysis of experiments on tempering wheat grains	68
5.3 Materials and methods	69
5.3.1 Characteristics of the organic wheat mill plant	70
5.3.2 Materials	70
5.3.2.1 Raw materials and wheat tempering	70
5.3.2.2 Predictive models	71
5.3.3 Method followed to develop a decision support tool in the wheat grain tempering process	73
5.3.3.1 Collection of relevant data	73
5.3.3.2 Preparation of the data	75
5.3.3.3 Creation of the predictive models	76
5.3.3.4 Validation of the models	77
5.4 Results	77
5.5 Discussion	80
5.6 Conclusion	83
5.7 Acknowledgements	84

CHAPITRE 6	ARTICLE 4: A DECISION SUPPORT TOOL TO ANALYZE THE PROPERTIES OF WHEAT, COCOA BEANS AND MANGOES FROM NIR SPECTRA	85
6.1	Introduction	86
6.2	Literature review	87
6.2.1	NIR spectroscopy in quality measurement	87
6.2.2	Analysis techniques of NIR spectra	89
6.3	Methods	95
6.3.1	Data preprocessing	97
6.3.2	Machine learning model	98
6.3.3	Feature selection	99
6.3.4	Model evaluation	100
6.4	Results	101
6.4.1	Dataset and Samples	101
6.4.2	Application of the methodology	102
6.4.2.1	Data preprocessing	102
6.4.2.2	Machine learning model	104
6.4.2.3	Feature selection	105
6.4.2.4	Model evaluation	106
6.5	Discussion	108
6.6	Conclusion	109
6.7	Acknowledgements	111
CHAPITRE 7	DISCUSSION GÉNÉRALE	112
7.1	Synthèse des travaux	113
7.2	Limites et perspectives	118
CHAPITRE 8	CONCLUSION	121
8.1	Synthèse des travaux	121
8.2	Limitations	123
8.3	Améliorations futures	123

RÉFÉRENCES	125
ANNEXES	141

LISTE DES TABLEAUX

Table 2.1	Summary of existing AI research on wheat selection	12
Table 2.2	Summary of existing AI research on wheat transport and storage conditions	15
Table 2.3	Summary of existing AI research on wheat transportation and storage conditions	20
Table 2.4	Summary of existing AI research on wheat tempering	24
Table 2.5	Summary of existing AI research on wheat milling	29
Table 2.6	Summary of existing AI research on organic wheat flour end product	33
Table 5.1	Performances of predictive models	78
Table 5.2	Variables used in the models (controllable parameters in red)	78
Table 6.1	BPSO parameters description	93
Table 6.2	Descriptive statistics of the spectra of each dataset	101
Table 6.3	Descriptive statistics of quality parameters for each dataset	102
Table 6.4	BPSO parameters values for the dataset on cocoa beans	106
Table 6.5	Model performance for the prediction of wheat (A) quality features based on pre-processing, machine learning techniques and the selection of wavelength.	107
Table A.1	Model performance for the prediction of Cocoa bean quality features based on pre-processing, machine learning technique and wavelength selection.	141
Table B.1	Model performance for the prediction of Wheat (B) quality features based on pre-processing, machine learning technique and wavelength selection.	142
Table C.1	Model performance for the prediction of Mango quality features based on pre-processing, machine learning technique and wavelength selection.	143

LISTE DES FIGURES

Figure 2.1	Diagram of the wheat flour milling production value chain	9
Figure 2.2	The anatomy of a grain of wheat (longitudinal section)	10
Figure 2.3	Different end-product utilization, depending on type of wheat, protein percentage, kernel hardness (Posner and Hibbs, 2005)	30
Figure 2.4	The types of wheat flour according to ash content	31
Figure 2.5	Mind map for future topics research centered around the optimization of organic wheat flour milling	35
Figure 3.1	Démarche du projet de recherche	45
Figure 4.1	Processus simplifié de l’approvisionnement en grain de blé	52
Figure 4.2	Construction de la cartographie de l’approvisionnement des grains de blé du Québec	58
Figure 4.3	Relation fictive entre la table fournisseur et la table analyse échantillon	59
Figure 4.4	Mouvements au sein d’un centre de grain	60
Figure 4.5	Exemple d’indicateurs visuels pour le suivi des inventaires d’un centre de grain	61
Figure 5.1	Diagram of the wheat grains’ tempering process	67
Figure 5.2	DBSCAN plot to detect potential outliers	76
Figure 5.3	Scatter plot of predicted value from a LASSO model against real values	80
Figure 6.1	Diagram of the method followed for the prediction of the quality properties of raw materials from the developed decision support tool . . .	96
Figure 6.2	(a) Raw NIR spectra of cocoa beans samples; (b) raw NIR spectra of the samples of grains of wheat (A); (c) raw NIR spectra of wheat grain (B) samples; (d) raw NIR spectra of mangoes samples.	103
Figure 7.1	Synthèse des travaux portant sur le développement d’outils d’aide à la décision visant l’ajustement des paramètres contrôlables dans la production agroalimentaire	114

LISTE DES SIGLES ET ABRÉVIATIONS

BEM	Brabender Energy Max
FAO	Food and Agriculture Organization
FSIN	Food Security Information Network
GNAFC	Global Network Against Food Crises
IAOM	International Association of Operative Millers
MAPAQ	Ministère de l'Agriculture, des Pêcheries et de l'Alimentation
NIR	Near-Infrared Spectroscopy
SVM	Support Vector Machines
RMSE	Root Mean Square Error

LISTE DES ANNEXES

Annexe A	Performance des modèles de prédiction pour les fèves de cacao	141
Annexe B	Performance des modèles de prédiction pour les grains de blé (B) . .	142
Annexe C	Performance des modèles de prédiction pour les mangues	143

CHAPITRE 1 INTRODUCTION

Sous l'effet de la convergence de multiples facteurs, le monde moderne est témoin d'un changement profond dans son approche de la production alimentaire. Avec l'accroissement de la population mondiale et l'amélioration du niveau de vie dans les pays en développement, le besoin de subsistance n'a jamais été aussi grand (FAO, 2016; Hobbs, 2007). Les projections indiquent que d'ici 2050, la demande mondiale de nourriture augmentera de 60 % ou plus par rapport aux niveaux de 2006 (FAO, 2016). Cette demande croissante s'accompagne par ailleurs d'une évolution des préférences des consommateurs, reflétant une tendance croissante à la consommation d'aliments biologiques. D'après le récent rapport de Willer et al. (2023), il est indiqué qu'entre 2012 et 2021 la taille du marché biologique a plus que doublé, atteignant 135,5 milliards de dollars US de ventes en 2021. Actuellement, l'Amérique du Nord et l'Europe dominent les ventes, représentant ensemble 90 % du marché. Cependant, la croissance la plus significative provient d'autres régions, notamment de l'Asie. Le marché des aliments biologiques gagne en importance dans des pays tels que la Chine, l'Inde et la Corée du Sud (Willer et al., 2023).

La transformation agroalimentaire est un secteur de l'industrie alimentaire qui consiste à transformer des produits agricoles bruts (tels que les cultures et le bétail) en divers produits alimentaires propres à la consommation. Ce secteur englobe un large éventail d'activités, notamment le nettoyage, le tri, l'emballage, la cuisson, la conservation et d'autres processus destinés à améliorer la sécurité, la durée de conservation, le goût et la qualité globale des produits alimentaires. La transformation agroalimentaire joue un rôle crucial dans la chaîne d'approvisionnement alimentaire, permettant la distribution des produits alimentaires aux consommateurs et la création de produits à valeur ajoutée.

Sous l'effet des conflits, des chocs économiques et des crises climatiques, l'insécurité alimentaire a de nouveau progressé dans le monde en 2022 (FSIN and GNAFC, 2023). La guerre en cours en Ukraine a provoqué une onde de choc sur les marchés des matières premières et de l'énergie, perturbant l'équilibre délicat de l'offre et de la demande. Ce conflit imprévu a eu des répercussions considérables, déclenchant des fluctuations des prix des produits de

base, en particulier dans les secteurs des céréales et de l'énergie. L'interaction complexe entre la guerre, la pandémie, la reprise économique et la flambée des prix des denrées alimentaires et des intrants agricoles engendre des répercussions diverses sur la faim et l'insécurité alimentaire selon les régions et les populations.

Alors que le monde doit faire face à ces multiples défis, il est devenu impératif d'accroître l'efficacité et la résilience du secteur agroalimentaire, non seulement pour faire face à la croissance de la demande mondiale de denrées alimentaires, mais aussi pour atténuer les répercussions de ces perturbations mondiales telles que le coût énergétique et le coût des matières premières.

La transformation numérique, souvent associée à l'industrie 4.0, offre des solutions pour améliorer l'efficacité des lignes de production. La transformation numérique se définit comme une stratégie d'intégration des technologies numériques, de l'analyse des données et de l'automatisation dans divers aspects des opérations, des processus et des produits d'une organisation (Danjou et al., 2017). Cette transformation met plus particulièrement l'accent sur la connectivité des appareils et des systèmes, l'analyse des données en temps réel et la capacité à s'adapter et à évoluer en fonction des données recueillies. Dans les milieux industriels, elle peut conduire à des usines intelligentes et à une prise de décisions plus agile et basée sur les données.

Exploiter le potentiel de la transformation numérique dans l'industrie agroalimentaire n'est pas seulement une option, mais un impératif stratégique. La transformation numérique offre une voie vers une plus grande efficacité dans l'utilisation des ressources, une réduction des déchets, une amélioration de la qualité et un renforcement de la durabilité de la production alimentaire. En ces temps tumultueux, l'appel à révolutionner la façon dont nous fabriquons les aliments n'est pas simplement une question d'innovation ; il s'agit d'une réponse aux changements climatiques, géopolitiques et économiques qui affectent les fondements mêmes de nos systèmes alimentaires mondiaux.

Face à cet impératif stratégique de transformation numérique du secteur agroalimentaire, cette thèse se positionne comme un élément essentiel de l'évolution vers une production alimentaire plus efficace et plus durable. Elle se distingue par sa contribution significative à

travers le développement d'outils d'aide à la décision pilotés par les données, spécifiquement orientés vers l'ajustement des paramètres contrôlables dans la production agroalimentaire. Ancré dans une approche de transformation numérique, ce travail s'appuie sur un cas d'étude réel au sein de l'industrie de la meunerie. En effet, la meunerie, en tant que secteur pivot dans la transformation du grain en farine, offre un terrain propice à l'application et à la validation de ces outils d'aide à la décision. Ainsi, cette recherche vise à relever les défis actuels en combinant la pertinence sectorielle avec les avantages tangibles de la transformation numérique, démontrant la capacité de l'innovation à façonner l'avenir de la production agroalimentaire.

La thèse, par article, est présentée comme suit : tout d'abord, le chapitre 2 présente une revue de littérature des méthodes utilisées dans le domaine de l'intelligence artificielle pour améliorer les processus de meunerie. Il offre une vue d'ensemble des lacunes et opportunités dans le secteur de la meunerie que l'intelligence artificielle peut combler dans une quête d'efficacité optimale des opérations de meunerie. La revue de littérature a fait l'objet d'un 1er article intitulé "Future trends in organic flour milling : The role of AI" publié dans *AIMS Agriculture and Food*. Sur la base de cette analyse et de notre expérience dans le domaine de l'industrie de la minoterie biologique, le chapitre 3 définit notre problématique, notre cadre de recherche et la démarche suivie pour ce travail. Les chapitres 4, 5 et 6 présentent les travaux scientifiques qui s'accordent chacun à répondre à l'un des 3 sous-objectifs énumérés au chapitre 3. Le chapitre 4 présente l'approche de cartographie du processus et de structuration des données en vue de développer un outil de suivi des grains de blé sur la chaîne d'approvisionnement. Ces travaux ont été présentés dans le cadre d'un article de conférence intitulé "Développement d'un outil de suivi des mouvements des grains de blé biologique au niveau de la chaîne d'approvisionnement". Le chapitre 5 présente un outil d'aide à la décision basé sur un modèle de régression Linéaire pour prévoir l'augmentation du taux d'humidité des grains de blé selon les facteurs et paramètres contrôlables au niveau du processus de mouillage. Ce travail a fait l'objet d'un troisième article intitulé "A decision support tool for the first stage of the tempering process of organic wheat grains in a mill", publié dans *International Journal of Food Science & Technology*. Le chapitre 6 présente une méthodologie pour prédire les attributs de qualité des grains de blé, fèves de cacao et

mangues, basée sur les mesures de longueurs d'onde issues d'appareils de mesure des proches infra-rouge (NIR). Il a fait l'objet d'un article soumis pour publication, intitulé "A Decision support tool to analyze wheat grain, cocoa bean and mango properties from NIR spectra". Les limites et perspectives des travaux menés sont finalement exposés dans le chapitre 7.

CHAPITRE 2 REVUE DE LITTÉRATURE - ARTICLE 1: FUTURE TRENDS IN ORGANIC FLOUR MILLING: THE ROLE OF AI

Parrenin, L., Danjou, C., Agard, B., and Beauchemin, R. (2023b). Future trends in organic flour milling: The role of ai. *AIMS Agriculture and Food*, 8(1):48–77

Submitted on august 16, 2022 and accepted on December 7, 2022

ABSTRACT

The milling of wheat flour is a process that has existed since ancient times. In the course of history, the techniques have improved, the equipment modernized. The interest of the miller in charge of the mill is still to ensure that a mill is functional and profitable, as well as to provide a consistent quality of flour. The production of organic flour means that methods of adding chemicals and unnatural agents are not possible. In organic flour production, it is necessary to work with the raw material. A grain of wheat is a living material, and its quality varies according to a multitude of factors. Challenges are therefore present at each stage of the value chain. The use of artificial intelligence techniques offers solutions and new perspectives to meet the different objectives of the miller. A literature review of artificial intelligence techniques developed at each stage of the value chain surrounding the issues of quality and yield is conducted. An analysis of a large number of variables, including process factors, process parameters and wheat grain quality from data collected on the value chain enables the development and training of artificial intelligence models. From these models, it is possible to develop decision support tools and optimize the wheat flour milling process. Several major research directions, other than constant quality, are to be studied to optimize the process and move towards a smart mill. This includes energy savings, resource optimization and mill performance.

Keywords: Industry 4.0, organic wheat grains, flour milling optimisation, flour quality, artificial intelligence, machine learning, data mining.

2.1 Introduction

Organic agriculture is a specific method of food production using natural substances and processes. This type of agriculture is gaining in popularity and supported by people who are concerned by agri-food, environmental and health issues (Dérignon, 2016; Hughner et al., 2007). According to Willer et al. (2021) global organic food sales grew by more than 15% between 2017 and 2019. They showed that in 2019 organic wheat was the most widely grown organic cereal in the world.

Flour milling refers to the industry of processing grains of cereal into flour. Wheat flour is produced by milling wheat. Milling is the process used to grind grains of cereal. This typically takes place in an industrial flour mill, which has replaced antique flour mills powered by wind or water. The objective of wheat flour milling is to separate the different parts of the wheat grain and reduce the grains into smaller particles. Various techniques and strategies have emerged over the years, two of which are still widely used today: stone milling and roller milling (Cappelli et al., 2020c; Doblado-Maldonado et al., 2012). Both techniques have advantages and disadvantages. However, cylinder milling is the most often used because of its high yield. It also results in very white flour. By using cylinder milling, production parameters must be adjusted according to the quality of the grains of wheat.

Wheat flour milling is one of the oldest processes that is still in operation today (Campbell, 2007). The methods and tools have improved over the centuries (Boudreau and Ménard, 1992; Campbell, 2007). The process of wheat flour milling follows a linear and continuous flow (Mateos-Salvador et al., 2011). It can be broken down into 3 phases, as follows: cleaning, tempering and milling. The cleaning process eliminates all impurities and foreign materials present in a wheat batch that can degrade the quality of flour or, at times, damage equipment. Then, the tempering process takes place to condition the wheat for optimal performance in the milling process. Finally, the grains of wheat are milled in a multistage environment that takes place successively until flour is obtained. The wheat milling process can generally be separated into 2 operations: breaking the grain into particles and reducing the particle size. At each stage, the particles are sifted through several sieves. The particle size will classify

the particle as flour. According to FAO (1985) flour should have a particle size less than 212 μm .

Although equipment for the wheat flour milling process has been modernized and some parts are now semi-automated, the miller is responsible for controlling the entire process. He starts by selecting the right kind of wheat to produce a specific quality of flour. A mix can be made with different qualities of wheat if necessary to balance and obtain the desired quality of flour. The wheat mix is then processed following the 3 operational phases, described previously, in the mill. During each phase, the miller adjusts the right process parameters to manufacture the quality of flour required, while reaching the highest achievable yield of flour. The choices and adjustments made by the miller during those procedures are often described as an art. The miller applies his accumulated knowledge and experience to make decisions and changes in the wheat flour milling process.

The industrial production of organic flour differs in some regards from the conventional flour production method. In conventional flour production, volumes are significantly greater. From the vast availability of different types of wheat and the use of improvers, fewer wheat mixtures are required. The use of improvers helps maintain a stable level of flour quality throughout the year. Among the improvers available, it is possible to use chemical agents such as chlorine for flour bleaching to improve the visual aspect of the flour, ascorbic acid to improve baking quality of the flour and enzymes to improve rheological performance of the wheat dough (Boudreau and Ménard, 1992; IAOM, 2018; Steffan, 2012). In contrast, organic wheat flour is produced from organic wheat, only natural ingredients explicitly authorized by organic norms and in certain proportions may be used. For those reasons, two main levers of action are possible, in addition to the use of natural ingredients, to improve and reach the organic flour quality desired: organic wheat mixing and adjustment of the production parameters. These levers of action exist in conventional flour production. However, the miller must deal with a greater variety of organic wheat from different regions compared to conventional. Moreover, the quantities of protein are lower but offer better protein quality than those present in the conventional environment (Vrček et al., 2014). Mixtures and adjustments must be constantly readjusted by the miller to reach a certain level of organic flour quality.

This adds complexity to the production process.

As wheat is a living raw material, every grain of wheat that is cultivated has unique, intrinsic properties. The intrinsic properties vary according to the variety of wheat, climate where it was grown, weather, harvesting period and crop rotation (Borghi et al., 1995; Johansson and Svensson, 1998; Triboi et al., 2000). Because of the variations in wheat quality, there is an impact on flour yield and flour quality if the same process parameter is applied to the same type of wheat mix. The miller's goal is to reduce this variability in the process to produce the best yield with the desired flour quality. By relying on his knowledge and experience, the miller learns through trial and error to select the right wheat mix and to adjust it to the right process parameters. However, many variables exist along the value chain that can affect the flour yield and quality of the flour output. For this reason, it is difficult to fully understand and measure the impact of each variable and its impact on the output.

This is where artificial intelligence can come in. Artificial Intelligence (AI) is a part of computer science. Many have tried to broadly define AI. The most recent definitions describe AI as computer programs or algorithms that imitate intelligent human behavior (Köksal et al., 2011; Misra et al., 2020). Machine Learning (ML) is a sub-field of AI. Kelleher (2019) explains that ML involves “the development and evaluation of algorithms that enable a computer to extract (or learn) functions from a dataset (set of examples)”. From the available data, it is then possible to identify and discover patterns that accurately associate relationships between different system variables without any prior knowledge of the system (Fisher et al., 2020).

This paper intends to review the state of the art of analytical tools and methods (such as AI algorithms) that could be applied to the organic flour industry sector to improve the quality and yield of organic wheat flour.

This paper is organized into four sections. Section 2.2 describes organic wheat flour milling more in depth and the different AI techniques used to improve flour quality and yield output based on the literature review. Section 2.3 covers the future direction that deserves to be studied to improve the flour milling process. It explores aspects other than flour quality and

yield output, such as energy savings, resource optimization and mill performance. Section 2.4 concludes the paper.

2.2 State of the art

The production process of wheat flour has existed for a long time and still applies to organic wheat flour milling. This section focuses on explaining the key steps in the value chain of wheat flour milling. An overview of the challenges regarding flour quality and yield are presented across the value chain at specific steps. Research that has been conducted to overcome those challenges using AI techniques and statistical tools is then outlined.

Figure 2.1 illustrates the workflow of the wheat flour milling process. In this diagram, the key steps of the processing of grains of wheat are shown. It begins with the quality of wheat selected by the miller. It ends in either a manufacturing process of an end-product made from the wheat flour, or by selling the wheat flour that is produced to wholesalers and grocery stores. Between those two starting and ending points in the value chain, the diagram details the different steps that the wheat grains follow in a factory mill.

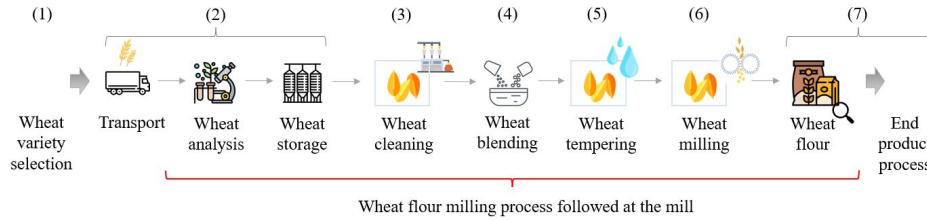


Figure 2.1 Diagram of the wheat flour milling production value chain

The grains of wheat received at the mill experience the transformation process illustrated in Figure 2.1. The objective of this process is to extract the maximum quantity of wheat flour at a desired quality level. The different steps numbered in Figure 2.1 will be explained in the following subsections.

2.2.1 Selection of the wheat variety

The selection of the wheat variety involves choosing the right quality and quantity of wheat to order from a mill. Different types of wheat varieties and levels of quality exist. A grain of wheat, as pictured in Figure 2.2, is located in a wheat ear. It is composed of four distinct parts: the germ (3% of the total mass of the grain), the endosperm (84%), the bran (13%) and the brush (negligible mass).

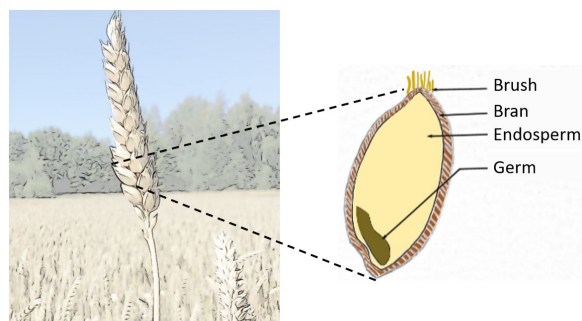


Figure 2.2 The anatomy of a grain of wheat (longitudinal section)

The *germ*, which is rich in fat and vitamins, represents the future plant. The *endosperm* contains mainly carbohydrates (starch) and proteins (including gluten). The starch is a complex carbohydrate (slow sugar) that is useful for the growth of the new plant. Gluten is a protein that forms a viscoelastic dough and gives dough sufficient elasticity to ensure that it will rise. The elasticity and the tenacity of the dough are explained by the creation of the gluten network, which is formed during the process of hydration and kneading dough made from wheat flour. network provides unique rheological properties to the wheat dough, allowing it to be stretched and deformed without breaking up with a certain level of force. The *bran* protects the components inside the wheat grain. It is rich in fiber, but nutritionally poor compared to the endosperm (Campbell, 2007). The *brush* at the top of the grain is characterized by thin and short bristles and contains no nutrients.

Depending on the quantity and quality of wheat available at the mill, the miller must select and order the appropriate quality of wheat to produce and deliver the desired quality of flour to customers. As previously explained, each batch of wheat that is cultivated at a certain time in a specific geographic area has unique intrinsic and physical properties. Those properties

vary according to genetic factors, variations in environmental conditions and crop rotation (Borghi et al., 1995; Branlard et al., 2001; Maphosa et al., 2014; Wrigley et al., 1994). The knowledge of those properties is therefore important, as it influences the adjustments made in the following steps at the mill and affects the final quality of the flour (Meffleh et al., 2020).

The miller wants the farmer to grow varieties of wheat that produce good flour. Depending on the kind of flour desired by the client, the miller will analyze the intrinsic and physical properties of the wheat grains that are the most suitable to mill. The intrinsic properties inform the chemical composition of the wheat grain. For example, *protein content* and *moisture content* are two main intrinsic properties among others, that are analyzed. The physical properties are more related to the appearance and texture of the wheat grain. A few examples of the physical properties of wheat grains are *hardness*, *dimension*, *shape* and *color*.

Challenges

In wheat fields, the quality of wheat (intrinsic and physical properties) fluctuate from one geographic area to another, as well as from one portion of a field to another. Climate change accentuates those variations resulting from more abrupt and less predictable changes in weather conditions. The variation in the quality of wheat grains from wheat fields affects the flour quality and yield produced at the mill. In addition to the fluctuating quality of wheat, there are many more varieties of wheat in the organic sector than in the conventional sector. This makes it more difficult to control and produce consistent quality flour throughout the year.

Progress of work research

To address these issues, some research focuses on improving the selection of wheat variety that will offer the best quality of wheat according to a miller's needs. To this end, some research has studied the physical properties of grains of wheat and its impact on the yield of flour milling. Aydin et al. (2020) have developed an artificial neural network model (ANN) capable of predicting the flour yield based on four physical properties of wheat grains. Their

study showed that wheat grain hardness plays an important role in determining flour yield. Dowell et al. (2006) have studied Near-Infrared Spectroscopy (NIRS) data from wheat grains to predict the quality of wheat grains, flour quality and flour milling yield. The results showed that NIRS data, such as intrinsic properties, were useful in predicting the quality of wheat. Today, most analysis of the intrinsic properties of wheat is conducted by NIRS as it is a reliable, quick and non-destructive method to test the quality of wheat (Caporaso et al., 2018). However, NIRS analysis has shown mixed results for identifying flour quality and flour milling yield. Assadzadeh et al. (2022) focused on predicting flour milling yield. They proposed a Gaussian process regression model based on grain image data and NIRS data. Image data allows physical properties to be described and interpreted, such as grain shape, color and morphology. This model was able to explain the variation of the milling yield variable better than previous work based on a multiple linear regression model (Delwiche et al., 2013). The Gaussian process regression model reached a R-squared equal to 0.61.

Table 2.1 Summary of existing AI research on wheat selection

Author(s)	Objective type	Model	Parameters	Output
(Dowell et al., 2006)	Predict flour quality and yield	PLS (Partial least square) regression	Near-Infrared Spectroscopy (NIRS)	<i>Wheat quality, flour quality and flour milling yield</i>
(Delwiche et al., 2013)	Predict flour yield	Multiple linear regression model	NIRS and physical grain properties	<i>Flour milling yield</i>
(Aydin et al., 2020)	Predict flour yield	ANN (Artificial neural network)	physical grain properties	<i>Flour milling yield</i>
(Assadzadeh et al., 2022)	Predict flour yield	Gaussian process regression model	NIRS and physical grain properties	<i>Flour milling yield</i>

Table 2.1 highlights some representative AI research on wheat selection. The studies relied on intrinsic and physical properties to predict the yield and quality of flour. The intrinsic properties are collected by NIRS. Physical properties are collected from physical tests and images. From the parameters studied, we may note that no author analyzed both the intrinsic properties and the hardness of wheat. Yet, Aydin et al. (2020) showed through his prediction

model that wheat hardness is an important factor in predicting flour yield. Moreover, no author has taken into account the variety and locality of wheat. Finally, the gluten quality of wheat grains measured from the rheological behavior of the dough deserves to be explored to improve the prediction results of the studies performed (Gucbilmez et al., 2019).

2.2.2 Transport / wheat analysis / wheat storage

After wheat grains are selected by the miller, they are transported from the farmer's land to the mill. Depending on the quantity ordered, batches of wheat are received at the mill throughout the year. After it is received, a sample test is conducted to analyze and measure the quality of the wheat grains received. The results of the analysis help to decide in which specific silos (storage tanks) they will be stored, according to their quality level. This grouping allows an homogeneity to be maintained in the quality of the grains in each silo. Before wheat grains are stored in silos, pre-cleaning is conducted to eliminate big pieces of waste (stones, mud, dirt, and so forth) as well as metallic objects that could damage production equipment (IAOM, 2018).

Challenges

The main challenges at this stage of the value chain are related to transportation time and conditions, as well as storage time and condition. Depending on initial location departures and circumstantial events (political, economic, etc.), the transport time varies. This makes it necessary to have a stock of wheat grains that can handle long transportation lead time (TLT) before replenishment. In addition, as wheat is harvested during summer and delivered throughout the year in lots, wheat grains can be stored in a farmer's silo for several months or years. In this context, several problems can arise. First, if wheat is stored for a long period of time in silos in an uncontrolled environment, the quality can fluctuate and deteriorate (Boudreau and Ménard, 1992). This deterioration causes issues in terms of flour quality and yield once the wheat is milled. Second, in terms of stock, as wheat storage is limited by the number of silos, it is necessary at times to load different wheat lots together in a silo due to space constraints. This loading mixture of different quality of wheat lots impacts the

resulting flour quality and yield. According to Boudreau and Ménard (1992), the grouping rule of wheat lots followed are based on wheat class, wheat grade and wheat protein content. A *wheat class* regroups wheat varieties. A *wheat grade* informs the quality of wheat and its overall visual quality. Based on these three pieces of information, a decision is made to store wheat in a specific silo. However, this raises questions about the relevance and appropriateness of these attributes. With today's rapid analysis of intrinsic and physical properties of grains of wheat, additional information could improve decision-making on the grouping of lots of wheat and, consequently, improve the flour manufacturing process.

Progress of work research

The aim of wheat storage is to preserve the quality of a wheat harvest until it is processed. Wheat grains are a living raw material. This means that their quality is affected by environmental factors that can stimulate the presence and growth of microorganisms and insects (Sisman and Ergin, 2011). According to Jia et al. (2001), storage temperatures lower than 15 degrees can prevent insect development. Mite and microorganisms such as fungi can be controlled by reducing wheat moisture content and relative humidity (Jia et al., 2001). In regards to wheat moisture, Campbell (2007) specifies that a moisture content lower than 14% allows the wheat grains to be stored for a long period of time. However, it is not always possible to store wheat in a controlled environment or to dry wheat to reduce the moisture content of wheat. This would lead, in most cases, to an increase in the cost of wheat storage. To overcome this, several studies have examined the effects of wheat properties under different storage conditions. Campbell (2007); Kibar (2015) showed that a combination of high temperatures and relative humidity during wheat grain storage had a significantly decreasing effect on potential flour yield, leading to economic loss. In addition to the flour yield performance metrics, they showed mixed results in terms of dough quality in the bread-making process. However, Posner and Deyoe (1986) showed, contrary to recent studies, an improvement in milling performance and flour quality based on storage time in a stable environment (relative humidity = 68% and temperature = 21 degrees) for the Hard Red Winter (HRW) wheat variety. The study shows that wheat goes through a "sweating process" after harvest that tends to change its intrinsic properties and improve milling and baking qualities. An

increase in flour extraction of 2-5% has been realized when milling HRW wheat after the first five months. As most wheat is not stored in a controlled environment, additional research must be done to evaluate the possibility of increasing the wheat quality for milling by taking into account wheat variety, wheat properties, storage time and fluctuation in environmental conditions.

Other research focuses on developing a method to improve the selection and separation of wheat batches to store in silos. The method is based on the physical properties of the wheat grains. Physical properties are important characteristics that directly influence the production yield and the quality of the flour (Campbell et al., 2012; Posner and Hibbs, 2005; Yoon et al., 2002). The knowledge of these characteristics is useful for the adjustment of the production parameters. A large dispersion of grain sizes from one silo implies a longer processing time, a higher energy cost and eventually a degradation of the quality of the flour produced (Yoon et al., 2002). Therefore, Yoon et al. (2002) used mathematical tools and a clustering technique to better sort the different batches of wheat based on their physical properties such as grain size, hardness and shape. Three cluster groups are used based on wheat grain size: small, medium, and large. The cluster analysis enabled an improvement in the flour yield by 0.13%, which represent a 5% increase in milling income per day, on average.

Table 2.2 Summary of existing AI research on wheat transport and storage conditions

Author(s)	Objective type	Model	Parameters	Output
(Posner and Deyoe, 1986)	Predict flour yield and quality	Regression analysis	<i>Storage time</i>	Milling characteristics and flour quality
(Yoon et al., 2002)	Improve flour yield	Clustering – KNN	<i>Grain size</i>	Wheat lot assign to a cluster
(González-Torralba et al., 2013)	Study input parameters on wheat quality and flour quality	PCA (Principal component selection) and ANOVA (Analysis Of Variance)	<i>Temperature, relative humidity, time of storage</i>	Wheat quality and flour quality
(Kibar, 2015)	Study input parameters on wheat quality	ANOVA and PCA	<i>Relative humidity and temperature in silo</i>	Wheat quality

Table 2.2 provides a summary of existing AI research on wheat transport and storage condi-

tions. It shows that research has focused on variables that consider storage time, environmental conditions, and grain quality during wheat storage. However, there is a noticeable lack in studies on the conditions of wheat grain transportation. Environmental conditions during transport could impact wheat quality. Moreover, the mix of wheat batch made at a grain terminal or in cargo storage to optimize space is another approach to take into consideration to improve flour yield and quality.

2.2.3 Cleaning

At the cleaning stage, a wheat recipe is created to select the different grains of wheat needed according to the miller for flour production. A milling run can begin. The milling run starts with a cleaning process with the purpose of eliminating all foreign materials and impurities remaining with the wheat grains, such as broken grains, dust and straw. Four different machines are usually set up in a flour mill to efficiently eliminate all kinds of waste present in a wheat grain batch (Catterall, 1998; IAOM, 2018). One cleaning machine is dedicated to eliminating foreign material according to its size. A second aims to sort foreign materials that have a mass density that is significantly different than wheat grains. A third seeks to sort broken grains and cereal grains from grains of wheat by their shape. Finally, a fourth uses aspiration to eliminate thin materials that are bonded to the grains of wheat, such as brush and dust. Sometimes, a fifth could be added in addition to the one already presents at the pre-cleaning stage (occurring at the step 2.2.2 in the value chain), to ensure that metal objects are properly eliminated.

Challenges

The main challenge in the wheat cleaning process is ensuring that no residual material remains with the grains of wheat. One control lever at the miller's disposal is to regulate and balance the airflow according to the percentage of foreign materials and impurities present with the wheat grains. The greater the quantity of foreign material and impurities, the more the air flow needs to be increased to efficiently clean the grains of wheat. However, this adjustment is done manually and is approximated by visualizing a sample of wheat grains. This can

lead to a fluctuation in the quality in the wheat flour milling process depending on a miller's expertise. Additionally, the miller can decide, by visualizing the output result of the cleaning stage, to start a second cleaning cycle if necessary. However, an additional cycle is time consuming and implicates additional energy costs.

To the best of our knowledge, no research paper has addressed those issues using artificial intelligence techniques to improve flour quality at the mill. The studies about cleaning grains of wheat are limited to being evaluated by a series of tests, according to particular interest in using a specific cleaning machine or a method for wheat cleaning (Bettge et al., 1989; Magyar et al., 2019; Tibola et al., 2016). Therefore, some ideas will be discussed in section 2.3 for research perspectives.

2.2.4 Wheat blending

Wheat blending is the process of mixing different qualities of wheat. During this process, grains of wheat that have been cleaned are mixed. Wheat blending acts as a lever action to improve the quality of the flour produced (Campbell, 2007). It is necessary when the desired quality of wheat is not directly available. It enables the use of low-grade wheat as a compensating material in response to baking demand when the supply of wheat is insufficient (Hayta and Çakmakli, 2001). To this end, a wheat mixture or wheat recipe is created at the beginning of a milling run. A milling run covers the main production process required to make flour. It includes the cleaning process, the tempering process and the milling process described in the following subsections 2.2.3, 2.2.5 and 2.2.6. The wheat blending is done between the cleaning and tempering stages. For each milling run, production parameters must be adjusted. The parameters are adjusted not only according to the quality of the selected wheat grain when mixing, but also according to the type of flour desired by the customer. The wheat mixes and the adjusted production parameters together form a *production recipe*.

Challenges

Wheat mixing is a good lever action for improving and reaching the quality of the flour

desired (Campbell, 2007). According to Posner and Hibbs (2005), it is estimated that 75% of flour quality is determined by the quality of the wheat and 25% by the milling technology, adjustment of mill parameters and environmental conditions in the mill. This makes wheat blending a necessary step in reaching the quality of flour desired. Many factors influence the decision of making one specific wheat mixture at a specific time. Those factors include wheat grain price, wheat grain quality, wheat grain availability and the type of flour desired. It is difficult to determine the best blend and mix ratio to use to produce the right quality of flour at the best output yield. The blend of certain types of wheat could negatively impact the flour quality and yield if, for example, certain grains of wheat need different treatment conditions. The miller, using their knowledge and expertise, is in charge of creating the right blend taking into consideration the different factors that were stated. The miller will adjust the milling parameters by trial-and-error during the mill run. This trial-and-error process occurs through the analysis of a machine's performance.

Progress of work research

The price of organic wheat grain varies according to quality and demand. Wheat recipes are therefore a variable unit production cost (\$/Ton) for a company. To improve the profitability of a company, wheat recipes are the subject of scientific research. Hayta and Çakmakli (2001) worked on a wheat recipe optimization model. The objective of their model is to minimize the cost of wheat recipes while respecting a satisfactory quality level of flour production adapted to the bread-making process. For this purpose, they build a linear optimization model. In their model, they select the volume of the bread as the variable measuring the bread quality. Depending on the volume of the bread, they perform a multiple linear regression analysis to select the variables that can best explain the volume of bread. To do this, they use the best subset regression method. Among the 8 possible explanatory variables describing characteristics of the quality of the wheat grains, 3 are selected. The 3 variables selected explain the bread volume relatively well, with an R-square equal to 0.77. More recently, Steffan (2012) focused his research on building an optimization model for blending wheat and additives that both meet the needs of its customers, while being the least expensive. His study is limited to minimizing the production cost of a single type of flour, intended for

the manufacture of French-style bread (baguettes). Since additives are present, this study only concerns the production of conventional flour. In his method, several linear regression models are used to predict different output variables that are associated with flour quality, which are specific to bread production and bread volume. The R-square value obtained by the linear regression models fluctuates between 0.41 and 0.93 depending on the variable output to predict. The linear regression equations will then be used as constraints in the optimization model to respect the quality of flour to be produced for baguettes. Finally, Elevi et al. (2017) seek to optimize the cost of wheat recipes for all types of flour. They estimate flour quality by calculating the weighted average of each wheat grain quality measurement based on the wheat ratios used in a future recipe. For example, the moisture content of the flour is estimated by the weighted average of the moisture content of the different batches of wheat grains used. However, they do not specify the performance or accuracy of their mathematical model to estimate the quality of the flour produced. Elevi et al. (2017); Steffan (2012) both studied wheat blending and flour blending. Milling wheat separately and mixing the final flour at the end of the process is a possibility. This method benefits the miller by responding quickly in a "just-in-time" mode for customer orders (Campbell, 2007). However, according to Campbell (2007), blending flour requires additional space and investment in equipment. He also explains that, in practice, it is much easier to blend wheat than flour.

Table 2.3 Summary of existing AI research on wheat transportation and storage conditions

Author(s)	Objective type	Work frame	Model	Parameters	Output
(Hayta and Çakmakli, 2001)	Minimizing cost (predict flour quality)	Flour for bread	Linear regression model	<i>Particle size index, dough volume and falling number</i>	<i>Bread volume</i>
(Steffan, 2012)	Minimizing cost (predict flour quality)	Conventional flour for baguette	Linear regression model	<i>Quantity of soft wheat, hard wheat, gluten, ascorbic acid and enzyme mixes</i>	<i>Flour quality and bread volume</i>
(Elevi et al., 2017)	Minimizing cost (estimate flour quality)	All types of flour	Average	<i>Moisture content, gluten, gluten index, sedimentation, delayed sedimentation, falling number</i>	Cost of wheat recipe

Table 2.3 highlights existing AI research on wheat transportation and storage conditions. It shows that studies have focused on minimizing the cost of wheat blending while resulting in a satisfactory flour quality. However, none of these studies included the adjustment of parameters at each step in a mill run. Those adjustments must be made according to the wheat blend.

2.2.5 Tempering

The milling run continues with a tempering process. Tempering is the process in which water is added to the grains of wheat, followed by a period of rest in an empty silo to let the water penetrate the wheat (IAOM, 2018). The tempering process takes place in two and sometimes three cycles for some types of grains. The number of cycles depends on several factors, such as wheat grain quality and environmental conditions. The tempering aims to moisten the wheat to facilitate the separation of the endosperm from the rest of the grain component in the milling process. This is explained by the fact that water absorption by wheat grains will increase wheat bran flexibility and soften the endosperm (Boudreau and Ménard, 1992). When manufacturing white flour, the goal is to extract as much endosperm as possible from the grain without any bran contamination. Tempering during milling helps

separate the endosperm from the bran, which as a result, improves flour quality. Moreover, tempering helps compensate for the loss in moisture that occurs during milling operations. As flour prices are fixed according to quality and weight, there is an economic interest in increasing the final flour moisture content.

Before milling, it is suggested that the wheat moisture should be raised to about 16% (Campbell, 2007). A tempering stage is thus necessary as the moisture content of wheat grains is mostly kept below 14% during storage for shelf-life reasons (Campbell, 2007). However, an increase in the moisture content of wheat grains is not proportional to the water quantity or water flow used. Depending on the quality of wheat, its specific variety and environmental conditions, grains of wheat do not all absorb water the same way. Therefore, adjustments must be made during the tempering process to reach the moisture content desired.

Challenges

The adjustment of the tempering process parameters is difficult as they are influenced by many factors. According to IAOM (2018) those factors include : wheat properties, ambient moisture, ambient temperature, resting time and tempering bin space available. Furthermore, the wheat mixes made before the tempering stage add complexity to the control of the process. The miller, according to his experience and knowledge, is in charge of adjusting tempering parameters. Depending on the type and quality of flour desired, the miller sets a target moisture content. This is the desired moisture content that needs to be reached by the end of each tempering stage. The target moisture content is used to calculate (by a mathematical formula) the water flow rate required to temper the wheat grains according to the wheat grain flow rate and the initial wheat moisture content (IAOM, 2018). Other parameters controlled by the miller are the flow of wheat grains, quantity of wheat, the duration of the resting time and the number of tempering stages.

In this context, many different factors must be considered to adjust the tempering parameters efficiently and to reach the wheat moisture content desired. Depending on the wheat moisture content reached at the end of the tempering stage, flour yield and flour quality are impacted

during the milling process. If wheat grains are not sufficiently moistened, the wheat grains are more crumbly during the milling process. This results in bran contaminating the flour at the end of the milling process. On the contrary, if the grains are too humid, the wheat grains are stickier. This will make flour extraction difficult as wheat grains will not flow well and will be inefficiently sifted (IAOM, 2018). Thus, an optimal wheat moisture content must be reached to efficiently extract quality flour with high yield.

Progress of work research

Several studies have investigated the wheat tempering stage to better understand the variables of interest in this stage. Hook et al. (1982a); Kweon et al. (2009); Warechowska et al. (2016) have studied different wheat properties and process parameters to evaluate their influence on flour yield and flour quality. Their method relies on statistical analysis to evaluate the influence and interaction of several independent variables on a single dependent variable. Using this method, they showed, for example, the existence of a relationship between moisture content and flour quality as well as flour yield. When the moisture content of the wheat increases, the yield decreases while the flour quality increases. Therefore, a good compromise has to be found, especially for the production of white flour which should not be contaminated with bran. Moreover, Hook et al. (1982a) show that hard wheat is more tolerant to the addition of water, which results in less of a decrease in flour yield compared to soft wheat. In addition to the impact of wheat tempering on flour quality and yield, Warechowska et al. (2016) show that an increase in wheat moisture content for hard wheat softens the wheat but leads to more energy required during the wheat flour milling step. This results in higher production costs. This variable must be considered in order to optimize a mill's profitability.

There is research that is limited to the production of whole wheat flour. Whole wheat flour differs from white flour because of its higher proportion of bran residue. Bran residue is made up of particles of wheat grains that come from the bran and the germ. In this context, Doblado-Maldonado et al. (2013) studied the relevance of the tempering stage, when there is no real interest in efficiently separating the different wheat components. To this end, a two-way ANOVA (analysis of variance) is used to compare and measure the effect of different

tempering treatments and wheat varieties on the wheat particle size distribution obtained from the milling process. In this study, the flour quality is based on particle size distribution. Significant differences are noted as a result of the tempering treatments prior to milling. Much larger particles are produced after one milling step when wheat grains have undergone a tempering stage. According to Doblado-Maldonado et al. (2013), a higher quantity of thin particles is related to good bread quality, which is explained by a higher bread volume. If wheat is tempered, it is necessary to reduce those large particles to produce whole wheat flour for bread-making by performing additional milling steps. Milling wheat grains without any tempering could avoid those additional milling steps that are time and energy consuming. Furthermore, the study showed that flour milled from wheat grains that did not undergo a tempering stage had the same level of quality as expected in commercially milled flour. However, the study is limited to one milling step, while in industrial wheat flour milling, several milling steps exist. Moreover, Noort et al. (2010) explains that a higher quantity of thin bran particles is not always related to good bread quality. Bread quality is influenced by the intrinsic and physical properties of wheat, wheat mixing and bread baking procedures (Noort et al., 2010). Additional research needs to be done to compare flour quality produced from wheat tempered grains that follow several milling steps in an industrial mill.

Other research has focused on the stone milling process, which is a different method of wheat milling. In this context, Cappelli et al. (2020a) studied the effect of two parameters, which include stone milling speed and wheat moisture, on the operative milling performance, flour quality and bread quality. A two-way ANOVA is used to measure the influence of the two parameters with four grades of moisture and three grades of rotational speed. They show that wheat tempering had a significant effect on flour quality and bread quality. However, wheat mill productivity decreases as the moisture content of wheat increases. They found that a 13% wheat moisture content is the best compromise between milling overall performance and flour quality for bread-making. On the other hand, at moderate or high moisture content, they found that rotational speed did not significantly influence flour yield, although it is maximized by the lowest rotational speed grade.

Table 2.4 Summary of existing AI research on wheat tempering

Author(s)	Objective type	Model	Parameters	Output
(Hook et al., 1982a)	Study the effect of added water on milling performance for different levels of wheat hardness	Multiple regression analysis	<i>Wheat hardness, wheat moisture content</i>	<i>Flour yield and Flour color</i>
(Kweon et al., 2009)	Study the variable effect on flour yield and flour quality	Statistics: Full factorial design	<i>Initial wheat moisture, tempered wheat moisture, tempering temperature, tempering time</i>	<i>Flour yield and flour quality</i>
(Doblado-Maldonado et al., 2013)	Study the impact of omitting tempering stage for wholegrain flour	Two-way ANOVA and One way ANOVA	<i>Wheat variety, tempering treatment (dried, as is, tempered)</i>	<i>Particle size distribution, flour quality and baking quality</i>
(Warechowska et al., 2016)	Study the variable effect and pattern between input and output variable	Analysis of variance	<i>Initial wheat moisture, tempered wheat moisture, wheat variety</i>	<i>Energy milling, flour yield and flour quality</i>
(Cappelli et al., 2020a)	Study the effect of stone milling speed and wheat tempering on milling efficiency, flour and bread quality	Two-way ANOVA	<i>Stone rotational speed and wheat tempering</i>	<i>Operative milling performance, flour quality and bread quality</i>

Table 2.4 highlights the existing AI research on wheat tempering. The research shows that wheat tempering has an impact on the milling process on flour yield and quality. This impact is reduced when milling whole wheat, as there is no need to efficiently separate the different wheat components. Whole wheat flour yield is therefore not affected but the quality is. This is explained by the fact that wheat tempering tends to produce large bran particles and small endosperm particles when milled (Campbell, 2007; Fang and Campbell, 2003b). The particle size dimension is a key characteristic of flour quality that will influence the product quality after baking or bread-making (Lin et al., 2020; Tóth et al., 2006).

Although every research shows interest in tempering wheat grains for flour yield and quality, none of the research has studied the process itself. None of the studies addressed the control

of tempering process parameters, which depends on several factors such as environmental conditions and input quality. Moreover, additional research has to be done concerning the relationship between the particle size dimension of wheat bran and whole wheat flour quality as several research studies have had contradictory results (Doblado-Maldonado et al., 2013; Noort et al., 2010).

2.2.6 Milling

The milling run ends in the milling process. The milling process breaks the grains, separates the different parts of the grain and reduces the endosperm into flour. The milling process is divided into 4 stages: grinding, sifting, sizing and reduction. *The grinding operation* is always the first stage of the milling process. It aims to break the grain into large particles by using fluted rolls. These large particles are called semolina. At the end of the grinding operation, the particles go to the *sifting operation*. At this stage, the particles are separated and sorted according to their size using several sieves located in a machine called a planischter. The objective of sorting the grain is redirecting the different particles to a new milling operation or to silos that store flour and bran. *The sizing operation* is to reduce the white semolina (large particles of the endosperm). *The reduction operation* aims to reduce the size of coated semolina. The coated semolina is described as large particles where most of the endosperm is still attached to the bran. The last two operations are realized by smooth cylinders. The particles created at the end of each grinding, sizing and reduction operation go once again to a new sifting operation. The operations follow one another until the entire set of particles is converted into flour or bran residue. The 4 stages are carried out successively, which is why several machines are present in the mill. The process follows a progressive reduction approach to extract the maximum amount of endosperm particles and to reduce them into flour.

The objective of wheat milling, as mentioned previously, is to extract the maximum quantity of flour of a specific quality from the wheat grains. *The quantity of flour extracted from the wheat grains represents the production yield*. It is most often referred to as the extraction rate in the flour industry. Formula (2.1) explains the method of calculating the production

yield:

$$Production\ yield = \frac{Flour\ Weight * 100}{Weight\ of\ clean\ and\ conditioned\ wheat} \quad (2.1)$$

Maximum production yield is sought to maximize profits due to a higher price of flour compared to its by-products, such as bran geared toward the animal industry (Pagani et al., 2014). Endosperm extraction reaches its theoretical maximum at around an 81-84% grain extraction rate as endosperm represents about 84% of the wheat grain weight. However, the variation in wheat grain quality and operating conditions make it difficult for the miller to achieve this extraction rate (IAOM, 2018). The yield generally achieved in the flour industry is actually around 72% (IAOM, 2018). Above this percentage, the flour becomes increasingly contaminated with bran residue. Therefore, a trade-off between flour production yield and quality exists to ensure the economic profitability of the mill.

Challenges

Based on the production equipment installed and the quality of the grain, production parameters must be controlled and adjusted according to terms of wheat mixture, cleaning, tempering, and milling process to produce organic flour. The adjustments ensure optimal production yield and quality when there is variation in the quality of the raw material. However, the number of parameters and their influence on the whole manufacturing process make each adjustment complex.

Progress of work research

Some authors have analyzed the production systems present in the flour mill (Campbell et al., 2001, 2012; Campbell and Webb, 2001; Cappelli et al., 2020b; Dal-Pastro et al., 2015, 2016; Fang et al., 1998a; Stefan et al., 2018). Production parameters influence the quantity and quality of flour extracted from wheat grains (Campbell, 2007). Among flour production operations, milling operations are at the heart of the production system. Most mills use rollers to grind wheat grains in order to achieve higher yields. However, rollers need to be well-configured. To this end, Campbell and Webb (2001) analyzed the operation of roller mills

and their parameters. They group the input factors and parameters affecting the grinding process into 3 sets. The 3 sets include: raw material properties, roll design and roll operation. The raw material properties explain the quality of the wheat grains. The roll design is defined at the time of purchase and is considered to be fixed production parameters. Roll operation factors account for the current state or settings of the rolls during production. They include roll arrangement, roll speed, roll speed differential, roll gap distance and roll wear.

The flour particle size distribution generated by the roller is an indicator of the flour quality and performance of the milling process (IAOM, 2018). To predict the particle size distribution at the first grind, Campbell and other co-authors developed a mathematical model (Campbell et al., 2001; Campbell and Webb, 2001; Fang and Campbell, 2003a). The equation calculates the particle size distribution of the particles generated at the end of the first grinding operation according to the wheat grain size and the roll gap distance. Campbell et al. (2007); Fang and Campbell (2003b) then improve the equation to take into account the moisture and the hardness of the wheat grains, which affect the flour particle size distribution and, consequently, the flour quality (Dal-Pastro et al., 2016).

Dal-Pastro et al. (2016) goes further and takes an interest in several crushing rollers as well as the plansichter. They seek to predict the particle size distribution of the flour obtained after milling operations, reduction operations and sorting operations. The prediction models are based on a partial least square regression method. They are based on 3 inputs: roll gap distance (μm), wheat grain flow (kg/h) and the moisture content of the wheat grains (%). The data is obtained using a factorial design of experiment 2^3 (DOE - Design of Experience) including the 3 parameters mentioned. Dal-Pastro et al. (2016) show from their results that the roll gap distance is the parameter with the greatest impact on the particle size distribution followed by moisture content and wheat grain flow.

Work has shown the relevance of using a neural network model to predict the quality of wheat particles produced after the first milling operation (Fang et al., 1998a). Fang et al. (1998a) confirm the importance of the 3 explanatory variables used by Dal-Pastro et al. (2016) for the prediction of flour particle size distribution. Furthermore, they show that average kernel

size, average kernel weight, and average kernel hardness are important variables for accurately predicting the quality of wheat particles obtained after the first milling. According to their results, the neural network model is able to correctly predict the average particle size at the first milling stage ($R^2 > 0.99$).

Table 2.5 highlights existing AI research on wheat milling. Several prediction models have been developed to predict the quality of particles generated at the output of production equipment. However, most authors only try to predict the quality of the flour particles generated from milling experiments in a laboratory. They rely on simulation data (Fang et al., 1998a) or experimental data (Campbell and Webb, 2001; Dal-Pastro et al., 2016; Fang and Campbell, 2003a). Analyzing real production data on the whole process to improve the quality of the flour would be an interesting approach in addition to the studies done. The quality of the flour could be influenced not only by the milling process, but also by the previous production steps including the wheat blending step, the cleaning step and the tempering step. Moreover, operating conditions and equipment wear have been infrequently studied in flour quality prediction studies. These fluctuating conditions in the mill have an impact on the quality of the flour and the yield of the production. Hook et al. (1982b) showed, by performing a series of milling experiments, an increase in flour yield as roll temperature increases due to successive millings. The use of data describing environmental and production constraints could provide additional and relevant information to the flour quality prediction model. The prediction models would therefore learn from real experiences and be able to adapt production parameters according to current operating conditions.

Table 2.5 Summary of existing AI research on wheat milling

Author(s)	Objective type	Model	Parameters	Output
(Fang et al., 1998a)	Predict the ground wheat quality (flour particles) and milling performance depending on wheat properties and operational parameters	ANN	<i>Wheat moisture content, hardness, size, weight, feed rate, roll speed, roll speed differential, roll gap</i>	<i>Geometric mean diameter (GMD), specific surface area increase (SSAI) and break release (BR)</i>
(Fang et al., 1998b)	Predict the power and energy requirements for size reduction of wheat (milling) depending on wheat properties and operational parameters	Stepwise regression	<i>Class of wheat, moisture content, feed rate, fast roll speed, roll speed differential, roll gap, wheat hardness, wheat weight</i>	<i>Fast roll speed, slow roll speed, net power, energy per unit mass and specific energy</i>
(Campbell et al., 2001; Campbell and Webb, 2001)	Predict particle-size distributions from First Break milling of wheat	Breakage function	<i>Roll gap and wheat size particle</i>	<i>Particle-size distributions of flour</i>
(Fang and Campbell, 2003a,b)	Predict the outlet particle-size distributions from First Break milling of wheat	Breakage function	<i>Roll gap, wheat size particle, Wheat varieties and roll disposition</i>	<i>Particle-size distributions of flour</i>
(Campbell et al., 2007)	Predict the outlet particle-size distributions from First Break milling of wheat	Breakage function	<i>Roll gap, wheat size particle, Wheat varieties, roll disposition and wheat hardness</i>	<i>Particle-size distributions of flour</i>
(Dal-Pastro et al., 2015)	Predict the particle size distribution of the milled material at the end of the second break stage	Partial least squares (PLS) regression	<i>Roll gap, Wheat flow rate, wheat moisture content, ambient humidity and ambient temperature</i>	<i>Particle-size distributions of flour</i>
(Dal-Pastro et al., 2016)	Improve process understanding and predict the particle size distribution of the milled material and of the sieved material	Multivariate statistical models	<i>Roll gap, Wheat flow rate and wheat moisture content</i>	<i>Particle-size distributions of flour</i>
(Kalitsis et al., 2021)	Optimize the process to produce flour to maximize flour yield while meeting quality constraints	Response surface methodology (RSM)	<i>Wheat conditioning level, first break roll gap and second break roll gap</i>	<i>Ash content, protein content, moisture content and flour yield</i>

2.2.7 Organic wheat flour end product

Different types of flour are available on the market. Each type of flour manufactured has specific qualities sought after by a client to produce a specific end product. A list of possible end-products are presented in Figure 2.3.

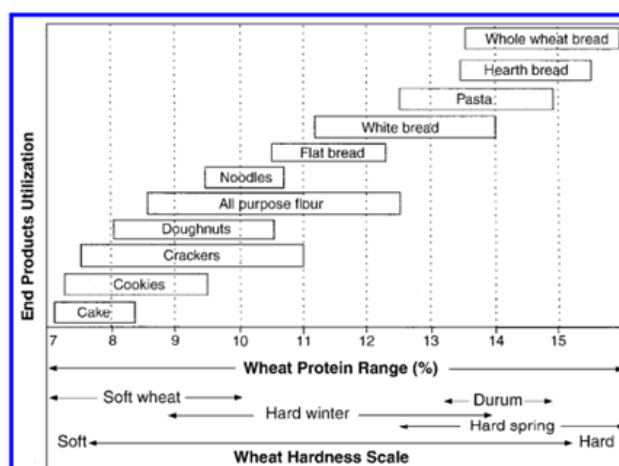


Figure 2.3 Different end-product utilization, depending on type of wheat, protein percentage, kernel hardness (Posner and Hibbs, 2005)

Figure 2.3 illustrates in a graph the different type of wheat and quality that are suitable for producing a specific food product. Depending on end product utilization, it is important to select wheat grains based on their intrinsic and physical properties such as protein and hardness. Then, from the selected grains, the production parameters will influence the types of flour that can be made.

Figure 2.4 illustrates in a diagram the different types of flour produced according to their ash content. The classification of wheat flour as described in Figure 2.4 extends from T45 to T150 flour. This number scale indicates the ash percentage of the wheat flour. It informs of the amount of mineral matter in the flour and influences its color (Oliver et al., 1993). A low ash content indicates a white flour, mainly composed of endosperm particles. As the ash content increases, the flour becomes more like integral flour, which includes all of the particles from the wheat grains. This becomes more nutritionally rich.

To ensure the client receives the right quality of flour as desired, a flour sample is collected

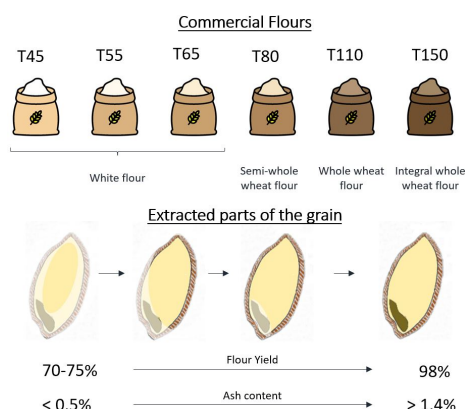


Figure 2.4 The types of wheat flour according to ash content

and analyzed at the end of the milling stage. The result of the analysis will control the quality and bring additional information about the behaviour of the future wheat dough. This information is useful to the client when adjusting the fermentation and bread-making process to produce a final product that has constant quality.

Challenges

Most of the tests are conducted at the end of the wheat milling process to measure wheat flour quality. By following the wheat flour milling process that was previously described, it is often too late and difficult to make changes to the production process to achieve the desired level of flour quality. In this context, data analysis approaches using AI tools have been used to improve the bread-making process and to produce an end product that has consistent quality.

Progress of work research

The flour is used as a raw material in many processing operations. The flour quality has an effect on the quality of the final processed product, especially in bakery products. Several studies have focused on the process of bread-making. They studied the quality of flour and its impact on the quality of bread manufactured (Banu et al., 2011; Dowell et al., 2008; Parenti et al., 2020a,b; Torbica et al., 2019). The quality of the flour influences the fermentation process as well as the rheological properties of the bread dough, which are specific to each

customer (Cappelli et al., 2019). Studies on enzymes that act as flour improvers will not be covered in this article. Although ingredients or improvers could be added during the bread-making process to improve the dough quality, we limit our field of study to the production of organic end products made from wheat flour.

To optimize the bread making process, some research focuses on the prediction of bread quality based on rheological quality measurement data of dough (Dowell et al., 2008; Ktenioudaki et al., 2010; Oliver and Allen, 1992). From the predictions made, flour mixture and bread making process methods seek to be improved to reach a consistent and higher quality of flour used. Three models use a regression method to evaluate the relevance of variables. Dowell et al. (2008); Ktenioudaki et al. (2010) based their regression model on flour properties and dough properties to predict the bread quality, which depends mainly on the volume of the loaf and crumb hardness. Their models reach a good predictive accuracy of bread quality with a R-square of 0.8. From their model, they were able to identify certain influential variables of flour properties, such as: protein content, falling number and ash content. Among dough properties, they found that dough tenacity and dough extension were highly relevant as well. Apart from the regression model, Abbasi et al. (2011) built a neural network model to predict dough properties based on seven input variables that describe flour properties. The artificial neural network (ANN) is optimized by a genetic algorithm. The model tries to reproduce the same results obtained from a *Farinograph* instrument. This instrument provides valuable information about flour quality and dough properties to the baker, who seeks to optimize production control parameters and the final product quality (IAOM, 2018). Their model obtains satisfactory results with a R-squared value higher than 0.96 and a normalized squared error lower than 0.09. Their work would allow flour manipulations to be avoided, and thus save time in the manufacturing process. Finally, Cappelli et al. (2019) limited their study to whole wheat bread. Whole wheat flour is nutritionally rich but leads to rheological problems such as: lower loaf volume and increased crumb density. To solve these issues, Cappelli et al. (2019) analyzed the bread making process. They conducted an analysis of variance to measure the effect of two controlled parameters during the bread making process on the dough and bread quality. The parameters included the percentage of bran and middling

content added, and the time it was added during the process. They found that the addition of bran and middling to white flour at different times in the process influences wheat dough and bread quality. From their results, the delayed addition of bran and middlings during the bread making process showed improvements. It led to better dough extensibility and a greater loaf volume of the bread. This is a good solution for bakers who need to improve their organic end product quality.

Table 2.6 Summary of existing AI research on organic wheat flour end product

Author(s)	Objective type	Model	Parameters	Output
(Dowell et al., 2008)	Predict bread quality	Regression model	Wheat, flour and dough properties	Bread properties
(Ktenioudaki et al., 2010)	Predict bread quality from different varieties of wheat	Stepwise multiple regression	Wheat varieties and rheological properties	Baking quality
(Różyło and Laskowski, 2011)	Predict bread quality	Stepwise multiple regression	Wheat flour and dough properties	Bread properties
(Abbasi et al., 2011)	Predict dough quality	Artificial Neural Network (ANN)	Wheat flour properties	Dough properties
(Cappelli et al., 2019)	Improve wheat dough and bread quality from kneading methods	ANOVA	Bran and middlings content and addition times	Dough and bread quality

Most of the studies base their mathematical model on either flour or dough properties as input variables. Only Dowell et al. take into consideration the wheat quality selected. Further research could be conducted to analyze the relationships between manufacturing process parameters, including wheat blends and their effect on wheat flour, dough and bread quality. According to Dowell et al. (2008), milling can affect dough and baking properties. Although the studies presented in Table 2.6 are useful in addressing variations in the rheological properties of wheat flour dough, it would be interesting to find solutions upstream in the value chain to achieve consistent and satisfactory flour quality from the milling process. Customers would then no longer have to mix different flours or adapt their bread making process each time.

2.2.8 Discussion

The literature review demonstrates that several authors have studied different parts of the wheat flour production value chain to improve quality or yield. The studies were grouped according to their intervention point in the value chain. We note that little research has focused on the impact of several processes in the value chain on flour quality and yield. Only Dal-Pastro et al. (2016) take into account several milling steps (grinding and sieving) in a data-driven approach. Moreover, Cappelli et al. (2020a); Kalitsis et al. (2021) are the only ones to consider both tempering and grinding parameters. Yet the parameters and factors of one process affect the next. Moreover, most research that focuses on milling operation is conducted using experimental data in a specific environment. Little research has been conducted considering environmental and production constraints as inputs to better control and improve flour yield and quality at the end of the wheat flour milling process.

From the literature review, it is possible to note that most authors limit their work to statistical analysis or prediction algorithms. The objective of their work is to better understand each process and the importance of specific parameters on the quality and yield of the flour. However, none have proposed decision support tools that could help the miller and operators control and improve flour quality and yield according to the configuration and factors present in the mill. Only 3 research studies were interested in proposing decision support and optimization tools. These studies that focus on wheat blending are found in section 2.2.4. They develop tools for suggesting the best wheat blend ratio at minimum cost while meeting the customer's desired quality. Further studies are still necessary as these models do not take into account the production process of a milling run that includes cleaning, tempering and milling. As a result, the expected quality and yield of the flour cannot be guaranteed.

With regard to the number of studies conducted on organic wheat flour milling at each point of the value chain, two processes are relatively unstudied, if at all. They include wheat blending and wheat cleaning. Both areas should be explored as they both impact flour quality. Wheat blending, which is considered a major lever for improving flour quality as well as mill profitability, is even more important.

Given the analysis conducted on quality and yield in this section, section 2.3 will address other aspects of optimizing the wheat milling process.

2.3 Research areas towards smart wheat milling

The aim of this section is to propose several research directions that combine AI and the organic wheat milling sector to optimize the wheat flour milling process. In the milling process, there are other dimensions besides flour quality to ensure optimal mill operation and profitability, such as energy savings, resource optimization and mill performance. Each dimension includes key objectives that can be achieved and solved using AI to optimize the milling of organic wheat flour. These key objectives are identified based on the literature review, observations and experiments conducted in an organic flour mill over a two-year period and inputs from wheat milling experts. Figure 2.5 shows a mind-map constructed from the 4 dimensions and the key objectives identified.

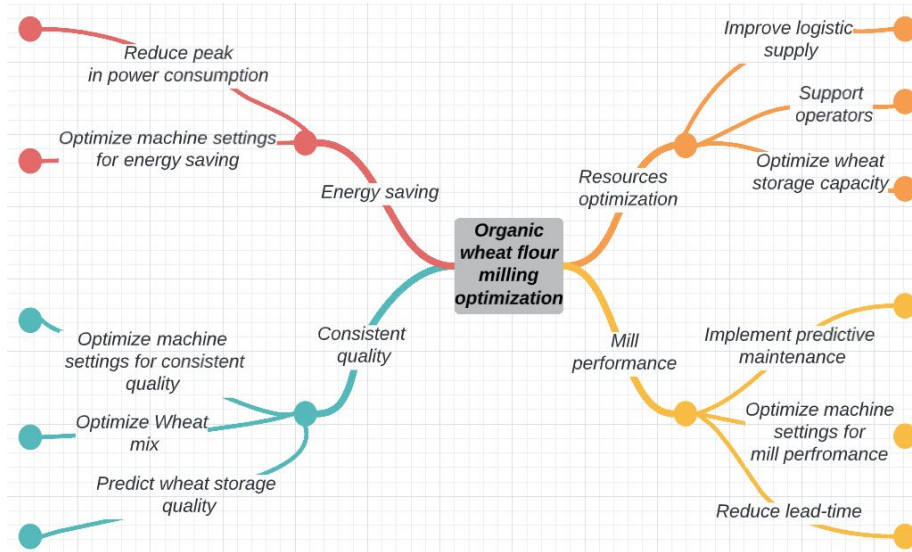


Figure 2.5 Mind map for future topics research centered around the optimization of organic wheat flour milling

Four dimensions are represented in Figure 2.5, illustrated by the different color branches. The sub-branches represent the key objectives to be addressed in a specific dimension. To ensure the consistent quality of organic flour, it is important to properly adjust machine parameters, wheat recipes and to know the quality of the raw material stored in inventory.

To save energy and increase mill profitability, the reduction of peak consumption and the adjustment of machines are avenues to explore. Resource optimization is possible through improved logistical efficiency, decision support for operators and better management of wheat inventories. Regarding mill performance, the implementation of a preventive maintenance approach based on data collected in real time, the adjustment of production parameters and strategies to reduce operating time using AI techniques offer solutions to optimize this dimension. The actions of the key objectives can be involved in several dimensions, depending on the final objective set. This is, for example, the case for the optimization of the machine set-up, in which we try to find the best parameters to meet our objective. Each of the four dimensions will be explained in more detail in the following subsections.

2.3.1 Energy savings

In order to reduce production costs and improve mill profitability, the authors have explored ways to optimize energy requirements to efficiently grind wheat grains (Fang et al., 1998b; Gómez Sarduy et al., 2018; Warechowska et al., 2016). The authors all rely on linear regression models to predict the energy consumption of the roller mill based on process parameters and wheat properties. Among the input variables studied, Fang et al. (1998b) notice that wheat hardness is the most useful variable from the physical properties variable to predict energy requirements for grinding. On the other hand, the feed rate of wheat grains and the roll gap are the most significant process parameters. In addition to those input variables, Warechowska et al. (2016) showed that the tempering process also influences the energy used for grinding. An increase in moisture content of the wheat grain results in a higher energy requirement. This is explained by the fact that a change in moisture content implies a change in wheat grain hardness. Although moisture content has been shown to influence yield and quality, no study has taken into account the energy required, yield and quality together to optimize the milling process. Moreover, no study has been done on the energy used according to the seasons and the external environments that influence the properties of wheat grains. Based on energy input, weather data and process parameter settings over time, AI can detect patterns and provide a decision support tool to **optimize machine settings** to efficiently grind grains of wheat and save energy. In addition, there are energy peaks in the wheat

mill when the grinding equipment starts. This causes significant energy consumption. With a better estimation of the operation time and the amount of flour required from AI models, it would be possible to schedule the production in order to **reduce peaks in energy consumption**. Finally, high energy consumption could be explained by poorly adjusted parameters or by equipment wear. This could be a good indication for preventive maintenance and should be explored further.

2.3.2 Consistent quality

Consistent wheat flour quality is the most overarching facet of the wheat flour milling process. It has a direct impact on a client's needs. Through the literature review, however, we note the absence of research on the wheat cleaning process. AI computer vision models could provide a solution to detect the percentage of foreign material and improve the cleaning process parameters and the number of cycles required. We also note that there are currently no decision support tools for either the milling process or the tempering process to assist the miller and operator in achieving their flour quality goals. One of our research objectives will be to develop a support tool from AI models to adjust the tempering parameters and to efficiently achieve the desired wheat moisture. Furthermore, none of the studies have attempted to predict flour quality based on both wheat blend ratio and process parameters. From historical production data, AI models could be trained to predict wheat flour quality at the end of milling. The objective would be to **optimize and suggest a wheat mix** to reach a specific quality of flour depending on the factors and parameters present at the mill. This will mark our second research axis. Similarly, AI models would provide solutions to **optimize milling parameters** based on grain quality, surrounding factors and desired quality while maximizing yield. In this context, our objective will be to develop a decision support tool that maximizes both the quality and yield of flour production. For the optimization of the milling process, other optimization models taking into account resources and energy related to quality would also be relevant to develop. Finally, a more accurate estimation of the quality of wheat grains stored for several days, weeks or months in a silo would offer more precision when blending wheat for a specific milling run and thus on the flour quality produced. In addition to the work already done on the subject, it would be interesting to **analyze and**

predict the quality of the grain stored in a silo using an AI model by taking into account wheat variety, the storage time and the indoor environmental conditions.

2.3.3 Resource optimization

One of the least discussed aspects is resource optimization. Resources include time, material, as well as labor. These resources must be optimized to reduce production cost and waste. To do this, several avenues can be explored. One is to examine how to properly set the equipment parameters for each milling run. This would save production time and material. A second is to look at the supply chain in order to transport the right amount and quality of wheat grains to produce a certain type of flour. This would reduce lead time production and optimize wheat storage capacity. Another research axis is to look at how to support workers. Workers' knowledge and skills are needed to handle this equipment and manage the mill operations. The technical nature of the operations makes the profitability of the mill dependent on expert and vulnerable to high worker turnover. AI could help optimize the different resources. Based on data collected, it could offer **decision support tools**. It would then help workers make better decisions to optimize resources. The AI models developed and used at the mill would help retain the knowledge within the mill. the data automatically acquired to feed the AI models would reduce the number of manual data collection tasks performed by workers to make decisions. This would optimize the time and number of workers managing milling operations. By analyzing travel times, economic and weather events, estimates of wheat grain supply could be made from AI models to better schedule production and **improve logistic supply**. Finally, with a better forecast of the demand for flour, information on inventories, information on wheat quality and estimates of grain arrivals, solutions to **optimize wheat storage capacity** at the mill would be possible.

2.3.4 Mill performance

Mill performance is primarily related to flour yield based on milling parameters and wheat properties. However, while the extraction rate is a good indicator of the mill's production performance, it is important to look at the mill's overall operation efficiency through the

OEE (Overall Equipment Effectiveness) performance indicator. The OEE indicator provides insights into machine availability, operation performance and production quality. The interest would then be to maximize the OEE indicator and use an AI model for this purpose. Another point is that the studies have been conducted in a controlled environment. No research has been done on how to optimize parameters to minimize operation time without affecting the quality. This would offer an opportunity to **reduce production lead time**. Also, to our knowledge, no research has been done on equipment maintenance. The equipment degrades after each wheat milling. This influences the adjustments of production parameters and can cause unplanned production stops. AI models could be used to learn **optimal adjustments** based on the condition of the tools and perform **predictive maintenance**. Finally, the limits on laboratory experimental data are that they do not take into account environmental conditions that impact flour yield. They influence the optimal parameters in producing a specific flour.

2.4 Conclusion

This paper reviews the state of the art of AI techniques applied in the organic flour industry to improve quality and yield. It aims to identify future trends to optimize the milling process. Some authors have focused their work on wheat flour milling. With the popularity of organic products among consumers and the emergence of data science over the past several years to address industrial issues, it seems important to review the literature on the use of AI techniques in the flour manufacturing process applicable to the organic sector. In view of the literature review, the tools used in AI are mainly centered around statistical analysis and regression models. Their models focus on understanding the system and their parameters in order to have better control. Deep learning models could offer good perspectives on optimizing several processes that no longer have linear relationships. In order to optimize the wheat milling process, four major research areas emerge: energy savings, consistent quality, resource optimization and mill performance. The new research areas identified in this paper are promising and future work will focus on these leads. This research will enable the development of wheat mills that are more connected and smarter through the use of AI

in optimizing the process for a specific purpose.

2.5 Acknowledgements

We would like to thank our industrial partner, La Milanaise, for their support and collaboration in this project as well as MAPAQ (project IA119053) for their financial support.

CHAPITRE 3 PROBLÉMATIQUE ET DÉMARCHE DU PROJET DE RECHERCHE

3.1 Problématique

Avec une demande alimentaire mondiale en constante augmentation, les conséquences des conflits, des crises économiques et des problèmes climatiques actuels engendrent une insécurité alimentaire significative dans diverses régions et populations. Il devient impératif d'améliorer l'efficacité et la résilience du secteur agroalimentaire afin qu'ils puissent continuer à répondre aux besoins alimentaires croissants. Parallèlement, la pénurie de main-d'œuvre dans le secteur manufacturier et la perte d'expertise due aux départs à la retraite posent des défis pour la maîtrise des processus de production. Dans les opérations de l'industrie alimentaire, la maîtrise des procédés est d'une importance majeure, servant de cadre général pour la gestion et l'optimisation des diverses variables inhérentes aux processus de production alimentaire, garantissant ainsi une qualité, une sécurité et une efficacité constantes.

La revue de la littérature effectuée dans le Chapitre 2 met en évidence la complexité du processus de meunerie, qui englobe plusieurs étapes. Il apparaît que chaque étape présente ses propres défis et complexités et nécessitent une gestion appropriée. L'analyse de la littérature souligne également l'importance d'assurer une production de farine de qualité constante, bien que la variabilité de la qualité des grains de blé a un impact sur le rendement et la qualité de la production. Ainsi, l'intervention d'un meunier expérimenté est essentielle pour ajuster chaque étape du processus et atteindre les rendements et la qualité de farine souhaités. Ces ajustements se basent sur les connaissances et l'expérience tacite des meuniers, qui les réalisent par essais et erreurs.

La nécessité d'une expertise spécialisée dans le contrôle du processus de meunerie, conjuguée à la pénurie de main-d'œuvre et à la perte de compétences lors du départ d'un employé expérimenté (Delorme, 2022; National Intelligence Council, 2021), constitue un défi pour garantir une production de farine de qualité constante. La variabilité de la qualité des grains de blé et des facteurs de production (tels que les conditions météorologiques, conditions

des machines, etc.) rendent les instructions de fonctionnement insuffisantes. En l’absence de meuniers ou en raison du manque d’expertise métier, il devient difficile d’effectuer les ajustements nécessaires pour produire une farine de qualité constante. C’est dans ce contexte complexe, où la stabilité de la qualité de la farine est cruciale, que se pose la question fondamentale : Comment contrôler le processus de meunerie pour garantir une production de farine de qualité constante?

Pour répondre à cette problématique nous proposons de développer des outils d’aide à la décision pour accompagner et améliorer les processus de meunerie. L’intérêt de notre recherche sera ainsi d’apprendre à contrôler le processus de meunerie à partir de données d’expériences réelles intégrant les comportements et contraintes présents sur la chaîne de production.

L’objectif général et les objectifs spécifiques définis pour répondre à la problématique soulevée sont définis dans la section suivante.

3.2 Cadre de recherche et contribution

En examinant la problématique, il est évident que le défi réside dans la manière de contrôler le processus de meunerie afin d’assurer une production de farine de qualité constante. En raison de la croissance du volume de données enregistrées sur les chaînes de production et du manque d’exploitation des techniques d’intelligence artificielle dans le domaine de la meunerie, il est proposé ici de résoudre cette problématique en développant des outils d’aide à la décision pour l’ajustement des paramètres contrôlables en production agroalimentaire.

Ainsi l’objectif global de ce projet est de **développer des outils d’aide à la décision visant l’ajustement des paramètres contrôlables dans la production agroalimentaire**. Dans un environnement de production agroalimentaire, les outils d’aide à la décision doivent prendre en compte les multiples facteurs et paramètres qui influencent la qualité et le rendement de la farine. Ils seront basés sur des modèles d’apprentissage automatique entraînés à partir des données présentes dans la chaîne de valeur.

L’objectif général se décline en trois objectifs spécifiques :

- **SO1 : Explorer les données dans un contexte de production agroalimentaire**

par l’expertise métier et la cartographie des processus;

- **SO2 : Développer un outil d’aide à la décision pour améliorer le contrôle d’un processus de transformation agroalimentaire;**
- **SO3 : Développer un outil d’aide à la décision pour l’analyse de la qualité des aliments en production.**

L’approche utilisée a été développée sur la base d’un contexte industriel, qui est présenté dans la section suivante.

3.3 Contexte industriel

Ce projet est développé en partenariat avec une minoterie biologique dans la province du Québec, au sud de Montréal. Le partenaire est une PME (Petite et Moyenne Entreprise) qui transforme et ensache différents produits biologiques tels que : de la farine, des flocons d’avoine, des légumineuses, de la semoule et du sucre. Le partenaire industriel est un acteur important dans le secteur de la transformation des grains de blé biologiques au Canada. L’entreprise se fournit auprès d’agriculteurs canadiens de blé certifiés biologiques et expédie ses produits transformés au Canada ainsi qu’aux États-Unis. Les produits transformés peuvent être livrés en vrac ou en sac de différents formats par palette.

L’usine du partenaire industriel représentera notre cas d’étude. L’objectif consiste à améliorer l’efficacité des processus de meunerie. Pour cela, nous exploiterons les données présentes sur la chaîne de production à l’usine. L’usine possède de nombreuses machines automatisées très performantes assurant le nettoyage, le mouillage et le broyage des grains de blé. Les machines, contrôlées depuis une interface SCADA (Supervisory Control And Data Acquisition), sont capables d’enregistrer de nombreuses informations lors d’une course de mouture. De plus, des machines d’analyses de la qualité sont également présentes dans la minoterie et au laboratoire. Les données de production et d’analyse de la qualité sont stockées dans leur département respectif. Ces données ne sont pas reliées entre elles et servent principalement de point de contrôle.

Les paramètres de production nécessitent des ajustements en fonction du type de farine à produire, de la qualité des grains utilisés et des recettes créées. Les ajustements sont actuellement

réalisés par le personnel de production ou par le chef meunier à l'aide de leurs connaissances et leur expérience. Les données présentes au niveau du plancher de production sont peu exploitées pour améliorer l'efficacité des processus de meunerie, elles sont davantage utilisées pour des exigences de traçabilité requises par les normes biologiques imposées. L'objectif du projet de recherche consiste à valoriser les données présentes sur la chaîne de valeur afin d'améliorer l'efficacité des processus de meunerie. Le développement d'outils d'aide à la décision facilitera la prise de décisions par le personnel de production lors de l'ajustement de certains paramètres de production pour améliorer la qualité et le rendement.

Basée sur ce contexte, l'approche pour atteindre l'objectif du projet est présentée dans la section suivante.

3.4 Méthodologie de recherche

Dans le but de répondre à l'objectif général et aux sous-objectifs, une approche empirique est suivie. La Figure 3.1 représente sous forme de diagramme cette démarche. Les contributions apportées à chacune des étapes sont présentées à la suite du diagramme.

La première étape consiste à effectuer une revue de littérature des travaux existants dans le domaine de la meunerie et qui utilisent l'intelligence artificielle. La revue littérature a fait l'objet d'un article publié dans une revue scientifique. Elle a permis de comprendre les étapes de la mouture, de lister la liste des problèmes liés au processus de meunerie biologique et d'identifier les différentes études existantes liées à l'intelligence artificielle qui traitent des enjeux de production. L'état de l'art a fait ressortir certaines lacunes qui offrent des opportunités de recherche que les algorithmes d'intelligence artificielle peuvent contribuer à combler. Ces opportunités sont regroupées sous différents axes de recherches, à savoir : les économies d'énergie, la production d'une farine de qualité constante, l'optimisation des ressources utilisées pour soutenir la production et l'amélioration des rendements de production.

En prenant connaissances des lacunes présentes dans la revue de littérature, nous avons ensuite cartographié le processus de meunerie et les données collectées auprès de notre partenaire industriel afin de structurer les données dans un contexte de production agroalimentaire par l'expertise métier et la cartographie des processus. Cet état des lieux nous permet de

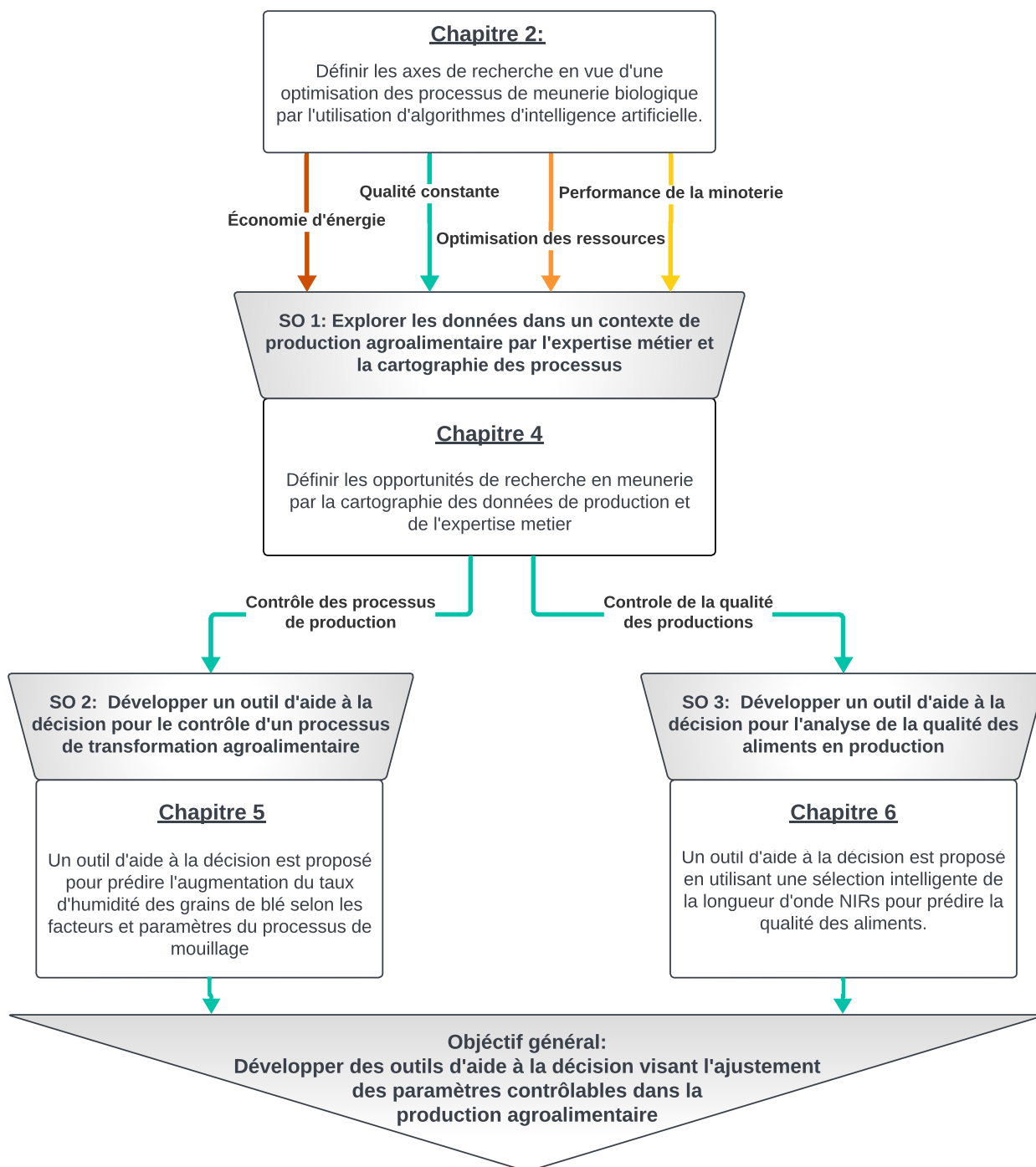


Figure 3.1 Démarche du projet de recherche

bien comprendre les différents enjeux sur site et d'évaluer les contributions possibles avec la valorisation des données de production. La cartographie a été réalisée en deux étapes. Une première cartographie a été construite sur la chaîne d'approvisionnement de la minoterie du partenaire industriel et a révélé des difficultés pour suivre le transport des lots de grains jusqu'à la minoterie. Elle a fait l'objet d'un article de conférence qui a été présenté au chapitre 4. La deuxième cartographie, ciblant plus spécifiquement la chaîne de transformation des grains de blé, a été conduite suivant la même approche qu'expliqué au chapitre 4. La cartographie de production a été élaborée et améliorée grâce à l'expertise du meunier. Le meunier qui est en charge des opérations de la minoterie pour la fabrication de la farine assure la confection des recettes et l'ajustement des paramètres des opérations selon la qualité des grains de blé réceptionnés et le type de farine à produire. Ses connaissances et son expertise métier nous permettent de contextualiser les données de production et d'identifier les variables non collectées qui, si elles l'étaient, offriraient de meilleures perspectives de contrôle de la production. La contextualisation des données joue un rôle central dans notre recherche afin de pouvoir cibler les opportunités possibles à partir des données disponibles. Cette dernière cartographie soutient la base de notre travail de recherche et répond au **SO 1: Explorer les données dans un contexte de production agroalimentaire par l'expertise métier et la cartographie des processus.**

À partir de la cartographie de production, qui fait état de la situation actuelle, le processus de mouillage a été ciblé car il présente un grand potentiel d'amélioration de la qualité et du rendement de la production de farine. D'après la revue de littérature, le contrôle de ce processus présente de nombreuses lacunes dans la recherche et représente une étape clé pour atteindre la qualité et la performance optimale d'une minoterie. Pour traiter cet enjeu, les données historiques de production du partenaire industriel ont été collectées. L'analyse de ces données réelles nous a permis de développer un outil d'aide à la décision pour le contrôle du processus de mouillage selon les différents facteurs et paramètres présents. Le développement de cet outil répond au **SO 2: Développer un outil d'aide à la décision pour le contrôle d'un processus de transformation agroalimentaire.** Ce travail a fait l'objet d'un article de journal présenté au chapitre 5.

La cartographie de la production a également identifié la mesure de la qualité des grains de blé comme un autre facteur critique influençant la qualité et le rendement de la farine. Les données historiques des analyses de la qualité des grains de blé du partenaire ont alors été collectées. Toutefois, les données obtenues se limitaient à certaines caractéristiques des propriétés intrinsèques du blé et n'incluaient pas les données acquises par des tests de laboratoire considérés comme des données de référence, relativement longues à obtenir. Différents ensembles de données disponibles en ligne ont été analysés. Ces données décrivent des aliments présentant des caractéristiques et des enjeux similaires à ceux que l'on retrouve dans le processus de transformation des grains de blé. À partir des données sur la qualité de grains de blé, un outil d'aide à la décision a été développé pour l'analyse de la qualité des grains de blé étudiés, selon une approche spécifique. Cet outil a ensuite été validé sur les données de fèves de cacao et de mangues. Le développement de cet outil répond au **SO 3: Développer un outil d'aide à la décision pour l'analyse de la qualité des aliments en production**. Ce travail a fait l'objet d'un article de journal présenté au chapitre 6.

En résumé, notre contribution prouve que sur la base de l'état de l'art et de la cartographie réalisée, nous sommes en mesure de cibler les opportunités d'amélioration dans un processus de production agroalimentaire, à partir du traitement des données historiques de production. La contextualisation et l'analyse des données ont permis de développer un outil d'aide à la décision pour paramétrer le processus de mouillage des grains de blé. Elles ont également été utilisées pour développer un outil d'aide à la décision pour mesurer la qualité des aliments dans un processus de production, validé sur différents types d'aliments, à savoir les grains de blé, les fèves de cacao et les mangues. Enfin, la généralité de l'approche est validée dans un secteur agroalimentaire différent en la reproduisant avec un autre partenaire industriel.

CHAPITRE 4 ARTICLE 2 : DÉVELOPPEMENT D’UN OUTIL DE SUIVI DES MOUVEMENTS DES GRAINS DE BLÉ BIOLOGIQUE AU NIVEAU DE LA CHAÎNE D’APPROVISIONNEMENT

Parrenin, L., Danjou, C., and Agard, B. (2020). Développement d’un outil de suivi des mouvements des grains de blé biologique au niveau de la chaîne d’approvisionnement. In *13ème conférence internationale de modélisation, optimisation et simulation (MOSIM2020)*, 12-14 Nov 2020, Agadir, Morocco

Soumis le 31 mai 2020 et accepté le 23 mars 2021

Résumé

Dans la production de farines biologique, les grains de blé suivent un processus logistique avec de nombreux contrôles en termes de qualité et de quantité. Différentes variétés de blé, chacune avec ses propres qualités, impactent directement la qualité de la farine produite. Le manque d’information sur la qualité des grains de blé entreposés ou mélangés dans des silos le long de la chaîne d’approvisionnement complexifie le suivi à destination de l’usine de transformation. Nous proposons de nouveaux outils pour le suivi des grains de blé. Ce suivi fournit une meilleure connaissance des qualités et des quantités de grains de blé en inventaire, d’où une meilleure réactivité sur la chaîne logistique. Face à l’augmentation des flux logistiques et aux imprévus météorologiques ou économiques, il est nécessaire d’avoir un processus d’approvisionnement en grain flexible et réactif pour assurer l’ensemble des commandes de chaque client.

Keywords: Grain de blé biologique, chaîne d’approvisionnement, transformation numérique, industrie 4.0.

4.1 Introduction

Dans le contexte socioéconomique actuel, plusieurs raisons poussent les consommateurs à valoriser des produits biologiques, que ce soit pour des choix de santé, un choix environnemental,

ou simplement pour soutenir l'économie locale (Hemmerling et al., 2015). Les habitudes des consommateurs ainsi que leurs besoins évoluent. Avec la croissance démographique et des habitudes de consommation différentes, la demande de produits alimentaires biologiques augmente (Aitken et al., 2020), en particulier pour le blé biologique. Dans le secteur industriel de transformation des grains de blé, de nombreux contrôles sont réalisés le long de la chaîne de valeur afin de garantir les quantités, la traçabilité et la qualité de la farine. Les données collectées ne sont pas forcément exploitées à leur plein potentiel et peuvent être sauvegardées par différents intervenants à plusieurs endroits le long de la chaîne de valeur. À partir des données collectées, l'objectif global est d'optimiser les paramètres de production d'une usine de transformation de grain biologique pour améliorer les rendements et maîtriser la variabilité de la qualité du produit fini. La collecte des données sur la chaîne d'approvisionnement et de production (qualité et quantité des grains de blé, paramètres de contrôle et d'état des machines de production) est la première étape pour atteindre l'objectif global. La deuxième étape consistera à structurer et traiter l'ensemble des données (collectées à l'étape 1) afin de construire un modèle de prédiction sur la qualité des produits finis en fonction des paramètres machines et de qualités des grains de blé utilisés. Finalement un modèle d'optimisation des réglages des machines de production sera élaboré à l'aide d'algorithmes d'apprentissage machine dans le but d'améliorer les rendements des moulins à farine. La chaîne d'approvisionnement a une influence importante dans la qualité des produits finis biologiques. En effet, le transformateur ne peut avoir recours à des agents chimiques pour corriger la qualité de ces intrants. La qualité des grains de blé réceptionnés à l'usine de transformation conditionne donc la qualité de la farine biologique pouvant être produite. Il est donc nécessaire de suivre de manière rigoureuse les qualités et quantités des lots de grain de blé arrivants de chaque fournisseur. Pour répondre à cette problématique, nous proposons un état de l'art sur les grains de blé et leurs spécificités (4.2.1), la chaîne d'approvisionnement de la récolte aux usines de transformation au Canada (4.2.2) et les points de contrôles des grains de blé au niveau de la chaîne d'approvisionnement au Canada (4.2.3). La section 4.3 présente la méthodologie de conception de l'outil de suivi pour les mouvements de grains de blé le long de la chaîne d'approvisionnement. La section 4.4 expose un cas d'étude qui a servi à la validation de la méthode. Finalement, la section 4.5 rappelle les résultats et propose

quelques perspectives pour l'utilisation de l'outil.

4.2 État de l'art

Le secteur de la transformation alimentaire est un domaine complexe notamment en raison de la manipulation de matières premières vivantes. La qualité des grains de blé peut varier en fonction des variétés, de leur provenance, des conditions d'entreposage, du processus de mouillage des grains de blé, etc. Cette section va décrire les principales caractéristiques des grains de blé, la chaîne d'approvisionnement, et les différents points de contrôles qualité du blé.

4.2.1 Caractéristiques des grains de blé

Deux principales sortes de blé sont plus largement cultivées (Boudreau and Ménard, 1992) : le blé dur (durum) et le blé tendre (hard et soft). Le blé dur au Canada est utilisé pour la fabrication du couscous et des pâtes alimentaires. Le blé tendre au Canada est employé dans les boulangeries pour la fabrication du pain (blé tendre vitreux dit blé dur ou de force), de pâtisserie (blé tendre dit blé faible ou mou) et de la pâte à pizza. Les blés sont semés à différentes périodes de l'année en fonction des zones climatiques. Le blé de printemps est semé généralement autour du mois de mars (entre avril et mai au Québec) dans les régions où les conditions climatiques durant l'hiver sont plus rudes. Le blé d'hiver est généralement semé autour du mois d'octobre (entre août et septembre au Québec) dans les régions méditerranéennes et tempérées. La récolte se déroule généralement en été par temps sec. Le grain de blé est situé dans un épi de blé et se compose de 3 parties distinctes (Campbell, 2007) : le germe (3% en masse) ; l'albumen ou amande farineuse ou endosperme (84%) et les enveloppes ou son (13%). Le germe est riche en matières grasses et en vitamines B et E, l'endosperme est principalement constitué de glucides (amidon) et de protéines (dont le gluten) et les enveloppes sont riches en fibres, mais nutritionnellement pauvres par rapport à l'endosperme (Campbell, 2007). La teneur en gluten est intimement corrélée avec la teneur en protéine. La teneur en protéine de blé varie en fonction des facteurs d'environnement alors que la qualité du gluten varie en fonction des facteurs héréditaires du blé (Boudreau

and Ménard, 1992). Pour classifier ces blés, les principales caractéristiques analysées lors de la réception sont : le taux d'humidité (quantité d'eau présente au niveau du grain), le taux de protéine (le gluten est à l'origine de l'élasticité de la pâte et le taux de protéine est un des critères principaux pour fixer le prix du blé), le taux d'amidon qui permet le gonflement de la pâte à pain (Vigni and Cocchi, 2013), les concentrations en toxines DON (Désoxynivalénol) et OTA (Ochratoxine A) - les concentrations de ces toxines produites par certaines espèces de moisissures peuvent nuire à la salubrité des grains et des aliments (Commission canadienne des grains, 2019d) - et le poids spécifique des grains (indicateur déterminant le prix du grain (Wilkinson et al., 2003)). Des essais technologiques sont réalisés pour obtenir de l'information plus précise sur la qualité de ces grains de blé. Lors des essais technologiques, un des indicateurs clés informant sur la qualité des grains de blé pouvant impacter le prix d'achat est le BEM (Brabender Energy Max), qui se mesure grâce au Glutopeak. Le Glutopeak est un appareil de mesure pour tester la qualité du gluten (Bakerpedia, 2019). Plus le BEM est élevé, meilleur est la qualité du gluten et donc de la farine.

4.2.2 Chaîne d'approvisionnement de la récolte du grain de blé aux usines de transformation

Entre la récolte du grain de blé et sa transformation en farine, le blé suit un long processus au cours duquel il sera manipulé et stocké. (Boudreau and Ménard, 1992) précise que « la généralisation du moissonnage-battage a conduit au développement d'une infrastructure, concrétisée par la mise en place des installations (bâtiments et équipement) en fonction de la réception, du séchage, du nettoyage, de la désinfestation et de l'expédition ». Le moissonnage-battage est l'opération qui consiste à récolter les grains de blé ou tout autre type de céréale dans un champ à l'aide d'une moissonneuse-batteuse. La moissonneuse-batteuse aura ainsi pour rôle de couper les tiges de blé et battre le blé pour extraire les grains de blé des épis de blé. Le processus pour acheminer ces grains à l'usine de transformation (Figure 4.1) débute aussitôt que les grains de blé sont récoltés (1). À l'étape 2, le blé est transporté jusqu'aux silos de collecte (3). Un échantillon est prélevé et analysé pour l'émission de certificat d'analyse (4). Cette étape permet d'identifier le type de blé et vérifier la qualité du blé. Le blé est transporté (5) dans un « silo terminus » (6) où il sera entreposé, nettoyé et séché avant d'être

expédié (7) à l'acheteur (8).

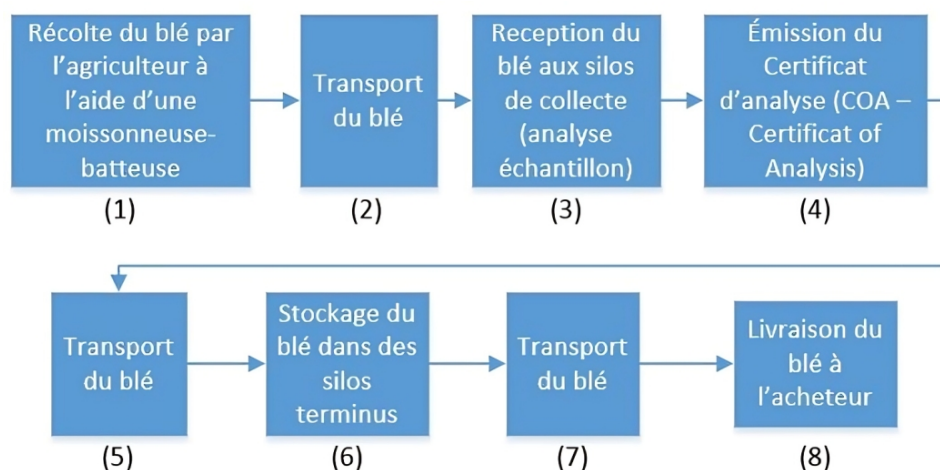


Figure 4.1 Processus simplifié de l'approvisionnement en grain de blé

Les commandes de blé passées à l'agriculteur se font par rapport aux besoins de l'usine de transformation de manière à couvrir les besoins en matières premières pour la production. Le transport des grains suit un processus rigoureux. Le blé peut transiter à travers un centre de grain, une plateforme de chargement ou être acheminé directement à l'usine de transformation. La Commission canadienne des grains précise les règlements à suivre pour le stockage du blé (Loi sur les grains du Canada, 2020b) et le transport des grains (Loi sur les grains du Canada, 2020c). La Commission canadienne des grains propose également un système de classement par grain et d'inspection permettant d'identifier la qualité du grain et assurer sa commercialisation (Loi sur les grains du Canada, 2020a).

4.2.3 Contrôle de la qualité du grain de blé au niveau du processus d'approvisionnement

D'après la Loi sur les grains (Loi sur les grains du Canada, 2020a), la commission canadienne des grains est responsable de l'approbation des normes de classification ainsi que des critères d'inspection (Commission canadienne des grains, 2019b). L'inspection est réalisée par plusieurs prises d'échantillons (Commission canadienne des grains, 2019c), en établissant un classement et en délivrant une certification d'analyse des lots de blé. Les analyses sont réalisées pour les grains manipulés aux silos de collecte et de terminus. Les échantillons

représentatifs du lot de blé réceptionné sont prélevés à l'entrée des silos directement depuis les véhicules de transport à l'aide d'une sonde mécanique. L'analyse regroupe de nombreuses informations telles que : le taux d'impuretés (pailles, pierres, brindilles, etc.); masse spécifique du grain; taux d'humidité; taux de protéine; la catégorie (printemps ou hiver); la présence de graines étrangères; la condition du grain (apparence, maturité, etc.); la classe (blé roux de printemps de l'Est canadien, blé roux de printemps Canada Prairie, etc.) (Commission canadienne des grains, 2019a); l'aspect (vitreux ou translucide). À partir de ces analyses, le responsable de silo détermine le grade du blé. Il est alors possible de déterminer le silo dans lequel le lot de blé sera entreposé. Avant d'être entreposé dans le silo approprié, le lot de blé réceptionné subira une étape de pré-nettoyage afin de retirer un maximum de matières étrangères. Un processus de séchage est parfois nécessaire lorsque le taux d'humidité dépasse 14,5% à des fins de conservation. Il faut toutefois surveiller à ne pas décolorer, fissurer ou endommager le grain lors du séchage (Boudreau and Ménard, 1992). Le long de la chaîne d'approvisionnement, le blé peut être stocké à divers endroits. Il est important de veiller aux conditions environnementales ainsi que les propriétés des grains de blé qui peuvent altérer leur qualité. Les moisissures, les insectes parasites, les acariens, les rongeurs, les altérations d'origine mécanique (fissure, cassure des grains) peuvent causer une altération de la qualité du grain lors du stockage (Boudreau and Ménard, 1992). Une bonne conservation du blé nécessite de préserver l'intégrité des protéines notamment le gluten et de maintenir le pouvoir germinatif du grain (un grain non germé). Avant toute commande, pour s'approvisionner en blé, le meunier réalise des analyses sur un échantillon provenant des silos afin de vérifier le type et la qualité du grain. Une fois l'échantillon accepté, le blé peut être livré. Une dernière inspection est réalisée lors de la réception des grains de blé à l'usine de transformation.

4.3 Méthodologie

De nombreuses données sont capturées le long de la chaîne d'approvisionnement informant sur la qualité des grains de blé, les quantités entreposées ou en mouvement. Pour une meilleure intégration et un meilleur suivi des flux d'approvisionnement des grains de blé entre les fournisseurs, les intervenants gérants les stocks de matière première et la production, nous

proposons de développer un outil intégrant l'ensemble des données pertinentes dans une base de données. Pour construire cet outil, permettant un meilleur suivi des grains de blé biologique à partir de la capture des données sur la chaîne de valeur, une méthodologie a été suivie rigoureusement. La méthode se compose de 3 phases détaillées ci-dessous. Le déroulement complet de la méthode permet d'avoir un meilleur suivi ainsi qu'une meilleure traçabilité de l'information.

4.3.1 Cartographie de l'approvisionnement

Dans cette première phase, il s'agit d'observer et de documenter les tâches effectuées par chaque individu participant dans le processus d'approvisionnement. Les symboles standardisés ANSI (institut national de normalisation américain) (Myler, 1998), sont utilisés pour la représentation graphique de la cartographie. Avec un flux du processus d'approvisionnement assez linéaire et impliquant assez peu d'intervenants, le formalisme de modélisation ANSI a été choisi pour réaliser la cartographie. La cartographie du processus permet de mieux comprendre les interactions et d'identifier les différentes données utilisées. La cartographie réalisée représente les processus actuels de niveau intermédiaire et montre la séquence des activités, mais ne décrit pas les tâches de chacune des activités à un niveau détaillé (HEC Montréal, 2013). Le but est d'analyser chaque processus et d'identifier les goulots d'étranglement potentiels, les tâches jugées répétitives, n'apportant aucune valeur ajoutée. La démarche suivie pour construire la cartographie a été de :

1. Récupérer des documents de cartographie existants au sein de l'entreprise (si disponibles) ;
2. Choisir une étape spécifique du processus ;
3. Observer et questionner sous forme d'entrevue semi-dirigée l'activité qui est réalisée par la personne ressource ;
4. Passer à l'activité suivante et répéter l'étape 3 pour obtenir une séquence d'activité jusqu'à l'état final du processus.

L'étape 3 a pour objectif de récupérer un maximum d'information. Les informations doivent permettre de connaître : l'action de la tâche effectuée ; les types de données manipulées ; les

documents créés ; les moyens de transmission des informations à partager ; les départements ou ressources externes à qui l'information est transmise.

4.3.2 Création de bases de données et d'interfaces

La cartographie réalisée à la phase 1 permet d'observer la séquence des activités et les types de données collectées le long du processus d'approvisionnement. Afin d'éviter toute redondance d'information et s'assurer de partager les bonnes informations à jour à d'autres utilisateurs, la création d'une base de données est nécessaire pour remplir les deux besoins précédemment évoqués. Une méthode de type agile a été appliquée afin de mettre en opération rapidement les outils développés tout en gardant une flexibilité sur la structuration et les différents besoins de chaque utilisateur. Les bases de données construites et prêtes à être utilisées seront améliorées de manière itérative à travers des retours d'expérience des utilisateurs. Il est important de se questionner sur l'utilité de chaque type de données et garder le strict minimum nécessaire comme donnée pour le bon suivi des grains de blé le long de la chaîne d'approvisionnement. L'approche proposée pour construire la base de données ainsi que l'interface se déroule comme ci :

1. Observer à partir de la cartographie réalisée la séquence d'activités et la simplifier si possible ;
2. Regrouper les types de données utilisées par les utilisateurs ;
3. Lister l'ensemble des utilisateurs futurs nécessitant un accès à la base de données ;
4. Communiquer avec chacun des utilisateurs concernant leurs besoins ;
5. Partager les responsabilités et définir les niveaux d'accès de chaque utilisateur ;
6. Modéliser l'outil et les différentes interfaces (formulaire, base de données, rapport)
7. Mettre en service l'outil ;
8. Améliorer l'outil en fonction des besoins et retours d'expériences des utilisateurs ;
9. Rédiger une documentation mentionnant les responsabilités de chacun et présentant l'utilisation des différentes interfaces créées.

4.3.3 Créer des indicateurs et des tableaux de bord

À partir de la base de données structurée, il est par la suite possible de créer des indicateurs de performance. Ces indicateurs vont permettre de faire un suivi, motiver le personnel, fixer des objectifs à atteindre tout en permettant de comprendre les causes de mauvaise performance. Les tableaux de bord sont des interfaces permettant de rassembler les indicateurs pertinents en une seule page et d'avoir une vue en temps réel de l'information sur des enjeux clés du processus. L'agrégation de données permet de prendre de meilleures décisions plus rapidement. Le développement d'indicateurs et de tableaux de bord s'appuie sur des besoins réels de visualisation et de synthétisation des informations à différents niveaux de l'organisation. La démarche suivie pour créer des indicateurs et des tableaux de bord a été de :

1. Comprendre les besoins des utilisateurs ;
2. Comprendre les données, les tables et les relations existantes (dans le cas d'une base de données relationnelle) dans la base de données étudiée ;
3. Créer des indicateurs qui répondent aux besoins de l'étape 1 et annoter les types de données manquantes nécessaires pour répondre aux besoins évoqués ;
 - (a) En cas de données manquantes, étudier les différentes possibilités pour collecter ces données sinon passer à l'étape 4 ;
 - (b) Intégrer ces données dans une ou plusieurs tables de la base de données ;
 - (c) Créer les relations dans la base de données entre les tables ;
4. Créer des tableaux de bord à partir des indicateurs créés à l'étape 3 (création d'une interface visuelle).
5. Valider les tableaux de bord auprès des utilisateurs Les résultats de chacune de ces étapes sont présentés dans le chapitre suivant.

4.4 Cas d'étude

4.4.1 Contexte industriel

L'outil développé pour le suivi des grains de blé a pour objectif de répondre à un besoin d'une entreprise de transformation de grain de blé, qui dans notre cas d'étude est La Milanaise. Cette PME québécoise se spécialise dans la production de farine biologique. Pour le stockage de ses données, notamment pour la chaîne d'approvisionnement qui nécessite une collaboration entre différents intervenants s'occupant de l'entrepôt des grains ou du transport, La Milanaise utilise le logiciel Smartsheet. Smartsheet est une plateforme hébergée sur le nuage informatique offrant la possibilité aux entreprises de planifier, capturer, gérer, automatiser et générer des rapports sur les travaux en cours de réalisation. Depuis les dernières années, La Milanaise connaît une forte croissance des ventes. Elle veut suivre de manière optimale les grains de blé sur sa chaîne d'approvisionnement pour être en mesure de gérer des flux plus importants et éventuellement améliorer son fonctionnement. Pour cela, une cartographie du processus d'approvisionnement des grains à l'usine a été réalisée.

4.4.2 Cartographie de l'approvisionnement

Pour débiter ou continuer notre cartographie, nous choisissons une activité spécifique dans le processus d'approvisionnement, par exemple une commande de grains de blé. Dans notre cas, pour les blés provenant du Québec, des ententes doivent être signées entre La Milanaise et l'agriculteur. De nouveaux agriculteurs peuvent contacter La Milanaise pour vendre leur blé ou La Milanaise peut contacter les agriculteurs présents dans son registre. Sous forme d'entrevue semi-dirigée, nous demandons au responsable acheteur :

- Par quels moyens se fait la prise de contact.
- Quelles informations sont collectées ainsi que les fichiers créés et utilisés pour sauvegarder ces informations.
- Quels utilisateur ou département ont besoin de l'information sauvegardée au sein de La Milanaise.
- Comment l'information est transmise.

Les réponses à ces questions aident à documenter l'activité représentée sur la cartographie

(Figure 4.2).

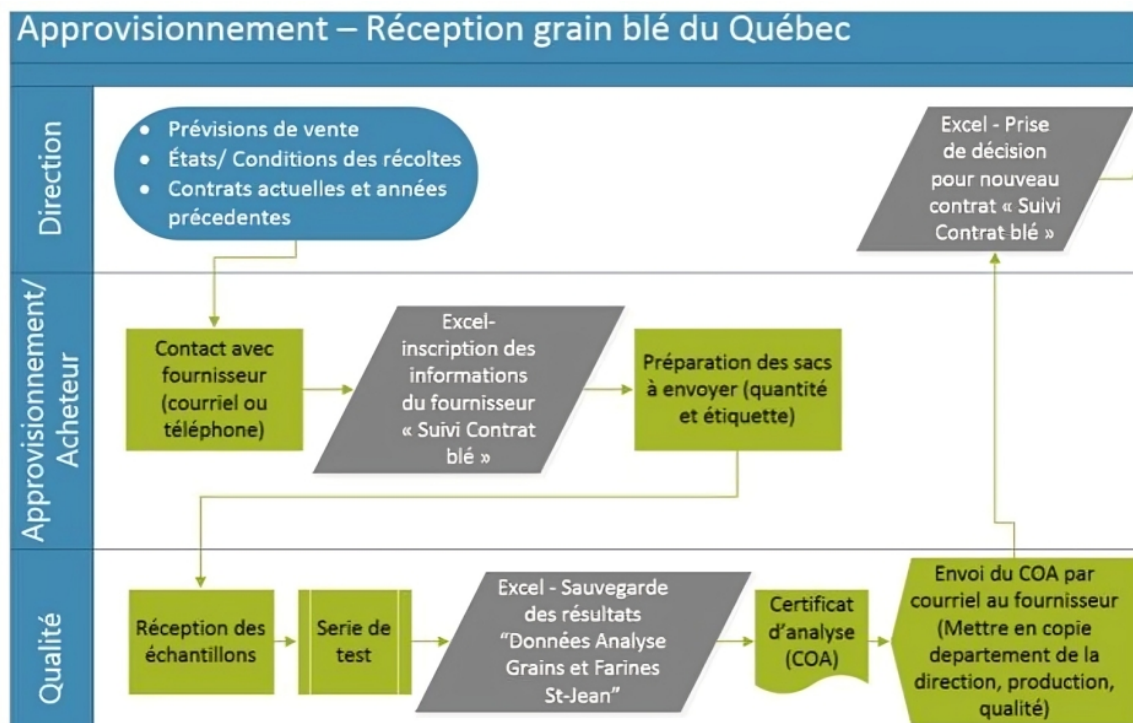


Figure 4.2 Construction de la cartographie de l’approvisionnement des grains de blé du Québec

Une fois la commande passée, on passe à l’activité suivante en interrogeant et observant ce que le responsable de l’approvisionnement doit réaliser. On continue ces étapes jusqu’à ce que la commande de grain de blé soit arrivée à l’usine de La Milanaise. La cartographie de la chaîne d’approvisionnement résume l’ensemble des opérations suivies par les grains de blé de la récolte au champ de blé jusqu’à l’usine de transformation, spécifiquement pour La Milanaise. On retrouve sur cette cartographie les points de collectes de données, les points de contrôles et les différents intervenants dans la chaîne d’approvisionnement. Cette cartographie est nécessaire pour regrouper par la suite les données collectées dans des bases de données.

4.4.3 Création de bases de données et d’interfaces

En analysant la cartographie, on constate que les résultats des analyses des échantillons et les contrats étaient inscrits sur un fichier Excel, sauvegardé sur un serveur de la compa-

nie. On souhaite créer une base de données pour centraliser, unifier et partager l'information aux utilisateurs. 5 informations (Nom fournisseur, Nom contact, Courriel, Numéro cellulaire, Adresse postale) étaient redondantes et pouvaient être considérées comme des objets statiques faisant partie d'élément organisationnel. Une première base de données regroupe les informations des agriculteurs. Une deuxième base de données regroupe les résultats d'analyses des échantillons de grains de blé provenant de chaque agriculteur. La base de données a été configurée sur Smartsheet. La plateforme n'étant pas une base de données relationnelle, un script Python a permis de faire ces relations fictives (Figure 4.3) permettant d'avoir une mise à jour rapide des listes déroulantes.

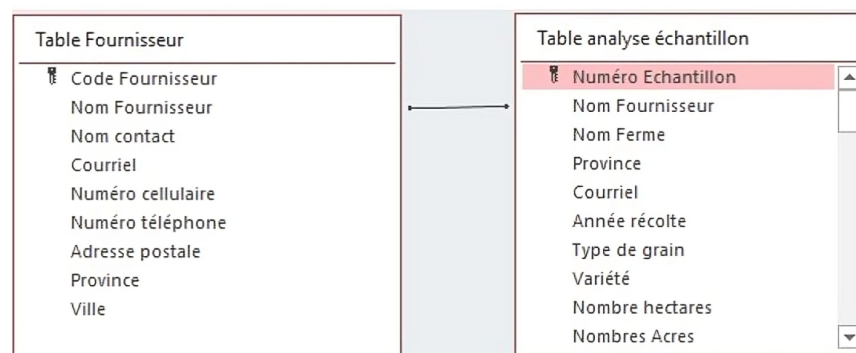


Figure 4.3 Relation fictive entre la table fournisseur et la table analyse échantillon

Les résultats d'analyses entrés et sauvegardés dans la base de données ont permis de semi-automatiser la production, la sauvegarde et l'envoi de certificats d'analyse. Une base de données a été construite au niveau d'un centre de grain (Figure 4.4), en s'appuyant sur les modèles précédents utilisés. Deux feuilles Excel étaient utilisées pour les transactions. Chaque feuille comprenait 15 colonnes. La base de données (Figure 4.4) a permis de regrouper l'ensemble des ces données sous 24 entités et de créer une relation avec la table analyse échantillon pour connaître la qualité de blé manipulé au centre de grain. 6 nouvelles entités ont donc été rajoutées en créant une relation unique à partir du numéro d'échantillon. En listant les utilisateurs à partir de la cartographie et en échangeant à travers des réunions impliquant les départements interne et intervenant externes, nous identifions 14 utilisateurs. Les réunions permettent de définir les besoins de chacun ainsi que leurs responsabilités. Les responsabilités définissent les droits d'accès de chaque utilisateur à la base de données. Des

formulaires sont créés pour permettre l'ajout de nouvelles transactions. La Figure 4.4 montre les lignes de transaction de grain de blé entrant (définis par une couleur verte) et sortant (couleur rouge) d'un centre de grain de blé. Des « Cut-off » peuvent intervenir pour remettre à niveau le Silo vide, dû à des mouvements de grains « non nettoyés » en tonne métrique. Finalement les retours d'expériences ont permis d'améliorer l'outil, s'assurer que l'ensemble des besoins importants des utilisateurs soit comblé et que le suivi des grains de blé soit possible.

Reference number	Projected IN/OUT date	Shipping/Delivery date	Scale tkt #	IN or OUT	Delivery/Shipping Weight (MT)	Silo number at NuVision	Silo Transferred to at NuVision	Inventory	Transport mode	Sample No
00010		31/10/19	25871	IN	43.92	Silo No 1		0.000		
00011	08/11/19	14/11/19	25967	IN	42.47	Silo No 1		0.000	Container	Échantillon xx-1
00012	08/11/19	14/11/19	25955	IN	44.32	Silo No 1		0.000	Container	Échantillon xx-1
00013	11/11/19	14/11/19	25956	IN	44.45	Silo No 1		0.000	Container	Échantillon xx-1
00014	11/11/19	14/11/19	25960	IN	43.88	Silo No 1		0.000	Container	Échantillon xx-1
00015		14/11/19	25968	OUT	20	Silo No 1		Done	Container	Échantillon xx-1
00016	12/11/19	15/11/19	25974	IN	44.37			0.000	Rail	Échantillon xx-2
00017	12/11/19	15/11/19	25977	IN	43.87			0.000	Rail	Échantillon xx-2
00018		15/11/19	N/A	OUT	88.24			Done	Rail	Échantillon xx-2
00019		15/11/19	25980	OUT	27.29	Silo No 1		Done	Container	Échantillon xx-1
00020		15/11/19	25983	OUT	26.18	Silo No 1		Done	Container	Échantillon xx-1
00021		18/11/19	25988	OUT	27.32	Silo No 1		Done	Container	Échantillon xx-1
00022		19/11/19	26001	OUT	26.32	Silo No 1		Done	Container	Échantillon xx-1
00023		26/11/19	26067	OUT	1.99	Warehouse		Done	Container	
00024		26/11/19	26068	OUT	25.19	Silo No 2		Done	Container	Échantillon xx-3
00025		27/11/19	26076	OUT	24.93	Silo No 2		Done	Container	Échantillon xx-3
00026		28/11/19	26078	OUT	10.58	Warehouse		Done	Container	
00027		28/11/19	26077	OUT	12.54	Silo No 2		Done	Container	Échantillon xx-3
		01/12/19		Cut-off	0.74	Silo No 1		Done		
		01/12/19		Cut-off	-0.15	Silo No 2		Done		

Figure 4.4 Mouvements au sein d'un centre de grain

4.4.4 Indicateurs et tableaux de bord

La cartographie a permis d'identifier les points de collectes de données ainsi que les étapes clés de transportations du blé de la récolte jusqu'à l'usine de transformation. Des bases de données ont ainsi pu être construites pour regrouper certaines informations collectées le long de la chaîne d'approvisionnement. À partir des bases de données, il est maintenant

possible de développer des indicateurs. La première étape énumérée dans la méthodologie consiste à comprendre les besoins spécifiques d'utilisateurs pour développer des indicateurs utiles. Un besoin de suivre les inventaires et les types de blé répertoriés, par numéro de lot d'échantillon, était souhaité par la direction, pour connaître la qualité des grains de blé en stock ou en mouvement. Les centres de grains n'indiquent pas la qualité de blé qu'il envoie à La Milanaise ou qu'il entrepose.

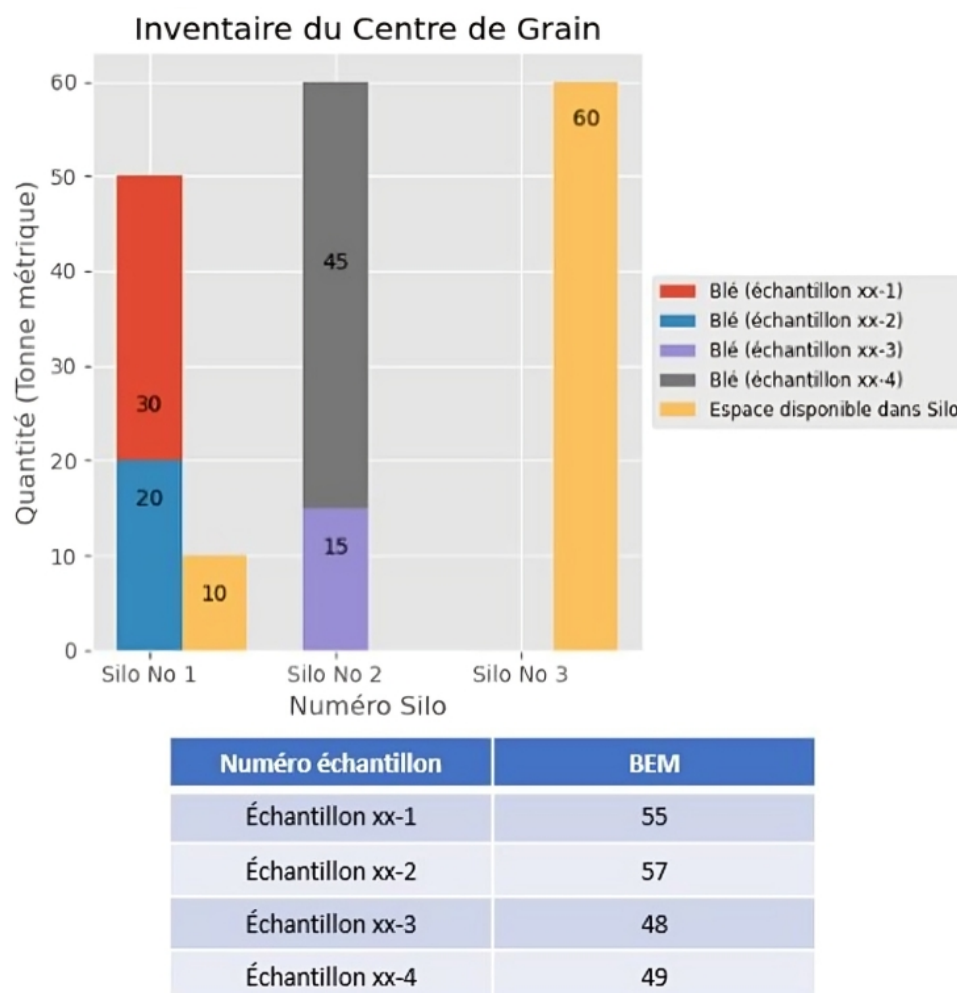


Figure 4.5 Exemple d'indicateurs visuels pour le suivi des inventaires d'un centre de grain

Des indicateurs ont donc été créés de manière à connaître les quantités de grains de blé restant et l'espace disponible dans chacun des silos. L'ensemble des informations nécessaires à la création des indicateurs se retrouve dans la base de données du centre de grain. Ces indicateurs sont construits par de simples calculs à partir des informations collectées dans la

base de données. Pour connaître les types et quantités de blés présents, un script Python a dû être développé. Ce script calcule l'inventaire toutes les 30 minutes à partir des transactions inscrites dans la base de données, dans la colonne « Inventory » de la base de données, par le concept « Dernier entré – Premier sorti » adapté à l'écoulement des grains de blé dans les silos. Ce script modifie les quantités restantes de chaque lot de blé. À partir des données enregistrées dans la base de données ainsi que des indicateurs créés, un tableau de bord a pu être construit spécifiquement pour un centre de grain (Figure 4.5). Ce tableau de bord permet de synthétiser l'information et partager celle-ci aux personnes concernées. Le tableau de bord du centre de grain met à disposition : des formulaires pour entrer de nouvelles transactions ; la base de données regroupant l'ensemble de transactions ; des indicateurs visuels sur l'inventaire des silos réservés à La Milanaise ; un tableau précisant le type de blé, sa quantité ainsi que sa qualité restant dans les silos. Finalement la synthèse de l'information de l'inventaire des grains de blé en temps réel permet de prendre des décisions plus rapides et donc une meilleure réactivité dans la chaîne d'approvisionnement.

4.5 Conclusion

Un outil pour le suivi des grains de blé au niveau de la chaîne d'approvisionnement a été développé, basé sur une méthode en 3 étapes : (1) la cartographie du processus d'approvisionnement, (2) la création de bases de données et d'interfaces et enfin (3) la création d'indicateurs et de tableaux de bord. La méthodologie a pu être appliquée sur un cas d'étude avec une PME spécialisée dans la production de la farine biologique. Les résultats obtenus ont permis d'avoir un meilleur suivi des grains de blé au niveau d'un centre de grain de blé et d'offrir une meilleure réactivité des différents intervenants sur la chaîne d'approvisionnement. Le tableau de bord précise à l'utilisateur du centre de grain le numéro de référence du blé qu'il envoie à l'usine. Le numéro de référence permet de suivre très simplement les types et qualités de blé en mouvement. Le suivi des types et qualité de blé est important pour la planification de la production, du réapprovisionnement des silos et l'achat de nouvelles matières premières. Le rendement et la qualité des produits finis obtenus suite au processus de transformation des grains de blé sont directement impactés par la qualité des matières

premières.

L'outil développé démontre la faisabilité et l'intérêt de suivre, partager les informations captées le long de la chaîne d'approvisionnement dans le secteur de la transformation des grains. Il devra cependant être étendu à d'autres centres de grain et entrepôts pour inclure l'ensemble des intervenants de la chaîne d'approvisionnement et prendre en compte les différents lieux de provenance des grains de blé jusqu'à l'usine de transformation. Le résultat final permettra d'avoir une vue globale des inventaires et une meilleure planification pour le réapprovisionnement des stocks.

4.6 Remerciements

Nous tenons à remercier notre partenaire industriel, La Milanaise, d'avoir participé au développement de cet outil ainsi que le MAPAQ (projet IA119053) pour leur soutien financier.

CHAPITRE 5 ARTICLE 3: A DECISION SUPPORT TOOL FOR THE FIRST STAGE OF THE TEMPERING PROCESS OF ORGANIC WHEAT GRAINS IN A MILL

Parrenin, L., Danjou, C., Agard, B., and Beauchemin, R. (2023a). A decision support tool for the first stage of the tempering process of organic wheat grains in a mill. *International Journal of Food Science & Technology*, 58(10):5478–5488

Submitted on January 16, 2023 and accepted on March 8, 2023

Abstract

Wheat tempering conditions grains before a milling process begins. Process adjustments must be made to reach a desired level of flour quality and yield, depending on multiple factors. This article aims to develop a decision support tool to help operators adjust the first stage tempering parameters. It is based on a regression model that predicts an increase in organic wheat moisture content according to the properties of the wheat (initial wheat moisture content, wheat protein content and wheat temperature), process parameters (targeted wheat moisture content, wheat flow rate, water flow rate, wheat quantity and resting time) and tempering conditions (water quantity, average day temperature, average day humidity). The increase in wheat moisture achieved during the first tempering stage varies between 0% and 5%. Five regression models were compared: OLS, LASSO, RIDGE, ElasticNet and XGBoost. The models have been developed and tested from a case study at an organic wheat mill plant for one wheat blend category. The results indicate that the LASSO model outperformed others, with an average prediction error of 0.428%. The model showed the importance of humidity and temperature factors during the tempering process. The flow of water and wheat were the most influential parameters for an increase in wheat moisture content.

Keywords: Industry 4.0, organic wheat, wheat tempering, decision support tool, data mining, multiple regression.

5.1 Introduction

Wheat is one of the most cultivated cereals in the world (FAO, 2021). Millers are responsible for transforming grains of wheat into types of flour that have different quality requirements (Posner and Hibbs, 2005). Quality requirements vary depending on the intended uses of the flour, the market, and whether it is a conventional or an organic flour. Organic flour is produced from organic wheat without any synthetic preservatives (Rana and Paul, 2017) and with strictly limited, pre-authorized ingredients (Government of Canada, 2022). The global organic food and beverage market is rapidly growing. Recent data has shown that the sales of organic food continue to grow and reached a value of over 120 billion Euros in 2020 (Willer et al., 2022). This growing demand is driven by consumer expectations (Rana and Paul, 2017; Willer et al., 2022).

At a mill, grains of wheat are cleaned, eventually blended, tempered, and finally, milled (Parrenin et al., 2023b). Cleaning and tempering grains of wheat conditions it for milling operations (IAOM, 2018). By cleaning the grains of wheat, all impurities and foreign materials that could damage machines and impact flour quality are removed. Then, tempering is done by adding water to the grains to increase their moisture content. This operation improves the efficiency of milling (Rana and Paul, 2017).

Many studies have shown the importance and impact of the tempering process on milling performance, flour quality and energy grinding (Cappelli et al., 2020a; Fang and Campbell, 2003b; Kweon et al., 2009). Depending on the quality of the wheat grains and the desired quality of the flour, the miller must adjust the process. Complexity increases when wheat blends are involved (Hook et al., 1984). As wheat water absorption is different for each batch of wheat, the miller relies on his knowledge and experience to correctly adjust the tempering parameters. Better knowledge of the different factors and parameters that influence the water absorption of wheat in a mill would allow for effective decisions on tempering adjustments, under specific manufacturing conditions.

The objective of this paper is to develop a decision support tool to predict an increase in wheat moisture content at the end of the first stage of tempering, based on wheat properties, operating conditions and process parameters. From the model that is built, it would then be

possible to acquire knowledge about the impact variables and their influence on the process. A literature review is presented in Section 6.2. Section 6.3 presents the proposed decision support tool. Section 6.4 presents the results obtained. The results are analysed and assessed in Section 7.

5.2 Literature review

5.2.1 Tempering process in a wheat mill plant

The tempering process conditions the wheat before milling. An optimal wheat moisture level for the milling process is necessary. If the grains of wheat are over-dampened, the milling operation becomes difficult, and the flour yield of the mill is immediately reduced. Too little tempering moisture makes the grain crumbly and leads to bran contamination of the flour. Although Gwartz (1998) shows the importance of the tempering process, this process remains difficult to control (Willm, 2009).

Figure 5.1 shows a diagram of the tempering process of the grains of wheat. The cleaned grain is stored in silos. Flow controllers on each silo control the flow of grains sent to the dampening machine. In the dampening machine, a constant flow of water is injected into the grains of wheat. The tempering screw in the dampening machine allows the movement of the wheat and efficiently mixes the wheat (Willm, 2009). Finally, the grains of tempered wheat are stored again in silos so that they can rest and absorb the water particles on their surface.

The tempering process acts in several stages. In each stage, a flow of water is sent to a moistening screw, through which the wheat grains pass. A rest period is then necessary for the grains in a silo to allow the water particles present on the surface to penetrate the grain. The number of stages depends on several factors, such as the quality of the wheat grains and environmental conditions. The tempering aims to moisten the wheat to facilitate the separation of the endosperm from the rest of the grain component in the milling process. This phenomenon is explained by the fact that the water absorption of the grains will increase the wheat bran's flexibility and soften the endosperm (Willm, 2009).

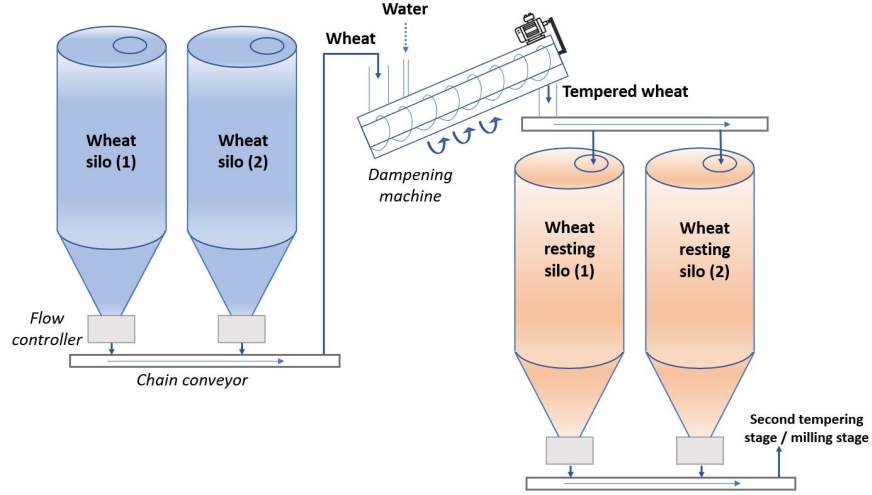


Figure 5.1 Diagram of the wheat grains' tempering process

Although this mechanism is the one that is the most frequently used in a wheat mill, it is worth mentioning that different tempering techniques and dampening machines exist. Bühler developed a vortex dampener to dampen the grains of wheat in a less brutal way (Willm, 2009). Other than tempering grains of wheat by a flow of water, steam tempering is possible (Chen et al., 2020).

In a mill, the miller learns to master the tempering process from their own knowledge and experience. To reach a certain level of wheat moisture at the end of the tempering stage, the miller fixes a target moisture content. This target changes depending on the type of flour desired. The type of flour varies based on its final purpose and its quality attributes that include protein content (%), moisture content (%), ash content (%) and falling number (FN). Considering the target moisture content (H_t), the flow of the wheat grains (F_{wg}) and the initial moisture content of the wheat grains (H_i), the miller calculates the flow of water required during the tempering process (F_w), as follows (IAOM, 2018; Willm, 2009):

$$F_w(L/h) = \frac{(H_t - H_i) * F_{wg}}{100 - H_t} \quad (5.1)$$

5.2.2 Analysis of experiments on tempering wheat grains

In the literature, there are various experiments that have been conducted to better understand wheat tempering in a controlled environment. These experiments could be grouped into two approaches: an analysis of the rate of water uptake into the grain, and an analysis of the influence of grain moisture content on milling operations.

The first approach analyzes and models the rate of moisture uptake in wheat grains when they are immersed in water for several hours. The objective is to better understand the biological characteristics of grains and the process variables that could influence the rate of moisture uptake by conducting experiments. Different techniques are developed using deep learning algorithms and diffusion equations. Kang and R. Delwiche (1999) showed that the water uptake is faster in the endosperm than in the bran layer, suggesting that the bran layer of the wheat acts as a barrier for water absorption. When analyzing the endosperm composition, Moss (1977) showed that the *protein content* and *initial moisture content* greatly influence water penetration into the wheat grain. Tagawa et al. (2003) determined different diffusion coefficients of water in a wheat grain for different water temperature levels, ranging from 10 to 50 degrees. They showed that the soaking temperature influences the water uptake of grains of wheat, and that higher temperature increases the rate of water uptake. Kashaninejad et al. (2009) confirmed this result and showed the relevance of an artificial neural network (ANN) for simulating the soaking behaviour and the effect of temperature and time on the hydration in a wheat grain. In this context, temperature and soaking time were used in the ANN model as input parameters, while water content was used as the output parameter that needed to be predicted. The application of higher temperatures during the tempering stage could reduce the necessary resting time for the grains of wheat.

The second approach is to study the different levels of grain moisture content and measure its effect on milling operations. Fang and Campbell (2003b) predict flour particle size distribution from an extended breakage function based on different roll gaps and moisture content. A mathematical equation estimates the milling performance according to an operative milling setup (roller gaps and disposition) and the grain quality (grain moisture and grain size). Kweon et al. (2009) analyzed the effect of tempering conditions for soft red winter wheat on

milling performance and flour quality. A full factorial design of experiment including initial wheat moisture, tempered wheat moisture, tempering temperature and tempering time at two levels for each factor was conducted. They found that the tempered wheat moisture had the largest effect on milling performance and flour quality. As the wheat tempering increases, flour yield decreases and flour quality increases. Similarly, Cappelli et al. (2020a) evaluated the effect of milling speed and wheat tempering on milling performance, energy consumption, dough quality and bread quality. They confirmed previous results, indicating that the moisture content had a significant influence on mill performance and flour quality related to the rheology of the dough and the quality of the bread. From the four tempering moisture levels studied, they found that 13% moisture was the best compromise between milling and bread-making performance.

Both approaches are based on experimental measurements in a controlled environment. However, there is still a need to focus on the tempering of wheat grains in the manufacturing process. Indeed, wheat grains in a mill plant are subjected to environmental factors and often include grain mixtures. In this context, the development of a decision support tool that can take these factors into account seems of interest. It would help overcome actual manufacturing challenges that are not present in a controlled environment. The following section explains the materials and methods used in the development of this tool.

5.3 Materials and methods

The objective of the proposed decision support tool is to help operators adjust the tempering parameters in their environment to condition the grains of wheat as desired. The aim of this tool is to predict the wheat moisture level that would be reached at the end of the tempering process, based on the factors and process parameter inputs. The tool is developed and tested in a case study at an organic wheat mill plant. First, the characteristics of the organic wheat mill are presented. Then, the materials used at the mill to analyze the wheat samples and to perform the tempering operation are described. The predictive models used to build the decision support tools are explained. Finally, the method followed in the development of the decision support tool for the wheat grain tempering process is detailed.

5.3.1 Characteristics of the organic wheat mill plant

The organic wheat mill is located in Saint-Jean-Sur-Richelieu, Quebec, Canada and is operated by the company La Milanaise. La Milanaise specializes in the transformation of organic cereals into organic flour in Canada. At the mill, up to 3 tempering stages can be performed to condition the grains of wheat. Four wheat resting silos, each with a capacity of approximately 80 metric tons are available for the first and second tempering stage. The third stage is achieved by recirculating the wheat at the beginning of the second stage. A second and third tempering stage is sometimes necessary when the desired moisture level of the wheat grains is not reached at the end of the first stage. This can be due to an excessive difference between the initial and the desired moisture level or to a relatively slow absorption rate of the wheat. This study focuses on the first stage, which is usually carried out, except when the moisture content of the grains is already sufficiently high (16% or above). The mill is located in a climate that is characterized by extremes of heat and cold throughout the year that tend to worsen with the effects of climate change (Almaraz et al., 2008). In Saint-Jean-Sur-Richelieu, wheat grains are subject to large temperature variations, from -32 degrees Celsius on the coldest days to 36 degrees Celsius on the warmest days of the year (Government of Canada, 2021). Based on historical weather data for the year 2021, large variations in relative humidity with a range extending from 36.5% to 95.5% are noted as well (Government of Canada, 2021). These conditions place this mill within a specific context that is different from other mills located in other regions of the world (e.g., Europe, North Africa or the United States). The production parameters of the tempering process must therefore be closely monitored throughout the year.

5.3.2 Materials

5.3.2.1 Raw materials and wheat tempering

Production data from the company, collected over an 18-month period from March 2020 to September 2021, was used to develop and analyze the prediction models. The data is separated into two sets, training and prediction.

The wheat that is studied was grown on organic farms in Canada. The wheat underwent a tempering stage before being milled. Upon arrival, organic wheat grains were pre-cleaned and stored in certain silos according to their quality, measured by the BEM (Brabender Energy Max) indicator (Rakita et al., 2018). Selected wheat grains are then cleaned and blended to be ready for the tempering process. At this stage, a sample of about 300g of wheat grains is taken at the beginning and at the end of the tempering process. A near infrared spectrometry (Pertten IM9500 NIR) is used to analyze the protein and humidity content of the grains of wheat. An infrared thermometer is used to control the temperature of the grains of wheat. The tempering process at this mill is realized with an intensive dampening machine. This machine sends a constant flow of water to the wheat grains, which is set up by the miller. The action of the tempering screw ensures that the wheat grains advance and are mixed well.

5.3.2.2 Predictive models

Despite the advances in deep learning models in terms of performance (Goodfellow et al., 2016) and their use in the field of wheat processing (Assadzadeh et al., 2022; Sabanci et al., 2020), models such as artificial neural network (ANN) are still difficult to interpret and require large amounts of data to be trained and to achieve good performance.

As a result of (i) the limited amount of data available in this study, (ii) the need to explain the model for the decision support tool and (iii) the prediction of a continuous output variable, regression models have been preferred. Several different regression models can be used in this specific case, such as linear regression, polynomial regression, RIDGE regression, LASSO regression, ElasticNet regression and regression trees. The goal is to produce a regression model that more accurately predicts the output variable, without overfitting the data. A multi-linear regression equation is mathematically represented as follows:

$$y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon \quad (5.2)$$

The output variable y will represent the percentage increase in the moisture content of the wheat that is achieved at the end of the first stage of tempering. Each independent variable

(X_i) is associated with a coefficient (β_i) that indicates the importance and impact of the variable on the output variable. Residual Sum of Squares (RSS) is used to fit the regression model (see Eq. 6.4). The β_i coefficients are adjusted to minimize this loss function. Ordinary least squares (OLS) is a statistical technique that best fits the coefficient of a linear regression model based on the RSS loss function (James et al., 2013).

$$RSS = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \quad (5.3)$$

If several independent variables are correlated, there is a risk of multicollinearity, which adds noise to the model. This implies that the estimated coefficient will not be able to generalize future data well, and the model will result in poor prediction performance. Regularization models such as LASSO, RIDGE and Elastic Net are used to reduce the risk of multicollinearity. The adjustment of the hyperparameters in these models penalizes specific correlated independent variables. The penalty terms added to the loss function aim to reduce or eliminate the influence of independent variables on the model depending on the regularization model used. The difference between regularization models is the penalty added to RSS :

$$L_{LASSO} = RSS + \lambda \sum_{j=1}^p |\beta_j| \quad (5.4)$$

$$L_{Ridge} = RSS + \lambda \sum_{j=1}^p \beta_j^2 \quad (5.5)$$

$$L_{ElasticNet} = \frac{RSS}{2} + \lambda \left(\frac{1-\alpha}{2} \sum_{j=1}^p \beta_j^2 + \alpha \sum_{j=1}^p |\beta_j| \right) \quad (5.6)$$

Linear regression is one of many algorithms that may be used to solve regression problems. Another method, among the supervised learning algorithms, that can be used as a regularization technique for a regression problem is XGBoost. XGBoost is a gradient boosted tree algorithm. Gradient boosted models are trees built sequentially. Each new model uses a gradient descent optimization to update the weights that are to be learned by the model to reach a local minima of the loss function. A new function is added in each step to predict the output. To fit the data, XGBoost uses a loss function and a regularization term that can be expressed mathematically as follows (Chen and Guestrin, 2016):

$$L_{XGBoost} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (5.7)$$

$$where \Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$$

l measures the difference between the prediction $\hat{y}_i$ of the t^{th} trees and the target y_i . $f_t(x_i)$ is the new function that is added to the previous prediction to minimize the loss function. The second term $\Omega(f_t)$ is the regularization function that penalizes the complexity of the model. In the regularization function T represents the number of leaves and w_j the weight of the leaf j . The terms γ and λ are two hyperparameters that act as penalty terms to reduce overfitting. The inclusion of a regularization function in the loss function distinguishes XGBoost from most tree sets.

5.3.3 Method followed to develop a decision support tool in the wheat grain tempering process

The general method used to develop the proposed support decision tool comprises the following four steps: (i) collect relevant data, (ii) prepare the data, (iii) create the regression models and (iv) validate the model.

5.3.3.1 Collection of relevant data

First, different variables that influence the tempering process according to the literature review are included. They are: wheat properties, tempering resting time and tempering bin space available. Then, based on interviews and observations at the mill, the different adjustments and decisions that are taken during the tempering process are added. These include: the flow of water, the flow of the grains of wheat, the quantity of wheat, the duration of the resting time and the number of tempering stages. The flow of water used in the tempering process is calculated from formula 6.2. Finally, millers and experts add that environmental conditions and wheat mixes are additional factors to consider to achieve a good amount of control in the process.

A mapping of the flow process on the production line was performed (Parrenin et al., 2020). The mapping identifies available data sources and the type of data recorded.

In the production process, most equipment at the mill is controlled from the Supervisory Control and Data Acquisition (SCADA) system. The SCADA system is linked to a relational database in which data is recorded at regular intervals every 10 minutes. The data that is stored includes information about the quantity of grains of wheat transferred from one silo to another and the quantity of water used. This provides the quantity of grains of wheat used at the beginning of each tempering stage, and the resting time of the grains of wheat that are tempered.

All of the production runs at the mill are saved in Excel files. An Excel file is created for each new production run that starts, referenced by a production run number. The file lists the quantity of wheat used, the quality of the grains and the different tempering conditions, such as the flow of water, the date and time of the beginning and end of the process, and the temperature of the grains. As each Excel file has the same structure, a Python program has been developed to efficiently collect the data.

Although the miller and experts take into account environmental conditions in their decision for the adjustment of the tempering parameters, no information regarding these attributes was stored in any of the various data sources. To obtain this information, a collection of meteorological data from a weather observation station closest to the mill was carried out.

If critical data is missing, dedicated sensor installation is considered. Different methods and protocols are used to collect the data, depending on the IT infrastructure and equipment present on the shop floor (Vermesan et al., 2014).

Finally, adjustable variables for the tempering process and non-adjustable variables are identified. Adjustable variables include: *target wheat moisture content*, *wheat flow rate*, *water flow rate*, *wheat quantity* and *resting time period* for grains of wheat. Non-adjustable variables include: *initial wheat moisture content*, *wheat protein content*, *wheat temperature*, *daily mean relative humidity*, *daily mean temperature* and *water quantity*. Among the given variables, water quantity is considered to be a non-adjustable variable because the quantity is theoretically defined by the flow of water, the flow of wheat and the quantity of wheat used.

This variable is nevertheless considered in the model because the water supply is cut off manually by operators and is not always proportionate to the three controllable variables.

5.3.3.2 Preparation of the data

14 different wheat categories are available with our partner and all wheat categories can be blended together, depending on the objective. Of all of the different wheat blend categories, we select the most popular blend, which is a mix of wheat called "Meunier" and "Force". This selection restricts our data to 266 examples out of the 873 available.

Regarding the SCADA database, the tables store many null values due to non-movement of wheat grains on the production line. These rows are deleted. Regarding Excel files, information is sometimes missing or incoherent due to manually completed information. Missing information such as wheat properties are in some cases completed by collecting additional information from other sources of data. In other cases, statistical averages are performed. For example, the average grain temperature over the week is used to fill in the missing data on this attribute. Sometimes, the attributes contain too much missing information that has been left blank by the user. Attributes like specific grain weight and waste quantity have been removed for this purpose.

Visualization techniques such as pair plots and DBSCAN (Density-Based Spatial Clustering of Applications with Noise) plots are used to detect outliers in the data. DBSCAN is a clustering algorithm (Burkov, 2019) based on density that is effective at detecting outliers in a dataset (Ester et al., 1996). Outliers are, most of the time, sources of inconsistent values or errors. To apply the DBSCAN algorithm and visualize outliers, the data is first scaled by standardizing the features variables. Two hyperparameters are defined, ϵ and n . ϵ represents the maximal neighborhood distance around one point, and n represents the minimum number of data points that are present in a neighborhood to form a cluster. Isolated points are then considered as outliers. Figure 5.2 shows a 3D graph in which the target wheat moisture increase (%), the real increase in wheat moisture (%) and the resting time period (hour) are plotted on the x, y and z axes, respectively. Summer and winter seasons are represented by squares and circles.

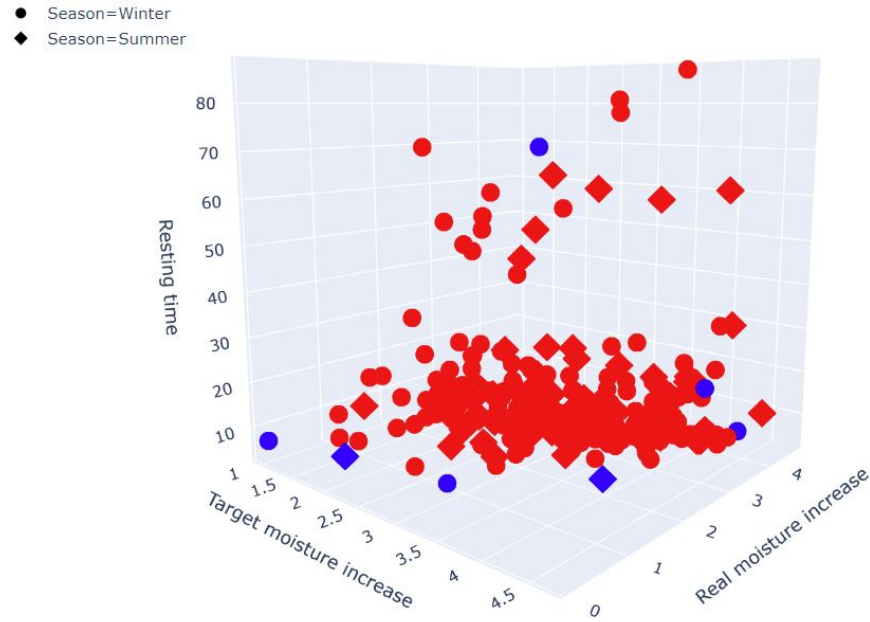


Figure 5.2 DBSCAN plot to detect potential outliers

On Figure 5.2, the potential outliers are identified in blue. Each potential outlier is specifically examined. In case of breakage or production shut down, the example is removed. If errors or changes in operations are written down, the values are corrected based on the comments and notes of the operators written in the Excel file. In other cases, the values are left as they are. This cleaning method led to the removal of 9 examples from the 266 examples available. Finally, the dataframe is divided such that 85% of the data is used for training and 15% is used for testing.

5.3.3.3 Creation of the predictive models

The different models are learned (LASSO, RIDGE, ElasticNet and a regression tree), such that the output variable is the increase in the moisture content of the grains of wheat after the first tempering stage. A resampling method is used, such as k-fold cross-validation with $k = 5$, and the average performance is evaluated. The hyperparameters are obtained using the *GridSearchCV* module available in the *Scikit-Learn* python library.

5.3.3.4 Validation of the models

Each model is evaluated on the test set using R^2 and Root Mean Square Error (RMSE). R^2 (equation 5.8) represents the degree of variation explained by the model (James et al., 2013). R^2 fluctuates between 0 and 1. A high value of R^2 shows a good fit of the model to the data, making it accurate for future predictions. A threshold, scaled to the output variable and applied to RMSE (equation 5.9), can indicate whether a model is accurate enough to make correct predictions, depending on a user's needs. The RMSE explains the average deviation between the predicted values and the actual values. These two performance indicators are used to evaluate and compare the performance of the regression models.

$$R^2 = 1 - \frac{\sum_{i=0}^N (y_i - \hat{y}_i)^2}{\sum_{i=0}^N (y_i - \bar{y}_i)^2} \quad (5.8)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^N (y_i - \hat{y}_i)^2} \quad (5.9)$$

N is the total number of examples (tempering production runs), y_i is the real output for a tempering production run (i), $\hat{y}_i$ is the corresponding predicted value and $\bar{y}_i$ is the average output across N examples. Each example represents a tempering production run using a different batch of wheat and different tempering conditions.

5.4 Results

Table 5.1 illustrates the performance of the different models. R^2 and RMSE are used to rank the models. The models were evaluated in a first approach without weather data, and then evaluated with them.

Table 5.1 shows that the model that performs the best for this problem is LASSO when weather data is included. It gets the highest R^2 score and the smallest RMSE value. We note that weather data improves the accuracy of predictions. Weather data, in our case, includes the average relative humidity and the average temperature of the day when the tempering process occurs. This validates the information given by the miller in past interviews so they

Table 5.1 Performances of predictive models

Model	With weather data		Without weather data	
	R ²	RMSE	R ²	RMSE
LASSO	0.647	0.428	0.600	0.456
RIDGE	0.639	0.433	0.601	0.455
Elastic Net	0.620	0.444	0.606	0.452
Regression tree (XGBoost)	0.578	0.469	0.574	0.471
OLS	0.585	0.571	0.567	0.586

know the relevant variables.

In the LASSO model, 9 variables remained from the 11 variables available. The two variables that have been left out by the regression model are *the target wheat moisture content* and *the initial wheat moisture content*. The LASSO regression equation that predicts the increase in the moisture content of wheat grains (y) is expressed as follows:

$$\begin{aligned}
 y = & 2.64 + 0.57(X5) - 0.27(X3) + 0.23(X1) - 0.21(X6) \\
 & + 0.19(X4) - 0.11(X8) + 0.09(X9) \\
 & + 0.04(X7) - 0.01(X2)
 \end{aligned} \tag{5.10}$$

with X_i defined in Table 5.2.

Table 5.2 Variables used in the models (controllable parameters in red)

X	Variable	Range of values (Min; Max)
X5	Water flow rate (L/h)	(100.0; 447.9)
X3	Wheat flow rate (Kg/h)	(5000.0; 8500.0)
X1	Water quantity (L)	(149.0; 3330.0)
X6	Wheat quantity (Kg)	(3610.0; 66071.8)
X4	Wheat temperature (°C)	(2.8; 31.2)
X8	Mean temperature weather (°C)	(20.2; 27.7)
X9	Mean humidity weather (%)	(42.5; 93.0)
X7	Resting time (h)	(4.50; 86.98)
X2	Wheat protein content (%)	(10.57; 13.82)

Table 5.1 presents the various variables used in the LASSO model by order of importance

to explain the percentage increase in wheat moisture content. The lines highlighted in red are parameters that can be controlled by the miller during the tempering process. For each variable in Table 5.1, the range of values is mentioned. Based on the data collected at the mill and used in this study, the resting times range from 4.5 to 86.98 h. A high resting time often corresponds to a wheat resting time covering the entire weekend and extends several hours before or after.

The coefficients in the equation 5.10 show the degree of influence of each feature variable on the increase in moisture content in the wheat. We note that among the four controllable parameters that the miller can adjust during the tempering stage, three have a significant impact: the flow of water and wheat that is sent to the dampening machine and the quantity of the wheat.

The flow of water and wheat could be adjusted based on the linear regression equation for a steep increase in the wheat moisture content of the grains of wheat. According to the standardization made, an increase in 14 liters of water per hour results in an increase of 0.1% in wheat moisture content at the end of the tempering stage. The quantity of wheat is mainly dependent on the wheat recipe, decided upon by the miller at the beginning of the production run and the number of resting silo available for the tempering stage. We observe that the weather features that include temperature and relative humidity have a significant role in the tempering process of the grains of wheat. Although the miller may be aware of these characteristics, the LASSO regression equation offers a way to estimate their degree of influence on the process.

Based on the OLS model, three variables were not found to be statistically significant. These include wheat protein content, resting time and mean weather temperature. However, these variables were nevertheless retained in the model mainly for two reasons. Firstly, based on the miller's expertise, protein content and the average weather temperature are two variables that influence the resting time required for wheat grains. Resting time is a key controllable parameter that would be interesting to optimise. Secondly, although they are not significant, they improve the performance of the model in this study. They reduce the average prediction error by 2.6

To get an overview of the performance of the LASSO model prediction, a graph representing the actual values versus the predicted values is presented in Figure 5.3.

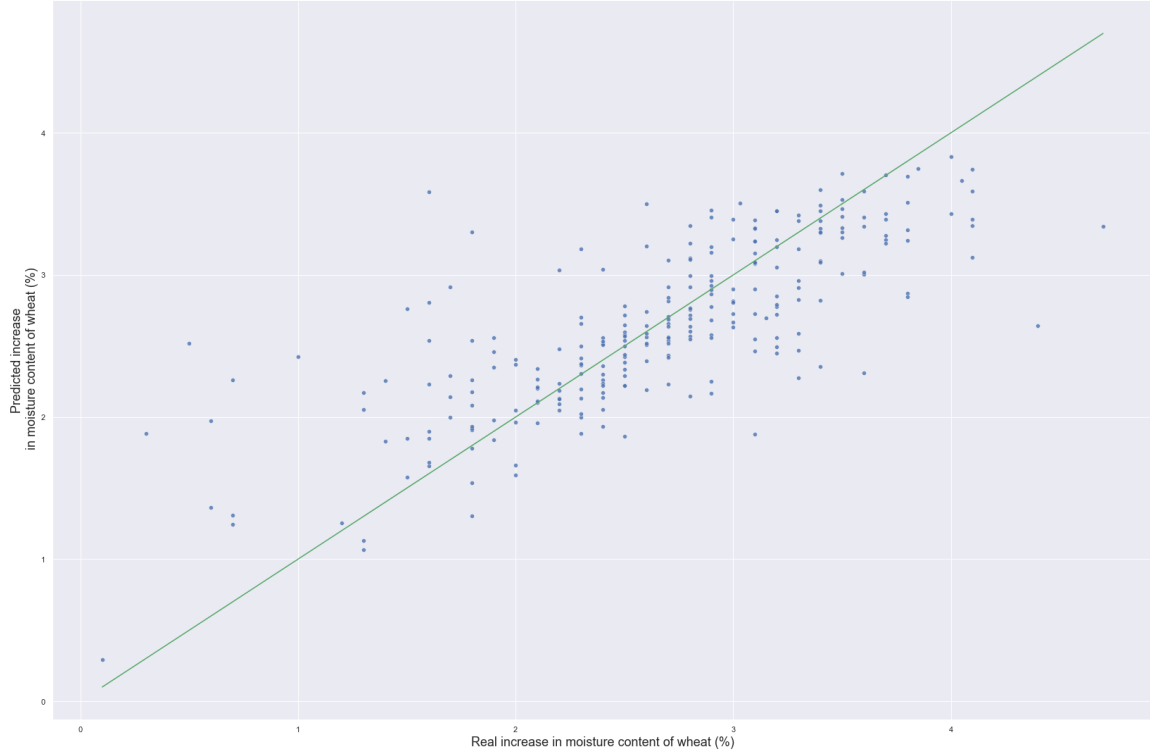


Figure 5.3 Scatter plot of predicted value from a LASSO model against real values

Although a linear trend can be observed, and most of the points are grouped around the green line, we notice that the model tends to predict values that are far from those expected. These values correspond to relatively low (between 0.5% and 1.5%) and very high (between 3.5% and 5%) moisture increases in the wheat grains.

5.5 Discussion

According to Table 5.1, the LASSO model offers the best performance, reaching a R^2 of 0.647 and a RMSE of around 0.428. This result is accurate enough to provide information about the tempering process and the adjustments that could be initiated by operators. With a RMSE of around 0.428, we obtained an average deviation error of 0.428% for the prediction of the increase in moisture content of the grains of wheat. Since most moisture content

increase values fluctuate between 0% and 5%, the decision support tool based on this model provides some level of control to the operator in the tempering process.

From Table 5.2, among the input variables selected to build the model, we see that the LASSO regression model does not rely on the variable specifying the initial moisture content of the wheat kernels, nor on the target moisture content. Yet, Almaraz et al. (2008); Moss (1977) explain that the initial wheat moisture content is important in the speed and quantity of water particle absorption in the grain of wheat. A dry grain, like a dry soil, will tend to absorb water slower than a humid grain or soil. The absence of these two input variables could be explained by information that was already entered when determining the water flow in equation 6.2, which does take them into account.

This decision support tool is based on a linear model that seeks to predict the rate with which moisture increases in wheat grains. However, a grain of wheat obviously cannot absorb an infinite amount of water and reacts differently depending on the moisture threshold reached. It is therefore important to remember that the tool is most useful in wheat grain tempering to condition it for milling. This conditioning aims to temper grains up to a maximum threshold of 17-18% depending on the type of flour that will be produced.

In Figure 5.3, we note that the model performs poorly for some ranges that predict an increase in wheat moisture, such as from 0.5% to 1.5% and 3.5% to 5%. This poor performance could be explained by the fact that few examples are available for these two ranges of increases in wheat moisture. Another reason for this low performance could be a lack of features variables in the set of input variables. Additional key feature variables such as wheat hardness, wheat texture or water temperature would provide information that could explain the reasons for a small or large increase at the end of the process that happens in specific cases. Even though Posner and Hibbs (2005) explain that wheat grain hardness affects the amount and rate of water uptake by a wheat grain, Butcher and Stenvert (1973) showed through a study of different Australian wheat varieties that hardness has no influence on the rate of water absorption in a grain. These contradictory results underscore the interest in analyzing the importance of hardness in Canadian wheat varieties and their influence in the tempering process.

In addition to wheat hardness, the vitreosity of the grain may also play an important role in water absorption. Vitreosity provides information about the appearance of a grain of wheat. Vitreous wheat grains have a dark, translucent and glassy appearance in contrast to mealy (non-vitreous) wheat grains (Butcher and Stenvert, 1973). According to Dobraszczyk et al. (2002), the vitreousness of the grains is correlated with the density of wheat grains. A denser grain with less porosity tends to be vitreous. High vitreosity makes the grain much more resistant to water penetration during tempering. This additional information would be useful in adjusting the controllable parameters of the tempering process to achieve the desired moisture content. In this study, this decision support tool is limited to one category of wheat mixtures, which concerns hard wheat varieties suitable for all-purpose flour. To generalise the model for all types of wheat, the other categories will have to be taken into account. For this, a rigorous approach would be to measure the hardness and vitreousness of wheat to distinguish soft wheat from hard wheat. Based on the literature review, the wheat hardness, the wheat vitreosity and water temperature are avenues that could explain better the water absorption of wheat grains in the tempering stage (IAOM, 2018; Moss, 1977). According to Moss (1977), the gradient of water migration in wheat grain is influenced by the bran morphology, endosperm structure, protein content and initial moisture content. In each variety of wheat, different bran morphologies and endosperm structures are found. Bran morphology is influenced by the vitreosity of the wheat grain (Moss, 1977). While the endosperm structure is influenced by the density and hardness of the wheat grain (Dobraszczyk et al., 2002). The wheat density and hardness are primarily determined by the wheat variety but are also strongly influenced by environmental factors during growth and development (Dobraszczyk et al., 2002). Collecting and adding variables that characterise wheat hardness, density and vitreosity, in addition to the protein content and initial moisture content of the wheat grain that we already have, would provide relevant information to the regression model about the water migration gradient for a wheat batch. This could improve the performance of the model in predicting the percentage increase in moisture content in a batch of wheat during the tempering process. Although these variables are not included in the regression model, the framework of the model presented in this paper is still applicable to other milling industries. For this, the LASSO regression model should be

trained with the data collected at the mill to find the relevant coefficients for each variable considered. This approach makes it possible to measure the degree of influence of each variable on the increase in wheat moisture content at the end of the tempering process in a specific mill. It provides millers with relevant information on how to adjust the various controllable tempering parameters. Based on the regression model developed from the data collected at the mill, the coefficients explain the sensitivity of a variable on the increase in wheat moisture content. In future work, one aspect that needs to be further investigated is the influence of the size and shape of water droplets used during the tempering process on wheat moisture increase. According to Chen et al. (2020), the use of steam water for wheat grain tempering accelerates the penetration of water particles into the grain compared to the use of water stream. Analysis of the size and shape of water droplets would be relevant to determine their influence on the increase in wheat moisture content and could further help predict this increase.

5.6 Conclusion

In this study, a decision support tool was built from data collected along the value chain in wheat milling. The goal is to offer better control of the tempering process that conditions the wheat grains before the milling process. The proper conditioning of the grain improves milling efficiency, which influences the mill's performance and the quality of the flour produced. This decision support tool will help new operators adjust the tempering parameters according to environmental factors, the properties of the grains of wheat, and the desired type of flour.

The proposed tool is based on a multiple regression model. Several regression models have been built and compared to determine the wheat moisture at the end of the first tempering stage. The models have been trained on one specific wheat blend category to limit our field of study. Among the different models, the LASSO regression model offers the best accuracy, with an average prediction error of 0.428% and a R^2 of 0.647. This model shows the relevance of using meteorological data such as average temperature and average relative humidity. Moreover, the mathematical equation of the LASSO model enables an understanding of the influence of each input variable used in the final increase in the moisture content of the grains

of wheat.

Although the LASSO model shows good results, more in-depth research is needed to improve the reliability of the model by trying to reduce the average prediction error. In this regard, further studies will be conducted to include wheat hardness, wheat vitreosity and water temperature. In addition, the analysis of the different categories of wheat mixtures will have to be taken into account in order to extend and generalize the model according to the type of wheat mixture made by the miller. Finally, to effectively control the entire tempering process, the second and third tempering stage will also need to be studied.

5.7 Acknowledgements

We would like to thank our industrial partner, La Milanaise, for their support and collaboration in this project, as well as MAPAQ (project IA119053) for their financial support.

CHAPITRE 6 ARTICLE 4: A DECISION SUPPORT TOOL TO ANALYZE THE PROPERTIES OF WHEAT, COCOA BEANS AND MANGOES FROM NIR SPECTRA

Parrenin, L., Danjou, C., Agard, B., Marchesini, G., and Barbosa, F. (2023c). A decision support tool to analyze the properties of wheat, cocoa beans and mangoes from NIR spectra.

Manuscript submitted for publication

Submitted on January 13, 2024 at Journal of Food Science

Abstract

Near infrared spectroscopy (NIRs) is an analytical technique that offers a real advantage over laboratory analysis in the food industry due to its low operating costs, rapid analysis and non-destructive sampling technique. Numerous studies have shown the relevance of NIR spectra analysis for assessing certain food properties with the right calibration. This makes it useful in quality control and in the continuous monitoring of food processing. However, the NIR calibration process is difficult and time-consuming. Analysis methods and techniques vary according to the configuration of the NIR instrument, the sample to be analyzed and the attribute that is to be predicted. This makes calibration a challenge for many manufacturers. This paper aims to provide a data-driven methodology for developing a decision support tool based on the smart selection of NIRs wavelength to assess various food properties. The decision support tool based on the methodology has been evaluated on samples of cocoa beans, grains of wheat and mangoes. Promising results were obtained for each of the selected models for the moisture and fat content of cocoa beans, acidity and vitamin C content of mangoes and protein content of wheat (B) with a RMSEP score of 0.354, 0.913, 20.028, 0.848 and 0.490 respectively. Moreover, the proposed methodology allows results to be obtained that are better than benchmarks for the moisture and protein content of wheat (A) with a RMSEP score of 0.337 and 0.177 respectively.

Keywords: Industry 4.0; decision support tool; near-infrared spectroscopy; food quality;

machine learning; optimization algorithm.

6.1 Introduction

Food products are living and perishable materials. Their quality varies throughout their life cycle, including when they arrive at the place of manufacture. The growing demands of food production require fast and efficient analytical methods for controlling the product quality on production lines. Quality assessment is an important part of the transformation process of raw living materials in order to meet the standard quality requirements of consumable products (Jespersen and Munck, 2009; Parrenin et al., 2023b). Traditional analytical methods are labor-intensive, time-consuming and costly. The use of connected equipment, data and analysis tools in the age of the fourth industrial revolution offers new perspectives on solving this type of problem.

One of the most common methods for assessing the quality of food that offers a simple, rapid and cost-effective alternative to traditional methods is near-infrared spectroscopy (NIRs). The use of NIR spectroscopy to measure food quality features has been well-studied in the literature (Huang et al., 2008). It allows for the real-time measurement of sample quality on the production line, with no sample preparation, enabling real-time process monitoring (Porep et al., 2015). However, as NIRs is based on indirect measurements and produces very complex and broad spectra, calibration is necessary. This NIR calibration process is difficult and time-consuming as many factors such as the type of food being analyzed and material of the equipment used will influence this process. This is a challenge for many manufacturers.

A large amount of research explains the theory, and demonstrates the relevance of using NIR spectra to analyze the properties of cocoa beans, grains of wheat and mangoes properties (Cen and He, 2007; Munawar et al., 2016; Mutlu et al., 2011; Teye et al., 2020). However, no research has focused on developing a methodology to develop a decision support tool to predict quality attributes with the selection of the features of those products. Therefore, to reduce this gap, this paper aims to provide a data-driven methodology for developing a decision support tool based on a smart NIRs wavelength selection to assess various food properties. The analysis of selected features of NIR spectra is intended to ensure the robustness

of the model by reducing its complexity. It could therefore limit the risk of model over-fitting and, moreover, provide explanatory results linked to the chemical properties of the elements. In this way, the manufacturer will be able to evaluate a product's quality features quickly and simply throughout the production chain in real-time. This paper is based on open-access datasets so that future research can compare and improve upon this approach.

This paper first presents a literature review of NIR technology and data analysis techniques in Section 6.2. Next, the proposed method for assessing food properties from NIR spectra is explained in Section 6.3. Then, Section 6.4 presents the results obtained from open-access datasets. Finally, Section 7 discusses the results obtained.

6.2 Literature review

In the field of food quality assessment, near-infrared spectroscopy (NIRs) has emerged as an essential analytical tool, as it has a significant impact on quality measurement and analysis in the food industry. This literature review section is structured into two key subsections: the first explores the wide-ranging applications of NIRs in quality assessments, while the second focuses on the analytical techniques employed for interpreting NIR spectra. This two-part approach provides an in-depth perspective on the role of NIRs in characterizing food quality and the methodologies used to harness its potential. We begin by examining NIRs applications in quality assessment, followed by a detailed analysis of the techniques used for interpreting NIR spectra, thereby offering a comprehensive overview of the subject's interdisciplinary dimensions and the interplay between spectroscopy and data analysis in the context of food quality assessment.

6.2.1 NIR spectroscopy in quality measurement

Spectroscopy is a technique used to study the interaction of matter with electromagnetic radiation. It provides valuable information about the structure, composition and properties of various substances. Near-infrared (NIR) spectroscopy is a tool that uses infrared rays that range 780 to 2500 nanometers in the electromagnetic spectra region (Nicolăi et al., 2007; Porep et al., 2015). On the electromagnetic spectra scale, this region lies between visible

light and mid-infrared (MIR) light.

Near-infrared (NIR) light was first discovered in the 19th century by Herschel (1800). However, it was not until the 1970s that the development and application of NIRs began to attract interest, particularly in the agri-food sector. The early work of Norris et al. (1976) and Norris (1982) demonstrated the potential of infrared reflectance for rapid quality assessment of forages and oat groats. This work demonstrated a strong linear relationship between the absorption intensity of certain wavelengths and certain food quality features. The use of NIR wavelengths has the advantage of being a non-destructive and non-invasive method of sample analysis. Applied to the food industry, this offers new solutions for monitoring production quality, while limiting the costs and losses associated with destructive laboratory testing and sample analysis. Although different spectroscopy techniques are used depending on the shape and material of the sample (Zhang et al., 2022), the principle of this method remains the same. Several infrared waves ranging from 780 to 2500 nm are sent toward the sample, and the reflected or transmitted waves are measured according to their intensity (Nicolăi et al., 2007). The absorption intensity of these waves fluctuates according to the sample's chemical and light-scattering properties. The absorption intensity of each wavelength forms the NIR spectra of the sample which provides useful information for sample quality analysis.

The NIR technique has been used in various industrial sectors, including food, petrochemicals and pharmaceuticals (Blanco and Villarroya, 2002). Its application in the food sector is becoming increasingly widespread, as highlighted in (Huang et al., 2008). In the milling industry, the NIR technique has even proven to be a reference for the rapid analysis of the intrinsic properties of wheat grains, given the speed, low cost and reliability of the results (Jespersen and Munck, 2009; Parrenin et al., 2023a). Research is being extended to other types of food that have certain key quality features that are difficult to measure from the NIR spectra with the same level of precision.

There are several vibrational spectroscopy techniques available for a real-time quality assessment of food samples on the production line. The best-known include the NIR, MIR and Raman techniques (Abbas et al., 2020). NIR light compared to MIR and Raman light has a greater penetration depth. This allows bulk and raw material to be analyzed without the

need for extensive sample preparation. Moreover, The NIR spectra light is strongly impacted by physical features (e.g. particle size, moisture content and protein content) and is therefore ideally suited for the capture of these characteristics. This is explained by the fact that most organic molecules usually have absorption bands in the NIR spectra region (Baianu and Guo, 2011). Therefore, this make NIR spectroscopy one of the most useful spectroscopic techniques for food analysis. However, these measurements are limited to features present in fairly large quantities and would probably not detect features that represent $<0.1\%$ of the grain mass (Dowell and Maghirang, 2002). The main drawback of NIR application is the need for robust calibration models and calibration transfer between NIR instruments to accurately analyze the desired food properties (Kondepoti and Heise, 2008).

Calibration of an NIR machine involves analysis of the NIR spectra. Depending on the physical state of the sample, the intrinsic properties of the sample, the type and configuration of the equipment, the manufacturing environment and the desired output feature, different pre-processing and analysis methods are required to accurately measure the output variable (Zhang et al., 2022). NIRs analysis relies on a secondary analytical approach, using conventional chemical methods to establish the relationship between the NIR spectra and the parameters of interest. This relationship is generally established using machine learning techniques and sometimes deep learning techniques (Yap et al., 2022). NIR spectral analysis techniques will be discussed in the following subsection.

6.2.2 Analysis techniques of NIR spectra

NIRs analysis is divided into several parts that include a pre-processing stage, a learning stage and a feature selection stage (Zhang et al., 2022). In most cases, a pre-processing step is required at the start of the analysis to reduce the noise present in the spectra data. Noise is an element of the environment that interferes with the measured signal. Its reduction makes spectral analysis and interpretation simpler. To achieve this, different signal pre-processing algorithms are used based on mathematical techniques. Popular approaches include smoothing, normalization and component-based methods (Jiao et al., 2020). The first and second derivatives of spectral data are often calculated and smoothed using the

Savitzky–Golay (SG) filter (Jiao et al., 2020). This method fits a polynomial of a given order in a sliding window across the spectra signal (Savitzky and Golay, 1964). A moving average would be a specific case of SG filter with a polynomial of order 0. Among standardization methods, the multiplicative scatter correction (MSC) is widely used to compensate for additive and/or multiplicative effects in spectral data due to the scattering of light (Jiao et al., 2020; Maleki et al., 2007). Similarly, standard normal variate (SNV) has the same effect. The main difference is that the SNV method does not require a reference spectra, which can be an advantageous technique if the typical NIR spectra profile for the sample being analyzed is unknown. It normalizes spectra by subtracting each spectra from its own mean and dividing by its own standard deviation (Sandak et al., 2016). In addition to these two normalization techniques, standardization and min-max normalization are two widely used techniques in data science and commonly used in the pre-processing of NIR spectra signals. Finally, component-based methods that include orthogonal signal correction (OSC) are designed to minimize in the spectral data the variability that is not related to the concentration of the compounds in the sample (Jiao et al., 2020).

In spectral analysis, effective pre-processing techniques are essential to refine the raw spectral data, enhancing the accuracy and interpretability of subsequent analyses (Oliveri et al., 2019; Rinnan et al., 2009). These techniques are an essential preparatory step, addressing inherent challenges such as noise, distortions and irregularities that may obscure the underlying signal. Through meticulous cleaning and smoothing procedures, pre-processing reveals the true essence of the spectral information, improving the quality of the dataset. As a result, the application of these techniques contributes significantly to the robustness of spectral analysis, providing researchers and practitioners with a clearer and more reliable basis on which to base their interpretations and draw meaningful conclusions. In the context of scientific research, the role of pre-processing techniques in improving the overall fidelity of spectral data cannot be overstated, acting as a catalyst for more accurate and insightful investigations.

The analysis of the processed NIR spectra is carried out by artificial intelligence algorithms, which seek to determine the relationship between the intensity of the NIR wavelengths and the content of the chemical compound present in the food sample. Partial Least Square Re-

gression (PLSR) and Support Vector Regression (SVR) are two machine learning algorithms widely used in NIR spectral analysis to determine this relationship (Porep et al., 2015; Zhang et al., 2022). These techniques are well suited to high-dimensional data problems such as NIR spectral analysis. PLSR is a multivariate regression technique used for modelling the relationships between input variables and response variables. It learns this relationship by constructing new variables, called components or latent variables, that are linear combinations of the original input variables. These components are created in a way that maximizes their covariance with the response variables. PLSR reduces the number of input variables to a smaller set of uncorrelated components and performs least squares regression on these components. Once the PLS regression model is trained on the data, it can be used to make predictions on new, unseen data. The model takes the input variables given and uses the learned relationships to predict the corresponding response variable. This is a well-known method of dimension reduction, useful when a dataset has more variables than observations, or when the input variables are correlated. Moreover, PLSR can handle situations with multicollinearity effectively, which is a common problem in traditional linear regression.

Support Vector Regression (SVR) is a machine learning algorithm used for regression analysis (Drucker et al., 1996). It is a variation of the popular Support Vector Machine (SVM) algorithm, which is primarily used for classification problems. SVR is designed to predict continuous numeric values rather than class labels. The SVR algorithm, like PLSR, is useful for solving high-dimensional feature regression problems (Thissen et al., 2004). It applies relatively well when the number of features is greater than the sample size and in cases of data non-linearity (Thissen et al., 2004). The SVR aims to find the best hyperplane in a high-dimensional feature space that best represents the relationship between the input variables and their corresponding output values (Thissen et al., 2004). To do this, it will try to find the hyperplane that will fit as many data points as possible within a threshold value (ϵ) while reducing the complexity of the model by minimizing the model coefficient (Zhang and O'Donnell, 2020). In SVR, the hyperplane is defined by two parallel boundaries, known as the "support vectors" which lie on either side of the hyperplane (Zhang and O'Donnell, 2020). The boundaries around the regression hyperplane are set by the ϵ hyper-parameter. The distance between the boundary and the regression hyperplane allows a certain tolerance

with regards to prediction errors. Data points within this distance do not contribute to the loss function, whereas data points (predicted values) that fall outside of the boundary reflect a loss that the algorithm attempts to minimize (Drucker et al., 1996).

Deep learning techniques have emerged from research into artificial intelligence and are a subfield of machine learning (Kelleher, 2019). They have shown their effectiveness in accurately analyzing NIR spectra in the presence of a large amount of features (Zhang et al., 2022). Deep learning techniques enable data-driven decisions by identifying and extracting patterns from large data sets that accurately map complex input variables to the right output outcome (Kelleher, 2019). Although they can detect patterns in large structured or unstructured data sets, a large amount of data is required for good performance. However, due to the high cost in terms of manpower resources and time required to label data, it is rare in the manufacturing environment to possess large quantities of labelled data for each NIR spectra. For this reason, the use of machine learning techniques is more appropriate for NIR spectroscopy analysis in the food processing environment and will be discussed in this paper.

The large number of spectral variables and the high collinearity between certain variables often make the prediction of a dependent variable difficult using multiple linear regression techniques (Mehmood et al., 2012). Although the PLS regression technique mitigates the risk of collinearity and reduces the number of input dimensions, it is essential to select the most useful spectral variables to improve the accuracy and robustness of the prediction model (Hong et al., 2020).

Mehmood et al. (2012) explored different techniques for selecting near-infrared spectral variables. These techniques can be grouped into 3 categories that include filter, wrapper and embedded methods. The filtering methods are independent of the learning algorithm and are based on the general characteristics of the data. The wrapping methods interact with the learning algorithms with an iterative loop. They are computationally expensive, but provide more accurate information than filtering methods. Finally, embedded methods are a combination of filtering and wrapping methods. Among these methods, metaheuristic algorithms have received a great deal of attention in the solution of many optimization problems. They

stand out for their ability to efficiently search large feature spaces and find optimal feature subsets. In feature selection, metaheuristic algorithms randomly search for the best subset of possible solutions. To do this, it interacts with the learning algorithm during the training phase. It is therefore considered to be a wrapping method (Agrawal et al., 2021).

In the family of metaheuristic algorithms based on swarm intelligence, a popular technique developed by Kennedy and Eberhart (1995) is particle swarm optimization (PSO). This is an optimization algorithm inspired by swarm behavior observed in nature. Kennedy and Eberhart (1997) later introduced a discrete binary version of PSO (BPSO) to solve discrete optimization problems. It aims to identify the best element among a set of possible solutions that minimize or maximize (depending on the problem) the objective function defined. To do this, the algorithm will search and test different solutions called "particles" in the defined space, using an iterative process as described by Marini and Walczak (2015). Initial particles are chosen, and at each iteration, the algorithm updates the new particle positions according to velocity vectors. The iterative process is terminated according to a defined stopping criterion. The solution that best optimizes the objective function is retained.

In the BPSO algorithm, a number of parameters presented in Table 6.1 must be set to configure the search space of possible solutions and the exploration of those solutions (Juneja and Nagar, 2016; Miranda, 2018). This parametrization is important since it affects the quality of the solutions and the speed of problem-solving.

The dimension of the problem represents the number of features. The number of particles (also called swarm size) defines the number of individuals in the swarm. The larger the

Table 6.1 BPSO parameters description

Parameter	Symbol	Description
Dimension of the problem	D	Number of features
Number of particles	Np	Number of individuals in the swarm
Cognitive parameter	C1	First acceleration constant
Social parameter	C2	Second acceleration constant
Inertia parameter	w	Speed for global and local search
Neighborhood size	k	Number of particles in a neighborhood
Vector norm function	p	Metric used for distance calculation

swarm size, the greater the initial diversity, meaning that more of the solution space can be explored per iteration. However, a large amount of particles increases the computational cost per iteration. A trade-off must therefore be made between resource cost and the quality of the optimal result found. Generally, the PSO algorithm is set empirically and is able to find optimal solutions with a swarm size composed of 20 to 50 particles depending on the dimensionality of the problem (Juneja and Nagar, 2016). The acceleration constants $C1$ and $C2$ each modulate the magnitude of the steps taken by the particle, $C1$ in the direction of its best personal performance and $C2$ in the direction of its best overall performance (Marini and Walczak, 2015). Kennedy and Eberhart (1995) explain in their work that both acceleration constants set at 2 generally works well in most cases. The inertia parameter later proposed by Shi and Eberhart (1998) in a modified version of the particle swarm optimizer plays the role of balancing global and local search. From their work, it has been shown that a PSO with an inertia weight between $[0.9 \text{ and } 1.2]$ will perform better, meaning that it is more likely to find the global optimum in a reasonable number of iterations. The size of the neighborhood defines the number of interactions performed between particles to discover the best solution in the neighborhood and obtain the best local solution (Suganthan, 1999). This parameter helps avoid falling into the trap of local minima. Smaller neighbourhoods converge upon a solution more slowly, but offer more reliable convergence and optimal solutions. The vector norm is an additional parameter that determines the metric to be used to minimize the loss function and update the search exploration. It is used to calculate the distance between two particles. This parameter proposes two metrics: L1 distance, which corresponds to the Manhattan distance, and L2 distance, which corresponds to the Euclidean distance (Miranda, 2018). Finally, in addition to the parameters listed in Table 6.1, a stopping criterion such as the maximum number of iterations is defined to stop the iteration loop and the search.

The BPSO algorithm explores solutions iteratively and selects near-infrared wavelengths based on a binary method. The binary number 0 in the solution vector indicates non-selection of the wavelength, while the number 1 indicates a selection.

The research aimed at forecasting food quality properties through NIR technology primarily relies on complete spectra (Huang et al., 2008; Porep et al., 2015). Various methods and

models are used, depending on the instrument configuration, the sample microstructure and the specific quality property to be predict. In the subsequent section, we will introduce a versatile approach capable of adapting to these distinct variables to predict quality properties via NIR technology, using the most relevant wavelengths.

A variety of methods have been investigated in the literature, including the pre-processing of NIR signals, the use of regression algorithms and the selection of attributes for the analysis of NIR spectra. However, no research has proposed a usable and reproducible framework combining these different methods to predict food quality properties from a selection of NIR wavelengths. We therefore are putting forth a framework for the development of a decision support tool for selecting the best machine learning model to predict the desired level of quality on the basis of an intelligent selection of NIR wavelengths and existing pre-processing methods.

6.3 Methods

The literature review in Section 6.2 demonstrates that NIR instruments are very useful in the analysis of food quality properties and are relevant for monitoring food processing on production lines. Despite the availability of numerous spectral analysis methods for analyzing NIR spectra wavelengths, calibrating NIR instruments in the industry continues to pose a substantial challenge for companies, resulting in notable time inefficiencies.

The objective of this work is to propose a data-driven methodology to develop a decision support tool based on the intelligent selection of NIRs wavelengths to assess different food properties. The decision support tool enables the NIR instrument to be calibrated efficiently and the quality of food properties to be measured quickly and simply on the production line in real time. This tool is based on a machine-learning analysis method combined with a BPSO algorithm to extract the most relevant NIR wavelengths. The methodology proposed to build the decision-support tool is illustrated in Figure 6.1, and is divided into 4 steps : (1) data preprocessing, (2) a machine learning model, (3) feature selection and (4) evaluation of the model. Each step in this methodology is explained in the following subsections.

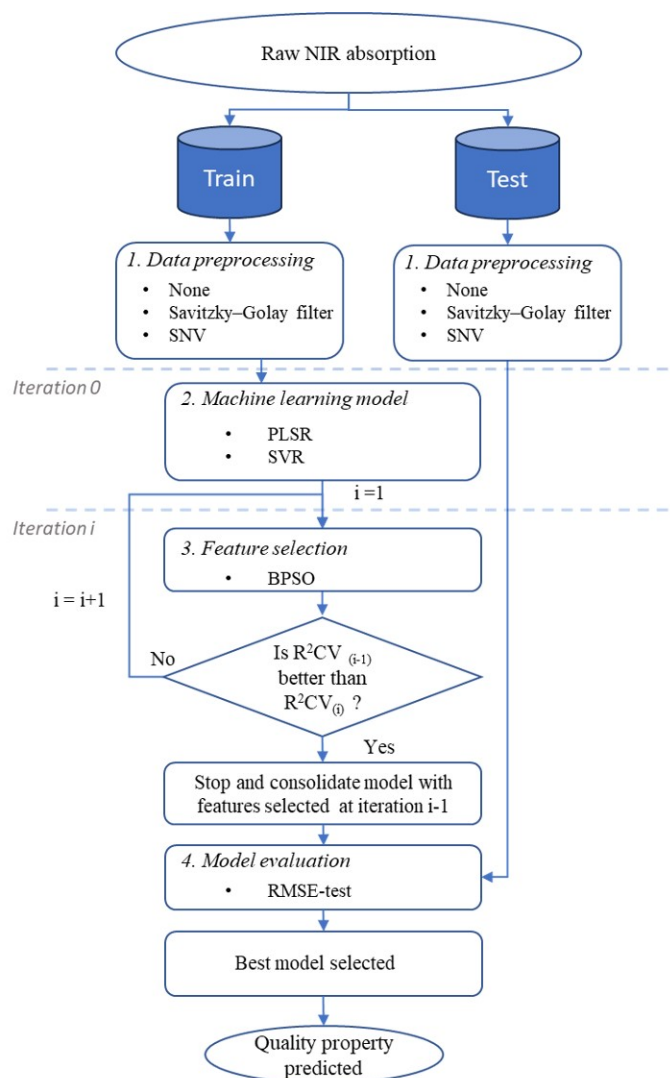


Figure 6.1 Diagram of the method followed for the prediction of the quality properties of raw materials from the developed decision support tool

6.3.1 Data preprocessing

Before pre-processing the input data, the NIR spectra data and the reference quality features data should be divided into two sets following an optimum split ratio for machine learning model development: a calibration set (train set) and a validation set (test set). The purpose of this split is to store data that is not seen by the model so it can be used later to evaluate its performance. When splitting the data into two sets, it is important to ensure that the entire range of values of the desired quality feature is well represented in the training dataset. This is achieved by classifying the continuous values according to certain value ranges labelled as categorical features. If the reference data follows a normal distribution, the data is grouped according to its integer value, with at least two examples in each class. Alternatively, a transformation of the reference variable can be envisaged to approximate a normal distribution. This grouping helps maintain data homogeneity during NIR calibration and reduces the level of uncertainty in the prediction model. Furthermore, the cleaner and more varied the data that is available, the more robust and accurate the prediction model will be as it is trained on a large and varied data sample.

The aim of the pre-processing stage is to reduce the noise present in the raw spectral data. The pre-processing step, which is sometimes necessary, improves the legibility of the information and therefore the accuracy of the quality features predicted based on the NIR spectra. To achieve this, pre-processing methods such as smoothing normalization and/or component-based methods presented in Section 6.2.2 are applied. Sometimes, a combination of these pre-processing techniques can improve the prediction accuracy. In other cases, no pre-processing is required, and pre-processing may even make interpretation more difficult. Choosing the right pre-processing techniques and combination is always a challenge. The choice of the most suitable pre-processing technique for the problem is made empirically and is only selected in the last stage when we can evaluate and compare the most appropriate pre-processing techniques.

6.3.2 Machine learning model

Once the data has been pre-processed and the training data is fully representative of the sample as a whole, it is now possible to process it using a machine-learning model. To do this, two machine learning algorithms, PLSR and SVR, are configured and fed with the training data.

Cross-validation is applied to the training data to avoid over-fitting the model. This method statistically divides the data sample into smaller subsets. During k-fold cross-validation, the training data is divided into k subsets. The model is then successively trained on the k-1 subsets and tested on the remaining set. The evaluation of the training model is calculated by the average performance over k iterations on the test sample. The k parameter is set according to the number of samples available and the user's experience.

For the SVR model, the default hyper-parameters are kept. For the PLSR model, an optimal number of components is chosen. This choice is based on an evaluation of the PLSR model's performance on an iteration loop of the number of components. Observations show that the model tends to overfit the training data beyond 8 components. In this case, accuracy is evaluated according to the cross-validated coefficient of determination from the training set, R^2_{cv} (6.1). R^2_{cv} (6.1) is calculated from the average R^2 (6.2) of each k-fold performed during the cross-validation. R^2 (6.2) is a performance metric that informs on the degree of variation explained by the model. This indicator varies between 0 and 1. A high number, close to 1, demonstrates a good fit of the model to the data.

$$R^2_{cv} = \frac{1}{k} \sum_{i=0}^k R_i^2 \quad (6.1)$$

$$R^2 = 1 - \frac{\sum_{i=0}^N (y_i - \hat{y}_i)^2}{\sum_{i=0}^N (y_i - \bar{y}_i)^2} \quad (6.2)$$

In equations (6.1) and (6.2), N represents the total number of observations in the training

dataset, y_i is the actual reference value of the food quality features (i), $\hat{y}_i$ is the corresponding predicted value from the machine learning model and $\bar{y}_i$ is the average output across N observations.

The optimal number of components for the PLSR model is selected during the next step at the same time the features are selected.

6.3.3 Feature selection

In this methodology, feature selection concerns the selection of the most relevant wavelengths from pre-processed NIR spectra to predict specific food quality features. To do this, the BPSO metaheuristic algorithm is used and applied to the training dataset. The algorithm tests and evaluates different subsets of binary-selected wavelengths. At each iteration, the algorithm compares the scores obtained and selects the best score that minimizes the objective function defined in equation (6.3). The best subset offering the best score of the objective function indicates the wavelengths to be selected for the machine learning model being studied. The objective function written in equation (6.3) is taken from the work of Vieira et al. (2013) and has been adapted to regression problems.

$$j = \alpha * (1.0 - P) + (1.0 - \alpha) * (1 - \frac{N}{Nt}) \quad (6.3)$$

$$where P = R^2cv$$

In equation (6.3), α represents a constant weight to define a compromise between regression performance and the number of features. P is the metric used to measure the performance of the machine learning model. For a regression problem, it was considered appropriate to measure it using the R^2cv , which has a scale of 0 to 1 and minimizes the function in the event of good model performance. Finally, N represents the number of wavelengths selected and Nt the total number of wavelengths available at the start of the iteration.

The chosen parameters of the BPSO algorithm influence the solution exploration space (speed and direction of search). The parameters are tuned by experiment (trial and error) based on previous work focusing on the optimal adjustment of these parameters (Juneja and Nagar, 2016; Kennedy and Eberhart, 1995; Shi and Eberhart, 1998) as specified in Section 6.2.2.

At this stage, two loops are performed. The first loop corresponds to the iterative process of the BPSO algorithm to test different feature solutions and stops according to a stopping criterion. The second loop corresponds to restarting the algorithm, starting with the most relevant features obtained from the first loop. This second loop runs until the R^2cv score of the learning model based on the selected subset of wavelengths no longer improves. Therefore, as soon as the R^2cv score is equal to or lower than the score of the previous iteration, the loop stops and the wavelengths selected from iteration $i-1$ are used.

With the best features of the NIR spectra selected by the algorithm at this stage, it is now possible to evaluate the performance of the different models in the next step.

6.3.4 Model evaluation

Based on the best features of the NIR spectra selected by the BPSO algorithm, the performance of the learning models is evaluated using new data from the test set. To achieve this, models are evaluated according to their predictive accuracy using the root mean square error (RMSE) performance indicator. The RMSE indicator is expressed mathematically in equation (6.4). It evaluates an error between the predicted ($\hat{y}_i$) and actual values (y_i). A low value of RMSE is therefore desired in order to express an accurate model that best fits the data.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^N (y_i - \hat{y}_i)^2} \quad (6.4)$$

In summary, the configuration of model hyperparameters and the selection of relevant wavelengths is assessed on the basis of the R^2cv from the training set, while the training models developed using different data pre-processing methods are compared and evaluated based

on the RMSE from the test set (also referred to in the literature as root mean square error prediction - RMSEP).

6.4 Results

A validation of the method presented in Section 6.3 for developing the decision support tool and assessing its robustness is performed on open-access NIRs data. The aim is to verify that the tool will be able to help calibrate the NIR instrument based on the extraction of relevant wavelengths specific to the food quality feature desired. To this end, results are presented, detailing each stage of the methodology.

6.4.1 Dataset and Samples

Data was collected from freely available datasets of NIR spectroscopy data associated with quality features. Four datasets are studied. Each one focuses on the analysis of quality features on cocoa beans, wheat grains and mangoes, respectively. Table 6.2 provides information on the characteristics of the NIR spectra of each data set, such as size and the range of NIR wavelengths studied. Table 6.3 presents statistical and descriptive information about the food quality features studied for each data set.

The open-access data sets are found online. Data on cocoa beans from Aceh province, Indonesia, provides information on moisture and fat content (Agussabti et al., 2020). Two sets of data concern the grains of wheat. The first wheat dataset (A) covers moisture content and protein content (Johnson et al., 2022b), while the second dataset (B) covers protein content only with a smaller-than-usual NIR wavelength range (Liu, 2016). Finally, data about mangoes imported from Brazil and Spain are collected in Germany, where their total acidity and vitamin C content is measured (Hayati et al., 2020).

Table 6.2 Descriptive statistics of the spectra of each dataset

Dataset	Cocoa	Wheat (A)	Wheat (B)	Mango
N. Wavelength	1557	1557	100	1557
Spectral region (nm)	1000-2500	1000-2500	850-1050	1000-2500

Table 6.3 Descriptive statistics of quality parameters for each dataset

Dataset	Cocoa		Wheat (A)		Wheat (B)	Mango	
Quality parameters	Moisture content (%)	Fat content (%)	Moisture content (%)	Protein content (%)	Protein content (%)	Total acidity (mg.100g ⁻¹ FM)	Vitamin C (mg.100g ⁻¹ FM)
N. Sample	72	72	65	65	523	58	58
Mean	9.03	40.31	11.85	14.11	9.95	495.48	32.20
Min	6.74	35.25	8.70	10.66	6.77	189.72	28.93
Max	12.08	45.74	14.5	17.46	16.95	772.77	35.66
Std. Deviation	1.27	2.15	1.35	2.02	1.60	131.07	1.30

6.4.2 Application of the methodology

This subsection will present the results obtained following the methodology step-by-step, using the data sets on the quality of the wheat (A). The results obtained by applying the same methodologies to the cocoa beans, wheat (B) and mangoes datasets are presented in Appendices A, B and C, respectively.

The data is first split according to the following proportions: 85% for the training set and 15% for the test set. As a result, there are 55 examples in the training set and 10 examples in the test set for cocoa bean data. To separate them in order to obtain a good representation of the training set's data, a new feature is added to categorize the examples in the sample studied according to the value ranges or data distribution of the output variable. Working with continuous output values, a simple and arbitrary technique is to transform these continuous variables into integers and assign them as a categorical class. As a result of an imbalance, they are grouped into 4 classes to obtain a Gaussian distribution according to the number of examples in each group. Using this new categorical variable, we perform stratified sampling at a split-off to ensure that we have at least one example of each category in the training set and to prevent the model from predicting unseen examples of a category.

6.4.2.1 Data preprocessing

The raw NIR absorbance spectra for each of the data sets studied are shown in Figure 6.2. More specifically, Figure 6.2(a) shows the raw NIR spectra for cocoa bean samples, Figure 6.2(b) shows those for samples of the wheat (A), Figure 6.2(c) for the wheat (B) samples and

Figure 6.2(d) for mangoes samples.

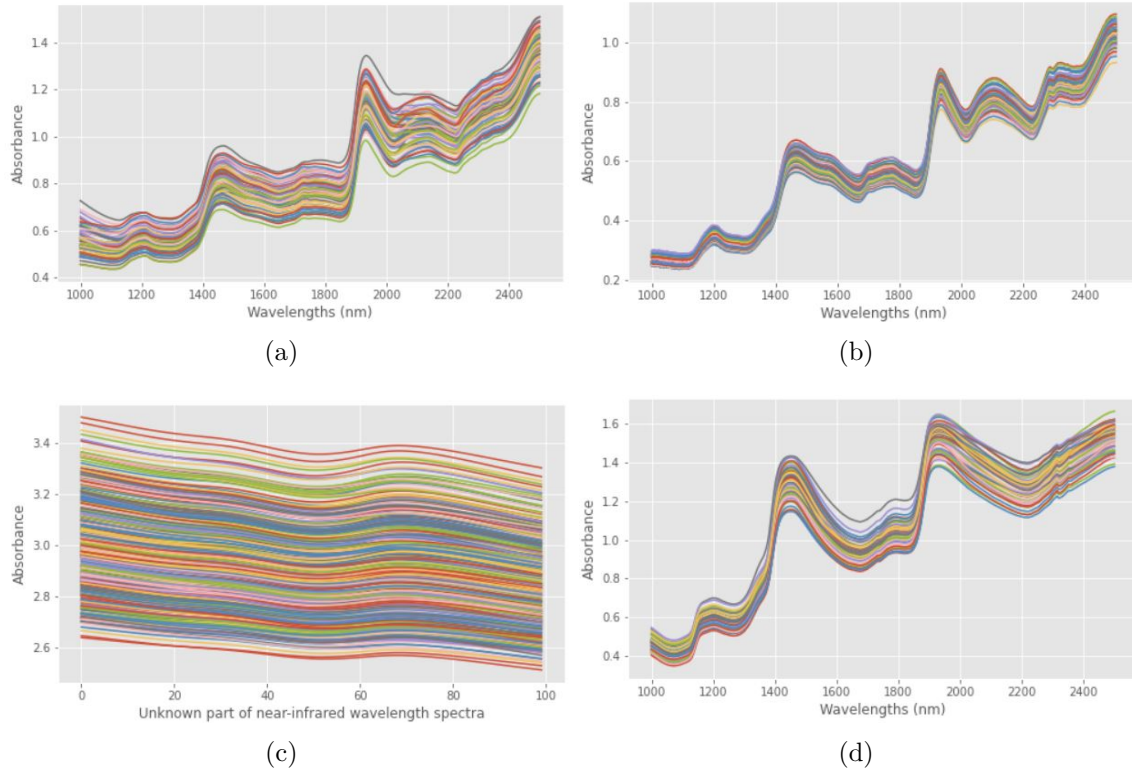


Figure 6.2 (a) Raw NIR spectra of cocoa beans samples; (b) raw NIR spectra of the samples of grains of wheat (A); (c) raw NIR spectra of wheat grain (B) samples; (d) raw NIR spectra of mangoes samples.

Figure 6.2(c) demonstrates that the spectra of the grains of wheat are nearly linear compared to the other 3. This is explained by a more restricted wavelength range and a lack of knowledge of the target wavelengths. This can be seen as a microcosm of the entire NIR spectra. Having no knowledge of the wavelength range for this data, an analysis may be simpler or more complicated, depending on the inference that can be drawn from the prediction of quality features. If we examine the four figures, we can see that the wavelength spectra all overlap. However, some have larger amplitudes, which provide different information, cause noise in the data analysis and will therefore have an impact on the prediction results. For example, if we compare figure 6.2(a) and 6.2(b) between 1500 nm and 1800 nm, we see a greater amplitude in figure 6.2(a), with the absorption intensity ranging from 1 to 1.4 in absorption intensity.

To reduce the level of noise present in the raw data, various pre-processing techniques that have already been proven effective in reducing noise in the NIR spectra are used. These include the Savitzky-Golay filter and the Standard Normal Variant (SNV). However, they are sensitive to the data and hyperparameters of use. This is why they must be tested so that the best configuration for the dataset at hand can be selected. In our case, 7 pre-processing techniques are applied in addition to a baseline without pre-processing. The combination of a pre-processing technique, the associated hyperparameters and a machine learning algorithm constitutes a model. In all, 8 pre-processing techniques with the hyperparameter specified in brackets are used here: **PT1**- None, **PT2**- SG Filter (3), **PT3**- SG Filter (5), **PT4**- SG Filter (7), **PT5**- SG Filter (9), **PT6**- SNV, **PT7**- SG Filter (3) with SNV and **PT8**- SNV with SG Filter (3). For example, **PT2**- SG Filter (3) means that the pre-processing technique SG Filter is applied to the raw NIR spectra with sliding windows of 3 wavelengths on each spectrum. For this pre-processing technique, the sliding window represents the hyperparameter. The larger the sliding window, the smoother the curve. The SG filter tends to fit a second-order polynomial with a specific sliding window to smooth the curve. In this research, when the SG filter is used, only the first derivative is applied to the NIR spectra. The combination of different pre-processing techniques can increase the performance of the overall models such as the preprocessing proposed in **PT7** and **PT8**. In this case, it is important to specify the order in which they are processed, as they will not provide the same results.

The output of the pre-processing step will be used as an input for the machine learning algorithms.

6.4.2.2 Machine learning model

In our research work, PLSR and SVR machine learning algorithms are used because of their power to handle high-dimensional data and their applicability in regression problems. For each algorithm, each wavelength of the spectra is scaled following a z-score normalization.

As a hyperparameter, the PLSR algorithm requires several components. To obtain this, an evaluation loop of the training data is performed for each component number between 1

and 8. The number of components that maximizes the R^2_{cv} metric is selected according to the number of wavelengths studied in step 3. On the other hand, the SVR algorithm that requires the epsilon (ϵ) hyperparameter is configured by default and set to 0.1.

For each food quality feature studied and each pre-processing technique applied to the raw NIR spectra, both machine learning algorithms are applied. The quality features studied, the machine learning algorithm selected and the pre-processing technique used form a specific machine learning model. Taking the example of wheat grain (A) data, two quality features are studied: moisture and protein. By applying both machine learning algorithms to each pre-processing technique, we end up with a total of 16 machine learning models for the quality features of each type of grain of wheat.

For each machine learning model, different wavelengths relevant to the model are selected.

6.4.2.3 Feature selection

A selection of wavelengths from the given NIR spectra is performed using the BPSO meta-heuristic algorithm. The BPSO algorithm seeks to minimize the objective function defined by equation (6.3) with α set to 0.88 to further penalize the objective function based on model performance, rather than reducing the number of features used in the solution search. The algorithm tests different wavelength solutions with the parameters specified in Table 6.4. Among the parameters, only parameter D varies according to the size of the raw data. With 1557 wavelengths covering the NIR spectra of grains of wheat (A), parameter D is set to 1557. The parameter p is set to 1 since we are selecting specific features. The other parameters are set based on the literature reviews and experience following several trials. The stopping criterion for stopping the navigation in the solution space by the BPSO algorithm is defined by a maximum number of iterations fixed at 100. After several manual trials with different numbers of iterations - between 20 and 1000 - the iteration number fixed at 100 offered the best ratio between the discovery of relevant solutions and computational cost.

For each solution, the BPSO algorithm evaluates the performance of the PLSR or SVR algorithm using the R^2_{cv} of the trained data set. Each solution provides the specific wavelengths to be selected. The best solution is selected at the end of the 100 iterations. In a loop, the

Table 6.4 BPSO parameters values for the dataset on cocoa beans

Parameter	D	Np	C1	C2	w	k	p
Values	1557	30	2	2	1	20	1

BPSO algorithm is applied several times, reducing the number of wavelengths to be selected at each stage. As soon as the R^2_{cv} metric no longer improves or the number of wavelengths no longer decreases, the loop is stopped and the best solution at iteration $i-1$ that maximizes R^2_{cv} is selected.

Once the attributes have been selected for each model, each one is evaluated to determine its robustness and predictive performance with new data.

6.4.2.4 Model evaluation

The 16 models developed to predict one quality attribute of grains of wheat (A) are evaluated based on test data. To do this, we calculate the RMSE-test using the trained models to estimate their accuracy in predicting new data. The best-performing of the 16 models is the one with the lowest RMSE-test.

Table 6.5 illustrates all of the results obtained by following the stated methodology for each machine learning model that was developed using the most relevant wavelengths.

The best-performing models according to the RMSE metric are *WMA14* and *WPA6* for predicting moisture content and fat content in cocoa beans, respectively. *WMA14* uses a SVR machine learning algorithm in which raw data is normalized by the SNV technique. Furthermore, the model is based on only 199 of the 1557 wavelengths present in the NIR spectra. On the other hand, *WPA6* uses a PLSR machine learning algorithm in which raw data is normalized by the SNV technique. *WPA6* is based on 462 of the 1557 wavelengths. Both models are highlighted in Table 6.5 in a bold font. Similarly, the performance of each model developed for the cocoa bean, grains of wheat (B) and mango datasets are presented in Tables A.1, B.1 and C.1, respectively.

Table 6.5 Model performance for the prediction of wheat (A) quality features based on pre-processing, machine learning techniques and the selection of wavelength.

Quality parameter	ML technique	Pre-processing technique	N. Wavelength	Metrics		Model
				R^2_{cv}	RMSE-test	
Moisture content	PLSR	None	284	0.966	0.582	WMA1
		SG filter [3,2,1]	222	0.976	0.765	WMA2
		SG filter [5,2,1]	186	0.980	0.699	WMA3
		SG filter [7,2,1]	137	0.983	0.699	WMA4
		SG filter [9,2,1]	192	0.980	0.690	WMA5
		SNV	194	0.987	0.368	WMA6
		SG filter [3,2,1]; SNV	181	0.970	0.641	WMA7
		SNV; SG filter [3,2,1]	245	0.957	0.818	WMA8
	SVR	None	177	0.757	0.871	WMA9
		SG filter [3,2,1]	115	0.857	0.743	WMA10
		SG filter [5,2,1]	156	0.878	0.772	WMA11
		SG filter [7,2,1]	153	0.896	0.594	WMA12
		SG filter [9,2,1]	155	0.912	0.543	WMA13
		SNV	199	0.905	0.337	WMA14
		SG filter [3,2,1]; SNV	118	0.856	0.861	WMA15
		SNV; SG filter [3,2,1]	93	0.861	0.908	WMA16
Protein content	PLSR	None	277	0.993	0.296	WPA1
		SG filter [3,2,1]	194	0.985	0.673	WPA2
		SG filter [5,2,1]	380	0.983	0.529	WPA3
		SG filter [7,2,1]	355	0.989	0.541	WPA4
		SG filter [9,2,1]	304	0.993	0.399	WPA5
		SNV	462	0.993	0.177	WPA6
		SG filter [3,2,1]; SNV	240	0.980	0.559	WPA7
		SNV; SG filter [3,2,1]	251	0.979	0.887	WPA8
	SVR	None	136	0.392	1.675	WPA9
		SG filter [3,2,1]	101	0.913	0.790	WPA10
		SG filter [5,2,1]	143	0.938	0.551	WPA11
		SG filter [7,2,1]	166	0.945	0.535	WPA12
		SG filter [9,2,1]	182	0.948	0.578	WPA13
		SNV	189	0.909	0.738	WPA14
		SG filter [3,2,1]; SNV	85	0.895	0.972	WPA15
		SNV; SG filter [3,2,1]	90	0.908	0.886	WPA16

6.5 Discussion

The robustness of NIR machine calibration lies in the quality of the NIR spectra analyses and the wide variety of samples. The 4 different datasets, featuring different dimension sizes, different numbers of examples and different spectra, allow the method to be tested on a variety of data. This provides a general method for the development of a support decision tool to help NIR machine calibration, which is applicable to the prediction of various food quality characteristics.

The best machine learning models for predicting food quality - highlighted in Tables 6.5, A.1 , B.1 and C.1 - demonstrates excellent performance. Of the 4 datasets, only the wheat dataset (A) has performance results for machine learning models developed from NIR spectra. If we compare this work with the work carried out by Johnson et al. (2022a) on wheat (A) dataset, we note that the *WMA14* and *WPA6* models provide better prediction performance with the right pre-processing and wavelength selection techniques. The *WMA14* model for predicting the moisture of wheat grains obtained an RMSE-test of 0.337 using 199 of the 1557 wavelengths, which is a better score than the RMSE-test of 0.62 obtained by Johnson et al. (2022a). Similarly, the *WPA6* model for predicting wheat grain protein content achieved an RMSE-test score of 0.177, which is better than the score obtained by Johnson et al. (2022a) with an RMSE-test of 0.22. However, these results will depend on how the dataset is split, given that we have no knowledge of the examples taken as training data and examples taken as test data by Johnson et al. (2022a).

For some of the models that were developed, there was a tendency to overfit, despite the use of the cross-validation technique. This is the case in the mango and wheat (A) data with models using the SG filter as a pre-processing technique and PLSR algorithms as a machine learning technique. There is a significant discrepancy between the RMSE train obtained from cross-validation on the training data and the RMSE test on the test data for these models. This demonstrates that the models not only learn the underlying patterns in the data, but also the noise or random fluctuations present in the training data set. As a result, the models work exceptionally well on the data on which they were trained but fail to effectively generalize to new, unseen data. To mitigate the risk of over-fitting, a regularization technique

and automatic selection of good representative data points can improve the model's reliability when applied with new data.

There are several limitations of this work, which should make it more generalizable. Firstly, as this document focuses on the selection of a method and features selection, it deals with a number of known pre-processing techniques. However, others exist and it would be of interest to explore a larger number of pre-processing techniques so that they could be tested automatically. Secondly, only the SVR and PLSR machine learning algorithms have been studied for their effectiveness in processing NIR spectra with small amounts of data. However, set models that combine several algorithms into a single model using the weight of importance can bring greater resilience to the prediction model for NIR spectra analysis. The study of weight and the algorithms to be used are topics that can be included in this approach. Thirdly, as we do not have very large samples of public data, it would be interesting to compare these regression techniques against deep learning algorithms, which do not necessarily require pre-processing with larger quantities of data. Finally, optimizing the hyper-parameters mentioned for SVR algorithms and BPSO metaheuristic algorithm according to context and data are ways of refining model performance that are outside the scope of this paper. These topics may be addressed in future work.

6.6 Conclusion

Fast, simple, and cost-effective analysis of NIR spectra requires no sample preparation, which offers a significant advantage over laboratory analysis on continuous food processing lines. However, as NIR spectra is based on indirect measurements and produces very complex and broad spectra, NIR machines require regular calibration. For manufacturers, NIR calibration is difficult and time consuming because it depends on the chemical composition of the samples and the configuration of the equipment.

This paper has introduced a four-step methodology and a decision support tool for analyzing grains of wheat, cocoa beans and properties of mangoes using NIR spectra. Through preprocessing of raw NIR spectral data and the development of machine learning models, we efficiently addressed the calibration of NIR machines in diverse food production scenarios. A

selection of relevant wavelengths in the NIR spectra was achieved using the BPSO algorithm, guided by preset parameters and an established objective function. The performance of a model was evaluated through the RMSE-test, allowing for a comparative assessment of various machine learning models. To validate this methodology, we employed four open-access datasets to evaluate the quality of different foods based on context-relevant quality features. Promising results have been observed on the moisture and fat content of cocoa beans, moisture and protein content of wheat (A), acidity and vitamin C content of mango, and protein content of wheat (B) with a R^2_{cv} score of 0.853, 0.727, 0.905, 0.993, 0.924, 0.663 and 0.905 respectively. The results of the selected models for wheat (A) demonstrate a better predictive performance than those obtained by Johnson et al. (2022a) using the same dataset with a RMSE-test of 0.337 and 0.177 for the prediction of moisture and protein content, respectively. However, some limitations remain. Some models show signs of over-fitting thus, caution is required when selecting the final models. Moreover, the scope of this research has been limited to certain pre-processing techniques and machine learning algorithms to focus on a framework and the selection of wavelength selection. More promising and relevant techniques for analyzing NIR spectra may be found, resulting in improved model performance. Hence, further research is needed to overcome these limitations. It would be of interest to examine how to more effectively limit the risk of over-fitting when developing machine learning models. Furthermore, it would be relevant to evaluate alternative pre-processing techniques and compare these machine learning models with ensemble models. Finally, optimizing the hyperparameters present in the SVR machine algorithm and during feature selection by the BPSO algorithm would be useful in navigating the solution space more efficiently. As the 4 datasets used to analyze food NIR spectra are freely available, they can be easily accessed and used to improve the performance and reliability of the models developed using this approach. The development of robust and flexible machine learning models will help put forth a more generalizable and applicable method for NIR calibration on food processing lines, paving the way for more effective real-time monitoring in agri-food production.

6.7 Acknowledgements

We would like to thank NSERC and MITACS Globalink for their financial support, and SENAI Innovation Institute for Embedded Systems for their help and involvement in this project. The code used for analysis is provided in a repository at https://github.com/Loicprn/NIR_food.git. The authors welcome requests for additional information regarding the material presented in this paper.

CHAPITRE 7 DISCUSSION GÉNÉRALE

Face à la demande alimentaire mondiale croissante et aux défis tels que les conflits, la crise économique et les problèmes climatiques, il devient impératif d'améliorer la résilience du secteur agroalimentaire. Cependant, les pénuries de main-d'œuvre et la perte d'expertise dans la production alimentaire constituent des obstacles cruciaux pour garantir la qualité, la sécurité, l'efficacité et la stabilité de la production, créant ainsi de nombreux défis dans le secteur de la transformation agroalimentaire. Plus spécifiquement, cette recherche s'est focalisée sur les processus de meunerie.

Dans ce contexte et afin de mieux comprendre ces défis, un état de l'art des techniques d'intelligence artificielles utilisées dans le domaine de la meunerie biologique a été réalisé. Il a fait l'objet d'un article intitulé "Future trends in organic flour milling : the role of AI" publié dans le journal *AIMS agriculture and Food* (Parrenin et al., 2023b) et constitue le chapitre 2 de cette thèse. L'état de l'art a révélé le potentiel d'amélioration de l'efficacité opérationnelle dans le domaine de la meunerie, une branche de la transformation alimentaire, grâce à l'exploitation des données. Notamment, il a souligné la sous-utilisation des techniques d'intelligence artificielle dans ce domaine spécifique. De plus, il a mis en évidence la complexité d'assurer une production de farine de qualité constante, en raison de la variabilité inhérente à la qualité du grain de blé, qui est la matière première du processus. Cette variabilité a un impact direct sur le rendement et la qualité de la production. Dans ce contexte, l'intervention d'un meunier expérimenté devient essentielle pour ajuster chaque étape du processus et atteindre les rendements et la qualité de farine souhaités. Ces ajustements se basent sur les connaissances et l'expérience tacite des meuniers, qui les réalisent par essais et erreurs.

La nécessité d'une expertise spécialisée dans le contrôle du processus de meunerie, associée à une pénurie de main-d'œuvre et à une perte de compétences, représente un défi pour la production de farine biologique. La variabilité de la qualité des grains de blé et des facteurs de production rendent les instructions d'exploitation inefficaces, surtout en l'absence de meuniers qualifiés pour les ajuster. Un défi majeur est donc de contrôler le processus de meunerie pour garantir une qualité de farine biologique constante. Pour relever ce défi, l'objectif général de

cette thèse est de développer des outils d'aide à la décision visant l'ajustement des paramètres contrôlables dans la production agroalimentaire.

Dans le but d'atteindre l'objectif général, trois sous-objectifs spécifiques ont été formulés :

- **SO1 : Explorer les données dans un contexte de production agroalimentaire par l'expertise métier et la cartographie des processus.**
- **SO2 : Développer un outil d'aide à la décision pour améliorer le contrôle d'un processus de transformation agroalimentaire.**
- **SO3 : Développer un outil d'aide à la décision pour l'analyse de la qualité des aliments en production.**

Les deux sections suivantes offrent une synthèse des travaux, exposant les contributions apportées et exposent les limites de ces contributions, ouvrant la voie à des perspectives d'amélioration et de futures recherches.

7.1 Synthèse des travaux

Les contributions de cette thèse s'insèrent dans une démarche de développement d'outils d'aide à la décision visant à ajuster les paramètres contrôlables dans la production agroalimentaire. Les différents articles ciblent certaines améliorations dans le processus de meunerie biologique qui fait l'objet de notre cas d'étude. Ils permettent d'atteindre l'objectif global qui est de développer des outils d'aide à la décision visant l'ajustement des paramètres contrôlables dans la production agroalimentaire, dans un contexte particulier. La Figure 7.1 met en évidence les contributions apportées dans cette thèse.

Pour contribuer à l'amélioration du processus de meunerie biologique grâce à l'exploration de données, il était impératif de procéder à une revue de littérature, en documentant et en synthétisant les recherches existantes. Cette étape a visé à établir un recensement des techniques d'intelligence artificielle employées dans le domaine de la mouture du blé biologique. Le chapitre 2 explique le processus global de la meunerie biologique, les difficultés rencontrées à chaque étape de la chaîne de valeur et les travaux qui ont été menés pour résoudre les problèmes à l'aide de modèles statistiques et parfois, de techniques d'intelligence artificielles. Les

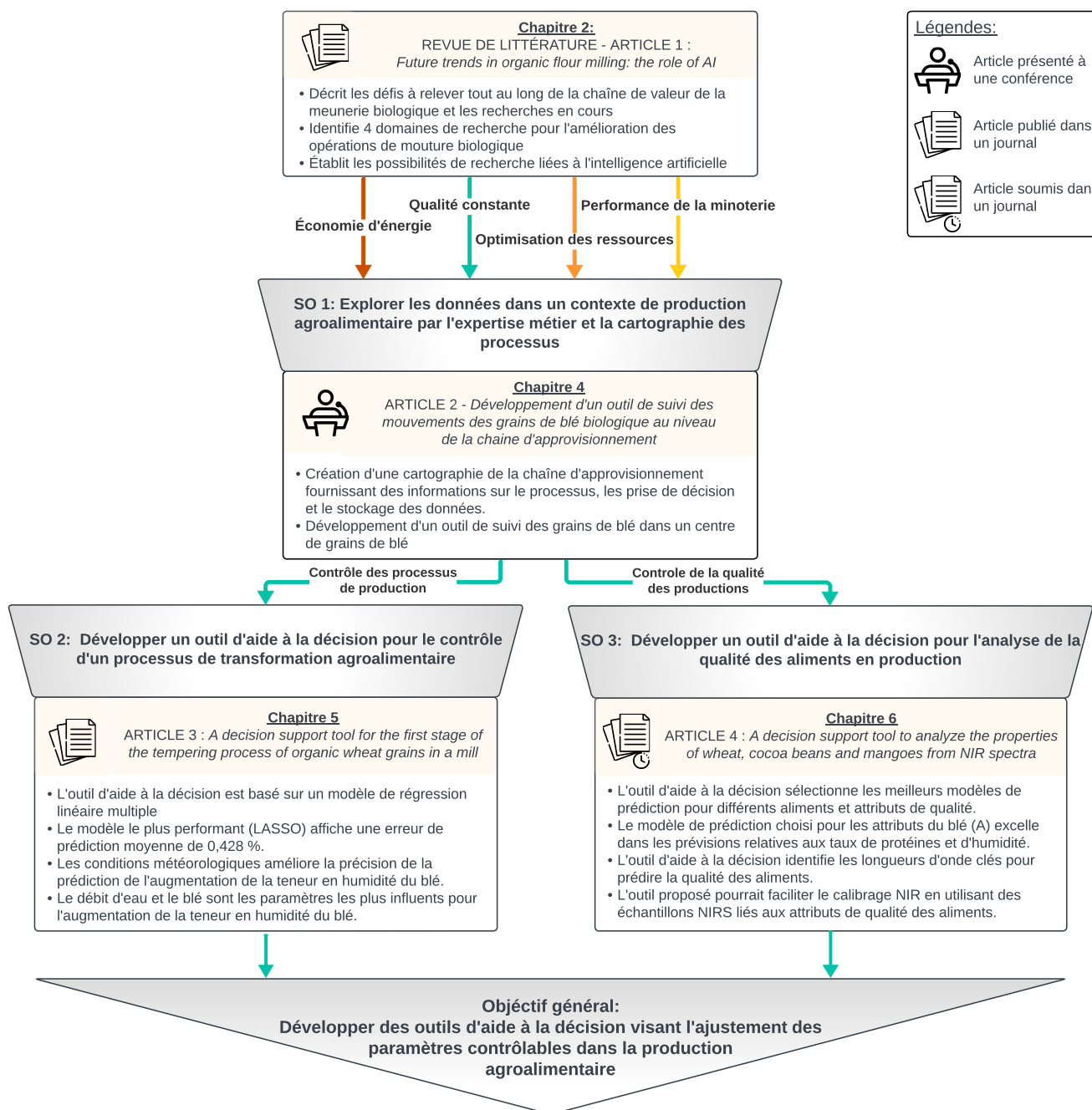


FIGURE 7.1 Synthèse des travaux portant sur le développement d'outils d'aide à la décision visant l'ajustement des paramètres contrôlables dans la production agroalimentaire

travaux de recherche qui contribuent à améliorer le processus global de la meunerie à partir de l'analyse de données sont répertoriés dans des tableaux pour chaque étape de la chaîne de valeur. Comme le montre la figure 7.1 dans l'encadré en haut, cette revue de littérature (article 1) fournit trois contributions principales. Elle décrit tout d'abord les défis présents tout au long de la chaîne de valeur de la meunerie biologique et les recherches en cours. Sur la base de ces défis et des recherches en cours, quatre domaines de recherche ont été identifiés pour améliorer les opérations d'une minoterie biologique, il s'agit des économies d'énergie, de la production de qualité constante, de l'optimisation des ressources et des performances des minoteries. Enfin, l'article a présenté les possibilités de recherche associées à l'intelligence artificielle dans le domaine de la meunerie biologique. Ce travail constitue une base de projets de recherche pour proposer des contributions scientifiques en fonction des lacunes identifiées dans chacun des domaines de recherche définis.

Dans le cadre du sous-objectif (**SO1**), la cartographie sur la chaîne d'approvisionnement d'une minoterie biologique a été réalisée et présentée dans l'article intitulé "Développement d'un outil de suivi des mouvements des grains de blé biologique au niveau de la chaîne d'approvisionnement" publié à la *13ème conférence internationale de modélisation, optimisation et simulation - MOSIM2020* (Parrenin et al., 2020). L'article de conférence, comme le montre la Figure 7.1, se compose de 2 principales contributions. En se concentrant sur le processus d'approvisionnement, la première contribution a concerné la réalisation d'une cartographie de la chaîne d'approvisionnement fournissant des informations sur le processus, la prise de décision et le stockage des données. À l'aide de cette cartographie et en fonction des besoins du partenaire industriel, il a ensuite été possible de développer un outil de suivi des grains de blé dans un centre de grains. Cet outil a été basé sur une interface cloud pour le stockage, la visualisation et la saisie des données. Cette expérimentation a suivi une approche systématique. Après avoir établi une cartographie complète du processus d'approvisionnement, l'attention s'est portée sur la contextualisation des données de production. Une cartographie détaillée a été réalisée sur la base d'entretiens semi-dirigés avec les responsables et le chef meunier. Cette cartographie, qui contient des informations sur le processus, la prise de décision et le type de données enregistrées par les machines ou saisies manuellement par les opérateurs, a été utile pour évaluer la valeur potentielle des données de production par rapport aux quatre

axes de recherche mis en évidence dans la revue de la littérature.

La revue de littérature et la contextualisation des données basée sur la cartographie et l'expertise métier constituent la base de notre travail de recherche pour atteindre l'objectif global. Au vu des données disponibles chez le partenaire industriel, de leurs besoins et des axes de recherche proposés, il était intéressant de se concentrer sur le contrôle de la production à qualité constante tout au long de l'année. Comme le montre la Figure 7.1, la contextualisation des données de production du partenaire industriel a conduit à deux autres travaux de recherche (SO2 & SO3).

Pour le sous-objectif (**SO2**), à partir des éléments précédents, un outil d'aide à la décision a été développé afin d'assurer un meilleur contrôle sur le processus de mouillage des grains de blé. Cet outil a fait l'objet d'un article intitulé "A decision support tool for the first stage of the tempering process of organic wheat grains in a mill" publié dans le journal "International Journal of Food Science & Technology" (Parrenin et al., 2023a).

Dans une minoterie, le contrôle du processus de mouillage est déterminant afin de broyer les grains de blé avec un rendement optimal et une qualité souhaitée. Ce contrôle est réalisé en ajustant les paramètres du processus de mouillage, qui dépendent de plusieurs facteurs tels que la qualité des grains de blé, les conditions météorologiques et opérationnelles, et le type de farine à produire. Le chapitre 5 propose donc le développement d'un outil d'aide à la décision pour aider les opérateurs dans le choix des réglages du processus de mouillage afin d'atteindre la teneur en humidité finale souhaitée pour les grains de blé (**SO2**). Ce chapitre comprend 4 contributions majeures. Tout d'abord, il s'agit du premier outil d'aide à la décision construit pour l'opération de mouillage des grains. Cet outil d'aide à la décision est basé sur un modèle de régression linéaire multiple. Deuxièmement, plusieurs modèles de régression ont été construits et validés sur des échantillons de données provenant de l'historique de production du partenaire industriel. Parmi les modèles de régression linéaire testés, le modèle le plus performant (LASSO) a obtenu une erreur de prédiction moyenne de 0,428%, un résultat convaincant étant donné que les grains de blé sont des éléments biologiques avec des facteurs d'absorption uniques. Troisièmement, il était intéressant de noter que la précision de la prédiction de l'augmentation de la teneur en humidité des grains de blé biologique est amé-

liorée en prenant en compte la température et l'humidité comme variables météorologiques indépendantes dans le modèle. Ces facteurs sont souvent négligés dans les expériences en laboratoire. Enfin, à partir des coefficients de régression obtenus à partir du modèle LASSO, le débit d'eau et du blé se sont avérés être les paramètres les plus influents dans l'augmentation de la teneur en humidité des grains de blé.

Pour le sous-objectif (**SO3**), en suivant la même approche que pour le SO2, un outil d'aide à la décision a été développé pour améliorer la précision d'analyse des grains de blé à partir du spectre NIR. Cette dernière contribution a été soumise pour publication. Une partie de ce travail a fait l'objet d'un article intitulé "A Decision Support Tool to Analyze Food Properties from Near Infrared Spectroscopy" publié à la *15th IEEE International Conference on Industry Applications (INDUSCON)* (Parrenin et al., 2023d). Le contrôle du processus de mouillage dépend de plusieurs facteurs, dont notamment la qualité des grains de blé biologique. En l'absence d'un contrôle précis de la qualité des grains, la fiabilité de l'outil d'aide à la décision serait compromise, car il reposerait sur des données imprécises. Le contrôle de la qualité des grains assure un ajustement efficace des lignes de transformation. Dans la transformation alimentaire, l'utilisation de machines NIR est avantageuse en raison d'une analyse non destructive, rapide, économique et en temps réel. Cependant, les machines NIR nécessitent des ajustements en fonction de leur configuration, du type d'aliment et des attributs de qualité à analyser. Le chapitre 6 se concentre donc sur le développement d'un outil d'aide à la décision pour analyser la qualité alimentaire en production (**SO3**). Il vise à fournir un outil simple et robuste pour ajuster les machines NIR, basé sur une sélection intelligente des longueurs d'onde NIR. Cette recherche a abouti à 4 contributions majeures, comme présenté dans la Figure 7.1. Tout d'abord, ce travail présente une méthodologie pour créer un outil d'aide à la décision qui utilise une sélection intelligente des longueurs d'onde NIRs pour prédire la qualité alimentaire. Plusieurs modèles sont testés en fonction de différents types de prétraitement, de techniques d'apprentissage automatique et de longueurs d'onde sélectionnées. L'outil sélectionne les meilleurs modèles de prédiction pour différents aliments et attributs de qualité. Deuxièmement, selon les résultats obtenus, le modèle de prédiction choisi pour les attributs des grains de blé (A) excelle pour les prédictions de protéines et d'humidité, comparé aux résultats obtenus par Johnson et al. (2022a). Cela démontre la robustesse et la pertinence de

cet outil. Troisièmement, l'outil d'aide à la décision identifie les longueurs d'onde clés pour prédire la qualité alimentaire. L'algorithme d'optimisation (BPSO) sélectionne les longueurs d'onde pertinentes pour le modèle, ce qui est particulièrement intéressant en chimie pour expliquer certaines relations ou prédire certaines propriétés chimiques. Enfin, l'outil proposé pourrait faciliter l'étalonnage NIR en utilisant des échantillons NIRs liés aux attributs de qualité alimentaire.

7.2 Limites et perspectives

Cette recherche présente plusieurs limites, exposées dans cette section. Pour pallier ces limitations, des propositions de recherche sont formulées en vue de les surmonter.

L'approche adoptée pour la conduite de la revue de littérature présente une première limite de cette recherche. Bien que l'étude synthétise les travaux sur les techniques d'intelligence artificielle utilisées dans la meunerie biologique et qu'elle englobe un large éventail d'études pertinentes, elle ne suis pas une méthodologie d'étude systématique. Cela peut influencer la portée et les résultats de la recherche. L'absence d'un cadre systématique peut introduire de la subjectivité et restreindre l'analyse approfondie de la littérature existante. De plus, le choix des termes dans la revue peut influencer l'interprétation des résultats, affectant ainsi la précision et la clarté des conclusions. Pour remédier à cette limitation, une méthodologie de revue systématique pourrait être envisagée dans les futurs projets de recherche pour renforcer la rigueur et l'objectivité du processus de synthèse littéraire.

L'atteinte de l'objectif général a permis de développer des outils d'aide à la décision visant l'ajustement des paramètres contrôlables dans la production agroalimentaire. Leur développement ont été présenté aux chapitres 5 et 6. Cependant leur implémentation et leur utilisation sur le plancher de production nécessite de développer une interface utilisateur qui n'a pas été abordée dans cette recherche. Bien que robustes, l'efficacité de leur application sur le terrain est entravée par le manque d'une interface optimisée et conviviale. L'interface joue un rôle crucial dans la facilitation de l'interaction entre les opérateurs et l'utilisation en temps réel de ces outils. Pour remédier à cette limitation, des efforts sont nécessaires pour affiner et optimiser l'interface utilisateur en fonction des besoins opérationnels spécifiques, maximisant

ainsi l'utilité pratique des outils d'aide à la décision développés.

Une autre limite de cette recherche est de ne pas avoir exploré les recettes de mélange de blé et leur optimisation pour obtenir une qualité de farine constante. Cet axe de recherche n'a pu être envisagé pour le moment en raison du manque et de la fiabilité des données collectées, telles que les quantités de blé utilisées ou les paramètres de production de mouture. Alors que les études existantes se concentrent uniquement sur le processus de mouture, l'omission des recettes de blé représente une lacune significative compte tenu de leur rôle dans l'ajustement des paramètres de production et, par conséquent, de la qualité de la farine produite. Remédier à cette limitation pourrait contribuer à une compréhension plus complète de l'interaction entre les recettes de blé et les résultats de production.

L'outil d'aide à la décision présenté dans le chapitre 5 n'a pas été soumis à des tests dans une autre minoterie. Le processus de développement et de validation de l'outil s'est appuyé sur des échantillons de données provenant exclusivement de la minoterie d'origine, ce qui peut soulever des inquiétudes quant à sa généralisation et à son applicabilité à divers environnements de mouture du blé biologique. L'absence de tests dans d'autres moulins limite la validité externe de l'outil et empêche une évaluation complète de son efficacité dans différents contextes opérationnels. Pour remédier à cette limitation, il faudrait que les recherches futures incluent des tests et une validation dans plusieurs moulins afin de garantir la robustesse et la transférabilité de l'outil d'aide à la décision dans différents scénarios de production.

Le développement d'outils d'aide à la décision repose sur les données disponibles, ce qui constitue une dépendance à l'égard de la qualité et de la quantité d'informations disponibles. Bien que des efforts aient été déployés pour assurer la fiabilité des données, la taille relativement modeste de l'ensemble de données dans les chapitres 5 et 6 pose des défis pour l'application de techniques analytiques avancées. L'inclusion d'un ensemble de données plus étendu pourrait permettre l'utilisation d'algorithmes d'apprentissage profond, améliorant ainsi la capacité prédictive. Reconnaisant cette contrainte, des ensembles de données plus vastes et diversifiés seraient essentiels pour explorer les capacités des techniques de modélisation avancées dans des études futures.

Enfin, une limite de cette recherche peut être observée au chapitre 6, en ce qui concerne l'hy-

perparamétrisation de l'algorithme d'optimisation d'essaim de particules binaires (BPSO). L'ajustement de ces hyperparamètres n'a pas été entrepris de manière systématique. Au lieu de cela, les ajustements ont été effectués par une approche essais et erreurs. Dans ce travail de recherche, l'accent a été mis sur la présentation de la démarche et de l'efficacité de la méthode BPSO elle-même, le réglage des hyperparamètres étant une considération secondaire. Bien que la contribution de la méthode soit importante, l'absence d'un processus complet de réglage des hyperparamètres constitue une limite. Les efforts de recherche futurs pourraient bénéficier d'une exploration et d'une optimisation plus rigoureuses des hyperparamètres BPSO avec des modules tels que "GridSearch" ou "Random Search". Cette approche permettrait non seulement d'améliorer la robustesse de l'algorithme, mais aussi de mieux comprendre son comportement dans différents scénarios. En s'attaquant à cette limitation dans des études ultérieures, la recherche peut parvenir à une application plus nuancée et plus raffinée de l'algorithme BPSO, améliorant encore son potentiel et élargissant son applicabilité dans divers contextes.

Les limites rencontrées dans ce travail de recherche représentent des améliorations qui peuvent être apportées pour obtenir un plus grand impact positif dans l'industrie agroalimentaire et de meilleurs résultats. Les outils d'aide à la décision développés n'ont pas encore été déployés. Leurs performances et leur acceptabilité pour une utilisation en temps réel par les opérateurs restent à déterminer et à évaluer.

CHAPITRE 8 CONCLUSION

Dans le contexte de la demande alimentaire mondiale croissante et des défis actuels, cette thèse vise à renforcer la résilience du secteur agroalimentaire. Axée sur le processus de meunerie, elle explore les opportunités d'amélioration de l'efficacité opérationnelle par l'utilisation judicieuse des données. La revue de littérature souligne le potentiel insuffisamment exploité de l'intelligence artificielle dans la meunerie biologique, identifiant quatre axes de recherche. Parmi ces axes, la production de farine de qualité constante émerge comme une préoccupation majeure, alignée avec la problématique centrale de la thèse.

Cette recherche vise à répondre à la question fondamentale de savoir comment contrôler le processus de mouture pour garantir une production constante de farine de qualité. Dans le cadre de cet objectif global, trois sous-objectifs spécifiques ont été définis. Le premier sous-objectif s'articule autour de la contextualisation des données, réalisée grâce à l'élaboration d'une cartographie des processus basée sur un cas d'étude industrielle et les connaissances d'experts. Cette cartographie a jeté les bases nécessaires pour atteindre le deuxième sous-objectif, qui consiste à développer un outil d'aide à la décision pour contrôler le processus de mouillage des grains de blé. De manière similaire, le troisième sous-objectif a visé à concevoir un outil d'aide à la décision pour améliorer l'analyse des grains par la spectroscopie dans l'infrarouge proche (NIR). Ces sous-objectifs, atteints au fil de la recherche, contribuent de manière significative à l'objectif global de développer des outils d'aide à la décision pour l'ajustement des paramètres contrôlables en production agroalimentaire.

8.1 Synthèse des travaux

Le **chapitre 2** constitue une première contribution en réalisant une revue de littérature des techniques d'intelligence artificielle appliquées à la meunerie biologique. Cette revue a permis d'identifier quatre axes de recherches, à savoir l'économie d'énergie, la production d'une farine de qualité constante, l'optimisation des ressources et l'amélioration de la performance de la minoterie. Parmi ces axes de recherches, plusieurs opportunités liées à l'intelligence

artificielle sont présentées. Le **chapitre 4** présente la deuxième contribution de cette thèse. Son objectif est de structurer les données de production agroalimentaire chez un partenaire industriel, en utilisant l'expertise métier et la cartographie des processus (**SO1**). Une cartographie de la chaîne d'approvisionnement du partenaire a été effectuée, facilitant la création d'un outil de suivi des grains de blé au niveau d'un centre de grains. Suivant cette même approche, une cartographie du processus de production a été élaborée pour évaluer les données collectées sur le plancher de production et explorer les opportunités de recherches parmi les quatre axes identifiés dans la revue de littérature. En se basant sur une étude de cas industriel dans une minoterie biologique, cet effort a permis de cibler un axe de recherche significatif, à savoir la production de farine de qualité constante, apportant une valeur ajoutée substantielle à notre partenaire industriel. Le **chapitre 5** découle du sous-objectif 1, visant à développer un outil d'aide à la décision pour le contrôle d'un processus de transformation agroalimentaire à partir des données de production disponibles (**SO2**). Cet outil, basé sur le modèle LASSO, a démontré des performances de prédiction de l'augmentation de la teneur en humidité suffisamment satisfaisantes pour être utilisé et amélioré, affichant une erreur de prédiction moyenne de 0.428%. Les résultats ont mis en évidence l'impact significatif des conditions météorologiques sur l'ajustement des paramètres du processus de mouillage des grains de blé, tout en soulignant l'importance d'un réglage minutieux des débits d'eau et de blé. Face à un deuxième défi crucial, le contrôle de la qualité des intrants, identifié dans le processus de production, le **chapitre 6** a été consacré à développer un outil d'aide à la décision pour l'analyse de la qualité des aliments en production à partir des données de spectroscopie NIR (**SO3**). Cet outil sélectionne les meilleurs modèles de prédiction pour différents aliments et attributs de qualité, identifiant les longueurs d'onde clés pour prédire la qualité des aliments, offrant ainsi une contribution significative dans le contexte de la calibration NIR.

Les deux outils d'aide à la décision développés dans SO2 et SO3 apportent un soutien important dans le contrôle et l'amélioration du processus de meunerie biologique.

8.2 Limitations

Quelques limites méritent d'être rappelées. Tout d'abord, la revue de littérature limitée à l'analyse des techniques d'intelligence artificielles visant à améliorer l'efficacité des processus de meunerie biologique peut avoir manqué certains éléments.

Au niveau des outils d'aide à la décision proposés, bien que ces outils aient été développés pour ajuster les paramètres de production, l'absence d'une interface utilisateur entrave leur implémentation pratique. Une interface adaptée aux besoins opérationnels spécifiques est cruciale pour faciliter l'interaction en temps réel avec les opérateurs. Également, ces outils d'aide à la décision bénéficieraient à être testés dans d'autres minoteries, afin d'en garantir la généralisation et l'applicabilité dans d'autres contextes industriels. Des tests et validations dans divers environnements sont nécessaires pour assurer leur robustesse et transférabilité.

Les modèles d'apprentissage machine utilisés dépendent des données réelles, souvent limitées ou mal comprises en milieu industriel, restreignant le développement de modèles sophistiqués. L'exploration de jeux de données plus vastes devrait permettre de bonifier le potentiel de l'intelligence artificielle.

Enfin, le choix méthodologique de ne pas explorer les recettes de mélanges de blé, en raison du manque de données fiables, représente certainement une lacune dans notre recherche et devrait offrir une alternative dans l'amélioration de la transformation des produits agroalimentaires.

Ces limites suggèrent des axes d'amélioration potentiels pour optimiser l'impact de cette recherche dans le secteur agroalimentaire présentés dans la section suivante.

8.3 Améliorations futures

Une analyse systématique de la littérature avec une justification de chaque mot-clé devrait permettre d'aborder une analyse plus large et plus complète des techniques d'intelligence artificielle dans le domaine de la meunerie biologique.

En parallèle, le développement d'une interface utilisateur adaptée faciliterait la mise en œuvre pratique des outils d'aide à la décision, en favorisant l'interaction en temps réel avec les

opérateurs. Une évaluation pratique des outils dans des conditions réelles de production agroalimentaire permettrait de mesurer leur efficacité en temps réel et d'identifier d'éventuels ajustements nécessaires. Des tests de validation des outils dans différentes minoteries pourraient également être conduits et examinés pour en déterminer la robustesse et leur applicabilité dans divers contextes industriels, tandis que la collecte d'ensembles de données plus importants permettra de maximiser le potentiel des modèles d'apprentissage automatique.

Il est suggéré d'explorer les recettes de mélange de blé en travaillant sur l'installation de capteurs performants permettant de collecter des données sur la quantité de chaque lot de grains utilisé pour chaque course de production. L'optimisation des quantités de différents lot de grains de blé mélangés à l'aide d'un outil d'aide à la décision en fonction de la qualité des grains, des paramètres de production et de la farine souhaitée offrirait un réel avantage concurrentiel pour optimiser les bénéfices et garantir une qualité constante de la farine.

Finalement, le travail présenté dans cette thèse ouvre la voie au développement d'outils d'aide à la décision dans le secteur agroalimentaire, ayant pour objectif d'optimiser l'efficacité du processus de production. Les prédictions offertes par les modèles simplifieraient l'identification des ajustements optimaux à effectuer sur les paramètres de contrôle à des étapes spécifiques du processus, en réponse à des contraintes particulières.

RÉFÉRENCES

- Abbas, O., Pissard, A., and Baeten, V. (2020). 3 - near-infrared, mid-infrared, and raman spectroscopy. In Pico, Y., editor, *Chemical Analysis of Food (Second Edition)*, pages 77–134. Academic Press.
- Abbasi, H., Emam-Djomeh, Z., and Seyedin, S. (2011). Application of artificial neural network and genetic algorithm for predicting three important parameters in bakery industries. *International Journal of Agricultural Science and Research*, 2(4) :51–64.
- Agrawal, P., Abutarboush, H. F., Ganesh, T., and Mohamed, A. W. (2021). Metaheuristic algorithms on feature selection : A survey of one decade of research (2009-2019). *IEEE Access*, 9 :26766–26791.
- Agussabti, A., Rahmaddiansyah, R., Satriyo, P., and Munawar, A. A. (2020). Data analysis on near infrared spectroscopy as a part of technology adoption for cocoa farmer in aceh province, indonesia. *Data in Brief*, 29 :105251.
- Aitken, R., Watkins, L., Williams, J., and Kean, A. (2020). The positive role of labelling on consumers’ perceived behavioural control and intention to purchase organic food. *Journal of Cleaner Production*, 255 :120334.
- Almaraz, J. J., Mabood, F., Zhou, X., Gregorich, E. G., and Smith, D. L. (2008). Climate change, weather variability and corn yield at a higher latitude locale : Southwestern quebec. *Climatic Change*, 88(2) :187–197.
- Assadzadeh, S., Walker, C. K., McDonald, L. S., and Panozzo, J. F. (2022). Prediction of milling yield in wheat with the use of spectral, colour, shape, and morphological features. *Biosystems Engineering*, 214 :28–41.
- Aydin, N., Sayaslan, A., Sönmez, M., Aslan, M., Demir, L., and Sabanci, K. (2020). Wheat flour milling yield estimation based on wheat kernel physical properties using artificial neural networks. *International Journal of Intelligent Systems and Applications in Engineering*, 8 :78–83.
- Baianu, I. and Guo, J. (2011). Nir calibrations for soybean seeds and soy food composition analysis : Total carbohydrates, oil, proteins and water contents. *Nature Precedings*, 6.

- Bakerpedia (2019). Glutopeak. <https://bakerpedia.com/processes/glutopeak/>.
- Banu, I., Stoenescu, G., Ionescu, V., and Aprodu, I. (2011). Estimation of the baking quality of wheat flours based on rheological parameters of the mixolab curve. *Czech Journal of Food Sciences*, 29(1) :35–44.
- Bettge, A., Rubenthaler, G., and Pomeranz, Y. (1989). Air-aspirated wheat cleaning in grading and in separation by functional properties. *Cereal Chem*, 66(1) :15–18.
- Blanco, M. and Villarroya, I. (2002). Nir spectroscopy : a rapid-response analytical tool. *TrAC Trends in Analytical Chemistry*, 21(4) :240–250.
- Borghi, B., Giordani, G., Corbellini, M., Vaccino, P., Guermandi, M., and Toderi, G. (1995). Influence of crop rotation, manure and fertilizers on bread making quality of wheat (triticum aestivum l.). *European Journal of Agronomy*, 4(1) :37–45.
- Boudreau, A. and Ménard, G. (1992). *Le blé : éléments fondamentaux et transformation*. Presses Université Laval.
- Branlard, G., Dardevet, M., Saccomano, R., Lagoutte, F., and Gourdon, J. (2001). Genetic diversity of wheat storage proteins and bread wheat quality. *Euphytica*, 119(1) :59–67.
- Burkov, A. (2019). *The Hundred-Page Machine Learning Book*. Burkov, A.
- Butcher, J. and Stenvert, N. L. (1973). Conditioning studies on australian wheat. iii. the role of the rate of water penetration into the wheat grain. *Journal of the Science of Food and Agriculture*, 24(9) :1077–1084.
- Campbell, G. M. (2007). *Chapter 7 Roller Milling of Wheat*, pages 383–419. Elsevier Science B.V.
- Campbell, G. M., Bunn, P. J., Webb, C., and Hook, S. C. W. (2001). On predicting roller milling performance : Part ii. the breakage function. *Powder Technology*, 115(3) :243–255.
- Campbell, G. M., Fang, C., and Muhamad, I. I. (2007). On predicting roller milling performance vi : Effect of kernel hardness and shape on the particle size distribution from first break milling of wheat. *Food and Bioproducts Processing*, 85(1) :7–23.
- Campbell, G. M., Sharp, C., Wall, K., Mateos-Salvador, F., Gubatz, S., Huttly, A., and Shewry, P. (2012). Modelling wheat breakage during roller milling using the double normalised

- kumaraswamy breakage function : Effects of kernel shape and hardness. *Journal of Cereal Science*, 55(3) :415–425.
- Campbell, G. M. and Webb, C. (2001). On predicting roller milling performance : Part i : the breakage equation. *Powder Technology*, 115(3) :234–242.
- Caporaso, N., Whitworth, M. B., and Fisk, I. D. (2018). Near-infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains. *Applied Spectroscopy Reviews*, 53(8) :667–687.
- Cappelli, A., Guerrini, L., Cini, E., and Parenti, A. (2019). Improving whole wheat dough tenacity and extensibility : A new kneading process. *Journal of Cereal Science*, 90 :102852.
- Cappelli, A., Guerrini, L., Parenti, A., Palladino, G., and Cini, E. (2020a). Effects of wheat tempering and stone rotational speed on particle size, dough rheology and bread characteristics for a stone-milled weak flour. *Journal of Cereal Science*, 91 :102879.
- Cappelli, A., Mugnaini, M., and Cini, E. (2020b). Improving roller milling technology using the break, sizing, and reduction systems for flour differentiation. *Lwt-Food Science and Technology*, 133.
- Cappelli, A., Oliva, N., and Cini, E. (2020c). Stone milling versus roller milling : A systematic review of the effects on wheat flour quality, dough rheology, and bread characteristics. *Trends in Food Science & Technology*, 97 :147–155.
- Catterall, P. (1998). *Flour milling*, pages 296–329. Springer US, Boston, MA.
- Cen, H. and He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology*, 18(2) :72–83.
- Chen, T. Q. and Guestrin, C. (2016). Xgboost : A scalable tree boosting system. In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, pages 785–794.
- Chen, Y., Guo, X., Xing, J., and Zhu, K. (2020). Effects of tempering with steam on the water distribution of wheat grains and quality properties of wheat flour. *Food Chemistry*, 323 :126842.
- Commission canadienne des grains (2019a). Classes de blé canadien.

Commission canadienne des grains (2019b). Guide officiel du classement des grains, blé : Classes et variétés.

Commission canadienne des grains (2019c). Inspection officielle du grain.

Commission canadienne des grains (2019d). Modifications au règlement sur les grains du canada (indice de chute (ic) et désoxynivalénol (don) comme facteurs officiels de classement des grains).

Dal-Pastro, F., Facco, P., Bezzo, F., Thomas, H., Zamproga, E., and Barolo, M. (2015). Data-based multivariate modeling of a grain comminution process. *Computer Aided Chemical Engineering*, pages 2219–2224.

Dal-Pastro, F., Facco, P., Bezzo, F., Zamproga, E., and Barolo, M. (2016). Data-driven modeling of milling and sieving operations in a wheat milling process. *Food and Bioproducts Processing*, 99 :99–108.

Danjou, C., Rivest, L., and Pellerin, R. (2017). Industrie 4.0 : Des pistes pour aborder l’ère du numérique et de la connectivité. Report 978-2-923852-71-3, Cefrio.

Delorme, I. (2022). La pénurie de main-d’oeuvre : le grand défi. <https://www.ledevoir.com/societe/757706/la-penurie-de-main-d-oeuvre-le-grand-defi>.

Delwiche, S. R., Souza, E. J., and Kim, M. S. (2013). Limitations of single kernel near-infrared hyperspectral imaging of soft wheat for milling quality. *Biosystems Engineering*, 115(3) :260–273.

Doblado-Maldonado, A. F., A.Flores, R., and Rose, D. J. (2013). Low moisture milling of wheat for quality testing of wholegrain flour. *Journal of Cereal Science*, 58(3) :420–423.

Doblado-Maldonado, A. F., Pike, O. A., Sweley, J. C., and Rose, D. J. (2012). Key issues and challenges in whole wheat flour milling and storage. *Journal of Cereal Science*, 56(2) :119–126.

Dobraszczyk, B. J., Whitworth, M. B., Vincent, J. F. V., and Khan, A. A. (2002). Single kernel wheat hardness and fracture properties in relation to density and the modelling of fracture in wheat endosperm. *Journal of Cereal Science*, 35(3) :245–263.

Dowell, F. and Maghirang, E. (2002). Accuracy and feasibility of measuring characteristics of single kernels using near-infrared spectroscopy. In *Proceedings of the ICC Conference*,

pages 313–320. Citeseer.

Dowell, F. E., Maghirang, E. B., Pierce, R. O., Lookhart, G. L., Bean, S. R., Xie, F., Caley, M. S., Wilson, J. D., Seabourn, B. W., Ram, M. S., Park, S. H., and Chung, O. K. (2008). Relationship of bread quality to kernel, flour, and dough properties. *Cereal Chemistry*, 85(1) :82–91.

Dowell, F. E., Maghirang, E. B., Xie, F., Lookhart, G. L., Pierce, R. O., Seabourn, B. W., Bean, S. R., Wilson, J. D., and Chung, O. K. (2006). Predicting wheat quality characteristics and functionality using near-infrared spectroscopy. *Cereal Chemistry*, 83(5) :529–536.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., and Vapnik, V. (1996). Support vector regression machines. *Advances in neural information processing systems*, 9.

Dérargon, F. (2016). *La ferme biologique, un espace d'éducation relative à l'éco-alimentation et de construction du rapport à la terre*. Master thesis.

Elevi, B., Öztürk, H., and Kacir, Z. (2017). Optimization of wheat and flour blending for cost minimization by using mathematical modelling. *Kastamonu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, pages 264–272.

Ester, M., Kriegel, H., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press.

Fang, C. and Campbell, G. M. (2003a). On predicting roller milling performance iv : Effect of roll disposition on the particle size distribution from first break milling of wheat. *Journal of Cereal Science*, 37(1) :21–29.

Fang, C. and Campbell, G. M. (2003b). On predicting roller milling performance v : Effect of moisture content on the particle size distribution from first break milling of wheat. *Journal of Cereal Science*, 37(1) :31–41.

Fang, Q., Biby, G., Haque, E., Hanna, M. A., and Spillman, C. K. (1998a). Neural network modeling of physical properties of ground wheat. *Cereal Chemistry*, 75(2) :251–253.

Fang, Q., Haque, E., Spillman, C. K., Reddy, P. V., and Steele, J. L. (1998b). Energy requirements for size reduction of wheat using a roller mill. *Transactions of the Asae*, 41(6) :1713–1720.

FAO (1985). Standard for wheat flour. Report.

- FAO (2016). *The state of food and agriculture*. FAO, Rome.
- FAO (2021). *World Food and Agriculture - Statistical Yearbook 2021*. FAO, Rome.
- Fisher, O. J., Watson, N. J., Escrig, J. E., Witt, R., Porcu, L., Bacon, D., Rigley, M., and Gomes, R. L. (2020). Considerations, challenges and opportunities when developing data-driven models for process manufacturing systems. *Computers & Chemical Engineering*, 140.
- Food Security Information Network and Global Network Against Food Crises (2023). Global report on food crises. Report.
- González-Torralba, J., Arazuri, S., Jarén, C., and Arregui, L. M. (2013). Influence of temperature and r.h. during storage on wheat bread making quality. *Journal of Stored Products Research*, 55 :134–144.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.
- Government of Canada (2021). Data retrieved from Historical Climate Data : "https://climate.weather.gc.ca/historical_data/search_historic_data_e.html".
- Government of Canada (2022). Organic claims on food labels. Retrieved from "<https://inspection.canada.ca/food-labels/labelling/industry/organic-claims/eng/1623967517085/1623967517522#s9c9>".
- Gucbilmez, C. M., Sahin, M., Akcacik, A. G., Aydogan, S., Demir, B., Hamzaoglu, S., Gur, S., and Yakisir, E. (2019). Evaluation of glutopeak test for prediction of bread wheat flour quality, rheological properties and baking performance. *Journal of Cereal Science*, 90.
- Gwartz, J. A. (1998). *Effects of tempering time, moisture and wheat protein level on milling and single kernel characterization system measurements*. PhD thesis, Kansas State University.
- Gómez Sarduy, J., Viego, P., Diaz Torres, Y., Alvarez Guerra Plasencia, M., Santos, V., and Haeseldonckx, D. (2018). A new energy performance indicator for energy management system of a wheat mill plant. *International Journal of Energy Economics and Policy*, 8.
- Hayati, R., Munawar, A. A., and Fachruddin, F. (2020). Enhanced near infrared spectral data to improve prediction accuracy in determining quality parameters of intact mango. *Data in Brief*, 30 :105571.

- Hayta, M. and Çakmakli, U. (2001). Optimization of wheat blending to produce breadmaking flour. *Journal of Food Process Engineering*, 24(3) :179–192.
- HEC Montréal (2013). Modélisation et diagnostic des processus (actuel). University Lecture.
- Hemmerling, S., Hamm, U., and Spiller, A. (2015). Consumption behaviour regarding organic food from a marketing perspective—a literature review. *Organic Agriculture*, 5 :277–313.
- Herschel, W. (1800). Xiii. investigation of the powers of the prismatic colours to heat and illuminate objects; with remarks, that prove the different refrangibility of radiant heat. to which is added, an inquiry into the method of viewing the sun advantageously, with telescopes of large apertures and high magnifying powers. *Philosophical Transactions of the Royal Society of London*, 90 :255–283.
- Hobbs, P. (2007). Conservation agriculture : what is it and why is it important for future sustainable food production? *The Journal of Agricultural Science*, 145(2) :127–137.
- Hong, Y., Chen, S., Chen, Y., Linderman, M., Mouazen, A. M., Liu, Y., Guo, L., Yu, L., Liu, Y., Cheng, H., and Liu, Y. (2020). Comparing laboratory and airborne hyperspectral data for the estimation and mapping of topsoil organic carbon : Feature selection coupled with random forest. *Soil and Tillage Research*, 199 :104589.
- Hook, S. C. W., Bone, G. T., and Fearn, T. (1982a). The conditioning of wheat. the effect of increasing wheat moisture content on the milling performance of uk wheats with reference to wheat texture. *Journal of the Science of Food and Agriculture*, 33(7) :655–662.
- Hook, S. C. W., Bone, G. T., and Fearn, T. (1982b). The conditioning of wheat. the influence of roll temperature in the bühler laboratory mill on milling parameters. *Journal of the Science of Food and Agriculture*, 33 :639–644.
- Hook, S. C. W., Bone, G. T., and Fearn, T. (1984). The conditioning of wheat. moisture migration between the components of a mixed grist, and its effect on milling performance. *Journal of the Science of Food and Agriculture*, 35(5) :584–590.
- Huang, H., Yu, H., Xu, H., and 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.
- Hughner, R. S., McDonagh, P., Prothero, A., Shultz Ii, C. J., and Stanton, J. (2007). Who

are organic food consumers? a compilation and review of why people purchase organic food. *Journal of Consumer Behaviour*, 6(2-3) :94–110.

IAOM (2018). *Fundamentals of Flour Milling*. International Association of Operative Millers, Lenexa.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning : with applications in R*. Springer, New York.

Jespersen, B. M. and Munck, L. (2009). Chapter 11 - cereals and cereal products. In Sun, D.-W., editor, *Infrared Spectroscopy for Food Quality Analysis and Control*, pages 275–319. Academic Press, San Diego.

Jia, C., Sun, D.-W., and Cao, C. (2001). Computer simulation of temperature changes in a wheat storage bin. *Journal of Stored Products Research*, 37(2) :165–177.

Jiao, Y., Li, Z., Chen, X., and Fei, S. (2020). Preprocessing methods for near-infrared spectrum calibration. *Journal of Chemometrics*, 34(11) :e3306.

Johansson, E. and Svensson, G. (1998). Variation in bread-making quality : effects of weather parameters on protein concentration and quality in some swedish wheat cultivars grown during the period 1975–1996. *Journal of the Science of Food and Agriculture*, 78(1) :109–118.

Johnson, J., Walsh, K., and Naiker, M. (2022a). Associated data for investigation of the phenolic and antioxidant content in australian grains using traditional and non-invasive analytical techniques. Master’s thesis, Central Queensland University.

Johnson, J., Walsh, K., and Naiker, M. (2022b). Dataset name : Wheat mean nir data 65 samples. Data retrieved from Figshare, https://acquire.cqu.edu.au/articles/dataset/Associated_Data_for_Investigation_of_the_phenolic_and_antioxidant_content_in_Australian_grains_using_traditional_and_non-invasive_analytical_techniques/20250117/1.

Juneja, M. and Nagar, S. K. (2016). Particle swarm optimization algorithm and its parameters : A review. In *International Conference on Control, Computing, Communication and Materials (ICCCCM)*, pages 1–5.

Kalitsis, J., Minasny, B., Quail, K., and McBratney, A. (2021). Application of response

surface methodology for optimization of wheat flour milling process. *Cereal Chemistry*, 98(6) :1215–1226.

Kang, S. and R. Delwiche, S. (1999). Moisture diffusion modeling of wheat kernels during soaking. *Transactions of the ASAE*, 42(5) :1359–1366.

Kashaninejad, M., Dehghani, A. A., and Kashiri, M. (2009). Modeling of wheat soaking using two artificial neural networks (mlp and rbf). *Journal of Food Engineering*, 91(4) :602–607.

Kelleher, J. (2019). *Deep Learning (MIT Press Essential Knowledge series)*. The MIT Press; Illustrated edition.

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948 vol.4.

Kennedy, J. and Eberhart, R. C. (1997). A discrete binary version of the particle swarm algorithm. In *IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, volume 5, pages 4104–4108 vol.5.

Kibar, H. (2015). Influence of storage conditions on the quality properties of wheat varieties. *Journal of Stored Products Research*, 62 :8–15.

Kondepati, V. R. and Heise, M. H. (2008). The potential of mid-and near-infrared spectroscopy for reliable monitoring of bioprocesses. *Current trends in biotechnology and pharmacy*, 2(1) :117–132.

Ktenioudaki, A., Butler, F., and Gallagher, E. (2010). Rheological properties and baking quality of wheat varieties from various geographical regions. *Journal of Cereal Science*, 51(3) :402–408.

Kweon, M., Martin, R., and Souza, E. (2009). Effect of tempering conditions on milling performance and flour functionality. *Cereal Chemistry*, 86(1) :12–17.

Köksal, G., Batmaz, , and Testik, M. C. (2011). A review of data mining applications for quality improvement in manufacturing industry. *Expert Systems with Applications*, 38(10) :13448–13467.

Lin, S., Gao, J., Jin, X., Wang, Y., Dong, Z., Ying, J., and Zhou, W. (2020). Whole-wheat flour particle size influences dough properties, bread structure and in vitro starch

digestibility. *Food & Function*, 11(4) :3610–3620.

Liu, W. (2016). Dataset name : Wheat kernel. Data retrieved from Figshare, https://figshare.com/articles/dataset/wheat_kernel_dataset/4252217/1.

Loi sur les grains du Canada (2020a). Partie 2 - classement et inspection des grains.

Loi sur les grains du Canada (2020b). Partie 5 - installations, négociants en grains et manutention du grain.

Loi sur les grains du Canada (2020c). Partie 6 - transport du grain.

Magyar, Z., Veha, A., Pepo, P., Bartok, T., and Gyimes, E. (2019). Wheat cleaning and milling technologies to reduce don toxin contamination. *Acta Agraria Debreceniensis*, 0(2) :89–95.

Maleki, M. R., Mouazen, A. M., Ramon, H., and De Baerdemaeker, J. (2007). Multiplicative scatter correction during on-line measurement with near infrared spectroscopy. *Biosystems Engineering*, 96(3) :427–433.

Maphosa, L., Langridge, P., Taylor, H., Parent, B., Emebiri, L. C., Kuchel, H., Reynolds, M. P., Chalmers, K. J., Okada, A., Edwards, J., and Mather, D. E. (2014). Genetic control of grain yield and grain physical characteristics in a bread wheat population grown under a range of environmental conditions. *Theoretical and Applied Genetics*, 127(7) :1607–1624.

Marini, F. and Walczak, B. (2015). Particle swarm optimization (pso). a tutorial. *Chemometrics and Intelligent Laboratory Systems*, 149 :153–165.

Mateos-Salvador, F., Sadhukhan, J., and Campbell, G. M. (2011). The normalised kumaraswamy breakage function : A simple model for wheat roller milling. *Powder Technology*, 208(1) :144–157.

Mefleh, M., Conte, P., Fadda, C., Giunta, F., and Motzo, R. (2020). From seed to bread : variation in quality in a set of old durum wheat cultivars. *Journal of the Science of Food and Agriculture*, 100(11) :4066–4074.

Mehmood, T., Liland, K. H., Snipen, L., and Sæbø, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 118 :62–69.

- Miranda, L. J. V. (2018). PySwarms, a research-toolkit for Particle Swarm Optimization in Python. *Journal of Open Source Software*, 3.
- Misra, N. N., Dixit, Y., Al-Mallahi, A., Bhullar, M. S., Upadhyay, R., and Martynenko, A. (2020). Iot, big data and artificial intelligence in agriculture and food industry. *IEEE Internet of Things Journal*, pages 6305 – 6324.
- Moss, R. (1977). The influence of endosperm structure, protein content and grain moisture on the rate of water penetration into wheat during conditioning. *International Journal of Food Science and Technology*, 12(3) :275–283.
- Munawar, A. A., von Hörsten, D., Wegener, J. K., Pawelzik, E., and Mörlein, D. (2016). Rapid and non-destructive prediction of mango quality attributes using fourier transform near infrared spectroscopy and chemometrics. *Engineering in Agriculture, Environment and Food*, 9(3) :208–215.
- Mutlu, A. C., Boyaci, I. H., Genis, H. E., Ozturk, R., Basaran-Akgul, N., Sanal, T., and Evlice, A. K. (2011). Prediction of wheat quality parameters using near-infrared spectroscopy and artificial neural networks. *European Food Research and Technology*, 233(2) :267–274.
- Myler, H. R. (1998). *Fundamentals of engineering programming with C and Fortran*. Cambridge University Press.
- National Intelligence Council (2021). Global trends 2040 : A more contested world. Report.
- Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., and Lammer-tyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of nir spectroscopy : A review. *Postharvest Biology and Technology*, 46(2) :99–118.
- Noort, M. W. J., van Haaster, D., Hemery, Y., Schols, H. A., and Hamer, R. J. (2010). The effect of particle size of wheat bran fractions on bread quality – evidence for fibre–protein interactions. *Journal of Cereal Science*, 52(1) :59–64.
- Norris, K. (1982). Use of near-infrared reflectance spectroscopy and dye-binding techniques for estimating protein in oat groats. *Cereal Chem*, 59(5) :333–335.
- Norris, K. H., Barnes, R. F., Moore, J. E., and Shenk, J. S. (1976). Predicting forage quality by infrared replectance spectroscopy. *Journal of Animal Science*, 43(4) :889–897.

- Oliver, J., Blakeney, A., and Allen, H. (1993). The colour of flour streams as related to ash and pigment contents. *Journal of Cereal Science*, 17(2) :169–182.
- Oliver, J. R. and Allen, H. M. (1992). The prediction of bread baking performance using the farinograph and extensograph. *Journal of Cereal Science*, 15(1) :79–89.
- Oliveri, P., Malegori, C., Simonetti, R., and Casale, M. (2019). The impact of signal pre-processing on the final interpretation of analytical outcomes – a tutorial. *Analytica Chimica Acta*, 1058 :9–17.
- Pagani, M., Marti, A., and Bottega, G. (2014). *Chapter 2 : Wheat Milling and Flour Quality Evaluation*, pages 20–49. John Wiley & Sons, Ltd.
- Parenti, O., Guerrini, L., Cavallini, B., Baldi, F., and Zanoni, B. (2020a). Breadmaking with an old wholewheat flour : Optimization of ingredients to improve bread quality. *Lwt-Food Science and Technology*, 121.
- Parenti, O., Guerrini, L., and Zanoni, B. (2020b). Techniques and technologies for the breadmaking process with unrefined wheat flours. *Trends in Food Science & Technology*, 99 :152–166.
- Parrenin, L., Danjou, C., and Agard, B. (2020). Développement d’un outil de suivi des mouvements des grains de blé biologique au niveau de la chaine d’approvisionnement. In *13ème conférence internationale de modélisation, optimisation et simulation (MOSIM2020)*, 12-14 Nov 2020, Agadir, Morocco.
- Parrenin, L., Danjou, C., Agard, B., and Beauchemin, R. (2023a). A decision support tool for the first stage of the tempering process of organic wheat grains in a mill. *International Journal of Food Science & Technology*, 58(10) :5478–5488.
- Parrenin, L., Danjou, C., Agard, B., and Beauchemin, R. (2023b). Future trends in organic flour milling : The role of ai. *AIMS Agriculture and Food*, 8(1) :48–77.
- Parrenin, L., Danjou, C., Agard, B., Marchesini, G., and Barbosa, F. (2023c). A decision support tool to analyze the properties of wheat, cocoa beans and mangoes from NIR spectra. Manuscript submitted for publication.
- Parrenin, L., Lorbieski, R., Jaraceski, J. C., Danjou, C., and Agard, B. (2023d). A decision support tool to analyze food properties from near infrared spectroscopy. In *2023 15th IEEE*

International Conference on Industry Applications (INDUSCON), pages 1498–1503.

Porep, J. U., Kammerer, D. R., and Carle, R. (2015). On-line application of near infrared (nir) spectroscopy in food production. *Trends in Food Science & Technology*, 46(2, Part A) :211–230.

Posner, E. and Deyoe, C. (1986). Changes in milling properties of newly harvested hard wheat during storage. *Cereal Chem*, 63(5) :451–456.

Posner, E. and Hibbs, A. (2005). *Wheat flour milling*. American Association of Cereal Chemists, Inc.

Rakita, S., Dokić, L., Dapčević H., T., Hadnađev, M., and Torbica, A. (2018). Predicting rheological behavior and baking quality of wheat flour using a glutopeak test. *Journal of Texture Studies*, 49(3) :339–347.

Rana, J. and Paul, J. (2017). Consumer behavior and purchase intention for organic food : A review and research agenda. *Journal of Retailing and Consumer Services*, 38 :157–165.

Rinnan, , Nørgaard, L., van den Berg, F., Thygesen, J., Bro, R., and Engelsen, S. B. (2009). Data pre-processing. *Infrared spectroscopy for food quality analysis and control*, pages 29–50.

Rózyło, R. and Laskowski, J. (2011). Predicting bread quality (bread loaf volume and crumb texture). *Polish Journal of Food and Nutrition Sciences*, 61.

Sabancı, K., Aydin, N., Sayaslan, A., Sonmez, M. E., Aslan, M. F., Demir, L., and Sermet, C. (2020). Wheat flour milling yield estimation based on wheat kernel physical properties using artificial neural networks. *International Journal of Intelligent Systems and Applications in Engineering*, 8(2) :78–83.

Sandak, J., Sandak, A., and Meder, R. (2016). Assessing trees, wood and derived products with near infrared spectroscopy : Hints and tips. *Journal of Near Infrared Spectroscopy*, 24(6) :485–505.

Savitzky, A. and Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8) :1627–1639.

Shi, Y. and Eberhart, R. (1998). A modified particle swarm optimizer. In *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*, pages 69–73.

- Sisman, C. B. and Ergin, A. S. (2011). The effects of different storage buildings on wheat quality. *Journal of Applied Sciences*, 11(14) :2613–2619.
- Stefan, E., Voicu, G., Constantin, G., Mariana Gabriela, M., and Mariana, I. (2018). Effects of wheat seeds characteristics on roller milling process -a review. In *7th International Conference on Thermal Equipment, Renewable Energy and Rural Development*.
- Steffan, P. (2012). *An optimization model : minimizing flour millers' costs of production by blending wheat and additives*. Master thesis.
- Suganthan, P. N. (1999). Particle swarm optimiser with neighbourhood operator. In *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, volume 3, pages 1958–1962 Vol. 3.
- Tagawa, A., Muramatsu, Y., Nagasuna, T., Yano, A., Iimoto, M., and Murata, S. (2003). Water absorption characteristics of wheat and barley during soaking. *Transactions of the ASAE*, 46(2) :361.
- Teye, E., Anyidoho, E., Agbemaflle, R., Sam-Amoah, L. K., and Elliott, C. (2020). Cocoa bean and cocoa bean products quality evaluation by nir spectroscopy and chemometrics : A review. *Infrared Physics & Technology*, 104 :103127.
- Thissen, U., Pepers, M., Üstün, B., Melssen, W. J., and Buydens, L. M. C. (2004). Comparing support vector machines to pls for spectral regression applications. *Chemometrics and Intelligent Laboratory Systems*, 73(2) :169–179.
- Tibola, C. S., Fernandes, J. M. C., and Guarienti, E. M. (2016). Effect of cleaning, sorting and milling processes in wheat mycotoxin content. *Food Control*, 60 :174–179.
- Torbica, A., Blazek, K. M., Belovic, M., and Hajnal, E. J. (2019). Quality prediction of bread made from composite flours using different parameters of empirical rheology. *Journal of Cereal Science*, 89.
- Triboi, E., Abad, A., Michelena, A., Lloveras, J., Ollier, J. L., and Daniel, C. (2000). Environmental effects on the quality of two wheat genotypes : 1. quantitative and qualitative variation of storage proteins. *European Journal of Agronomy*, 13(1) :47–64.
- Tóth, , Prokisch, J., Sipos, P., Széles, , Mars, , and Gyóri, Z. (2006). Effects of particle size on the quality of winter wheat flour, with a special focus on macro- and microelement

- concentration. *Communications in Soil Science and Plant Analysis*, 37(15-20) :2659–2672.
- Vermesan, O., Friess, P., Guillemin, P., Sundmaeker, H., Eisenhauer, M., Moessner, K., Arndt, M., Spirito, M., Medagliani, P., Giaffreda, R., Gusmeroli, S., Ladid, L., Serrano, M., Hauswirth, M., and Baldini, G. (2014). *Internet of Things Strategic Research and Innovation Agenda*, book section Chapter 3, pages 7–142. River Publisher, Denmark.
- Vieira, S. M., Mendonça, L. F., Farinha, G. J., and Sousa, J. M. C. (2013). Modified binary pso for feature selection using svm applied to mortality prediction of septic patients. *Applied Soft Computing*, 13(8) :3494–3504.
- Vigni, M. L. and Cocchi, M. (2013). Near infrared spectroscopy and multivariate analysis to evaluate wheat flour doughs leavening and bread properties. *Analytica Chimica Acta*, 764 :17–23.
- Vrček, I. V., Čepo, D. V., Rašić, D., Peraica, M., Žuntar, I., Bojić, M., Mendaš, G., and Medić-Šarić, M. (2014). A comparison of the nutritional value and food safety of organically and conventionally produced wheat flours. *Food Chemistry*, 143 :522–529.
- Warechowska, M., Markowska, A., Warechowski, J., Miś, A., and Nawrocka, A. (2016). Effect of tempering moisture of wheat on grinding energy, middlings and flour size distribution, and gluten and dough mixing properties. *Journal of Cereal Science*, 69 :306–312.
- Wilkinson, J. M., Miller, H. M., McCracken, K., Knox, A., McNab, J., and Rose, S. (2003). Effect of specific weight on the metabolizable energy content of four cultivars of wheat grain in ewes. *Animal feed science and technology*, 105(1-4) :215–224.
- Willer, H., Lernoud, J., Meier, C., and Bernhard, S. (2021). The world of organic agriculture - statistics & emerging trends 2021. Report 978-3-03736-394-2, Research Institute of Organic Agriculture FiBL, Frick, and IFOAM – Organics International.
- Willer, H., Schlatter, B., and Trávníček, J. (2023). *The World of Organic Agriculture Statistics and Emerging Trends 2023*. Research Institute of Organic Agriculture FiBL, Frick, and IFOAM – Organics International, Bonn.
- Willer, H., Trávníček, J., Meier, C., and Schlatter, B. (2022). *The World of Organic Agriculture Statistics and Emerging Trends 2022*. Research Institute of Organic Agriculture FiBL, Frick, and IFOAM – Organics International, Bonn.

- Willm, C. (2009). *La mouture du blé*. Editions CEMP.
- Wrigley, C. W., Blumenthal, C., Gras, P. W., and Barlow, E. W. R. (1994). Temperature variation during grain filling and changes in wheat-grain quality. *Functional Plant Biology*, 21(6) :875–885.
- Yap, X. Y., Chia, K. S., and Suarin, N. A. S. (2022). Adaptive artificial neural network in near infrared spectroscopy for standard-free calibration transfer. *Chemometrics and Intelligent Laboratory Systems*, 230 :104674.
- Yoon, B. S., Brorsen, B. W., and Lyford, C. P. (2002). Value of increasing kernel uniformity. *Journal of Agricultural and Resource Economics*, 27(2) :481–494.
- Zhang, F. and O'Donnell, L. J. (2020). Chapter 7 - support vector regression. In Mechelli, A. and Vieira, S., editors, *Machine Learning*, pages 123–140. Academic Press.
- Zhang, W., Kasun, L. C., Wang, Q. J., Zheng, Y., and Lin, Z. (2022). A review of machine learning for near-infrared spectroscopy. *Sensors*, 22(24).

ANNEXE A PERFORMANCE DES MODÈLES DE PRÉDICTION POUR LES FÈVES DE CACAO

Table A.1 Model performance for the prediction of Cocoa bean quality features based on pre-processing, machine learning technique and wavelength selection.

Quality parameter	ML technique	Pre-processing technique	N. Wavelength	Metrics		Model
				R^2_{cv}	RMSE-test	
Moisture	PLSR	None	191	0.811	0.5079	CM1
		SG filter [3,2,1]	226	0.903	0.406	CM2
		SG filter [5,2,1]	197	0.893	0.432	CM3
		SG filter [7,2,1]	113	0.894	0.461	CM4
		SG filter [9,2,1]	181	0.863	0.469	CM5
		SNV	143	0.729	0.597	CM6
		SG filter [3,2,1]; SNV	289	0.890	0.449	CM7
		SNV; SG filter [3,2,1]	164	0.904	0.354	CM8
	SVR	None	225	0.597	1.274	CM9
		SG filter [3,2,1]	156	0.858	0.7	CM10
		SG filter [5,2,1]	177	0.849	0.699	CM11
		SG filter [7,2,1]	231	0.841	0.707	CM12
		SG filter [9,2,1]	199	0.845	0.75	CM13
		SNV	328	0.796	0.734	CM14
		SG filter [3,2,1]; SNV	196	0.861	0.567	CM15
		SNV; SG filter [3,2,1]	256	0.853	0.588	CM16
Fat	PLSR	None	161	0.642	1.528	CF1
		SG filter [3,2,1]	180	0.954	1.441	CF2
		SG filter [5,2,1]	191	0.957	1.230	CF3
		SG filter [7,2,1]	121	0.966	1.513	CF4
		SG filter [9,2,1]	115	0.959	1.279	CF5
		SNV	450	0.584	1.518	CF6
		SG filter [3,2,1]; SNV	149	0.965	1.285	CF7
		SNV; SG filter [3,2,1]	157	0.961	1.276	CF8
	SVR	None	215	0.182	1.547	CF9
		SG filter [3,2,1]	118	0.734	1.121	CF10
		SG filter [5,2,1]	119	0.721	0.942	CF11
		SG filter [7,2,1]	115	0.727	0.913	CF12
		SG filter [9,2,1]	137	0.703	1.05	CF13
		SNV	182	0.607	1.353	CF14
		SG filter [3,2,1]; SNV	113	0.758	1.094	CF15
		SNV; SG filter [3,2,1]	129	0.778	0.988	CF16

ANNEXE B PERFORMANCE DES MODÈLES DE PRÉDICTION POUR LES GRAINS DE BLÉ (B)

Table B.1 Model performance for the prediction of Wheat (B) quality features based on pre-processing, machine learning technique and wavelength selection.

Quality parameter	ML technique	Pre-processing technique	N. Wavelength	Metrics		Model
				R^2_{cv}	RMSE-test	
Protein content	PLSR	None	80	0.862	0.624	WPB1
		SG filter [3,2,1]	92	0.866	0.708	WPB2
		SG filter [5,2,1]	83	0.873	0.680	WPB3
		SG filter [7,2,1]	92	0.867	0.691	WPB4
		SG filter [9,2,1]	87	0.866	0.682	WPB5
		SNV	75	0.851	0.618	WPB6
		SG filter [3,2,1]; SNV	94	0.905	0.490	WPB7
		SNV; SG filter [3,2,1]	79	0.859	0.604	WPB8
	SVR	None	100	0.438	1.307	WPB9
		SG filter [3,2,1]	31	0.661	1.093	WPB10
		SG filter [5,2,1]	35	0.638	1.093	WPB11
		SG filter [7,2,1]	34	0.639	1.107	WPB12
		SG filter [9,2,1]	32	0.629	1.112	WPB13
		SNV	53	0.443	1.376	WPB14
		SG filter [3,2,1]; SNV	56	0.812	1.018	WPB15
		SNV; SG filter [3,2,1]	51	0.703	0.971	WPB16

ANNEXE C PERFORMANCE DES MODÈLES DE PRÉDICTION POUR LES MANGUES

Table C.1 Model performance for the prediction of Mango quality features based on pre-processing, machine learning technique and wavelength selection.

Quality parameter	ML technique	Pre-processing technique	N. Wavelength	Metrics		Model
				R^2_{cv}	RMSE-test	
Total acidity	PLSR	None	387	0.924	20.028	MTA1
		SG filter [3,2,1]	224	0.984	39.400	MTA2
		SG filter [5,2,1]	271	0.982	40.821	MTA3
		SG filter [7,2,1]	280	0.983	48.393	MTA4
		SG filter [9,2,1]	268	0.983	52.437	MTA5
		SNV	255	0.917	23.290	MTA6
		SG filter [3,2,1]; SNV	242	0.985	41.810	MTA7
		SNV; SG filter [3,2,1]	217	0.986	44.855	MTA8
	SVR	None	851	0.002	105.924	MTA9
		SG filter [3,2,1]	240	0.017	104.901	MTA10
		SG filter [5,2,1]	318	0.015	104.960	MTA11
		SG filter [7,2,1]	392	0.011	105.225	MTA12
		SG filter [9,2,1]	287	0.017	105.051	MTA13
		SNV	403	0.009	104.701	MTA14
		SG filter [3,2,1]; SNV	314	0.003	105.652	MTA15
		SNV; SG filter [3,2,1]	242	0.010	105.226	MTA16
Vitamin C	PLSR	None	66	0.663	0.848	MV1
		SG filter [3,2,1]	207	0.962	0.859	MV2
		SG filter [5,2,1]	140	0.965	1.372	MV3
		SG filter [7,2,1]	186	0.959	1.413	MV4
		SG filter [9,2,1]	146	0.971	1.383	MV5
		SNV	96	0.589	0.911	MV6
		SG filter [3,2,1]; SNV	291	0.945	0.965	MV7
		SNV; SG filter [3,2,1]	262	0.957	0.977	MV8
	SVR	None	260	0.075	1.028	MV9
		SG filter [3,2,1]	63	0.793	0.947	MV10
		SG filter [5,2,1]	103	0.681	1.012	MV11
		SG filter [7,2,1]	122	0.655	0.976	MV12
		SG filter [9,2,1]	114	0.644	0.897	MV13
		SNV	185	0.375	0.993	MV14
		SG filter [3,2,1]; SNV	96	0.733	0.979	MV15
		SNV; SG filter [3,2,1]	66	0.802	0.927	MV16