



# **Cadre méthodologique pour la gestion des opérations manufacturières avec prise en compte de la variabilité intrinsèque en contexte 4.0 : Application dans l'industrie textile et l'industrie agroalimentaire**

**Mémoire**

**Patrice Lajoie**

**Maîtrise en génie mécanique - avec mémoire**

Maître ès sciences (M. Sc.)

Québec, Canada

© Patrice Lajoie, 2020

**Cadre méthodologique pour la gestion des opérations  
manufacturières avec prise en compte de la variabilité  
intrinsèque en contexte 4.0 : Application dans l'industrie  
textile et l'industrie agroalimentaire**

**Mémoire**

**Patrice Lajoie**

Sous la direction de :

Jonathan Gaudreault, directeur de recherche  
Nadia Lehoux, codirectrice de recherche

# Résumé

Dans le contexte industriel actuel, plusieurs entreprises éprouvent des difficultés à améliorer leur performance, et ce, en raison d'une problématique de variabilité intrinsèque de certains procédés industriels. Cette variabilité peut ainsi provenir de diverses sources, que ce soit au niveau de la matière première, l'environnement du procédé ou le procédé en soi. Dans l'objectif de mieux comprendre et gérer les impacts de cette variabilité intrinsèque, un cadre méthodologique générique et flexible a été développé en se basant sur certains principes et certaines méthodes reconnues de la littérature. Composé de quatre étapes, soit l'analyse de la situation initiale, la modélisation prédictive, la modélisation prescriptive et enfin, une étape d'implantation, ce cadre méthodologique a par la suite été appliqué à deux reprises en entreprise pour des problématiques industrielles réelles. Ainsi, l'application de ce cadre pour un procédé de teinture de tissu aux prises avec une variabilité intrinsèque qui complexifie la standardisation de la couleur finale a permis le développement d'un modèle prédictif présentant une performance de prédiction de 92%, fournissant du même coup beaucoup d'information sur les sources de cette variabilité. Une seconde application du cadre pour un procédé de découpe de fromage a permis à l'entreprise de mieux connaître les variables affectant la performance du procédé. En terminant, malgré certaines limites associées à l'application du cadre telles que des contraintes de temps et une dépendance à la quantité et à la qualité des données utilisées, il a été possible de conclure que le cadre méthodologique répondait aux objectifs initiaux.

# Table des matières

Résumé .....	ii
Table des matières .....	iii
Liste des figures.....	v
Liste des tableaux.....	vi
Liste des abréviations, sigles, acronymes .....	vii
Liste des sigles.....	vii
Remerciements.....	viii
Introduction .....	1
Chapitre 1 Revue de littérature .....	3
1.1. Industrie 4.0.....	4
1.2. Notion de variabilité et d'incertitude.....	8
1.3. Méthodes.....	10
1.4. Applications industrielles .....	14
1.5. Forage de données et ses applications.....	15
Chapitre 2 Méthodologie de conception du cadre méthodologique .....	18
1.1. Méthodologie de la conception du cadre méthodologique .....	18
Chapitre 3 Cadre méthodologique proposé .....	21
3.1. Représentation conceptuelle.....	21
3.2. Présentation de haut niveau du cadre méthodologique .....	23
Étape 1 : Analyse de la situation initiale.....	24
Étape 2 : Conception du modèle prédictif .....	29
Étape 3 : Conception du modèle prescriptif .....	32
Étape 4 : Implantation, contrôle et suivi .....	34
Cadre méthodologique itératif .....	36
Chapitre 4 Mise en œuvre en contexte industriel – Cas Duvaltex .....	38
4.1. Étape 1 : Analyse de la situation initiale .....	39
Recommandation n° 1 : Mesure structurée de la couleur finale .....	52
Recommandation n° 2 : Étude et analyse de l'eau utilisée pour la teinture .....	52
Recommandation n° 3 : Structure de contrôle du tissu.....	53
Recommandation n° 4 : Récolte systématique de données .....	53
4.2. Étape 2 : Modélisation prédictive.....	54

Amélioration du modèle prédictif .....	63
4.3. Étape 3 : Modélisation prescriptive.....	65
Option 1.....	66
Option 2.....	66
Option 3.....	66
Option 4.....	67
4.4. Étape 4 : Implantation, contrôle et suivi.....	67
Chapitre 5 Mise en œuvre en contexte industriel – Cas Fromagerie L’Ancêtre .....	68
5.1. Étape 1 : Analyse de la situation initiale .....	69
Recommandation n° 1 : Mesure et utilisation des caractéristiques du lait .....	83
Recommandation n° 2 : Optimisation de la découpe .....	83
Recommandation n° 3 : Amélioration du contrôle du procédé de mise en moule.....	83
5.2. Étapes 2 à 4 : Modélisation prédictive, modélisation prescriptive et implantation, contrôle et suivi	84
Chapitre 6 Discussion et limites de la recherche .....	85
Conclusion .....	87
Références .....	89
Annexe 1 Équations des données de modélisation .....	93
Annexe 2 Liste des variables du cas d’étude Duvaltex.....	95
Annexe 3 Détails de la régression par la méthode des moindres carrés .....	97
Annexe 4 Détails de la régression nominale logistique.....	99
Annexe 5 Détails des modèles prédictifs du cas d’étude Duvaltex .....	101
Modèle 1.....	101
Modèle 2.....	103
Modèle 3.....	105
Modèle 4.....	107
Annexe 6 Article de conférence (Manufacturing Modelling, Management and Control – 9th MIM 2019 Conference).....	109

## Liste des figures

Figure 1: Les différentes phases du processus « CRISP-DM » (d'après Nisbet et al., 2009).....	12
Figure 2: Représentation visuelle des couches du processus « CRISP-DM » (d'après Nisbet et al., 2009) ....	13
Figure 3: Représentation conceptuelle d'un procédé industriel .....	22
Figure 4: Aperçu du cadre méthodologique.....	23
Figure 5: Modélisation du procédé.....	25
Figure 6: Aperçu du modèle prédictif .....	30
Figure 7: Aperçu du modèle prescriptif .....	33
Figure 8: Diagramme représentant le cadre méthodologique ainsi que ses boucles rétroactives .....	37
Figure 9: Diagramme des processus avec flux d'information.....	42
Figure 10: Représentation des sources de variabilité dans un diagramme de causes et effet pour le cas Duvaltex.....	43
Figure 11: Tableau représentant l'observabilité, la contrôlabilité ainsi que la connaissance des impacts de chaque source de variabilité du projet Duvaltex.. .....	51
Figure 12: Nomenclature des lots de tissu.....	56
Figure 13: Diagramme des processus avec flux d'information.....	71
Figure 14: Carte conceptuelle représentant les causes indirectes de la perte de fromage du procédé de découpe .....	72
Figure 15: Représentation des sources de variabilité dans un diagramme de causes et effet pour le cas L'Ancêtre.....	74
Figure 16: Tableau représentant l'observabilité, la contrôlabilité ainsi que la connaissance des impacts de chaque source de variabilité du projet L'Ancêtre. ....	82

## Liste des tableaux

Tableau 1: Identification des mesures de performance .....	26
Tableau 2. Matrice de contrôlabilité, observabilité et évaluation des impacts.....	27
Tableau 3: Exemple d'un tableau pouvant servir à classer les différentes sources de variabilité .....	28
Tableau 4: Exemple de tableau utilisé pour recenser les différentes variables associées aux sources de variabilité.....	32
Tableau 5: Présentation des sources de variabilité de même qu'une courte description, son état de mesure et contrôle et son importance relative évaluée qualitativement. ....	45
Tableau 6: Résultat des différents modèles testés dans l'étape de la modélisation prédictive .....	58
Tableau 7: Liste des facteurs en ordre décroissant d'importance du modèle 4 .....	60
Tableau 8: Différentes options possibles pour la poursuite du projet Duvaltex.....	65
Tableau 9: Résumé de la découpe des blocs de fromage.....	69
Tableau 10: Présentation des différentes sources de variabilité, d'une courte description de même que de certaines précisions à savoir si elles sont ou peuvent être mesurées et contrôlées.....	76

# Liste des abréviations, sigles, acronymes

## Liste des sigles

ACP : Analyse des composantes principales  
CRIQ : Centre de recherche industrielle du Québec  
*CRISP-DM : Cross Industry Standard Process for Data Mining.*  
*DMADV : Define, Measure, Analyse, Design, Verify*  
*DMAIC: Define, Measure, Analyse, Improve, Control*  
*EDD : Earliest Due Date*  
*FIFO : First In, First Out*  
*IoT : Internet of Things*  
*KWIC : Key Word in Context*  
PID : proportionnel, intégral, dérivé  
PME : Petites et moyennes entreprises  
*SIPOC : Supplier, Input, Process, Output, Customer*  
*SPC : Statistical Process Control*  
*SPT : Shortest Processing Time*  
*VSM : Value Stream Mapping*



# Remerciements

J'aimerais tout d'abord remercier mes directeurs de recherche, Jonathan Gaudreault et Nadia Lehoux, pour le support qu'ils m'ont offert tout au long de ma maîtrise. Je les remercie particulièrement d'avoir cru en moi et d'avoir cru que j'étais en mesure de réussir un tel projet malgré un court détour sur le marché du travail. L'achèvement de cette maîtrise n'aurait jamais été possible sans leurs précieux conseils, particulièrement lorsque j'étais en période de doutes.

Par la suite, j'aimerais aussi remercier grandement toutes les personnes du Centre de recherche industrielle du Québec (CRIQ), plus précisément Jean Wery et Corinne Chabot, pour leur aide inestimable, autant d'un point de vue technique que morale et surtout de m'avoir accueilli aussi chaleureusement. Je souhaitais aussi remercier François Gingras d'avoir cru en mes capacités à effectuer ce projet de recherche en collaboration avec le CRIQ. De surcroît, sans l'aide financière apportée par le CRIQ, ce projet de recherche n'aurait jamais vu le jour.

Les résultats obtenus dans ce projet et présentés dans ce mémoire n'auraient jamais été possibles sans la contribution plus qu'importante de l'entreprise Duvaltex, particulièrement Sandee Noonan, Mathieu Lachance, Mathieu Brousseau, Marie-Claude Côté ainsi que Marcel Côté. Leur expertise, leur temps et leur intérêt pour le projet de recherche ont fait en sorte que ce dernier fut un succès.

De la même façon, je souhaite saluer et remercier l'apport considérable de la Fromagerie L'Ancêtre pour la seconde étude de cas, plus précisément Jean Bolduc et Johanne Lemay. L'intérêt qu'ils ont porté au projet de recherche ainsi que le temps qu'ils ont accepté d'y allouer furent grandement appréciés.

Enfin, l'appui financier et l'encadrement de l'organisme Mitacs pour mon stage au sein du CRIQ ont été grandement précieux et m'ont permis de développer mes compétences dans le domaine de la recherche appliquée.

# Introduction

Considérant la compétition de plus en plus féroce, agressive et globale dans le milieu manufacturier, les entreprises québécoises, particulièrement les petites et moyennes entreprises (PME), sont forcées de redoubler d'efforts afin de maximiser la performance et la qualité de leurs produits et services tout en minimisant les coûts et les délais de commercialisation (Hofmann and Rüsçh, 2017). Ce défi d'optimisation de la performance nécessite de bien exploiter les données de production tout en ayant une meilleure connaissance vis-à-vis des processus. Pour ce faire, les nouvelles technologies développées dans les dernières années peuvent pratiquement être considérées comme des prérequis et c'est exactement dans ce contexte que la quatrième révolution industrielle, aussi appelée « industrie 4.0 », est apparue. Bien que l'étude de cette révolution industrielle ne constitue pas l'objectif principal de ce projet de maîtrise, il est pertinent d'établir sa définition dans le contexte du secteur manufacturier.

Ainsi, dans les dernières années, nous avons vu la concrétisation de projets d'envergure de transition numérique. Cependant, il semble qu'il n'y ait aucun consensus sur la méthodologie à suivre dans l'implantation des solutions de tels projets. En addition à ce manque du point de vue méthodologique, les entreprises manufacturières font face à une panoplie de problématiques dues à la nature même de leurs procédés de production. Parmi ces problématiques, on retrouve les impacts dus à la présence d'une variabilité intrinsèque des procédés. En effet, pour plusieurs entreprises, la variabilité de leurs procédés n'est pas attribuable à des conditions sous-optimales de production, mais bien à des éléments hors du contrôle de l'organisation. À titre d'exemple concret, on peut penser à l'industrie du bois d'œuvre où les entreprises doivent travailler avec une matière première qui est hautement variable lors de la découpe (Morin et *al.*, 2015). Que cette variabilité intrinsèque provienne de la nature et des caractéristiques de la matière première, de l'environnement ou bien du procédé en soi, elle affecte inévitablement la performance et ce n'est pas simplement par l'excellence opérationnelle qu'on peut y remédier. Pour le moment, il est encore incertain si le personnel technique de ces entreprises manufacturières comprend pleinement les impacts négatifs de cette variabilité de même que leur origine. La résolution de cette problématique relève d'une importance majeure : en plus de miner la performance des entreprises manufacturières québécoises, elle constitue un frein à l'évolution technologique de façon générale et nuit à la compétitivité du secteur manufacturier québécois dans son ensemble.

Dans cette optique, le but de ce projet de maîtrise est de développer un cadre méthodologique formel afin de cibler les sources de variabilité d'un procédé industriel et de déployer les solutions appropriées permettant de mieux comprendre et de gérer les impacts de cette variabilité.

Afin d'atteindre ce but, nous avons procédé en quatre étapes. Tout d'abord, une revue de littérature sur les différentes méthodes existantes se devait d'être conduite afin de connaître ce qui a été fait similairement. Ainsi, la proposition d'un cadre méthodologique préliminaire a enchaîné, pour ensuite conduire à une validation et à une vérification de ce dernier avec des cas réels. On retrouve donc deux cas réels qui seront détaillés davantage au chapitre 4, soit une application du cadre méthodologique à un procédé de teinture du secteur textile de même qu'une application à un procédé de découpe de blocs de fromage. L'application du cadre méthodologique a permis, dans le contexte de l'étude de cas portant sur le procédé de teinture, de dégager les sources de variabilité qui sont les plus statistiquement significatives sur la variabilité du procédé, permettant ainsi d'identifier des pistes d'amélioration des produits et du procédé. En addition à ces connaissances acquises lors de la modélisation du procédé, un modèle permettant de classer correctement la conformité du procédé de teinture dans 92% des cas et de détecter une non-conformité dans 73% des cas a été développé. Enfin, la dernière étape a consisté à effectuer la révision du cadre et à la mise au point de sa version finale qui sera présentée dans le chapitre 3 de ce mémoire.

En termes de contribution, ce projet de recherche se veut d'une certaine importance du point de vue de la contribution académique et industrielle. Au niveau de sa contribution académique, le cadre méthodologique développé dans le contexte de ce projet de maîtrise permet de formaliser le processus de développement et d'implantation de modèles basés sur l'utilisation de données (« *data-driven models* »). Au niveau de l'aspect industriel, le cadre méthodologique constitue un outil de grande valeur pour les entreprises en les aidant et en les encourageant à améliorer leurs façons de faire, tout en permettant un apprentissage opérationnel. Cet apprentissage permet à l'entreprise qui applique le cadre de mieux comprendre la variabilité de ses procédés et ses origines, de réfléchir à comment elle peut agir pour mieux gérer ou diminuer cette variabilité, de même qu'aider à identifier des opportunités d'améliorations au niveau des procédés ou des produits, voire de nouvelles façons de faire ou de nouvelles technologies. De plus, l'utilisation de ce dernier cadre peut servir de premières démarches dans le contexte d'une transition vers l'industrie 4.0.

Le reste de ce mémoire débutera par une revue de la littérature portant sur les sujets appropriés à la mise au point d'une méthodologie générique d'analyse de la gestion des opérations avec prise en compte de la variabilité dans un contexte 4.0. Par la suite, les objectifs, les hypothèses de même que la méthodologie employée dans le développement du cadre méthodologique seront présentés. Le troisième chapitre sera consacré à la présentation du cadre en tant que tel pour ensuite présenter les preuves de concept conduites afin de confirmer la validité et la fiabilité de ce dernier cadre. Enfin, ce document terminera avec une discussion sur les résultats et sur les limites du travail.

# Chapitre 1 Revue de littérature

Pour mieux comprendre le concept d'industrie 4.0 de même que la notion de variabilité et les méthodes employées par les experts pour mieux gérer ses impacts, une revue de la littérature basée sur la méthode du cercle herméneutique, développée par Boell et Cecez-Kecmanovic (2010), a été effectuée. Cette méthode « s'attarde au processus de création d'une compréhension interprétative. La compréhension d'un article n'est jamais isolée. L'interprétation est faite dans le contexte de plusieurs articles et documents de la littérature. La compréhension d'une littérature pertinente est par la suite influencée par chaque nouvel article lu et interprété » (traduction libre de Boell et Cecez-Kecmanovic, 2010). Elle se résume à 7 étapes :

1. La recherche avec l'utilisation d'opérateurs et de bases de données ;
2. Le tri en fonction de la pertinence, des dates et des citations ;
3. La sélection en fonction du titre, des résumés, des mots-clés and des « *KWIC* » (« *Key Word in Context* ») ;
4. L'acquisition via les différentes sources disponibles ;
5. La lecture afin d'augmenter la compréhension, la prise de notes et le référencement;
6. L'identification des termes centraux, des auteurs principaux et les principaux journaux ;
7. Et finalement, le raffinement de la recherche en utilisant deux stratégies distinctes :
  - a. le fractionnement successif : stratégie qui consiste à retirer graduellement et successivement des parties non pertinentes de la littérature ;
  - b. la construction de blocs : stratégie qui consiste à combiner des termes de recherche simples pour construire une recherche plus complexe et spécifique.

Pour ce qui est des bases de données explorées, la plus utilisée dans le cadre de cette recherche littéraire fut Web of Science de l'organisation Clarivate Analytics. Parallèlement, Google Scholar, Engineering Village d'Elsevier et la base de données ABI/Inform Global ont aussi été utilisées pour élargir le nombre d'articles disponibles. En premier lieu, les mots-clés utilisés pour démarrer la recherche concernaient principalement des composantes de la quatrième révolution industrielle tels que : « *process control* », « *intelligent process control* », « *industrie 4.0* », « *industry 4.0* », « *predictive modelling* » et plusieurs autres termes utilisés en combinaison ou individuellement. À partir des articles dégagés, les citations utilisées et les articles similaires des mêmes

auteurs furent utilisés pour explorer de nouveaux articles ou chapitres pertinents. Puisque l'objectif du cercle herméneutique n'est pas d'effectuer une revue systématique et complète de la littérature, l'idée ici était davantage d'évaluer si une méthodologie générique était employée par les experts des différents domaines de connaissance pour le développement et l'implantation de solutions technologiques dans une ère 4.0.

Le reste de cette section présente les résultats de la recherche littéraire, du plus général au plus précis. Ainsi, cela débute par une revue des différents concepts et technologies de l'industrie 4.0, pour ensuite aborder les notions d'incertitude et de variabilité. Une revue sur les méthodologies employées en industrie pour l'implantation de solutions technologiques enchaînera ensuite pour enfin terminer avec une revue de différents cas d'étude de la littérature.

## **1.1. Industrie 4.0**

Avant toute chose, il est important de mentionner que la définition de l'industrie 4.0 est relativement floue et variable selon les auteurs. Du fait de la multitude de sous-concepts et de technologies que le 4.0 sous-tend, il est très difficile d'en déduire une définition claire et concise (Hofmann et Rüscher, 2017). Hofmann et Rüscher (2017) amènent même la discussion à savoir si l'industrie 4.0 ne pourrait pas être considérée comme une sorte de mode de gestion mise de l'avant par certaines organisations pour des raisons financières plutôt que logiques et techniques (concept développé par Abrahamson, 1996). Afin d'avoir une définition qui concorde avec les réalités de l'industrie manufacturière québécoise, la définition du Ministère de l'Économie, de la Science et de l'Innovation (MESI) sera considérée pour l'ouvrage qui suit : « L'industrie 4.0, appelée également usine du futur ou quatrième révolution industrielle, se caractérise fondamentalement par une automatisation intelligente et par une intégration de nouvelles technologies à la chaîne de valeur de l'entreprise. Il s'agit d'une transformation numérique qui bouleverse l'entreprise manufacturière en apportant des changements radicaux non seulement aux systèmes et processus, mais également aux modes de gestion, aux modèles d'affaires et à la main-d'œuvre » (Ministère de l'Économie, de la Science et de l'Innovation, 2017).

Les termes « industrie 4.0 » et la « quatrième révolution industrielle » réfèrent sensiblement au même concept au point tel que plusieurs auteurs et organisations les utilisent de façon interchangeable (Hofmann et Rüscher, 2017; Ministère de l'Économie, de la Science et de l'Innovation, 2017); cependant, une distinction est à considérer. « Industrie 4.0 » renvoie en quelque sorte à une marque, à une signature publicitaire initialement employée par les Allemands, tandis que la quatrième révolution industrielle renvoie davantage à la position de ce courant dans l'histoire de l'industrialisation de l'ère moderne. Dans leur rapport sur l'industrie 4.0, la firme Roland Berger va jusqu'à affirmer que l'industrie 4.0 est en fait une réponse à la quatrième révolution, sous-entendant donc que la révolution industrielle a précédé le concept de l'industrie 4.0 (Berger, 2014).

Par sa définition, l'industrie 4.0 sous-tend plusieurs sous-concepts et technologies clés. Les concepts représentent en quelque sorte des éléments de définition de l'industrie 4.0 tandis que les technologies sont des éléments facilitateurs permettant le fonctionnement et l'implantation des concepts. Parmi les concepts clés de l'industrie 4.0, on retrouve :

- Des systèmes basés sur la collecte et l'utilisation de données ainsi que sur l'utilisation importante de modèles de simulation et d'optimisation (Gilchrist, 2016; Kusiak, 2017; Zhong et al., 2017), le tout en temps réel ;
- L'intégration horizontale et verticale (Kagermann et al., 2013) ;
- La numérisation intégrale (« end-to-end digitalization ») de la chaîne de valeur et une fusion des systèmes de production physiques avec les systèmes d'information (Kagermann et al., 2013) ;
- Un très haut niveau de collaboration et de transparence des données et de l'information à différents degrés de l'organisation (entre les machines et personnes, entre les unités d'affaires ou usines et entre les organisations) (Kagermann et al., 2013) ;
- La décentralisation des décisions (Hofmann and Rüsçh, 2017; Moeuf et al., 2017) ;
- Une personnalisation de masse de biens et services (Berger, 2014; Hofmann and Rüsçh, 2017; Zhou et al., 2015) ;
- Un haut niveau d'autonomie des équipements et des systèmes (Kagermann et al., 2013).

Les différents sous-concepts présentés ci-haut forment en quelque sorte la base de l'industrie 4.0. En revanche, l'implantation et l'intégration de ces sous-concepts ne sont pas chose facile et c'est pourquoi la quatrième révolution prévoit des technologies facilitantes dans l'atteinte des objectifs de cette transition :

- L'Internet des objets (« *Internet of Things* » ou « *IoT* »), soit l'ensemble des objets connectés ainsi que des réseaux de télécommunication et des plateformes de traitement des informations collectées qui leur sont associés (Commission d'enrichissement de la langue française, 2018; Kagermann et al., 2013) ;
- L'Internet des services, soit une « technologie qui fournit un cadre de référence pour l'utilisation de l'infrastructure, des plateformes, des applications et des services de l'Internet dans le but d'offrir des services » (traduction libre de Bertin et al., 2013) ;

- L'infonuagique (« *cloud computing* »), soit l'utilisation d'un « modèle informatique qui, par l'entremise de serveurs distants interconnectés par Internet, permet un accès réseau, à la demande, à un bassin partagé de ressources informatiques configurables, externalisées et non localisables, qui sont proposées sous forme de services, évolutifs, adaptables dynamiquement et facturés à l'utilisation » (Kagermann et al., 2013; Office québécois de la langue française, 2017a) ;
- La fabrication additive, soit la « méthode de fabrication d'un objet consistant à superposer des couches de matière à partir des données d'un modèle numérique tridimensionnel » et d'autres nouvelles technologies de fabrication (Kusiak, 2017; Office québécois de la langue française, 2017b) ;
- Les systèmes cyberphysiques, soit l'intégration des machines, des systèmes de gestion d'entrepôt et des systèmes de production dans le monde virtuel permettant l'échange d'information, le déclenchement d'actions et l'intercontrôle, le tout de façon autonome et indépendante (Kagermann et al., 2013) ;
- Les données massives ou mégadonnées (« *big data* »), soit « l'ensemble des données produites en temps réel et en continu, structurées ou non, et dont la croissance est exponentielle » (Office québécois de la langue française, 2017c) ainsi que le forage de données (« *data mining* »), soit le « processus de recherche et d'analyse de données qui permet de trouver des tendances ou des corrélations cachées souvent parmi des mégadonnées, de détecter des informations stratégiques ou de découvrir de nouvelles connaissances en s'appuyant sur des méthodes statistiques » (Office québécois de la langue française, 2017d) ;
- L'intelligence artificielle, soit le « domaine d'étude ayant pour objet la reproduction artificielle des facultés cognitives de l'intelligence humaine dans le but de créer des systèmes ou des machines capables d'exécuter des fonctions relevant normalement de celle-ci » (Office québécois de la langue française, 2017e).

Il est ainsi possible de conclure que l'industrie 4.0 consiste en une transition importante pour les entreprises manufacturières et que plusieurs organisations et centres d'expertise (autant publics que privés) tentent de faciliter ce changement global du secteur manufacturier. Dans ce contexte, Liao et al. (2017) ont recensé les avancements de la recherche sur le sujet de l'industrie 4.0 et de ses composantes. Dans un premier lieu, les auteurs ont fait un aperçu des trois premières révolutions industrielles de même que des efforts mis de l'avant par différents gouvernements et organismes œuvrant dans le domaine du développement industriel. En utilisant la méthode de « *Preferred Reporting Items for Systematic review and Meta-Analysis (PRISMA)* » (Moher et al.,

2009), les auteurs ont effectué une étude qualitative et quantitative des travaux. C faisant, ils ont tenté de répondre à quatre questions pertinentes :

- « (1) Which are enabling features of Industry 4.0?
- (2) Who is working on Industry 4.0, when and where?
- (3) What are the main research directions and the current research efforts?
- (4) What are the existing Industry 4.0 application fields? » (p.3611, Liao et al., 2017).

Les auteurs ont ainsi identifié et classifié les efforts de recherche portant sur l'industrie 4.0, par exemple le domaine de la standardisation et des architectures de référence ainsi que l'étude de la productivité et l'efficience des ressources, puis ont dégagé les lacunes sur lesquelles les efforts de recherche et de développement devraient s'orienter. Parmi ces lacunes, les chercheurs ont identifié le cadre de régulation ainsi que la numérisation intégrale des processus (« *end-to-end digitalization* »). Les auteurs ont aussi abordé le fait que certaines directions de recherche qui ne figurent pas parmi les priorités identifiées par Kagermann et al. (2013) ont tout de même fait l'objet d'efforts de recherche importants. À titre d'exemple, les auteurs mentionnent la science des données (« *Data Science* »), l'analyse des données en temps réel (« *Real Time Data Analysis* »), l'intégration des données (« *Data Integration* ») ou bien l'analytique des données massives (« *Big Data Analytics* »). Le projet de maîtrise décrit dans ce mémoire s'imbrique d'ailleurs dans plusieurs avenues de recherche citées par ces auteurs telles que la numérisation intégrale des processus, la science des données, l'intégration des données ainsi que la gestion des systèmes complexes.

Sur une autre note, plusieurs auteurs de la littérature abordent les implications et les impacts de cette nouvelle révolution industrielle qu'est l'industrie 4.0. Moeuf et al. (2017) ont étudié ce phénomène qu'est l'industrie 4.0 du point de vue des PME (les petites et moyennes entreprises) . Considérant la multitude des concepts et technologies compris dans cette tendance, aussi appelés « moyens de réalisation » dans cet article, les auteurs ont présenté des moyens (concepts et technologies) qui auront le plus d'impact et des bénéfiques pour ces organisations, souvent délaissées par les courants et technologies industriels, et ce, en se basant sur des cas d'application de la littérature. De façon similaire, Kusiak a présenté le futur de l'industrie manufacturière au sein de la quatrième révolution industrielle, dans une perspective plus générale et plus large (Kusiak, 2017). Plus précisément, l'auteur a divisé le concept de l'usine intelligente en six piliers principaux :

1. Les technologies et les processus ;
2. Les matériaux ;



3. Les données ;
4. L'ingénierie prédictive ;
5. La durabilité et le partage des ressources ;
6. Le réseautage.

L'auteur a aussi discuté des développements dans le domaine du transport et de la manutention du matériel, de même que de l'intégration du développement durable dans les processus manufacturiers.

## **1.2. Notion de variabilité et d'incertitude**

Le Larousse en ligne propose plusieurs définitions du mot « variable » (considérant que le mot « variabilité » est un dérivé du mot « variable »). Cependant, une des définitions qui semble correspondre à la problématique étudiée serait « qui change dans un temps assez court, qui n'est pas stable » (Larousse, 2018a). Or, cette définition est évidemment très vague et générale; selon les situations, la variabilité sera de nature différente. Pour ce qui est de l'incertitude, le Larousse définit le terme « incertain » comme étant « qui n'est pas établi avec exactitude, connu avec certitude » (Larousse, 2018b). Sans prétendre à l'exhaustivité, cette section se veut davantage une revue exploratoire sur les différentes facettes et perspectives de la variabilité et de l'incertitude afin de mieux camper la problématique de recherche.

Traditionnellement, dans les méthodes et techniques relatives au Lean Six Sigma, on retrouve plusieurs outils développés dans l'objectif de réduire la variabilité des procédés industriels ou, du moins, de mieux la contrôler (Pyzdek, 2003). Parmi ces outils, on peut penser aux outils relatifs au « *Statistical process control* » (« SPC ») tels que les cartes de contrôle de procédés. Ces derniers outils sont souvent utilisés pour détecter l'occurrence d'événements anormaux tels que des bris d'équipements, l'usure d'outils ou les dérives des procédés (Pyzdek, 2003). D'autres outils tels que les plans d'expérience, permettent, quant à eux, de mieux comprendre les relations entre des variables et une variable de contrôle, outil qui s'avère particulièrement utile dans des contextes industriels. Pyzdek (2003) a abordé dans l'un de ses chapitres la découverte des connaissances (« *Knowledge discovery* »), un ensemble d'outils et de techniques qui permettent de développer la compréhension d'une problématique, compréhension nécessaire avant l'étape de résolution. Plusieurs de ces outils et techniques consistent en des méthodes de visualisation de données qui permettent de faciliter la transmission d'informations. Évidemment, toutes ces techniques et outils tirés de la méthode Lean Six Sigma sont très orientés vers la pratique et l'amélioration continue des procédés industriels. Une autre avenue pour le développement de la compréhension de la variabilité et de l'incertitude repose sur les statistiques descriptives.

Le fait d'étudier la dispersion ou la forme de la distribution des données peut permettre de déceler d'importants indices sur la source de cette variabilité.

Certains auteurs s'intéressent également à la variabilité et à l'incertitude au niveau de la chaîne d'approvisionnement des organisations. Entre autres, Van der Vorst et Beulens (2002) ont abordé les stratégies de reconfiguration de la chaîne d'approvisionnement. Dans leurs démarches, les auteurs ont défini l'incertitude de la chaîne d'approvisionnement comme étant « les situations de prise de décision dans lesquelles le preneur de décision ne sait pas quelle décision prendre puisqu'il est indécis par rapport aux objectifs, il manque d'information (ou de compréhension) sur la chaîne d'approvisionnement ou son environnement, il manque d'information sur la capacité de traitement et enfin, il est incapable de prédire précisément l'impact possible des différentes actions alternatives ou manque d'actions de contrôle efficaces (incontrôlabilité) » (Van der Vorst and Beulens, 2002). Les auteurs ont également parlé de « *unchangeable external givens* », soit des acquis externes inchangeables, des facteurs externes considérés comme des contraintes par les gestionnaires. On voit ici un lien clair avec la problématique de variabilité des procédés manufacturiers où plusieurs gestionnaires et opérateurs de production considèrent « subir » certaines sources de variabilité.

Dans ce sens, un concept très important de la problématique illustrée précédemment dans l'introduction est la présence d'une variabilité intrinsèque dans certains procédés. Dans plusieurs secteurs industriels, la variabilité des procédés n'est pas le résultat d'un procédé sous-optimal, mais plutôt un élément intrinsèque au procédé, élément avec lequel les industriels doivent composer du mieux qu'ils peuvent. À titre d'exemple, Morin et al. (2015) ont conçu un métamodèle permettant de prédire le panier de produits en fonction des caractéristiques de la matière première, soit les billes de bois qui sont hautement variables. Cet exemple illustre que peu importe le contrôle qui est mis en place pour diminuer la variabilité du procédé, la haute variabilité de la matière première fait en sorte que les entreprises du secteur du bois d'œuvre doivent adapter leurs procédés à cette matière première, d'où l'utilisation d'un métamodèle. En revanche, pour maximiser la performance de ces recettes, il est important pour ces entreprises de bien comprendre et de caractériser les origines de cette variabilité. Cette variabilité intrinsèque pourrait être définie comme étant une forme de variabilité dont il est quasi impossible de totalement éliminer et dont l'entreprise qui y fait face doit trouver des façons d'adapter leurs procédés pour diminuer leurs impacts. D'ailleurs, ce phénomène de variabilité intrinsèque est rencontré dans plusieurs secteurs industriels :

- Secteur de la transformation alimentaire où la matière à transformer est hautement variable ;
- Secteur des pâtes et papier où la qualité et les caractéristiques des copeaux de bois utilisés varient ;

- Secteur des plastiques, par exemple la fabrication de baignoires en acrylique, où les caractéristiques de la résine utilisée pour le renforcement sont très variables ;
- Secteur du textile où plusieurs procédés sont fortement impactés par les caractéristiques du fil et du tissu de même que par les paramètres du procédé ;

### 1.3. Méthodes

Si on jette un coup d'œil aux méthodologies existantes pour instrumentaliser les données dans l'amélioration des processus, un consortium d'entreprises comprenant Daimler-Benz, SPSS et NCR ont mis au point un processus soutenant les projets de forage de données, le « *CRISP-DM* » (Nisbet et al., 2009) qui, selon Nisbet et al. (2009), serait la méthode disponible la plus complète à ce jour. La méthode se divise en six phases principales qui se chevauchent et qui s'appliquent de façon itérative :

- La compréhension de l'entreprise : cette étape du processus consiste à définir les objectifs d'affaires du modèle de forage de données, à évaluer l'environnement de modélisation (par exemple, évaluer la qualité et l'intégrité des données de même que les outils analytiques à la disposition de l'analyste de données) ainsi qu'à formuler les objectifs et les buts du modèle de forage de données ;
- La compréhension des données : cette seconde étape du processus a pour objectif de déterminer comment trouver et extraire les données nécessaires à la modélisation de même que l'intégration des données. En addition à cela, c'est aussi à cette étape que la description des données et l'évaluation détaillée de la qualité seront effectuées ;
- La préparation des données : une fois la compréhension des données complétée, il apparaît nécessaire de préparer les données pour le forage. Plus particulièrement, on considère dans cette étape tout ce qui touche au nettoyage, à la transformation, à l'imputation, à l'affectation de pondération, au filtrage, à l'abstraction, à la réduction et à la création de données ;
- La modélisation : cette prochaine étape du processus consiste à effectuer la modélisation des données et peut être séparée en quatre sous-étapes :
  1. La sélection d'une technique de modélisation en choisissant un algorithme, une architecture et en spécifiant les hypothèses de base ;
  2. La création du concept expérimental ;

3. La construction du modèle en question en testant différents types de modèles et différents paramètres ;
  4. L'analyse du modèle en comparant les résultats de son utilisation avec les résultats actuels du système ou en utilisant d'autres méthodes statistiques.
- L'évaluation : une fois le meilleur modèle respectant les contraintes trouvées, il s'agit d'évaluer le modèle à l'aide d'indicateurs de performance tels que le taux de bonne prédiction. Cette étape du processus est parfois intégrée à l'étape précédente de modélisation des données ;
  - Le déploiement : enfin, cette dernière étape du processus a pour but de planifier le déploiement du modèle de forage de données et d'en assurer le suivi et la maintenance. Finalement, cette étape prévoit aussi la production d'un rapport final présentant la version finale du modèle et une revue du projet de forage de données.

Afin de faciliter son fonctionnement, voici une représentation visuelle du processus « *CRISP-DM* » tel que présenté par Nisbet et *al.* (2009) :

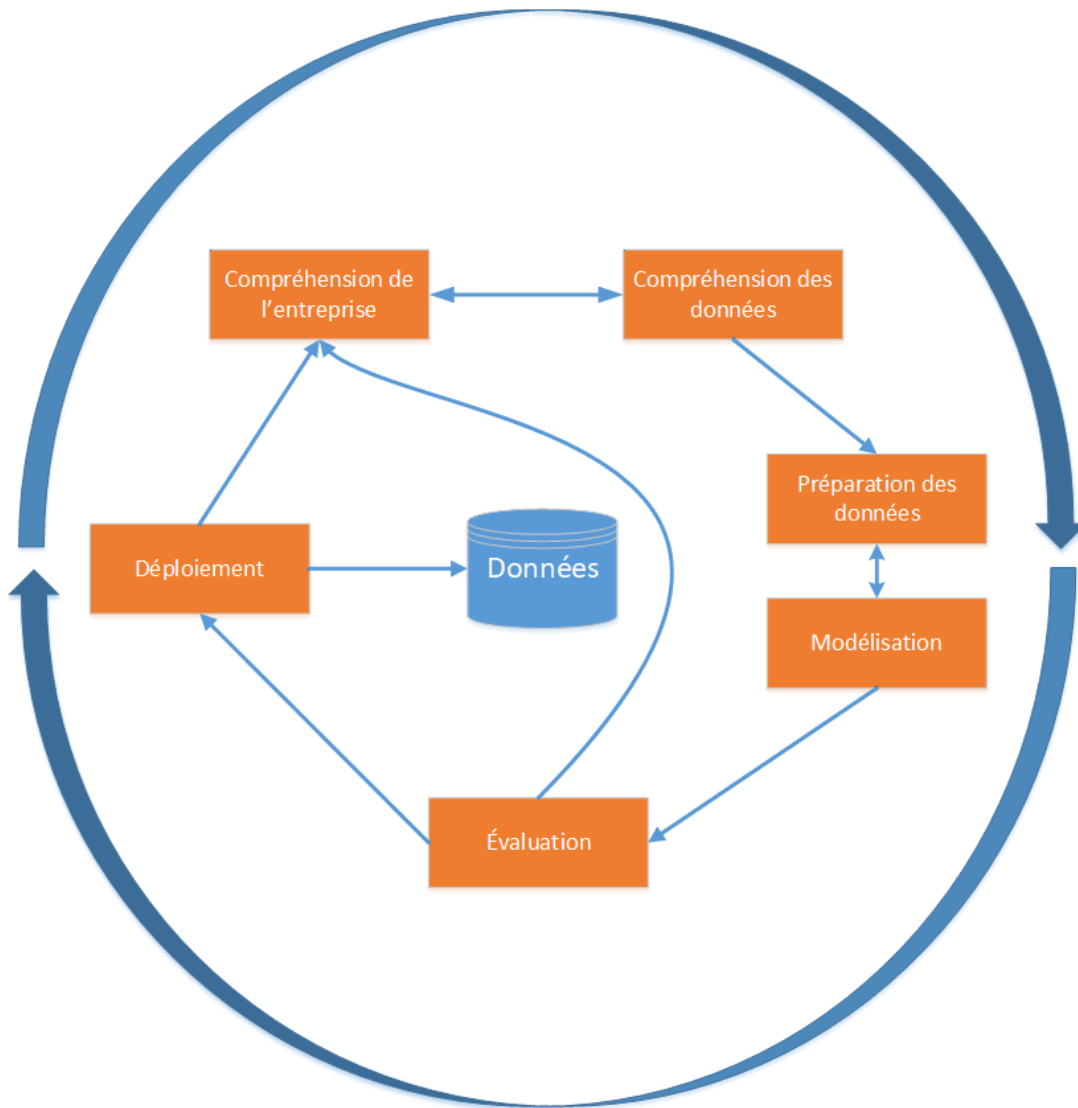


Figure 1: Les différentes phases du processus « CRISP-DM » (d'après Nisbet et al., 2009)

Selon Nisbet et al. (Nisbet et al., 2009), l'application de cette méthode est un mélange d'art et de sciences, en fonction des phases et des tâches effectuées. La méthode est composée de quatre couches hiérarchiques : la couche des phases présentées ci-haut, la couche des activités, la couche des opérations et enfin la couche des tâches. Ces couches se superposent et sont de plus en plus détaillées en termes d'actions à entreprendre. En voici une représentation visuelle :

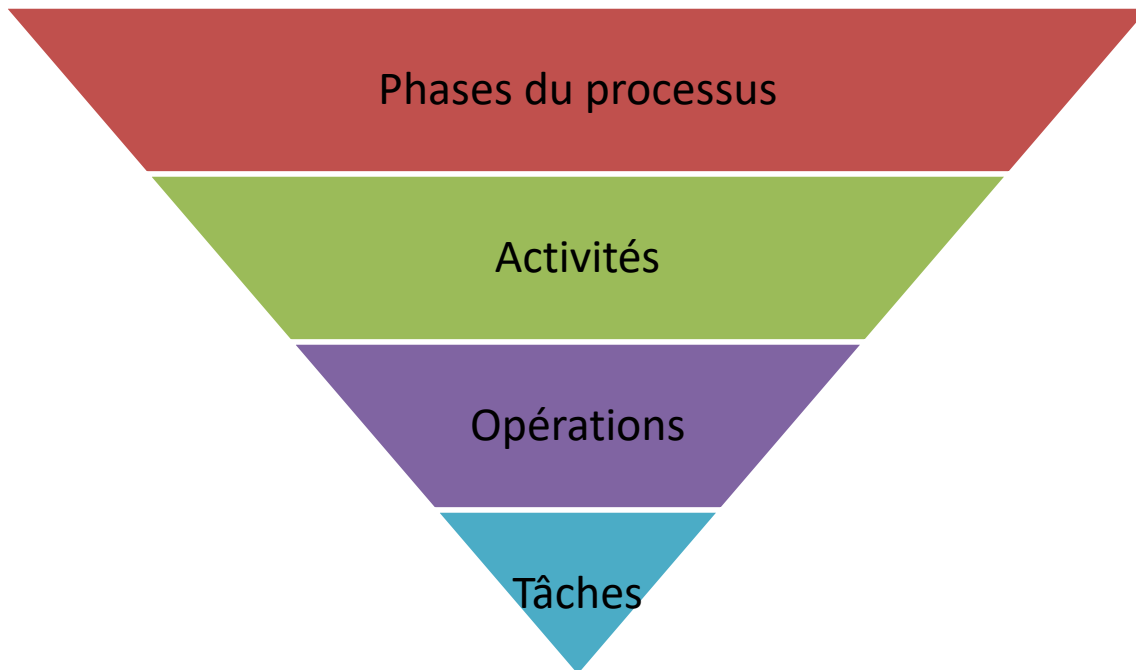


Figure 2: Représentation visuelle des couches du processus « CRISP-DM » (d'après Nisbet et al., 2009)

La méthode « CRISP-DM » ne prend toutefois peu ou pas en compte l'exploration et la compréhension des sources de variabilité des procédés manufacturiers. En effet, la première phase de compréhension de l'entreprise relève davantage de l'art selon Nisbet et al. (2009) et s'attarde au contexte général de la problématique plutôt qu'à identifier les sources de variabilité. À titre d'exemple, Nisbet et al. (2009) affirment sur le sujet de l'acquisition des données : « *First, you have to identify the various data sources available to you. These data may be nicely gathered together in an enterprise data warehouse. While this situation is ideal, most situations will be quite different. Data may be scattered in many business units of the company in various data "silos," spreadsheets, files, and hard-copy (paper) lists. Your next challenge is to put all this information together* ». Or, cette affirmation implique que les données se retrouvent du moins quelque part et qu'elles ont été récoltées pour une utilisation ultérieure, ce qui n'est pas toujours le cas dans des milieux manufacturiers, où parfois les données ne sont tout simplement pas récoltées. Sans données sur lesquelles s'appuyer, il devient très difficile de développer une bonne compréhension de la variabilité d'un procédé quelconque.

Une autre méthode très répandue et popularisée est celle utilisée dans les démarches d'amélioration continue et de « Lean Six Sigma », soit la méthode DMAIC (« Define, measure, analyze, improve, control » pour l'amélioration d'un processus) ou DMADV (« Define, measure, analyze, design, verify » pour la conception d'un processus) (Pyzdek, 2003). Malgré le fait que cette méthode a largement fait ses preuves dans les dernières décennies en ayant été utilisée à différents niveaux des organisations, il s'agit d'une méthode qui demeure générique et qui ne vise pas systématiquement à s'attaquer aux problématiques de variabilité décrites

précédemment. Dans ce contexte, De Mast and Lokkerbol (2012) ont démontré les limites de cette méthode du point de vue de la résolution de problèmes. En effet, les auteurs soulignent le fait que la méthode *DMAIC* n'est pas adaptée aux environnements industriels où la dynamique humaine, les valeurs personnelles et les perceptions subjectives sont parfois importantes, ce que les auteurs appellent un problème mal structuré (« *ill-structured problem* »). Ces problèmes subjectifs sont aussi caractérisés par un manque de clarté sur comment le problème devrait être abordé. De la même façon, les auteurs discutent du phénomène appelé « *power/generalizability trade-off* » qui fait en sorte que plus une méthode est générique, moins elle est utile pour résoudre une problématique spécifique à un domaine (par exemple, une problématique de variabilité dans l'industrie du bois d'œuvre). De Mast et Lokkerbol mentionnent aussi que la méthode *DMAIC* propose un ensemble d'outils qui ne prend pas suffisamment en compte des outils ou techniques issus d'autres domaines de connaissance (par exemple, l'utilisation de contrôleurs PID plutôt que d'utiliser la méthode *SPC* préconisée par la méthode *DMAIC* pour mieux contrôler les procédés industriels). Cependant, les concepts sous-jacents à cette méthode de même que la logique de ces étapes pourraient servir de base à la conception d'une méthodologie générique permettant de réduire les impacts négatifs de la variabilité qu'il serait possible de qualifier d'inhérente des procédés manufacturiers.

#### **1.4. Applications industrielles**

Plusieurs auteurs se sont intéressés aux procédés manufacturiers complexes dans un contexte de forte variabilité. Notamment, Morin et al. (2015) se sont penchés sur le procédé de sciage des billes dans l'industrie du bois d'œuvre qui repose sur un flux divergent, c'est-à-dire qu'une seule bille peut donner en extrants plusieurs planches de dimensions et de qualité différentes. À l'aide d'un équipement de simulation qui permet de prédire pour une bille donnée le panier de produits qui sera obtenu, Morin et al. (2015) ont utilisé une méthode de métamodélisation afin de prédire les résultats de la simulation à partir de quelques caractéristiques initiales de la bille. Cette méthode permet de prédire le panier de produits beaucoup plus rapidement que par l'utilisation d'un logiciel de simulation.

En utilisant une démarche plus exploratoire, Salim et al. (2016) se sont penchés sur l'influence de la matière première sur un procédé de fabrication de panneaux de bois, et ce, en termes de productivité et d'efficience. Les auteurs s'intéressent à l'impact du niveau de la qualité de la matière première sur l'efficacité du procédé et comparent le temps de production à valeur ajoutée pour chacun de ces niveaux en fonction de certaines causes d'arrêt opérationnelles.

Dans un même ordre d'idées, Noyel (2015) s'est intéressée à une entreprise spécialisée dans la fabrication de mobiliers d'agencement sur mesure (présentoirs pour événements, salles de bains, cuisines, etc.). Elle a constaté un taux de reprise très élevé sur un de leurs procédés de laquage qui avait beaucoup d'impacts sur les coûts et la productivité. Après avoir étudié et analysé les différentes méthodes existantes, l'étudiante et l'entreprise ont implanté un algorithme d'apprentissage automatique afin de modéliser ce procédé qui comportait des facteurs contrôlables et non contrôlables variables. Ce modèle leur a permis de prédire la qualité de leur produit en fonction de ces facteurs. En utilisant une carte de contrôle, ils peuvent aujourd'hui surveiller la dérive du système de prédiction et relancer l'entraînement de ce dernier (Noyel, 2015). En addition à cette fonction de prédiction pour le procédé de laquage, ils ont développé un outil qui permet de déterminer le mode de fonctionnement de la production (*EDD*, *FIFO* ou *SPT*) en fonction du nombre de défauts détectés et du nombre de produits en-cours.

Sur une même note, Konrad et *al.* ainsi que Lieber et *al.* (2013a; 2013b; 2012) ont travaillé sur l'implantation d'un système de contrôle intelligent de procédé manufacturier (« *Intelligent Manufacturing Process Control* » ou « *IMPC* ») pour un procédé de production de barres d'acier. Ce système permet, à chaque étape du procédé, de prédire un certain niveau de qualité du lot en cours en fonction des paramètres de production avec l'utilisation de méthodes de forage de données appliquées à l'historique de données de production. Si le niveau de qualité prédit ne respecte pas un certain seuil minimal, le lot est automatiquement interrompu pour éviter des pertes de temps et de ressources supplémentaires. De plus, les analyses statistiques effectuées préalablement à la prospection de données ont permis aux chercheurs et aux industriels de dégager les paramètres de production qui avaient le plus d'impacts positifs et négatifs sur la qualité finale d'un lot d'acier produit.

## **1.5. Forage de données et ses applications**

Une autre avenue d'instrumentalisation des données proposée dans la littérature porte sur le forage de données (« *data mining* ») dans le domaine manufacturier (Kusiak, 2017). Choudhary et *al.* (Choudhary et *al.*, 2009) se sont intéressés à ce domaine de recherche et plus précisément aux types de connaissances (« *Kind of knowledge* ») que l'on souhaite obtenir par l'utilisation du forage de données. Les auteurs dégagent cinq fonctions au forage de données :

- la description de concept ;
- le partitionnement ;
- la classification ;
- la prédiction ;



- l'association.

En plus de s'intéresser aux différents domaines de connaissances du « *data mining* », les auteurs ont exploré les différentes méthodes (algorithmes, techniques, etc.) employées pour certains secteurs manufacturiers, et ce, pour chacune des fonctions identifiées ci-haut. L'article de Choudhary et al. (2009) se veut un point de départ pour la recherche faisant appel au forage de données dans un contexte manufacturier, puisque les auteurs citent plus d'une centaine de cas d'application des méthodes.

Plusieurs exemples d'utilisation de techniques de forage de données afin de prédire la performance d'un procédé dans un contexte manufacturier peuvent être trouvés dans la littérature. Des chercheurs ont par exemple développé une méthode hybride (utilisation de plusieurs algorithmes de nature et de fonctions différentes) afin de prédire les extrants d'un procédé de cokéfaction (Lei et al., 2016). Ce procédé dépend grandement des paramètres de production, particulièrement de la température qui varie énormément et de l'approvisionnement en carburant pour obtenir une coke de bonne qualité. Dans un autre cas, une entreprise avait de la difficulté à prédire la qualité d'un procédé de fabrication de fil de tissu (Z. J. Lu et al., 2015). La variabilité de la matière première, les multiples étapes du procédé et un manque de contrôle sur le procédé demeuraient un frein à cette prédiction. Les chercheurs ont utilisé une méthode d'apprentissage afin de prédire la qualité de ce procédé de fabrication à partir d'un ensemble de données. À titre de dernier exemple, une entreprise avait des difficultés à prévoir le résultat d'un procédé de forgeage dû aux comportements rhéologiques complexes et variables ainsi qu'aux formes irrégulières des produits. Des chercheurs ont utilisé une machine d'apprentissage (méthode appelée « *extreme learning machine* ») afin de modéliser les forces de déformation de pièces forgées à l'aide d'un moule. Cette modélisation leur a permis d'obtenir un intervalle de confiance visuel du procédé (X. Lu et al., 2015).

Cook et Shannon (Cook and Shannon, 1992) ont employé des réseaux de neurones multicouches avec apprentissage par rétropropagation (« *back-propagation neural networks* ») pour prédire les paramètres d'un procédé de fabrication de panneaux de bois composites. Après avoir tenté de modéliser le procédé à l'aide d'une régression multivariée ainsi que d'une analyse de séries temporelles (« *time series analysis* »), les auteurs n'ont pas été en mesure de prédire efficacement les paramètres de production désirés et se sont donc tournés vers la théorie des réseaux de neurones. L'analyse et la visualisation de la surface de réponse faites par les auteurs ont permis de mettre en lumière la complexité du comportement du procédé et le fait que les variables dépendantes ne pouvaient pas être expliquées à l'aide d'une régression des variables indépendantes du problème. Ainsi, les auteurs ont utilisé certaines caractéristiques de la matière première telles que la densité et la teneur en humidité à divers endroits du procédé ainsi que certains paramètres des équipements pour prédire la teneur en humidité à la sortie du procédé, de même que la classification du traitement par résine. Au niveau

des performances, les modèles prédictifs développés en utilisant des réseaux de neurones ont obtenu entre 55 et 71% de taux de prédiction (selon le paramètre prédit et le nombre d'exemples utilisés), soit des performances modestes, mais largement supérieures aux résultats de la régression employée précédemment.

Certains chercheurs ont intégré ces méthodes de forage de données aux techniques traditionnelles de contrôle de procédé telles que les chartes de contrôle de procédé. Par exemple, Guh (2005) a utilisé des réseaux de neurones afin de détecter rapidement les tendances anormales des chartes de contrôle (Guh, 2005). En combinant une série de réseaux de neurones dédiés à la détection d'un type de dérive, le système développé est en mesure de détecter les différents problèmes pouvant apparaître sur le long terme dans les chartes de contrôle de procédés. Cette méthode s'avère être nettement plus rapide et plus efficace que les méthodes traditionnelles de détection des tendances. Ce dernier exemple ainsi que les autres énumérés précédemment témoignent de l'importance qu'occupe et qu'occupera dans le futur les méthodes relatives à l'intelligence artificielle et au forage de données dans les organisations.

Considérant les différentes avenues de recherche détaillées dans cette présente section, il semble que l'élaboration d'un cadre méthodologique permettant de mieux comprendre et de mieux gérer les impacts de la variabilité intrinsèque des procédés manufacturiers relève d'une pertinence notable, autant d'un point de vue académique qu'industriel. Bien que diverses méthodes existent afin de diminuer la variabilité des procédés ou tout simplement explorer des données pour en soutirer de l'information, ces méthodes démontrent souvent un manque du point de vue méthodologique en ce qui a trait à l'exploration et la compréhension des sources de variabilité. Le cadre méthodologique qui a ainsi été conçu dans le contexte de ce projet de recherche est détaillé dans le prochain chapitre.

# Chapitre 2 Méthodologie de conception du cadre méthodologique

À la lumière de ce qui précède, il est possible de définir les objectifs généraux et spécifiques de ce projet de recherche. L'objectif général se définit comme la mise au point d'un cadre méthodologique fonctionnel, éprouvé et original, permettant d'améliorer la gestion des sources de variabilité et de leurs impacts sur les procédés manufacturiers dans un contexte industriel. Les objectifs spécifiques rattachés à la recherche sont les suivants :

1. Déterminer si une méthode permettant d'identifier des sources de variabilité et de minimiser l'impact de cette variabilité sur la performance des procédés existe. Dans ce contexte, il est pertinent de se pencher autant sur la littérature portant sur les méthodologies dans les milieux manufacturiers que sur les études de cas et, pour ces dernières, se pencher sur la façon dont les solutions ont été développées.
2. Concevoir un cadre méthodologique préliminaire qui permettra, en premier lieu, de poser un diagnostic de la situation initiale d'une entreprise donnée et de la problématique de variabilité à laquelle elle fait face, pour ensuite proposer une démarche de modélisation prédictive et prescriptive, ainsi qu'une démarche d'implantation pour mieux gérer les sources de cette variabilité et leurs impacts sur la performance et la qualité du procédé.
3. Effectuer une preuve de concept de l'application de ce dernier cadre, et ce, en collaboration avec des entreprises manufacturières aux prises avec une problématique de variabilité dans leurs procédés.
4. Enfin, suite à la preuve de concept, réviser le cadre méthodologique préliminaire et mettre au point sa version finale.

Il est à noter que dans le contexte actuel, une différence est faite entre mesures et indicateurs de performance afin d'éviter la confusion. Les indicateurs de performance sont les façons dont un procédé est évalué quantitativement et comment sa performance est suivie, tandis que les mesures de performance font référence aux extrants d'un procédé qui permettent de mesurer les indicateurs de performance.

## 1.1. Méthodologie de la conception du cadre méthodologique

Cette sous-section décrit la méthodologie employée pour la conception du cadre méthodologique, ce dernier étant présenté au chapitre suivant.

Pour débiter, une recherche dans la littérature a été menée afin de déterminer si une méthodologie formelle avait été utilisée dans les études de cas retrouvés dans la littérature scientifique. De la même façon, une

exploration de la littérature portant sur les méthodologies d'implantation ou d'amélioration de procédés industriels avec l'aide de nouvelles technologies a été conduite. L'idée centrale de cette recherche littéraire était de comprendre la façon dont les auteurs ont abordé l'incertitude et la variabilité dans les procédés manufacturiers. À ce propos, une hypothèse a été élaborée, soit que tout procédé possède une ou plusieurs sources de variabilité dont le degré varie de très faible à très fort. En effet, dans le contexte du projet actuel, nous considérons que, peu importe le procédé à l'étude, il est possible de retrouver une ou plusieurs sources de variabilité qui ont un impact sur les mesures de performance du procédé. La revue a ainsi permis de conclure qu'aucune méthode formelle ne semble avoir été utilisée. Cependant, comme il a été mentionné précédemment, la méthode « *CRISP-DM* » s'apparente dans une certaine mesure à la méthodologie recherchée, tout en étant différent sur quelques points cruciaux, plus précisément sur l'étude et la compréhension des sources de variabilité, d'où la pertinence de mettre au point un cadre méthodologique permettant de remédier à ce manque.

La conception d'un cadre méthodologique (décrit au chapitre 3) a donc par la suite été réalisée. Ce cadre repose sur 4 étapes majeures et plusieurs sous-étapes, soit l'analyse de la situation initiale, la conception d'un modèle prédictif, la conception d'un modèle prescriptif et enfin, l'implantation et le suivi des modèles. D'ailleurs, il est à noter que deux hypothèses ont été posées lors de la mise au point de ces étapes. Tout d'abord, il existe une méthode de modélisation prédictive parmi celles que l'on retrouve dans la littérature qui obtiendra un taux de prédiction acceptable pour chaque problématique. Ainsi, suivant cette hypothèse, si un modèle prédictif n'est pas en mesure d'atteindre les objectifs en termes de taux de prédiction, le problème réside dans les données, le nombre de données ou tout simplement dans la méthode employée. Pour ce qui est de la seconde hypothèse, on considère que pour chaque modèle de prédiction, il existe une fonction approximative ou bien un modèle prescriptif qui permet de prescrire des valeurs de paramètres. Davantage de détails concernant ces hypothèses seront fournis dans le chapitre suivant.

La première étape a été conçue en se basant sur des conversations et des discussions avec des entreprises manufacturières qui faisaient face à des problèmes de variabilité au niveau de leurs procédés. Dans plusieurs de ces situations problématiques, l'organisation concernée considérait cette variabilité comme une contrainte à considérer, un élément impondérable auquel il fallait s'adapter. C'est ainsi qu'une première étape du cadre qui se veut plus exploratoire et plus orientée vers l'investigation fut créée. Pour ce qui est des étapes subséquentes du cadre (étapes 2 à 4), elles ont été conçues, basées sur le paradigme « descriptif/prédictif/prescriptif » que l'on retrouve dans le domaine de l'analytique décisionnelle ou de l'analyse des données, paradigme que l'on retrouve d'ailleurs dans plusieurs autres domaines de connaissance. Ainsi, Gilchrist (2016) définit quatre niveaux ou types de méthodes analytiques :

- l'analyse descriptive, qui cherche à comprendre ce qui est arrivé et comment, où ainsi que quand cela est arrivé ;
- l'analyse diagnostique, qui cherche à comprendre où chercher et comprendre pourquoi quelque chose est arrivé ;
- l'analyse prédictive, qui cherche à trouver des motifs dans les données et à prévoir ce qui arrivera dans le futur ;
- l'analyse prescriptive, qui cherche à valider des scénarios et à déterminer les meilleures actions ou décisions à prendre.

Au départ, l'intérêt pour les outils avancés d'analyse de données se limitait à certains secteurs d'affaires tels que l'assurance, le marketing et les finances (Gilchrist, 2016). Cependant, au fil des années, cet intérêt s'est aussi développé pour divers secteurs tels que le domaine de la santé et le secteur industriel. Dans le contexte du projet de recherche faisant l'objet de ce mémoire, ce seront les analyses prédictives et prescriptives qui seront d'intérêt.

Suite à la mise au point du cadre méthodologique théorique, la rigueur scientifique exige de tester ce dernier afin de vérifier s'il satisfait les exigences et les objectifs déterminés lors de la conception du cadre. Cette vérification/validation fut faite sous la forme de preuves de concepts qui feront l'objet des chapitres 4 et 5 du mémoire. Plus précisément, ces preuves de concept ont été conduites en appliquant le cadre méthodologique directement en industrie pour résoudre des problématiques de variabilité. La première preuve, qui comprend l'analyse de la situation initiale et la mise au point de modèles prédictifs et prescriptifs, a été conduite sur un procédé de teinture dans une entreprise textile, la compagnie Duvaltex, tandis que la deuxième preuve de concept, qui elle ne comprend que l'analyse de la situation initiale, a été conduite avec une entreprise agroalimentaire, soit une fromagerie possédant la certification biologique, la Fromagerie l'Ancêtre. Cette validation a par la suite permis d'apporter des améliorations à la version préliminaire du cadre méthodologique et ainsi parvenir à la version finale présentée au prochain chapitre de ce mémoire.

## Chapitre 3 Cadre méthodologique proposé

Compte tenu de la revue de littérature complétée au premier chapitre ainsi que de la méthodologie et des hypothèses discutées au second chapitre, ce chapitre du mémoire est dédié à la présentation d'un cadre méthodologique développé afin de répondre aux objectifs de recherche reliés à la problématique introduite précédemment. Il s'agit de la version finale du cadre à la suite des modifications et ajouts apportés au cadre préliminaire.

Il est question ici d'un cadre méthodologique, mais il semble pertinent d'expliquer ce qu'est un cadre et en quoi un tel outil permettra de faire face à la problématique décrite à l'introduction. Dans le contexte de ce projet, le sens prêté au terme « cadre » s'apparente à la signification que l'on retrouve dans le domaine du développement logiciel, soit un ensemble d'outils génériques et réutilisables permettant de supporter le développement d'un système, par exemple un logiciel (inspiré de Larman (2004)). Le cadre méthodologique qui sera présenté dans ce chapitre va même au-delà d'un ensemble d'outils en présentant aussi une ligne directrice à suivre, d'où l'emploi du terme « cadre méthodologique ».

### 3.1. Représentation conceptuelle

Avant de présenter les différentes étapes du cadre méthodologique, il est pertinent d'introduire la représentation d'un procédé industriel considérée tout au long du projet de recherche. Plus précisément, il est possible de représenter un procédé industriel  $F$  tel que  $F(X)=Y$ , où l'ensemble  $X$  est constitué des vecteurs  $x_{i=1..m}$  qui représente les paramètres et variables d'entrée que l'on appellera sources de variabilité et où l'ensemble  $Y$  est constitué des vecteurs  $y_{j=1..n}$  qui représente les mesures de performance du procédé. Cette représentation, retrouvée à la Figure 3, servira de base pour la définition du cadre méthodologique.

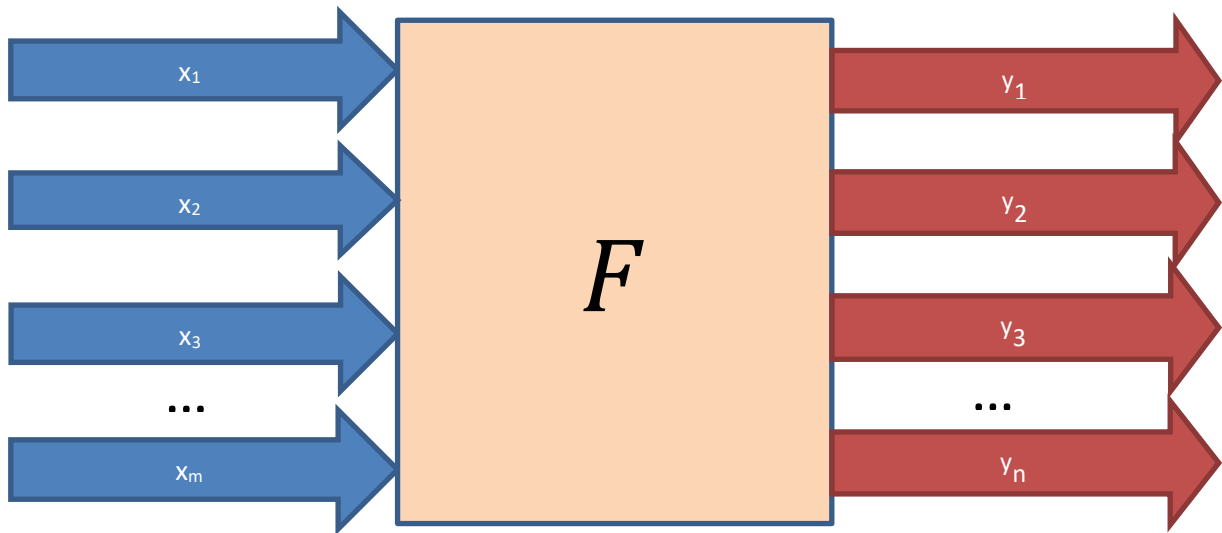


Figure 3: Représentation conceptuelle d'un procédé industriel

Ainsi, il est important de préciser que dans le contexte de ce cadre méthodologique générique, l'extrait considéré n'est pas le produit fini tel que l'on retrouve dans les représentations traditionnelles d'un procédé industriel, mais bien un ensemble de mesures de performance. Cette particularité s'avèrera très importante dans les étapes de modélisation.

### 3.2. Présentation de haut niveau du cadre méthodologique

La Figure 4 illustre le cadre méthodologique et ses différentes étapes.

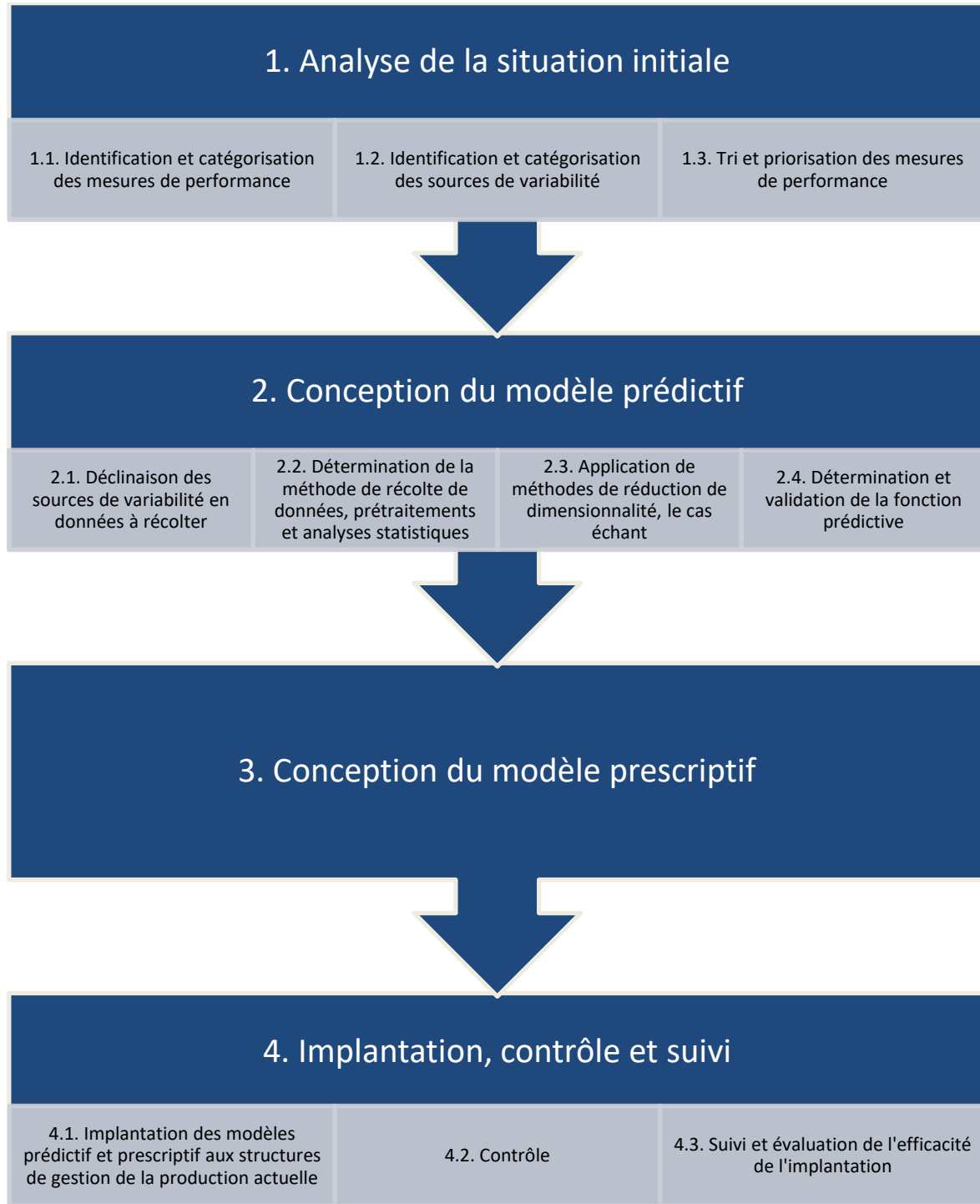


Figure 4: Aperçu du cadre méthodologique



## Étape 1 : Analyse de la situation initiale

### *Partie 1.1 : Identification et catégorisation des mesures de performance*

Cette première étape du cadre a pour objectif l'identification et la catégorisation des mesures de performance importantes. Pour que l'analyse de la situation initiale soit représentative et efficace, il est important de bien identifier et de formaliser les valeurs de l'ensemble  $Y$  que l'on cherche à mieux gérer et à mieux comprendre. L'ensemble  $Y$  peut être composé d'une multitude de  $y_i$  (mesures de performance), d'où l'importance de bien comprendre les relations entre les différents  $y_i$  ainsi que les effets d'interaction. Les effets d'interaction définissent un phénomène mathématique entre les variables dépendantes et indépendantes d'une régression. Un exemple d'un effet d'interaction serait que deux variables indépendantes  $A$  et  $B$  font varier positivement une variable dépendante  $C$ . Cependant, prises individuellement, ces deux variables indépendantes n'ont que très peu d'effet ou aucun effet sur  $C$  (Jaccard and Turrisi, 2003).

### *Partie 1.2 : Identification et catégorisation des sources de variabilités*

Par la suite, cette seconde étape du cadre méthodologique permet d'identifier et de catégoriser convenablement les sources de variabilités. Pour une meilleure formalisation du cadre, on introduit quatre (4) classes distinctes de sources potentielles de variabilité :

- Les caractéristiques de la matière première telles que sa qualité, ses caractéristiques physico-chimiques, etc. ;
- Les paramètres opérationnels tels que les choix d'équipements, le mode de gestion de la production, le nombre d'opérateurs, la vitesse ou le débit d'un équipement, etc. ;
- Les paramètres environnementaux tels que la température ambiante, l'humidité relative de l'air, la saison, etc. ;
- Les éléments aléatoires, une catégorie englobant des variables aléatoires non définies qui explique le caractère stochastique du procédé.

En introduisant ces catégories au modèle original, on obtient, à la Figure 5, la modélisation du procédé.

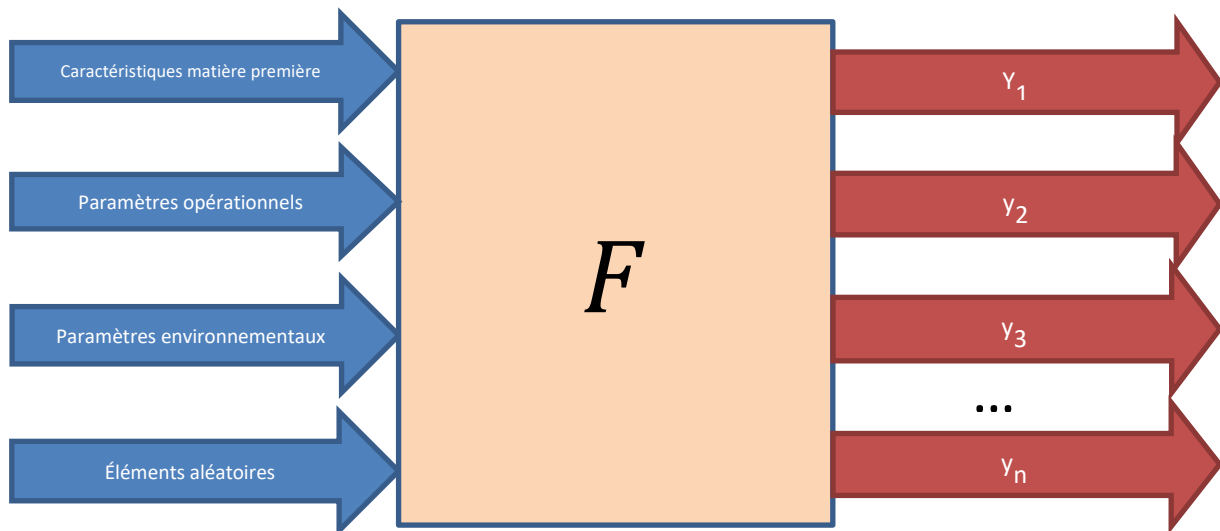


Figure 5: Modélisation du procédé

Ainsi, il est possible de classer chaque intrant  $x_i$  parmi ces quatre catégories d'intrants. De surcroît, on s'intéresse aussi à la nature contrôlable de ces intrants de même qu'à leurs impacts et à leur caractère observable (« observabilité »). Par contrôlable, on entend tout paramètre qui, par une quelconque façon ou méthode, peut être contrôlé à la guise du personnel de l'entreprise avec ou sans délai de réponse. L'observabilité pourrait être définie comme étant le caractère mesurable des paramètres en fonction des technologies et équipements disponibles; il n'est pas obligatoire qu'un paramètre soit mesuré, mais s'il est possible de le faire, on considère ici que ce dernier est observable. À titre d'exemple, la température d'un four industriel en temps réel est un paramètre observable, tandis que la température interne en temps réel d'une planche de bois n'est pas observable. Enfin, pour ce qui est de l'impact, on considère qu'un impact connu correspond au fait que l'entreprise connaît la relation quantitative entre un paramètre et une ou plusieurs mesures de performance.

### *Partie 1.3 : Tri et priorisation des mesures de performance*

À la lumière des étapes précédentes, on effectue un tri et une priorisation des mesures de performance  $y_i$  en fonction des conséquences et des impacts de la variabilité de chacun sur l'entreprise. Il s'agit en fait de cibler les mesures de performance dont la variation est la plus problématique pour le procédé et sur les indicateurs de performance globaux. Le nombre de mesures de performance à prioriser reste cependant variable en fonction des situations et des besoins. De façon générale, ce tri et cette priorisation sont très intuitifs et les organisations ont de prime à bord déjà une certaine priorisation des mesures de performance. Cette partie de la première étape du cadre méthodologique a donc pour but de formaliser cette priorisation, et ce, afin de faciliter l'élaboration des modèles prédictif et prescriptif des étapes subséquentes.

### Application de l'étape 1

La première étape du cadre méthodologique débute par l'identification et la catégorisation des mesures de performances du procédé problématique. Pour ce faire, il est possible d'utiliser un tableau pour colliger les informations. Voici un exemple simple au tableau 1.

Tableau 1: Identification des mesures de performance

$Y_i$	Définition	Unité	Mesurable	Priorité (1 à n)	Pondération
$Y_1$	Mesure de la perte (ex.)	%	Oui/Non	3	20%
...	...	...	...	...	...
$Y_n$	Temps d'arrêt	En minute	Oui/Non	7	5%

En pratique, cette étape consiste à mener, à l'aide de questions, discussions et présentations des objectifs du projet avec le personnel technique et de production de l'entreprise, une investigation sur les différentes sources de variabilité potentielles du procédé manufacturier à l'étude. L'information reçue du personnel aura indubitablement une certaine valeur subjective et c'est pourquoi il est pertinent à ce stade-ci de varier les sources d'information afin d'avoir un portrait le plus global et objectif possible.

Pour ce qui est de l'identification et de la catégorisation des sources de variabilité, l'utilisation de la matrice présentée au tableau 2 pour classer les intrants du procédé qui sont problématiques permet une meilleure compréhension des influences de ces derniers sur les extrants. Tel que mentionné dans les sections portant sur les étapes 1.1 à 1.3, à ce niveau du cadre méthodologique, on s'intéresse à déterminer l'importance relative de chacune des sources de variabilité de même que de séparer ces dernières selon leur caractère contrôlable, observable et selon leurs impacts, plus précisément à savoir si l'entreprise connaît ces impacts sur les processus.

Tableau 2. Matrice de contrôlabilité, observabilité et évaluation des impacts

	Contrôlable		Non Contrôlable	
Observabilité	Valeur observable	Valeur non observable	Valeur observable	Valeur non observable
Impact				
Impact(s) connu(s)				
Impact(s) inconnu(s)				

Ultimement, de par l'application de la méthodologie présentée, la situation désirée est que chacune des sources de variabilité devienne connue et observable (la contrôlabilité peut dépendre de plusieurs facteurs externes à l'organisation). L'utilisation de cette matrice sous-tend l'hypothèse que les sources de variabilités sont, pour la plupart, connues de l'entreprise. Dans la pratique, certaines pourraient ne pas être connues; l'utilisation d'outils tels que les diagrammes de causes à effet, ou plus communément appelés les arbres d'Ishikawa, peut alors aider à cibler de telles sources et à ainsi compléter le tableau.

D'autres outils déjà existants dans la littérature et dans les entreprises peuvent aussi être utilisés pour approfondir la compréhension de la problématique étudiée. On peut penser à la modélisation *SIPOC* (Pyzdek, 2003) qui permet de représenter les intrants et extrants du procédé tout en considérant son environnement. De la même façon, il est possible de concevoir un diagramme des processus ou un diagramme *VSM* (« *Value stream mapping* » ou cartographie de la chaîne de valeur), diagrammes représentant les processus d'affaires de l'entreprise ainsi que les intrants et extrants d'information pour chaque processus (Pyzdek, 2003). En fonction des circonstances, il est possible d'adapter cette sous-étape et d'utiliser d'autres outils reconnus dans l'industrie et dans la littérature tels que les méthodes de réduction de la variabilité relatives au *Lean Six Sigma*. À titre d'exemple, dans l'un des cas de preuve de concept en entreprise effectué, il s'avérait que le problème principal avait des causes directes de même que des causes indirectes.

En ce qui a trait au tri et à la priorisation des sources de variabilité, l'utilisation d'un tableau pour répertorier toutes les sources de variabilité permet d'avoir une vue d'ensemble et de faciliter la priorisation de ces dernières. Le Tableau 3 suivant est proposé pour l'analyse de la situation initiale de l'étape des preuves de concept avec les entreprises. Ce dernier permet de colliger les informations pertinentes reçues sur les différentes sources de variabilité; c'est à l'aide de l'importance relative estimée par rapport à la variation du procédé qu'il est possible de faire un tri et une priorisation. Cette importance relative estimée se base sur une échelle à cinq échelons

(très faible, faible, moyenne, forte, très forte), estimation basée sur l'information obtenue au fil des discussions, des questions et des observations.

Tableau 3: Exemple d'un tableau pouvant servir à classer les différentes sources de variabilité

#	Sources de variabilité	Description	Mesuré (oui ou non)	Contrôlé (oui/non) et comment	Importance relative estimée p/r à la variation	Actions à poser	
						Pour mesurer	Pour contrôler

Pour ce qui est des mesures de performance, l'attribution de pondérations estimées ainsi que l'utilisation d'une analyse Pareto avec ces pondérations peut être un bon moyen de voir visuellement et rapidement quelles mesures devraient être priorisées. Une analyse Pareto se définit comme étant le « processus de classement des opportunités afin de déterminer laquelle des nombreuses opportunités devrait être priorisée en premier » (Pyzdek, 2003). Évidemment, de nombreuses autres méthodes de la littérature peuvent être employées pour effectuer cette priorisation. De surcroît, cette sélection des sources de variabilité permettra aussi de comparer l'état actuel à la situation finale après implantation de la solution. Cette sous-étape est cruciale : elle permettra de bien cibler les objectifs de la solution développée et de mesurer l'efficacité de la méthodologie.

Cette étape du cadre méthodologique se conclut avec la rédaction et la remise d'un résumé d'observations, de même que la transmission de recommandations. Ce résumé a pour objectif, en plus d'un retour sur le contexte, la problématique et la méthodologie employée, de présenter les observations, les discussions faites au fil des visites de l'entreprise ainsi que les analyses effectuées à partir des observations et des informations obtenues. Le résumé termine en proposant des recommandations portant sur les observations et analyses présentées, et ce, dans une optique de transition vers l'industrie 4.0 (utilisation des concepts et des technologies importantes). Ce sont d'ailleurs ces recommandations qui permettront d'orienter la suite de l'application du cadre méthodologique, plus précisément la conception d'un modèle prédictif nécessitant la détermination des méthodes de récolte de données, un prétraitement de l'information et des analyses statistiques.

## Étape 2 : Conception du modèle prédictif

### *Partie 2.1 : Déclinaison des sources de variabilité en données à récolter*

D'emblée, après la complétion de la première étape de ce cadre méthodologique, il est nécessaire de décliner les sources de variabilité en données réelles à récolter. Plus précisément, les sources de variabilité identifiées et caractérisées à l'étape précédente ne sont en fait que des causes conceptuelles de la variabilité, mais ces dernières doivent tout de même être traduites en données pratiques et tangibles qu'il est possible de récolter. C'est ainsi à cette étape que la transition du conceptuel à la pratique se fait. La première étape étant une analyse conceptuelle de la problématique, les parties subséquentes de cette étape présentent plutôt des activités pratiques et concrètes à réaliser.

### *Partie 2.2 : Détermination de la méthode de récolte de données, prétraitement et analyses statistiques*

Cette seconde partie de la deuxième étape consiste à répertorier les besoins en termes d'acquisition et de stockage de données. Tout d'abord, il est nécessaire de recenser l'existant. Par exemple, dans cette étape, on s'intéresse aux capteurs, aux mesurages, aux inspections ou à toute autre information captée par une technologie ou par un opérateur à des fins de stockage.

Au niveau du stockage, on s'intéresse aux historiques de production, qu'ils soient stockés dans une base de données (relationnelles ou non) ou dans tout autre support. Une représentation visuelle (diagramme du processus d'affaires, par exemple) peut aider à bien comprendre les flux d'information et les données captées. L'accès à des données fiables et pertinentes relève d'une grande importance : la méthodologie repose sur l'information que l'on peut tirer des données de production.

À ce stade-ci du projet, il est pertinent de se questionner à savoir s'il est préférable d'utiliser les historiques de production pour effectuer les modélisations prédictive et prescriptive qui suivront ou bien de récolter des données expérimentales, c'est-à-dire des données spécifiquement récoltées pour la modélisation. Une autre possibilité serait de faire les deux, c'est-à-dire utiliser les historiques de production et enrichir ces dernières avec des données expérimentales.

Suite à la l'analyse de l'existant, tout en considérant ce qui a été fait à l'étape 1 du cadre méthodologie, il s'agit maintenant de proposer des modifications à apporter, s'il y a lieu, afin de répondre aux besoins en termes d'acquisition et de stockage de données et d'être ainsi en mesure de poursuivre le processus suggéré à l'intérieur du cadre méthodologique.

### Partie 2.3 : Application de méthodes de réduction de dimensionnalité, le cas échéant

Par la suite, tout dépendant des connaissances que l'entreprise possède sur le procédé problématique en question et les impacts des intrants sur ce dernier, il peut s'avérer nécessaire d'appliquer une méthode de réduction de dimensionnalité. Les méthodes de réduction de dimensionnalité permettent, comme le nom l'indique, de réduire la dimension du problème afin de simplifier le problème. À titre d'exemple, il peut s'agir d'une méthode permettant de sélectionner un sous-ensemble d'intrants qui ont un impact significatif sur les extrants (« *feature selection* ») ou bien d'une analyse des composantes principales (ACP), une méthode qui permet de transformer des variables corrélées entre elles en de nouvelles variables non corrélées (Ghods, 2006). Pour faire un lien avec le domaine de l'optimisation, la résolution de programmes mathématiques comportant des vecteurs de variables à haut niveau (avec plusieurs dimensions) devient très complexe, générant ainsi le phénomène du « fléau de la dimension » (« *curse of dimensionality* » en anglais) (Shapiro et al., 2009).

### Partie 2.4 : Détermination et validation de la fonction prédictive $F'$

La prochaine étape implique la conception, la vérification et la validation de la fonction prédictive  $F'$ . On définit cette fonction comme étant une fonction qui approxime la fonction  $F$  présentée antérieurement. Considérant que cette fonction est inconnue, on tente de la substituer en créant une approximation fiable, appelée  $F'$ . De plus, pour la suite des choses, une distinction sera faite entre les sources de variabilités contrôlables et les sources de variabilités incontrôlables (à l'exception des éléments aléatoires). Une représentation graphique de la fonction  $F'$  avec ce nouvel élément est présentée à la Figure 6.

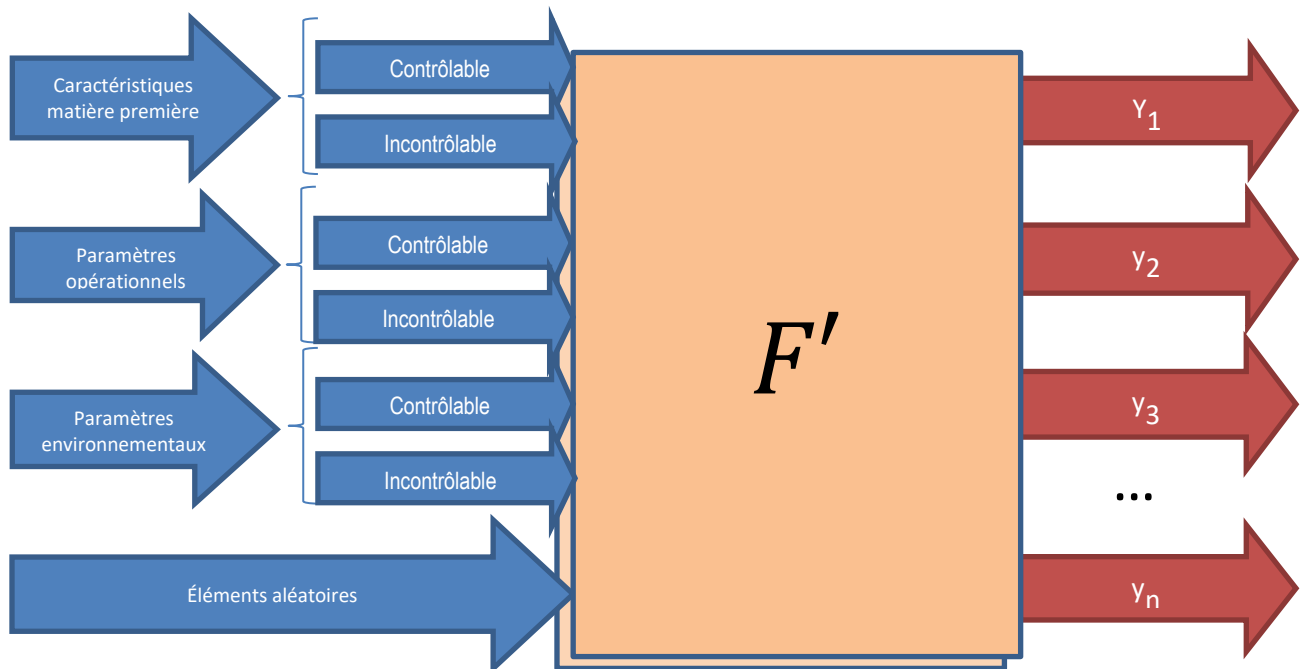


Figure 6: Aperçu du modèle prédictif

L'objectif de cette fonction est d'être en mesure de prédire les mesures de performance à partir des sources de variabilité présentées précédemment qui ont été déclinées en variables. Parmi les différentes méthodes pour concevoir ce système prédictif, on peut penser à (Kleijnen, 2017)<sup>1</sup> :

- des modèles de régression multivariée (aussi appelés modèles de régression multiple) ;
- des méthodes de machine d'apprentissage (réseaux de neurones artificiels, machine à vecteurs de support, arbres de décision, etc.) ;
- des méthodes d'estimation linéaire telle que le krigeage (« *Kriging modeling* »).

Ces techniques sont connues et il existe une littérature étendue sur ces sujets. C'est pourquoi les méthodes de modélisation prédictive ne seront pas développées davantage dans ce mémoire, à l'exception des méthodes qui seront utilisées lors des preuves de concept présentées au chapitre 4.

Dans le contexte de l'industrie 4.0, la méthode développée devrait être en mesure de s'adapter par elle-même, c'est-à-dire de relancer une phase d'apprentissage ou de calcul afin de mettre à jour les paramètres de la méthode, en fonction des dérives ou des modifications du procédé. Cette étape permet donc à l'entreprise d'avoir une meilleure compréhension du procédé problématique, de sa variabilité intrinsèque ainsi que du lien qui relie les intrants aux extrants. Suite à la détermination de la bonne méthode et de sa mise au point, il est primordial d'effectuer une vérification et une validation de la méthode. Si c'est possible, il apparaît pertinent de tester dans le monde réel la méthode développée pour la problématique, lorsque la complexité de cette dernière permet une implantation rapide. Avant de poursuivre à l'étape 3, l'utilisateur du cadre méthodologique doit avoir la certitude de la validité du modèle prédictif.

### *Application de l'étape 2*

Tel que présenté ci-haut, la première tâche à effectuer dans cette seconde étape du cadre méthodologique est de décliner les sources de variabilité identifiées et priorisées dans l'étape précédente en données réelles qui peuvent ou pourraient être récoltées. Un outil proposé pour faciliter l'accomplissement de cette tâche est l'utilisation d'un tableau recensant les différentes sources de variabilité et quelles variables pourraient y être associées. Le tableau 4 présente un exemple d'un tel outil.

---

<sup>1</sup> Liste non-exhaustive



Tableau 4: Exemple de tableau utilisé pour recenser les différentes variables associées aux sources de variabilité

Source de variabilité	Nom de la variable associée	Description	Fichier ou stockage source	Format	Calcul, formatage ou nettoyage à faire
Source de variabilité A	Variable X	...	Fichier W	Numérique	...
Source de variabilité A	Variable Y	...	Fichier V	Texte	...

Il est à noter qu'une source de variabilité peut être encodée avec plusieurs variables. Cette recension des variables devrait normalement être effectuée en collaboration avec un expert du procédé ou de la problématique ciblée par l'application du cadre. De plus, certaines variables peuvent être le résultat d'un calcul ou d'une équation fait à partir d'autres variables.

Une fois cette recension terminée, il est nécessaire de consolider les données au même endroit pour former une matrice qui pourra être manipulée et utilisée pour les étapes de modélisation qui suivront. C'est aussi à ce moment que la décision doit être prise à savoir si sont utilisées des données historiques ou des données récoltées spécifiquement via un plan d'expérimentation ou des équipements spécialisés de prise de données. Parallèlement à cette consolidation, il est nécessaire d'effectuer un formatage et un nettoyage des données. Un exemple de ce processus de formatage et de nettoyage sera détaillé dans la section « Étape 2 : Modélisation prédictive » de la première étude de cas du chapitre 4 de ce mémoire.

Une fois la matrice finale de données constituée, une période exploratoire des différentes méthodes de modélisation prédictives doit être menée. Dans ce sens, il existe un grand nombre de méthodes de modélisation prédictive dans la littérature. La présentation de ces méthodes n'étant pas l'objectif de ce projet de recherche, il est possible de se référer à Tufféry (2011) pour davantage d'information.

### Étape 3 : Conception du modèle prescriptif

Suite à la conception du modèle prédictif, la prochaine étape consiste à la mise au point d'un modèle prescriptif, c'est-à-dire d'un modèle qui proposera des plages de valeurs optimales ou quasi optimales pour les variables de contrôle ou qui permettra d'identifier des pistes d'améliorations au procédé ou aux produits fabriqués. Dans le cadre méthodologique actuel, on considère que si une méthode de modélisation prédictive existe (étape 2), il existe aussi une méthode de modélisation prescriptive. Selon la représentation de la Figure 6, ces variables de contrôle correspondent aux intrants contrôlables, qu'ils soient catégorisés comme des caractéristiques de la matière première, des paramètres opérationnels ou des paramètres environnementaux.

L'objectif de cette étape est en quelque sorte de déterminer une fonction « inverse » de celle développée à l'étape précédente. La Figure 7 résume l'idée derrière cette étape :

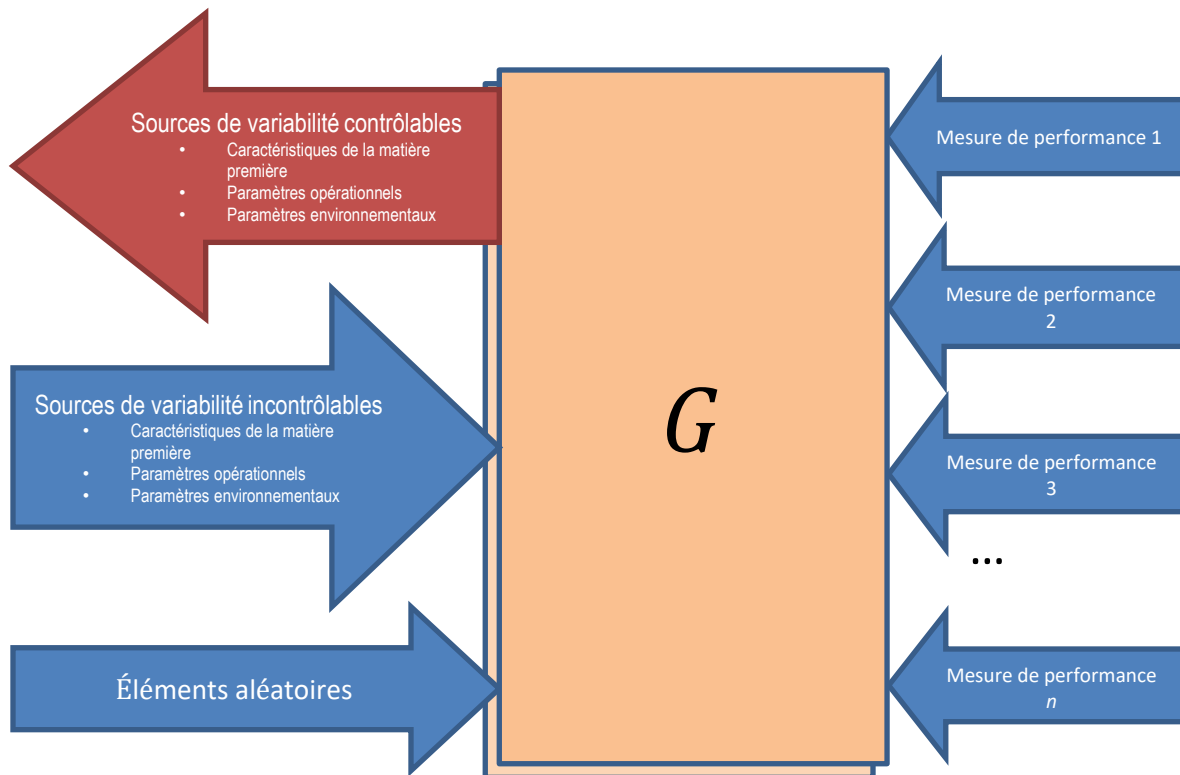


Figure 7: Aperçu du modèle prescriptif

Pour ce qui est de la méthode à employer afin de concevoir ce modèle prescriptif, elle dépendra grandement de la méthode choisie pour l'étape précédente (modèle prédictif). Selon les méthodes, la littérature présente plusieurs cas typiques et similaires qui clarifieront cette étape de la méthodologie. De plus, ce modèle devrait, à l'instar du modèle prédictif de l'étape 2, être adaptatif en permettant des réapprentissages ou calculs de mise à jour dans les situations qui l'exigent. Puisque le modèle prescriptif dépend du modèle prédictif, cette mise à jour devrait s'effectuer en même temps que ce dernier. De la même façon qu'avec le modèle prédictif, il est primordial d'effectuer les vérifications et validations qui s'imposent avec ce modèle prescriptif. L'étape suivante qui concerne l'implantation donnera l'opportunité de tester en temps réel ce nouveau modèle.

### Application de l'étape 3

Dans un même ordre d'idées, une fois le modèle prédictif conçu, il est possible de débiter la modélisation prescriptive telle que décrit précédemment. Concrètement, cette troisième étape du cadre méthodologique consiste à identifier, à partir du modèle prédictif ainsi que de ses résultats, des pistes d'améliorations ou bien d'identifier un système permettant de proposer des actions qui augmentent l'efficacité du procédé et la qualité des produits.

Une fois les différentes options identifiées, il est possible d'effectuer une estimation des efforts et des ressources nécessaires pour chaque option pour ensuite mener une analyse comparative des options. À l'aide de cette analyse, une sélection doit être faite. Le développement de la solution ou du modèle prescriptif peut ensuite débiter.

## Étape 4 : Implantation, contrôle et suivi

### *Partie 4.1 : Implantation des modèles prédictif et prescriptif aux structures de gestion de la production actuelle*

Par la suite vient l'étape d'implantation des solutions développées jusqu'à maintenant. Comme précisé plus tôt dans cette méthodologie, la conception et l'utilisation de diagrammes logiques et de schémas représentant la structure technologique de l'existant permet de mieux comprendre les interactions que cette dernière aura avec la nouvelle solution développée. Il est particulièrement important que les deux systèmes (ancien et nouveau) soient interoperables ou intégrés. C'est aussi à cette étape que s'effectueront les vérifications et validations finales du nouveau système en temps réel avec le procédé problématique. S'il s'agit d'une solution logicielle tel un outil d'aide à la décision, ce dernier doit être intégré parfaitement au processus de contrôle du procédé afin qu'il n'y ait pas de perte d'efficacité. Cette nouvelle mesure d'efficacité fera l'objet d'une étude approfondie à la dernière sous-étape du cadre méthodologique.

### *Partie 4.2 : Contrôle*

Suite à l'implantation de la solution, cette sous-étape a pour objectif d'en faire le contrôle et d'apporter toute modification nécessaire au bon fonctionnement (« *fine-tuning* » ou peaufinage). Ces modifications devront être documentées afin d'assurer une continuité du fonctionnement de la solution.

### *Partie 4.3 : Suivi et évaluation de l'efficacité de l'implantation*

Dans cette dernière sous-étape du cadre méthodologique, la rétroaction sur l'efficacité de la nouvelle solution ainsi que les livrables post-implantations sont abordés. Tout d'abord, il est important d'effectuer une étude comparative sur l'efficacité et la productivité du procédé avec l'utilisation du nouveau système prescriptif versus l'ancien système via l'utilisation d'historiques de production. Cette rétroaction comparative doit être analysée et calculée pour chacune des mesures de performance identifiées à la première étape du cadre.

Enfin, pour conclure, il ne reste plus qu'à effectuer un suivi continu du nouveau système. Afin de faciliter le suivi à long terme de la solution développée et son appropriation par les experts du procédé industriel, il est important de bien documenter le fonctionnement des modèles et de la solution obtenue de même que leur utilisation.

#### *Application de l'étape 4*

Afin de mieux définir les interactions qu'aura le système développé lors de la modélisation prescriptive avec les systèmes actuels, une modélisation de l'architecture du système de gestion de la production ou du système de contrôle du procédé devra réalisée, et ce, advenant que le résultat de la modélisation prescriptive soit un système (logiciel, application, technologique, etc.). Ce faisant, il est possible de refaire une modélisation de cette architecture telle qu'elle devrait être en intégrant ce nouveau système.

De plus, que le résultat de la modélisation prescriptive soit un système ou des pistes d'amélioration de produits et procédés, il est possible de s'inspirer des étapes du cycle de vie d'un projet classique pour définir la marche à suivre de l'implantation (Lock, 2007) :

1. définir le projet d'implantation ;
2. préparer et planifier ;
3. concevoir (cette étape devrait cependant être complétée en partie ou en totalité) ;
4. acquérir les ressources nécessaires ;
5. faire le projet ;
6. compléter le projet et transfert du projet (incluant un retour sur les objectifs de performance et un suivi au fil du temps de ces derniers).

Pour terminer, un rapport sur le déroulement du projet doit ensuite être rédigé. Ce rapport devrait inclure :

- un résumé de la nouvelle structure technologique ;
- une description des méthodes employées ;
- un mode d'emploi du nouveau système ;
- les gains obtenus avec les nouveaux modèles prédictif et prescriptif ;
- les difficultés rencontrés dans le déroulement du projet ;
- ce qui reste à faire dans le cadre du projet, s'il y a lieu ;
- des points d'amélioration pour le cadre méthodologique et pour les modèles développés, s'il y a lieu.

## Cadre méthodologique itératif

Bien que les étapes du cadre méthodologique aient été présentées de façon linéaire et directe, en pratique, l'application du cadre est un processus itératif comportant plusieurs boucles de rétroaction. Cette caractéristique a pour objectif d'accroître la flexibilité du cadre en fonction des problématiques rencontrées. À titre d'exemple, la deuxième partie de l'étape 2 du cadre (Partie 2.2 : Application de méthodes de réduction de dimensionnalité, le cas échéant) peut s'avérer inutile tant qu'une modélisation n'a pas été tentée avec les données à portée de main (on doit modéliser pour découvrir qu'il y a trop de variables dans le modèle). De la même façon, tel que représenté dans la Figure 8, si une modélisation prédictive ne permet pas d'obtenir les taux de prédiction désirés en raison du manque de données sur les sources de variabilité (ou tout simplement que des « mauvaises » données ont été utilisées), il est possible de revenir en arrière à la première étape du cadre pour refaire un tri et une priorisation de ces sources de variabilité et ainsi obtenir davantage de données.

Un parallèle pourrait être fait avec le processus unifié dans le domaine du développement logiciel. En effet, ce dernier est caractérisé comme étant flexible, itératif et ouvert à l'utilisation de pratiques empruntées à d'autres méthodes (Larman, 2004), à l'instar du cadre méthodologique présenté dans ce troisième chapitre. Tout comme le processus unifié qui présente un ensemble d'outils pour la modélisation des cas d'application et le développement des objets du logiciel, le cadre méthodologique abordé dans ce mémoire propose un ensemble de méthodes pour modéliser les procédés variables et ainsi mieux comprendre ces derniers.

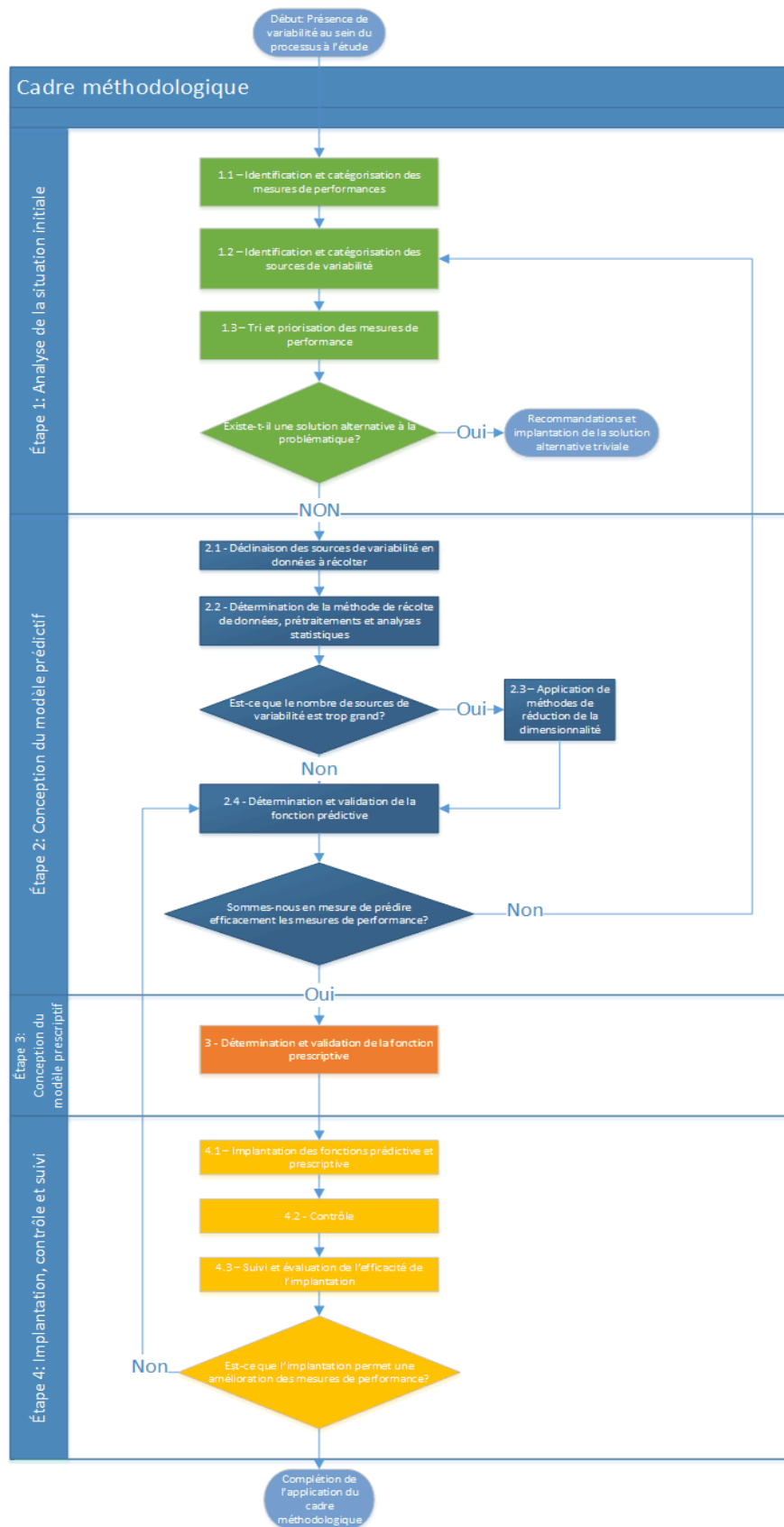


Figure 8: Diagramme représentant le cadre méthodologique ainsi que ses boucles rétroactives

## Chapitre 4 Mise en œuvre en contexte industriel – Cas Duvaltex

Dans les deux prochains chapitres, les preuves de concept effectuées afin de valider le cadre méthodologique seront présentées et discutées. Pour faciliter le déroulement, la preuve de concept a été séparée en deux phases, soit une première où seule l'étape 1 du cadre méthodologique est testée afin de valider que l'analyse de l'existant permet bien d'avoir une image fiable de la situation de départ et d'obtenir une meilleure compréhension des sources de variabilité de même que des mesures de performances, et une deuxième phase où les modélisations prédictive et prescriptive sont conçues et implantées. Les preuves de concepts ont été conduites sous la forme d'applications en entreprise à l'aide de partenaires industriels et la suite de ce chapitre présentera donc la première étude de cas qui a été faite en collaboration avec l'entreprise Duvaltex.

Duvaltex est un leader en textile à usage commercial (institutionnel et industriel) basé dans la région de la Capitale-Nationale (siège social), dans la région de Beauce-Appalaches (3 usines) ainsi qu'aux États-Unis où l'entreprise possède quelques installations fournissant les marques acquises au fil du temps. Sa mission se définit comme suit : « Chez Duvaltex, notre passion est de concevoir, de fabriquer et d'offrir des solutions textiles novatrices qui améliorent l'environnement » (Duvaltex, 2018). Pour ce qui est de ses valeurs, Duvaltex valorise l'approche client, l'innovation, le travail d'équipe et l'engagement (Duvaltex, 2018). L'entreprise effectue elle-même la majeure partie des processus de fabrication (acquisition du fil, tissage, teinture, livraison et vente). Dans le cas de ce projet de maîtrise, la problématique étudiée se situe au niveau du procédé de teinture des tissus, soit à l'usine de Saint-Victor dans la région de Beauce-Appalaches.

La problématique rencontrée dans le cas Duvaltex porte sur la répétabilité et la standardisation de la couleur obtenue grâce au procédé de teinture. Avec une panoplie de combinaisons de styles/couleurs de tissus, reproduire la même couleur sur deux lots de tissu peut s'avérer difficile.

La mesure d'une couleur peut être faite selon deux méthodes : une comparaison visuelle avec la couleur voulue (c.-à-d. avec le « master », la couleur de référence) en suivant certaines normes ou avec l'utilisation d'un spectrophotomètre. La première méthode, qui est actuellement la méthode préconisée et décisive, est largement subjective d'un inspecteur à un autre. En revanche, la deuxième méthode ne semble pas être parfaitement fiable due aux variations de mesure pouvant être induites par le style du tissu (le relief du tissu par exemple aura un impact sur la mesure).

Le fait de produire un lot de tissu dont la couleur n'est pas conforme à ce que le client veut peut être très dispendieux pour l'entreprise. À titre d'exemple, on peut penser à un client qui, à la suite d'une commande précédente, désire avoir la même couleur et le même tissu pour fabriquer des séparateurs. Si la couleur obtenue

avec le second lot ne correspond pas à celle du premier lot, c'est-à-dire si la différence est visible à l'œil, le lot ne peut être accepté. Les options en cas de non-conformité sont limitées : reteindre le tissu avec une couleur plus foncée ou, si cela n'est pas possible, vendre le tissu tel quel avec une marge de profit nettement moindre, voire inexistante. Aucun chiffre ne sera avancé quant au montant perdu chaque année à cause de lots de tissu dont la couleur n'est pas conforme; cependant, le montant représente une somme très importante et la moindre amélioration devrait avoir des impacts importants sur la performance de l'entreprise.

#### **4.1. Étape 1 : Analyse de la situation initiale**

Jusqu'à maintenant, les efforts faits par l'entreprise afin de remédier au problème consistent principalement à mieux contrôler la matière première et à établir des standards au niveau de cette dernière ainsi qu'au niveau de certains paramètres opérationnels. Aussi, il y a une volonté de récolter et de sauvegarder les données de production afin de pouvoir les analyser ultérieurement et en dégager des tendances et des pistes d'amélioration.

Ainsi, l'application de la première étape du cadre méthodologique a permis de relever des conclusions qui ont permis d'orienter les recommandations soumises à l'organisation. Tel qu'il est mentionné au chapitre précédent, la procédure suivie lors des visites et des rencontres se devait d'être flexible : en fonction de la complexité du problème et du nombre de sources de variabilité à l'étude, le temps nécessaire avant d'avoir une compréhension adéquate peut varier d'une situation à l'autre. Cependant, certaines activités de la procédure resteront constantes d'un problème à l'autre. Voici ces activités que l'on retrouvera dans la procédure suivie avec Duvaltex :

1. Présentation du projet de recherche ;
2. Discussion sur la problématique rencontrée par l'entreprise ;
3. Questionnement du personnel technique et des opérateurs par rapport à la problématique et par rapport aux sources de variabilité ;
4. Confirmation des observations et des informations récoltées ;
5. Analyse qualitative des observations et informations recueillies (classement de l'information et élaboration de diagrammes et tableaux pour mieux comprendre la problématique étudiée) ;
6. Élaboration d'un résumé d'observations et de recommandations préliminaire ;
7. Révision du résumé d'observations et de recommandations à l'aide de la rétroaction de l'entreprise partenaire ;



8. Élaboration du résumé d'observations et de recommandations finales puis transmission à l'entreprise partenaire.

Plus précisément, la visite a débuté avec une courte présentation du projet de recherche de même que de ses objectifs. Durant cette présentation, les questions et discussions étaient encouragées afin de s'assurer que le partenaire industriel ait une compréhension approfondie du projet de maîtrise. Suite à la présentation, une discussion spécifiquement sur la problématique rencontrée s'en est suivie. Le but de cette discussion était de confirmer que cette dernière problématique cadre bel et bien avec le projet de recherche. À titre de contre-exemple, une problématique qui nécessiterait une solution triviale telle qu'un simple ajustement de machine ou le changement d'une matière première serait à éviter. Ensuite, une visite du plancher de production a eu lieu afin de voir la réalité de production et être en mesure de relier la problématique expliquée précédemment à des phénomènes réels. Par la suite, une série de discussions ainsi qu'une période de questions avec des opérateurs et le personnel technique ont été préconisées pour aller chercher le plus de détails pertinents possible. Il s'en est suivi une activité de confirmation des observations et des informations récoltées afin de s'assurer que les informations utilisées pour la poursuite de l'analyse de la situation initiale étaient justes et représentatives de la problématique étudiée.

Une fois les observations confirmées, une analyse qualitative de ces observations a suivi. Durant cette analyse, l'objectif principal était de classer l'information recueillie par l'utilisation de tableaux et d'élaborer des diagrammes, et ce, afin de mieux comprendre la problématique étudiée. L'information recueillie de même que les différents tableaux et diagrammes utilisés pour comprendre la problématique ont par la suite été colligés dans un résumé d'observations révisé par le partenaire, pour enfin mener à l'élaboration d'un rapport final. En plus de présenter les constatations perçues jusqu'à maintenant, ce rapport comprend aussi un ensemble de pistes de recommandation visant à aider le partenaire à minimiser la variabilité de son procédé et à améliorer la performance de façon générale.

Il est important de considérer que cette première étape du cadre méthodologique consiste à une évaluation qualitative de la situation existante et à émettre des recommandations quant à la poursuite du projet (recommandations de solutions d'implantation, procéder à la seconde étape du cadre, etc.). Dans ce sens, la présente section introduit quelques points observés lors des visites ainsi que des diagrammes et tableaux permettant de faciliter la compréhension de la problématique. Il est à noter que les analyses et représentations qui suivent sont basées uniquement sur les discussions qui ont eu lieu de même que sur les réponses aux questions posées au personnel technique et aux opérateurs de l'entreprise.

Suite à ces observations et discussions, il apparaît que les caractéristiques de l'eau utilisée pour le procédé de teinture ainsi que les caractéristiques physiques du tissu utilisé sont deux pistes d'exploration critiques au

meilleur contrôle de la couleur finale du tissu. D'après les observations effectuées, il semble que les autres matières premières ainsi que la plupart des opérations sont relativement bien contrôlées et mesurées.

Dans le cas de l'eau, l'entreprise utilise un bassin de décantation afin d'épurer l'eau et d'y retirer les matières étrangères puisque cette dernière est puisée directement de la rivière Le Bras. En plus de contrôler l'acidité de l'eau, l'entreprise utilise des agents coagulants ainsi que des polymères afin de forcer l'agglutination des corps étrangers pour ainsi les faire décanter dans le fond du bassin. Cependant, le désavantage lié à ce procédé est que des traces de ces agents coagulants ainsi que des traces de polymères semblent demeurer lors du procédé de teinture. Ceci fait en sorte d'augmenter la dimension des molécules de teinture en les coagulant, sachant que plus les molécules de teinture sont de fortes dimensions, plus elles ont de la difficulté à pénétrer le tissu. Ceci est un exemple de lacune présente au niveau du contrôle de la qualité de l'eau utilisée dans la teinture, donc une piste de solution à explorer pour aider à diminuer la variabilité.

Pour ce qui est du tissu, il semble que la microstructure du tissu pourrait avoir un impact important sur la couleur finale du tissu. Cette microstructure se sépare en deux types de structure : la structure cristalline et la structure amorphe. Plus le polymère du tissu est recyclé, plus la proportion de la structure cristalline augmente. Cependant, seule la structure amorphe absorbe la teinture lors du procédé de teinture. Donc, plus un tissu est recyclé, moins il absorbera la teinture, modifiant ainsi la couleur finale. En plus de la microstructure, il semble que d'autres caractéristiques telles que la finesse du fil et sa rétractation ont un impact sur la couleur finale de teinture. Cette piste est aussi un élément important que l'entreprise aimerait pouvoir caractériser.

Enfin, les autres composantes utilisées dans le procédé de teinture (additifs chimiques, paramètres de contrôle du procédé, etc.) ont un impact possible qui est à évaluer. Entre autres, la fiabilité des équipements de distribution (« dispenser ») de même que les risques de contamination reliés à la réutilisation des contaminants sont des pistes d'exploration qui pourraient présenter des améliorations.

Afin de bien représenter la compréhension de la problématique, un diagramme des processus a été bâti. Ce diagramme (Figure 9) présente, en plus des étapes du procédé de teinture, les flux d'information et de données pour chacune de ces étapes.

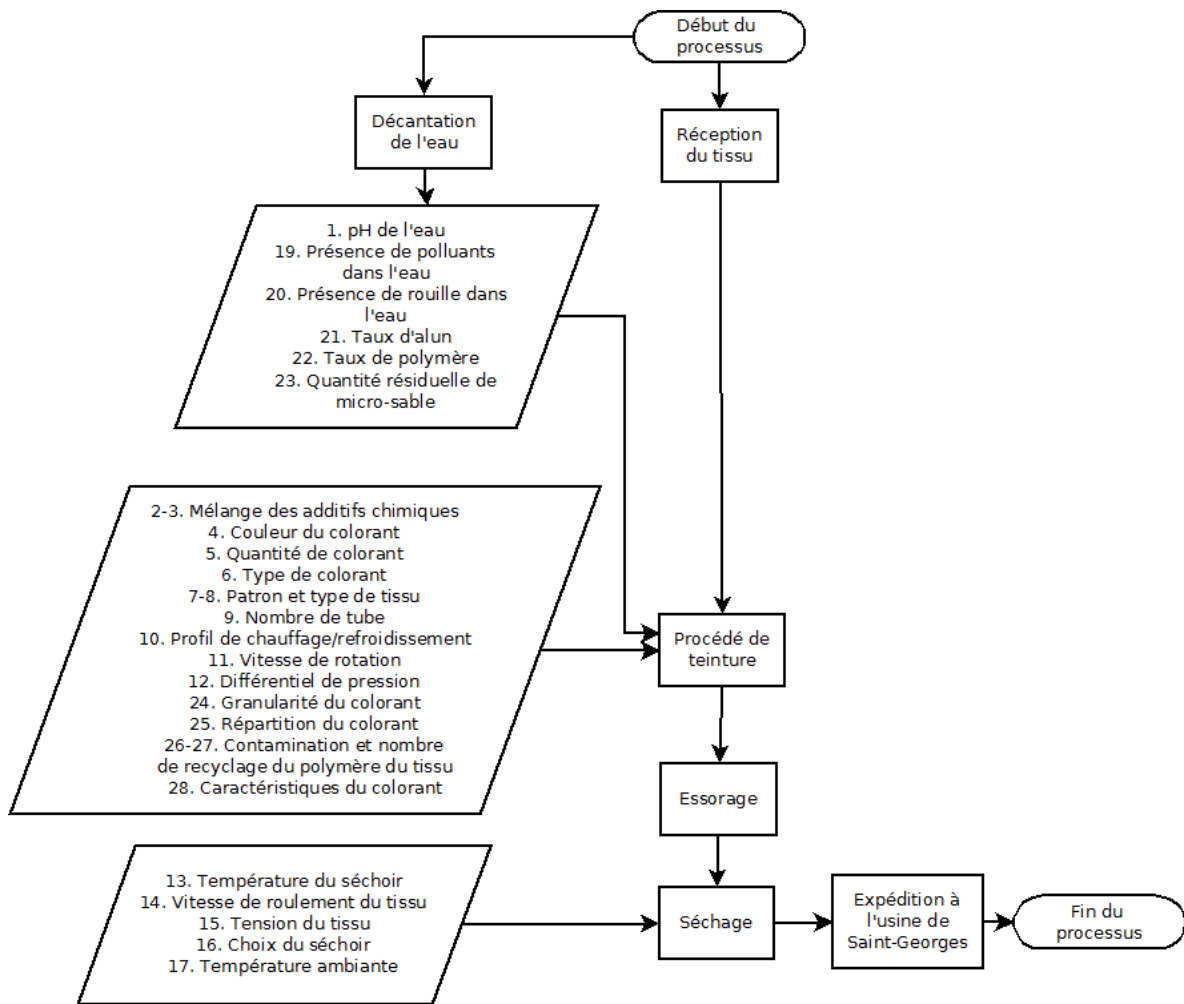


Figure 9: Diagramme des processus avec flux d'information.

Dans ce diagramme, les rectangles représentent les étapes du processus tandis que les parallélogrammes représentent les entrées ou sorties d'information ou de données. L'idée est de représenter les étapes critiques du procédé qui nécessitent une quantité importante de paramètres à contrôler.

Un autre diagramme (Figure 10) qui est utile à la compréhension de la problématique consiste à utiliser un diagramme de causes et effet (plus communément appelé diagramme ou arbre d'Ishikawa). Il permet de mettre en perspective les sources de variabilité en fonction des 6M : la main d'œuvre, les machines, les méthodes, le milieu, les mesures et le matériel. Pour simplifier la représentation, certaines sources de variabilité ont été regroupées dans des groupes logiques.

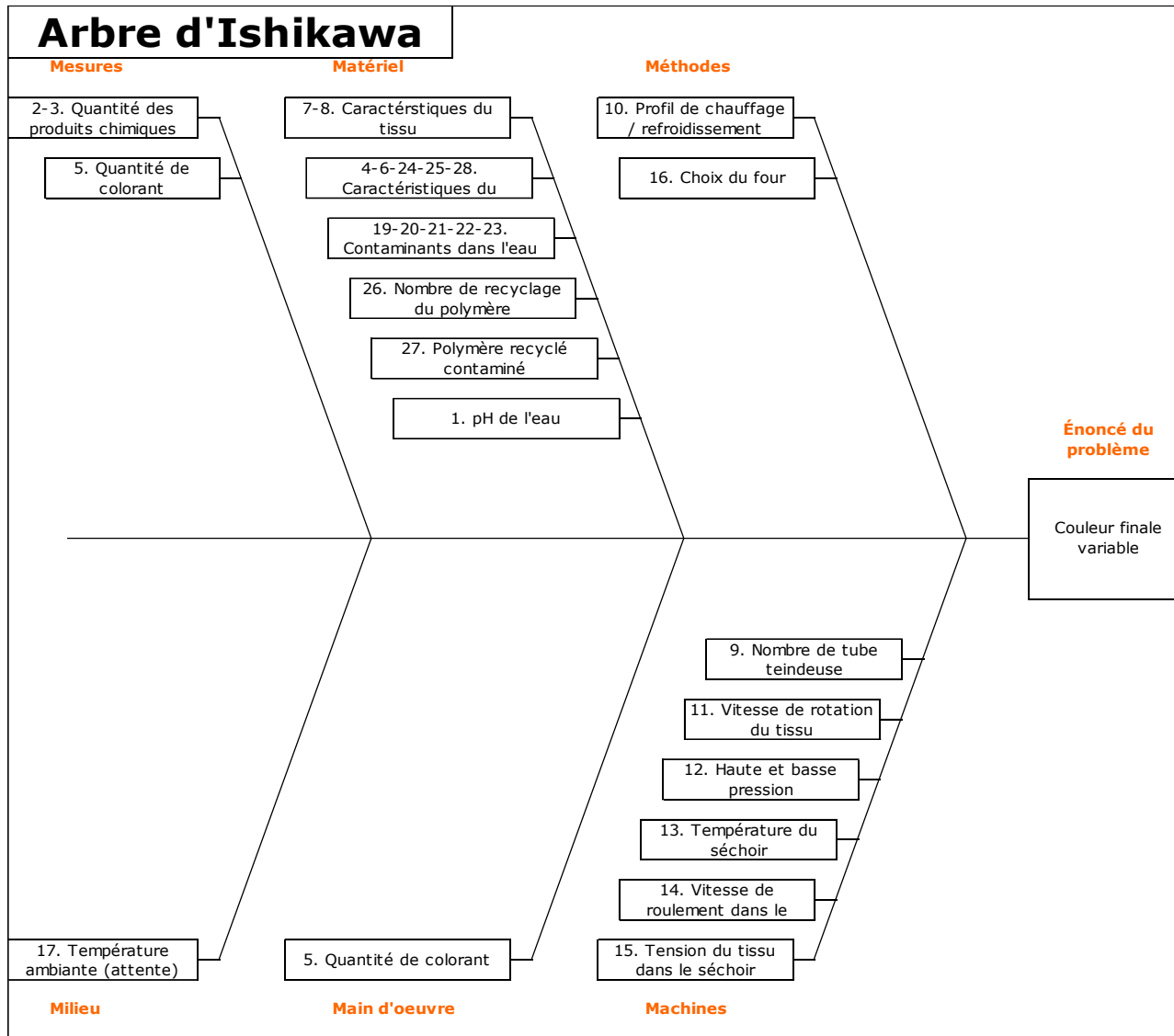


Figure 10: Représentation des sources de variabilité dans un diagramme de causes et effet pour le cas Duvaltex.

Avec ce diagramme, on remarque que ce sont les catégories de sources de variabilité « matériel » et « machine » qui comportent le plus de ces sources de variabilité. Il est à noter que cette représentation ne traite aucunement de l'importance relative de ces sources de variabilité. Dans ce sens, un groupe de sources de variabilité présentant moins de sources peut tout de même avoir un impact plus important qu'un autre groupe.

Le tableau 5 présente la liste des sources de variabilité identifiées lors des visites et rencontres avec les employés de Duvaltex. Sans dire qu'elle est exhaustive, elle contient les principales sources qui peuvent affecter la couleur finale du tissu, bien que beaucoup d'entre elles ne soient pas problématiques, puisqu'un processus de contrôle y est implanté. Pour l'importance relative évaluée par rapport à la variation, l'échelle suivante sera utilisée : très faible, faible, moyenne, forte et très forte.

Il est à noter qu'il s'agit d'une estimation préliminaire qualitative, l'objectif étant d'abord d'identifier des pistes à prioriser. Cette évaluation est basée sur les discussions et indications du personnel technique et du personnel de gestion de l'entreprise. Le terme « teindeuse » est utilisé pour faire référence aux équipements utilisés lors du procédé de teinture puisqu'il s'agit du terme couramment utilisé. De plus, il est à noter que la source de variabilité n°18 a été retirée suite aux propositions des dirigeants Duvaltex.

Tableau 5: Présentation des sources de variabilité de même qu'une courte description, son état de mesure et contrôle et son importance relative évaluée qualitativement.

#	Sources de variabilité	Descriptions	Mesuré (oui ou non)	Contrôlé (oui/non) et comment	Importance relative estimée p/r à la variation	Actions à poser	
						Pour mesurer	Pour contrôler
1	pH de l'eau	Il s'agit de la mesure d'acidité de l'eau.	Oui, avant chaque lot	Oui, à l'aide d'additifs chimiques.	Forte	Mettre en place des prises de mesure plus fréquentes.	Augmenter la fréquence de contrôle.
2	Composition du mélange d'additifs chimiques	Présence ou non des différents additifs ajoutés lors du procédé de teinture.	Oui	Oui, chaque recette de teinture a son mélange d'additifs chimiques selon le résultat voulu.	Faible	N/A	N/A
3	Quantité de chaque additif	Quantité de chaque additif dans le mélange, proportion de chaque additif par rapport aux autres.	Oui	Oui, à l'aide de débitmètres ou pesées manuellement, selon l'additif.	Forte	Une amélioration serait d'installer des débitmètres pour tous les additifs et d'investiguer la fiabilité des systèmes de distribution. Il appert que ce dernier pourrait entraîner des variations de la quantité d'additifs ajoutés.	Une amélioration serait d'installer des débitmètres pour tous les additifs et d'investiguer la fiabilité des systèmes de distribution. Il appert que ce dernier pourrait entraîner des variations de la quantité d'additifs ajoutés.
4	Couleur des colorants	Il s'agit de la sélection des couleurs de colorant pour le procédé de teinture.	Oui	Oui, chaque recette de teinture a son mélange de colorants à utiliser. Cependant, l'entreprise « subit » la couleur puisque c'est le client qui décide du patron qu'il veut. La couleur est connue avant une recette.	Faible	N/A	N/A
5	Quantité de colorant	Il s'agit, à l'instar des produits chimiques, de la quantité de chaque colorant utilisé.	Oui	Oui, à l'aide d'une balance, un opérateur mesure au millième de gramme chaque colorant du mélange de colorant.	Moyenne	N/A	Un projet en cours consiste à utiliser un sac de plastique hydrosoluble pour minimiser les pertes de colorant. La complétion de ce projet pourrait avoir des impacts positifs.
6	Type de colorant	Il s'agit de la marque du colorant, du fournisseur, du format, etc.	Oui	Oui, par la sélection du type lors de l'achat et lors de la mixture des recettes.	Très faible	N/A	N/A
7	Patron de tissu	Il s'agit de la texture du tissu, son relief, ce dernier a un impact sur la réflexion de la lumière et donc de la couleur apparente du tissu <sup>2</sup> . La variabilité provenant de cette source concerne davantage la différence entre les patrons différents.	Oui	Oui, en quelque sorte; l'entreprise « subit » le patron puisque c'est le client qui décide du patron qu'il veut. Cependant, le patron est connu avant une recette.	Faible	Faire un plan d'expérience ou une analyse de données (si disponibles) afin de mesurer l'impact de chaque patron sur les couleurs et documenter le tout.	Un meilleur contrôle de la variation reliée à cette source provient d'une meilleure compréhension (voir ci-gauche).
8	Type de tissu	Il s'agit des caractéristiques du tissu utilisé (rétractation, finesse du fil, provenance du fil d'un « merge », etc.)	Oui (à l'usine de Saint-Georges)	Oui, à l'usine de Saint-Georges, à l'aide de tests normalisés.	Forte	N/A	N/A

<sup>2</sup> Il est à noter que la luminosité utilisée pour faire les tests est normalisée, mais, advenant des changements, cette dernière pourrait aussi affecter la réflexion de la couleur sur le patron.

#	Sources de variabilité	Descriptions	Mesuré (oui ou non)	Contrôlé (oui/non) et comment	Importance relative estimée p/r à la variation	Actions à poser	
						Pour mesurer	Pour contrôler
9	Nombre de tubes de la teindeuse	Chaque teindeuse (équipement utilisé pour le procédé de teinture) peut comporter jusqu'à 3 tubes (un lot de tissu par tube).	Oui	Oui, l'entreprise sélectionne quelle teindeuse utiliser pour chaque lot.	Faible	Faire une analyse de données (si disponibles) afin de mesurer l'impact du nombre de tubes sur les couleurs et documenter le tout.	Un meilleur contrôle de la variation reliée à cette source provient d'une meilleure compréhension (voir colonne « Pour mesurer » de cette ligne).
10	Profil de chauffage et refroidissement	Il s'agit des paramètres de chauffage et de refroidissement du procédé de teinture; gradient de chauffage, température atteinte, durée du plateau de chauffage, gradient de refroidissement, etc.	Oui	Oui, il s'agit de paramètres principaux lors du départ des cycles de teinture.	Faible	N/A	N/A
11	Vitesse de rotation du tissu dans la teindeuse	Il s'agit de la vitesse à laquelle le tissu effectue sa rotation dans la teindeuse; calculée en verges/minute. Ce paramètre aura en fait un impact sur le nombre de cycles du tissu.	Oui	Oui, il s'agit en fait d'un paramètre de contrôle de la teindeuse.	Faible à moyenne	N/A	Peut-être qu'ajuster la vitesse de rotation en fonction du nombre de cycles et de la longueur du tissu pourrait être bénéfique. De plus, il semble que ce soit souvent des erreurs humaines qui causent des erreurs de teinture à ce niveau.
12	Différentiel de pression	Il s'agit de la différence de pression entre la basse pression et la haute pression de la teindeuse. La haute pression est la pression à laquelle le mélange (eau, teinture, additifs chimiques et autres) est remis dans la teindeuse via la buse de Venturi.	Oui	Oui, il s'agit d'un paramètre de contrôle.	Faible	Effectuer une analyse de sensibilité sur les données pour évaluer l'impact du différentiel de pression (en conservant les autres variables et paramètres le plus constants possible).	N/A
13	Température du séchoir	Il s'agit de la température de séchage du tissu après teinture.	Oui	Oui, il s'agit d'un paramètre de contrôle du procédé de séchage.	Faible	N/A	N/A
14	Vitesse de roulement du tissu dans le séchoir	Il s'agit de la vitesse à laquelle déroule le tissu dans le séchoir.	Oui	Oui, il s'agit d'un paramètre de contrôle du séchoir.	Faible	N/A	N/A
15	Tension du tissu dans le séchoir	Il s'agit de la tension du tissu lorsque ce dernier passe dans le séchoir, à la suite du procédé de teinture.	Oui	Oui, un contrôle en temps réel se fait afin d'obtenir une tension longitudinale constante dans le tissu; le type de tissu ainsi que le patron aura un impact sur la tension. Une tension trop grande étirera le tissu et ce dernier paraîtra plus pâle et vice-versa. Cependant, pour ce qui est de la tension transversale, elle est ajustée manuellement sur certains tissus.	Moyenne	N/A	Une amélioration serait de déterminer, à l'aide d'un plan d'expérience, les conditions optimales d'installation des différents tissus qui permettent de maintenir la tension voulue.
16	Choix du séchoir	L'entreprise possède deux séchoirs; ce faisant, elle a le choix du séchoir, l'un étant plus court que l'autre pour les lots plus petits.	Oui	Oui, il s'agit d'un choix que l'entreprise fait pour chaque lot de tissu.	Très faible	N/A	N/A

#	Sources de variabilité	Descriptions	Mesuré (oui ou non)	Contrôlé (oui/non) et comment	Importance relative estimée p/r à la variation	Actions à poser	
						Pour mesurer	Pour contrôler
17	Température ambiante	Il s'agit de la température ambiante dans l'usine; il peut y avoir formation de bactéries et dénaturation de la couleur lorsque le tissu est laissé trop longtemps humide à une certaine température.	Non	Non	Très faible	Mesurer <sup>3</sup> la température ambiante est très facile; probablement que l'entreprise a déjà sur place un thermomètre. Il s'agirait seulement de la mesurer de temps à autre pour éviter la formation de bactéries en période chaude.	Un système de climatisation dans un milieu manufacturier est très complexe et dispendieux lorsque ce n'est pas nécessaire.
19	Présence de polluants dans l'eau	Il s'agit en fait de la présence de polluants non identifiés et inconnus dans l'eau utilisée pour la teinture; il pourrait s'agir de purin, de composantes chimiques provenant d'engrais, déversement quelconque, etc.	Non	Non	Faible à moyenne	Investiguer les différentes techniques qui pourraient permettre de mesurer cette source de variabilité. L'utilisation de la spectroscopie comparative (comparer le profil d'un échantillon avec un échantillon témoin) pourrait permettre de déceler la présence de corps étrangers anormaux.	Des équipements ou un procédé de filtration d'eau plus pointu devraient être implantés afin de pouvoir contrôler convenablement cette source de variabilité.
20	Présence de rouille dans l'eau	Il s'agit de la présence de molécules d'oxyde de fer (rouille) dans l'eau utilisée pour la teinture. Cette rouille proviendrait, d'après les informations obtenues, des conduits entre le procédé de décantation et la teinture.	Non	Non	Moyenne	Investiguer les moyens de mesurer la présence d'oxyde de fer dans de l'eau. L'échantillonnage en aval des conduits suivi d'une analyse en laboratoire pourrait être une solution.	On pourrait penser à faire la réfection des conduits d'eau qui cause la présence de rouille dans l'eau, ou un système de filtration supplémentaire en aval des bassins de décantation.
21	Taux d'alun	Il s'agit d'un agent coagulant utilisé lors du procédé de décantation de l'eau; combiné avec le microsable et le polymère, il peut se créer une réaction de coagulation du colorant dans le procédé de teinture si les trois composés sont encore présents.	Oui	Non, il est possible de mesurer le taux d'alun en laboratoire, mais il n'est actuellement pas contrôlé.	Faible	Le taux d'alun est déjà mesuré lorsque des problèmes surgissent en production. Cependant, il serait intéressant de mettre en place une procédure de récolte systématique de la donnée <sup>4</sup> .	Des équipements plus performants ou un procédé de filtration d'eau plus pointu devraient être implantés afin de pouvoir contrôler convenablement cette source de variabilité.
22	Taux de polymère	Il s'agit de la quantité de polymère que l'on retrouvera dans l'eau utilisée pour la teinture; dans le procédé de	Non	Non	Moyenne	Investiguer si des méthodes telles que la chromatographie gazeuse	N/A

<sup>3</sup> La mesure systématique de la température ambiante pourrait permettre ultimement de mesurer son impact sur la couleur finale.

<sup>4</sup> Une récolte de données systématique permettra ultimement d'utiliser des outils statistiques et des méthodes d'analyses avancées. La récolte systématique des données consiste à la pierre angulaire de l'industrie 4.0 puisque les données fournissent les informations nécessaires pour la prise de décision optimale.



#	Sources de variabilité	Descriptions	Mesuré (oui ou non)	Contrôlé (oui/non) et comment	Importance relative estimée p/r à la variation	Actions à poser	
						Pour mesurer	Pour contrôler
		décantation, des polymères sont utilisés pour coaguler les corps étrangers dans l'eau. Certaines de ces molécules de polymères peuvent rester et augmentent la dimension des granules de colorant.				et la spectroscopie de masse (laboratoire externe) pouvaient permettre de mesurer cette source de variabilité.	
23	Taux de microsable	Il s'agit d'une sorte de sable utilisée pour la coagulation des corps étrangers dans le bassin de décantation de l'eau. Combiné avec l'alun et le polymère, il peut se créer une réaction de coagulation du colorant dans le procédé de teinture si les trois composés sont encore présents.	Non	Non	Très faible	Investiguer si les méthodes de microfiltration d'une large quantité d'eau pouvaient permettre d'obtenir une mesure de la quantité de microsable résiduelle dans l'eau.	N/A
24	Granularité du colorant	Il s'agit de la grosseur des molécules de colorant; plus les molécules sont grosses, plus elles sont difficiles à intégrer au tissu et sont faciles à retirer (les molécules restent en surface) et vice-versa.	Oui, lorsqu'il y a des problèmes, la granularité est mesurée (microscopie électronique en balayage).	Oui, mais seulement dans la sélection du colorant; la granularité sera impactée par la présence de polymères et d'alun.	Faible	La granularité est déjà mesurée lorsque des problèmes surgissent en production. Cependant, il serait intéressant de mettre en place une procédure de récolte systématique de la donnée.	Diminuer la quantité d'alun et de polymère dans la teinture pour avoir la granularité voulue.
25	Répartition du colorant	Il s'agit tout simplement de la répartition du colorant dans l'eau du procédé de teinture.	Non	Non	Très faible <sup>5</sup>	Une procédure de mesure de couleur sur un groupe d'échantillons d'un même lot pourrait permettre de mesurer la répartition du colorant.	Avoir un meilleur contrôle sur la granularité et les agents coagulants permettrait probablement d'avoir une répartition plus uniforme du colorant.
26	Nombre de recyclages du tissu	Il s'agit du nombre de fois que le polymère utilisé pour fabriquer le tissu a été recyclé. Le nombre de recyclages a un impact sur la structure moléculaire du tissu et ultimement sur l'absorption et la rétention du colorant.	Non	Non	Moyen	Pour être capable de mesurer cette source de variabilité précisément, l'entreprise doit développer un système de traçabilité de la matière première très complexe et coûteux. Cependant, elle peut estimer cette mesure à partir des caractéristiques mécaniques du tissu.	L'entreprise ne peut pas contrôler cette source de variabilité puisqu'elle utilise des polymères recyclés pour la fabrication de ses fils. Elle peut cependant en prendre compte lors du paramétrage du procédé.
27	Contamination du polymère du tissu	Il s'agit de la présence possible de contaminants dans le polymère recyclé qui pourrait affecter ses caractéristiques ou modifier la coloration du tissu. Cette source de variabilité a un lien avec les	Non	Non	Moyenne à forte	Investiguer la possibilité de mesurer la contamination du polymère utilisé pour la fabrication du fil.	Améliorer le processus de nettoyage du polymère recyclé en amont de la fabrication du fil.

<sup>5</sup> L'impact de la répartition du colorant est très faible sur les « off-shades ». Une mauvaise répartition du colorant peut cependant et possiblement causer des taches sur le tissu.

#	Sources de variabilité	Descriptions	Mesuré (oui ou non)	Contrôlé (oui/non) et comment	Importance relative estimée p/r à la variation	Actions à poser	
						Pour mesurer	Pour contrôler
		effets de « <i>merge</i> », où la contamination varie d'un lot de polymère à l'autre.					
28	Caractéristiques du colorant	Il s'agit des caractéristiques physico-chimiques du colorant utilisé (brassage adéquat, bonne couleur, humidité, lot de production, etc.). Ces caractéristiques ont parfois une incertitude.	Oui, sur certains éléments	Pour les éléments mesurés, il y a un processus de contrôle.	Moyen	Une amélioration serait de cibler les caractéristiques du colorant qui influencent le plus, statistiquement, la couleur finale du tissu et de mettre en place un processus rigoureux de contrôle à la réception.	Une amélioration serait de cibler les caractéristiques du colorant qui influencent le plus, statistiquement, la couleur finale du tissu et de mettre en place un processus rigoureux de contrôle à la réception.

La dernière perspective intéressante avant d'entamer la discussion et les recommandations est de classer ces sources de variabilité dans une matrice présentant le caractère contrôlable des sources en fonction de leurs impacts (connus ou inconnus) et de leur observabilité (voir Figure 11). Cette forme de matrice permet de déceler quelles sources de variabilité risquent d'être plus problématiques que les autres, par exemple celles dont l'impact est inconnu. Il est à noter que dans cette matrice, lorsque l'on fait référence au caractère contrôlable d'une source de variabilité, il s'agit de l'état actuel des choses et non de la faisabilité technique de ce même contrôle.

Identification des sources de variabilités	Contrôlable		Non Contrôlable	
	Valeur observable	Valeur non observable	Valeur observable	Valeur non observable
<b>Observabilité</b>				
<b>Impact</b>				
<b>Impact(s) inconnu(s)</b>	1. pH de l'eau		17. Température pièce ambiante	24. Granularité du colorant
	2. Composition du mélange d'additif		19. Présence de polluants dans l'eau	25. Répartition du colorant
	3. Quantité de chaque additif		20. Présence de rouille dans l'eau	26. Nombre de recyclage du tissu
	4. Couleur du colorant		21. Taux résiduel d'alun dans l'eau	27. Contamination du polymère
	5. Quantité de colorant		22. Taux résiduel de polymère dans l'eau	
	6. Type de colorant (marque/fournisseur)		23. Taux résiduel de micro-sable	
	7. Patron du tissu			
	8. Type de tissu			
	9. Nombre de tube de la teindeuse			
	10.1. Gradient de chauffage (procédé)			
	10.2. Température du plateau (procédé)			
	10.3. Durée du plateau (procédé)			
	10.4. Gradient de refroidissement (procédé)			
	11. Vitesse de rotation du tissu dans la teindeuse			
	12.1. Haute pression de la teindeuse			
	12.2. Basse pression de la teindeuse			
	13. Température du four de séchage			
	14. Vitesse de roulement du tissu dans le séchoir			
	15. Tension du tissu dans le séchoir			
	16. Choix du four (nouveau ou ancien)			
28. Caractéristique du colorant				
<b>Impact(s) connu(s)</b>				

Figure 11: Tableau représentant l'observabilité, la contrôlabilité ainsi que la connaissance des impacts de chaque source de variabilité du projet Duvaltex..

Toutes ces observations, discussions, analyses et représentations ont permis de dégager des recommandations d'amélioration qui ont été soumises à l'entreprise Duvaltex. Il est important de mentionner que ces recommandations sont basées uniquement sur des conclusions d'un processus d'investigation très qualitatif, tel que mentionné antérieurement. Pour cette raison, une rétroaction de la part de l'entreprise a été nécessaire afin de confirmer la validité et la faisabilité de ces recommandations.

### Recommandation n° 1 : Mesure structurée de la couleur finale

En premier lieu, il est important pour Duvaltex de mettre en place une procédure standard, fiable et répétitive pour mesurer la couleur du tissu. Actuellement, l'inspection de la couleur du tissu se fait avec un test visuel par des responsables qualité et qu'au besoin, ces derniers peuvent utiliser un spectrophotomètre pour valider leurs inspections. Selon les informations obtenues, l'utilisation du spectrophotomètre seule pour mesurer la couleur ne serait pas assez fiable puisque la mesure serait trop impactée par la texture du tissu. De plus, dans le cas de tissus à multiples couleurs, la mesure par spectrophotométrie devient plus complexe. Parmi les solutions possibles, on peut penser à ces solutions :

- Échantillonnage multiple de la couleur : par exemple, prendre 20 mesures différentes, prendre la moyenne et mesurer la variance ;
- Utilisation d'un autre procédé d'inspection de la couleur : par exemple, utiliser des techniques avancées d'imagerie. On peut penser à prendre une photo de l'échantillon à mesurer à côté du tissu original. Un logiciel analyse les deux tissus et donne le différentiel de couleur ;
- Possibilité d'acquérir un équipement de spectrophotométrie plus précis et mieux adapté à la réalité de l'industrie textile.

Enfin, peu importe la méthode utilisée, la mesure de la couleur finale doit être conservée afin d'être en mesure d'effectuer des analyses statistiques ultérieurement.

### Recommandation n° 2 : Étude et analyse de l'eau utilisée pour la teinture

Un élément important de la problématique à approfondir est l'impact sur les variations de couleur du taux d'alun, du taux de polymère, du taux de microsable, de la quantité de polluants ainsi que de la quantité de rouille dans l'eau utilisée pour la teinture. Cette analyse pourrait être faite sous la forme d'une étude de sensibilité afin de déterminer lesquelles parmi ces sources de variabilité ont un impact non négligeable, de même qu'une analyse statistique afin d'y dégager une corrélation. Dans le contexte d'une transition vers un procédé intelligent et flexible, il est recommandé d'implanter une récolte de données systématique à ce niveau du procédé. Une étude de faisabilité pourrait être menée préalablement, mais une première étape possible consisterait à faire la prise

d'échantillons d'eau pour chaque lot de teinture et d'envoyer ces échantillons à un laboratoire spécialisé externe afin d'obtenir les caractéristiques spécifiques de l'eau utilisée qui sont impossibles à obtenir avec les équipements de laboratoire à l'interne. Plus précisément, on parle ici de la quantité de polymère résiduelle dans l'eau ainsi que de la quantité de microsauble résiduelle présente. Si certaines caractéristiques sont difficiles à mesurer, une possibilité pourrait être de trouver un lien de causalité avec d'autres facteurs. À titre d'exemple, la présence de polluants dans l'eau pourrait s'avérer difficile à mesurer. Or, il est possible de mesurer la turbidité, soit une mesure de la quantité de matières qui troublent l'eau, pour évaluer indirectement la quantité de polluants dans l'eau. En revanche, il serait possible d'utiliser des capteurs ou d'autres équipements permettant d'obtenir ces mesures en temps réel. Encore une fois, pour atteindre cette situation désirée, on doit d'abord valider l'existence de tels équipements, la faisabilité de leur installation, leur coût d'installation et leur coût d'utilisation.

### **Recommandation n° 3 : Structure de contrôle du tissu**

De surcroît, il est suggéré de continuer à développer les processus de contrôle qualité du tissu à tous les niveaux de l'entreprise. Actuellement, les caractéristiques du tissu telles que le nombre de recyclage qu'a subi le fil ainsi que la contamination du polymère ont probablement un impact sur la couleur finale de ce dernier. Cependant, il faut considérer qu'il n'y a aucune façon dans le procédé de teinture de contrôler ces caractéristiques, sauf de les considérer lors de la création des recettes. C'est pour cette raison que le développement des connaissances et un meilleur contrôle du tissu seraient importants. Il a été mentionné précédemment que le nombre de fois que le tissu a été recyclé pouvait être estimé grâce à ses propriétés mécaniques. Dans ce contexte, il serait important de mettre en place un processus d'estimation des propriétés du tissu et de conserver l'information à des fins d'analyse ultérieure.

### **Recommandation n° 4 : Récolte systématique de données**

Une autre recommandation serait, de façon générale, d'instaurer des récoltes et sauvegardes systématiques des paramètres de production, des variables de production ainsi que des mesures de performance du processus de teinture, et ce, afin d'avoir une certaine traçabilité des lots de production de même que l'opportunité d'employer des méthodes avancées d'analyse de données pour relever et identifier des problèmes, mieux comprendre leurs causes, identifier certains liens de cause à effet ainsi qu'avoir la possibilité d'améliorer les situations problématiques d'une autre manière (par l'analyse de données) plutôt que par d'autres méthodes traditionnelles plus difficiles à mettre en œuvre. L'analyse des données peut également permettre d'identifier de nouvelles opportunités, d'identifier des problématiques de production, de prévoir des tendances et d'en faire les recommandations associées. Entre autres, il semble que des investigations soient nécessaires au niveau de l'ajout d'additifs chimiques aux recettes de teintures; la récolte et l'analyse de données sur ces derniers permettraient sans aucun doute d'identifier des pistes de solutions. Il s'agit en fait d'un pilier très important de

l'industrie 4.0 et d'un premier pas vers un processus manufacturier intelligent et hautement performant (Kagermann *et al.*, 2013).

#### *Recommandation n° 5 : Poursuite du projet de recherche*

La dernière recommandation qui a été soumise à Duvaltex portait sur la poursuite du partenariat dans la seconde phase du projet de recherche, plus précisément la phase de conception d'un modèle prédictif puis d'un modèle prescriptif. Certes, plusieurs investigations seraient à entamer, mais la problématique cadre adéquatement avec les objectifs et les contraintes du projet de recherche. Cette dernière recommandation qui a été mise en place fera l'objet de la prochaine section.

## **4.2. Étape 2 : Modélisation prédictive**

À l'instar de la dernière recommandation soumise lors de l'étape d'analyse de la situation initiale (Recommandation n° 5 : Poursuite du projet de recherche), les étapes 2 et 3 du cadre méthodologique ont été mises en œuvre en collaboration avec l'entreprise Duvaltex.

Tel qu'illustré dans la Figure 4, cette étape du cadre méthodologique débute avec la déclinaison des données à récolter à partir des sources de variabilité identifiées durant l'étape 1, soit l'analyse de la situation initiale. Plus précisément, cette tâche consiste à déterminer, pour chaque source de variabilité priorisée et sélectionnée pour la modélisation prédictive, quelles données sont disponibles ou peuvent être récoltées afin de convertir les sources de variabilité en valeurs réelles.

Une fois cette déclinaison faite et avant d'être en mesure de débiter la modélisation du procédé de teinture, il a été nécessaire d'effectuer l'acquisition, le formatage et le nettoyage des données fournies par l'entreprise Duvaltex, et ce, dans le cadre de la sous-étape « 2.2. Détermination de la méthode de récolte de données, prétraitements et analyses statistiques ». De façon générale, plusieurs options s'offrent aux utilisateurs du cadre méthodologique quant à l'acquisition des données utilisées pour la modélisation :

- Utilisation de données historiques accumulées par l'entreprise ;
- Mise au point d'un plan d'expérience pour récolter manuellement les données désirées ;
- Installation d'équipements spécialisés pour récolter automatiquement les données voulues (en continu ou durant une période prédéterminée).

Dans le cas de Duvaltex, l'option d'utiliser des données historiques de production semblait être la plus adaptée en raison de contraintes de temps et de budget. Ce processus d'acquisition, formatage et nettoyage des données se divise en sept (7) étapes :

### **1. Copie de différents fichiers sources contenant les données nécessaires**

La première étape de ce processus consistait à localiser à l'aide du partenaire industriel les différents fichiers sources de données et d'en faire une copie dans un dossier prévu à cet effet. Pour la plupart de ces données, il s'agissait d'un simple fichier de format « Excel ». La seule exception était les données contenues dans les fichiers « .dbf », soit un format de fichier utilisé principalement dans les systèmes de gestion de base de données.

### **2. Importation non triée des données de la base de données (fichiers de format « .dbf »)**

Tel que mentionné au point précédent, une partie des données était contenue dans des fichiers de format « .dbf », soit des fichiers utilisés pour des systèmes de gestion de base de données. Ces données sont en fait des séries temporelles du procédé de teinture<sup>6</sup>; pour chaque lot de teinture, ces séries temporelles de données sont générées par l'ordinateur de l'équipement et stockées dans une base de données « dBase ». Parmi ces données, on retrouve des temps de cycle du procédé, la température, la vitesse et plusieurs autres données qui peuvent s'avérer très utiles pour l'entreprise. Cette étape du processus d'acquisition, de nettoyage et de formatage des données a ainsi pour objectif d'importer de façon massive toutes les données contenues dans ces fichiers « .dbf » pour les mettre en commun dans un même tableur « Excel » afin d'être en mesure d'effectuer des analyses. La difficulté réside plus précisément dans le fait que chaque fichier « .dbf » contient une série temporelle de données pour un seul lot. Pour des fins d'analyse, il est nécessaire d'ouvrir individuellement chaque fichier et d'extraire l'information. Pour y arriver dans un délai raisonnable, un programme en langage VBA a permis d'automatiser cette étape du processus.

### **3. Tri des données pour n'obtenir que les styles de tissu 2334 et 2335**

Considérant certaines contraintes de temps du projet, il était nécessaire de limiter les données et la modélisation à un style de tissu en particulier. Cette limitation est due au fait que le style<sup>7</sup> du tissu peut grandement influencer la façon dont le procédé de teinture se déroule (taux d'absorption des tissus, caractéristiques du polymère utilisé différentes, équipements de tissage différents, etc.) et qu'une multitude de styles différents existent. Dans le cas actuel, cette limitation a un impact considérable, puisque dans le contexte du nettoyage et du formatage des données, chaque source de données doit être triée afin d'obtenir seulement les styles recherchés.

---

<sup>6</sup> Les temps de cycle et les vitesses sont calculés à l'aide d'un aimant inséré à un endroit précis dans le lot de tissu ainsi que de l'utilisation d'un capteur magnétique.

<sup>7</sup> Le style correspond en quelques sortes à la façon de fabriquer et tisser le tissu; le patron de tissage. À titre d'exemple, certains styles ont des fils de tissu de couleur différente pour créer un effet esthétique.



#### 4. Consolidation des données dans une matrice commune

Une fois les données triées par style et extraites de leurs fichiers sources, la prochaine étape consiste à consolider ces données afin d'obtenir une matrice unique de données. Ainsi, une ligne de cette matrice contient toutes les données nécessaires pour une pièce de tissu; une pièce correspond à l'unité de tissu la plus petite utilisée par l'entreprise. La Figure 12 représente la hiérarchie<sup>8</sup> des quantités de tissu :

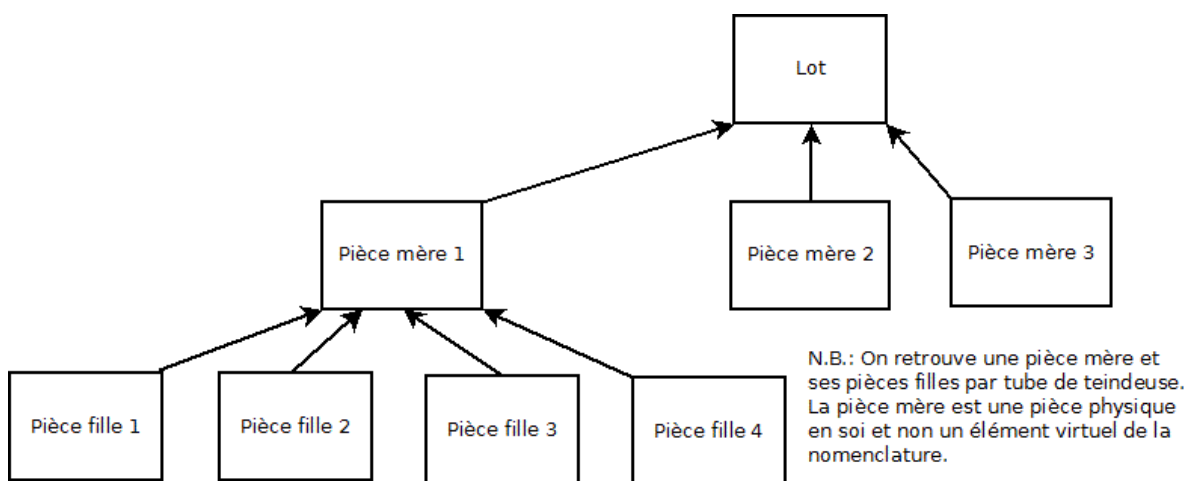


Figure 12: Nomenclature des lots de tissu

Pour ce qui est des données qui sont propres à un lot et non à une pièce, elles se répèteront pour chaque pièce du lot (par exemple, la vitesse de rotation dans la teindeuse est la même pour toutes les pièces d'un même lot).

#### 5. Calcul de certaines données manquantes

Plusieurs données identifiées lors de la première étape du cadre méthodologique ne se retrouvaient pas dans les fichiers sources de données. Ces données, lorsque cela est possible, doivent être calculées, extrapolées ou estimées à l'aide des autres données existantes (voir le tableau exhaustif de l'Annexe 1 présentant les différentes données calculées et la façon dont elles ont été calculées).

#### 6. Agrégation des données par lots de tissus

Une fois les données consolidées dans une matrice unique, il s'avère essentiel d'agréger les données par lot plutôt que par pièce, et ce, grâce à l'outil de tableaux croisés dynamiques du logiciel Excel. Cette décision d'utiliser le lot comme base de référence provient du fait que les non-conformités de la couleur (« off-shade ») affectent un lot au complet et sont identifiées de la même manière. Pour être en mesure de faire cette agrégation, il a été nécessaire d'identifier la façon d'agréger chaque type de données. Par exemple, les quantités de produits chimiques, les quantités de colorants et la plupart des paramètres de la teindeuse étaient communs pour toutes

<sup>8</sup> Il est à noter qu'un lot peut avoir jusqu'à l'équivalent de 3 tubes de teindeuse de tissu, et que chaque tube de teindeuse peut avoir jusqu'à 9 pièces incluant la pièce mère.

les pièces et tous les tubes de teindeuse, permettant ainsi d'utiliser une moyenne. En revanche, certaines données telles que le poids des pièces devaient être additionnées lors de l'agrégation pour obtenir le poids total par lot.

## 7. Élimination des lignes avec données manquantes

La dernière activité de nettoyage et de formatage des données consistait à éliminer les lignes possédant des données manquantes, considérant que l'approximation de ces données entraîne des risques d'affecter l'importance relative de chaque source de variabilité sur l'extrait du procédé. En contrepartie, considérant la quantité limitée des données, les données statistiquement aberrantes ont été conservées. En effet, la littérature démontre qu'il est possible de conserver ces données aberrantes par l'utilisation de méthodes de modélisation assez robustes qui sont insensibles à ces données (Kantardzic, 2011).

Une fois le processus de formatage et de nettoyage des données complété, une matrice contenant toutes les données dans un fichier Excel était ainsi disponible pour la modélisation prédictive du projet (une liste complète des différentes variables utilisées pour la modélisation prédictive de même que les sources de variabilité auxquelles les variables se rattachent est présentée dans l'Annexe 2).

Cette deuxième étape du cadre méthodologique a donc été poursuivie avec une période exploratoire où l'objectif principal était de tester diverses méthodes. Pour cette période exploratoire, seules les mesures d'ajustement (« *fit* ») étaient utilisées pour comparer les méthodes entre elles. Pour cette comparaison, l'indicateur d'erreur appelé « RMSE » (« *Root Mean Square Error* ») a été utilisé pour les modèles présentant une variable de sortie binaire, tandis que l'indicateur « PRESS RMSE » (« *Predicted Residual Error Sum of Squares* ») et « *Root Mean Square Error* » a été utilisé pour les variables de sortie continues (non binaires). À titre de clarification, on entend par « variable de sortie » (« *output variable* ») la mesure exprimant la présence d'une non-conformité de la couleur (« *off-shade* »), soit 0 si le lot est conforme et 1 si le lot présente un « *off-shade* ».

Une fois cette période exploratoire terminée, les méthodes présentées dans le Tableau 6 ont été retenues pour une analyse comparative plus complète. Puisque selon les méthodes utilisées, la variable de sortie pouvait être continue ou binaire (nominal), l'indicateur de comparaison commun utilisé a été le taux de prédiction de la méthode sur des données non entraînées, donc des données de validation. Les différents indicateurs de performance de la prédiction sont expliqués plus en détail à la suite du Tableau 6.

Le tableau ci-dessous propose un résumé des résultats obtenus avec ces quatre modèles :

Tableau 6: Résultat des différents modèles testés dans l'étape de la modélisation prédictive

Type de variable de sortie		Continue		Binaire		
Méthode de modélisation		SLS		NLR		
Interactions	Simples	Précision globale	Modèle 1	<b>94.87% ± 1.45%</b>	Modèle 2	<b>94.87% ± 1.48%</b>
		Fraction de faux négatifs		<b>61.05% ± 15.08%</b>		<b>51.58% ± 15.95%</b>
		Fraction de faux positifs		<b>0.36% ± 1.51%</b>		<b>1.17% ± 1.75%</b>
	Simples, croisées et quadratiques	Précision globale	Modèle 3	<b>95.62% ± 1.05%</b>	Modèle 4	<b>92.47% ± 2.14%</b>
		Fraction de faux négatifs		<b>50.53% ± 12.32%</b>		<b>27.37% ± 16.44%</b>
		Fraction de faux positifs		<b>0.45% ± 1.43%</b>		<b>5.83% ± 2.49%</b>

Pour toutes les méthodes testées, il était possible de séparer l'approche de modélisation en deux parties. La première partie consistait à appliquer une technique de réduction de variables pour ainsi créer un sous-ensemble de variables (à titre d'exemple, tiré du projet Duvaltex, cette méthode de sélection de variables permet de passer de plusieurs centaines de variables à 44 variables). À partir de ce sous-ensemble de variables, les différents modèles ont été entraînés et validés selon la validation croisée. Le concept de validation croisée consiste tout simplement à appliquer un modèle ou une équation entraîné ou calculé à partir d'un ensemble de données, appelé ensemble de données d'entraînement (« *training* »), sur un autre ensemble de données qui n'a pas été utilisé pour l'entraînement, pour ainsi mesurer la capacité de prédiction dudit modèle ou équation sur de nouvelles données (Jain et *al.*, 2018). Ainsi, le modèle est entraîné sur 10 ensembles de données d'entraînement pour ensuite être validé sur un ensemble de données de validation correspondant (chaque ensemble de données d'entraînement est complémentaire à un ensemble de données de validation). Enfin, des indicateurs de capacité de prédiction sont calculés à partir des ensembles de données de validation et ce sont ces indicateurs qui permettent de comparer les différents modèles entre eux. Toutes les étapes présentées ci-haut ont été effectuées à l'aide du logiciel JMP.

L'utilisation d'une méthode de réduction de variables permet de diminuer le nombre de variables dans le modèle de prédiction, et ce, pour améliorer la qualité et la précision de la prédiction. Un trop grand nombre de variables peut amener un surentraînement qui fait en sorte que les données expliquent parfaitement les résultats des données d'entraînement, mais démontrent une mauvaise capacité de prédiction sur de nouvelles données. Dans le cas actuel, la méthode pas à pas (« *stepwise regression* ») a été sélectionnée. Plus précisément, cette

méthode consiste à ajouter ou à retirer<sup>9</sup> itérativement une variable de prédiction et de comparer un indicateur statistique, souvent via l'utilisation d'un test F (test de Fischer) ou d'un test T (test de Student) (Nisbet et *al.*, 2009). Les itérations sont par la suite arrêtées à l'aide d'une règle d'arrêt prédéterminée. Dans le cas de Duvaltex, la direction avant (« *forward* ») a été choisie, la valeur-p a été utilisée pour comparer et la règle d'arrêt consistant à une valeur-p de 0.25 retenue (toute variable ayant une valeur-p supérieure à cette limite n'était pas incluse dans le modèle). La valeur-p est communément utilisée pour quantifier la significativité statistique d'un résultat lors de tests d'hypothèse statistique. De cette façon, telle qu'illustrée dans le Tableau 7, plus cette valeur est faible, plus un facteur est significatif (Hines et *al.*, 2008).

---

<sup>9</sup> Le logiciel JMP offre trois directions pour effectuer la méthode pas à pas; la direction ascendante (« *forward* ») ne fait qu'ajouter des variables, la direction descendante (« *backward* ») ne fait que retirer des variables et la direction mixte (« *mixed* ») peut faire les deux.

Tableau 7: Liste des facteurs en ordre décroissant d'importance du modèle 4

Facteurs	Valeur-p
Niveau d'eau moyen	0.0462
Pression différentielle maximale* <sup>10</sup> Débit moyen de tissu dans la teindeuse	0.04962
Poids de la pièce	0.07199
Mesure L du « DIANIX BLACK CCR »	0.08338
Quantité de « TRISÉTILE BLUE CM »	0.09004
Pourcentage de rétraction longueur dans la teindeuse	0.09723
(Pourcentage de rétraction longueur dans la teindeuse) <sup>2</sup>	0.10367
Niveau d'eau moyen*Niveau d'eau maximal	0.10466
Pression différentielle moyenne	0.10519
Niveau d'eau moyen*Niveau d'eau moyen	0.10585
Niveau d'eau maximal*Pression différentielle moyenne	0.10679
Mesure A du « LUMACRON BLUE ERD »	0.12125
Mesure L du « DIANIX YELLOW BROWN CC »	0.1218
Durée approximative du plateau de chauffage*Pression différentielle maximale	0.12338
Durée approximative du plateau de chauffage*Pourcentage de rétraction longueur dans la teindeuse	0.12663
Mesure B du « DIANIX RED CC »	0.12818
Longueur initiale demandée pour le lot	0.1316
Durée approximative du plateau de chauffage*Turbidité résiduel	0.13184
Niveau d'eau moyen*Masse linéique	0.13594
Niveau d'eau maximal*Pourcentage de rétraction longueur dans la teindeuse	0.1383
(Température du plateau de chauffage) <sup>2</sup>	0.14092
Température du plateau de chauffage*Vitesse moyenne	0.14167
Présence d'un nœud lors de la teinture	0.14205
Durée approximative du plateau de chauffage	0.14212
Niveau d'eau moyen*Pourcentage de rétraction longueur dans la teindeuse	0.14235

La création des différents modèles est basée sur deux paramètres de modélisation :

1. la méthode de modélisation ;
2. les interactions considérées dans le modèle.

<sup>10</sup> Les astérisques représentent une multiplication.

Ainsi, deux méthodes de régression ont été utilisées :

- La première méthode, la méthode des moindres carrés (« *Standard Least Square Regression* » ou SLS), est une méthode de régression largement utilisée pour modéliser une variable continue (pour les modèles employant cette méthode, la variable de sortie était donc une valeur numérique continue). Cette méthode consiste à effectuer une régression multiple en minimisant la somme des carrés des erreurs où l'erreur représente l'écart entre ce que la régression prédit et la vraie valeur (le détail de cette méthode de régression est présenté dans l'Annexe 3). Il s'agit de la méthode qui a été employée pour les modèles 1 et 3 (voir Tableau 6).
- La seconde méthode a pour objectif de modéliser une variable binaire ou nominale. La régression logistique nominale (« *Nominal Logistic Regression* » ou NLR) est une méthode qui permet de prédire la probabilité que la variable de sortie soit 1 ou 0, selon ce que l'on cherche à prédire. Dans le contexte actuel, la méthode NLR permet d'obtenir une probabilité qu'un lot ait un « *off-shade* » (par exemple, un certain lot a une probabilité de 92,51% d'être non conforme) (le détail de cette méthode de régression est présenté dans l'Annexe 4). Il s'agit de la méthode qui a été employée pour les modèles 2 et 4 (voir Tableau 6).

Dans un même ordre d'idées, le second paramètre de modélisation concerne les interactions considérées. On retrouve ainsi deux cas d'interactions considérées (Hines *et al.*, 2008):

- Les effets simples, c'est-à-dire que les variables sont ajoutées au modèle telles qu'elles sont et les effets croisés entre les variables et les effets quadratiques (2<sup>e</sup> degré) ne sont pas considérés. Il s'agit des effets considérés pour les modèles 1 et 2 (voir Tableau 6) ;
- Les effets simples, croisés et quadratiques sont inclus dans la modélisation, (par exemple, l'ajout de  $X*Y$  et de  $X^2$ ). Il s'agit des effets considérés pour les modèles 3 et 4 (voir Tableau 6).

Le détail de chacun de ces modèles est présenté dans l'Annexe 5.

Dans le domaine de la classification binaire, c'est-à-dire lorsqu'on tente de prédire une variable binaire, quatre termes sont utilisés pour qualifier le résultat de la prédiction :

- Un vrai positif, un résultat positif qui a été prédit correctement ;
- Un faux positif, un résultat négatif qui a été prédit comme positif ;
- Un vrai négatif, un résultat négatif qui a été prédit correctement ;

- Un faux négatif, un résultat positif qui a été prédit comme négatif.

De plus, quelques indicateurs ont été utilisés pour comparer les différents modèles. La précision globale (« *overall accuracy* »), la fraction de faux négatifs (« *false negative fraction* ») et la fraction de faux positifs (« *false positive fraction* ») sont exprimées par les équations ci-dessous (Metz, 1978):

$$\text{Précision globale} = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total d'observations}}$$

$$\text{Fraction de faux négatifs} = \frac{\text{Nombre de faux négatifs}}{\text{Nombre de prédictions réellement positives}}$$

$$\text{Fractions de faux positifs} = \frac{\text{Nombre de faux positifs}}{\text{Nombre de prédictions réellement négatives}}$$

Dans le contexte actuel du procédé de teinture, la fraction de faux négatifs se rapporte aux situations où la conformité a été prédite, mais où le tissu échoue véritablement le contrôle qualité de la couleur, tandis que la fraction de faux positifs se rapporte aux situations où une non-conformité (« *off-shade* ») a été prédite, mais où, en réalité, le tissu réussit le contrôle qualité de la couleur. Il s'agit donc de deux types d'erreurs différents.

De plus, il est possible d'obtenir une liste de variables ayant le plus d'impact statistique sur le procédé de teinture, et ce, par ordre d'importance avec le logiciel JMP. À titre d'exemple, le Tableau 7 ici-bas présente les 25 premières variables ayant le plus d'impact sur le procédé de teinture suivant le modèle utilisant une régression nominale logistique et incluant les effets simples, croisés et quadratiques (modèle 4). Ces variables proviennent évidemment du sous-ensemble de variables déterminé à l'aide de la méthode de régression pas-à-pas (« *stepwise regression* »).

Il est à noter que cette liste présente les 25 variables ayant les plus fortes corrélations statistiques avec la probabilité d'obtenir un « *off-shade* », mais elle ne prouve pas nécessairement l'existence d'une relation de cause à effet.

Avec les résultats des quatre modèles présentés précédemment, l'identification du meilleur modèle dépend du type d'erreur que l'entreprise partenaire souhaite minimiser; est-il préférable de réduire l'erreur de prédiction des non-conformités de la couleur, la fraction des faux négatifs, ou l'erreur de prédiction des lots conformes, la fraction des faux positifs? Dans le premier cas, le modèle 4 serait le plus approprié tandis que dans le deuxième cas, il s'agirait du modèle 1.

### Amélioration du modèle prédictif

Bien que les résultats obtenus avec le modèle prédictif retenu soient utiles, il serait possible d'y apporter des améliorations pour potentiellement augmenter sa capacité de prédiction. Voici quelques avenues possibles pour parfaire ce modèle :

1. Ajouter certaines sources de variabilité. En effet, certaines sources de variabilité identifiées en début de projet comme pouvant être relativement importantes n'ont pas pu être intégrées au modèle dû à l'absence de données sur ces sources en question. Par exemple, certaines caractéristiques de l'eau utilisée pour la teinture avaient été identifiées, mais l'entreprise n'avait pas de données fiables sur ces caractéristiques, résultant ainsi à l'omission de ces facteurs dans la modélisation prédictive. On peut aussi penser aux caractéristiques du fil composant le tissu; un certain manque de connaissance sur la relation entre ces caractéristiques et le procédé de teinture ainsi que le manque de données sur ces dernières ont fait en sorte que ces facteurs ont eux aussi été omis volontairement. Ce constat est de surcroît appuyé par la présence de plusieurs facteurs relatifs aux caractéristiques du tissu dans le Tableau 2, par exemple le poids du tissu et le pourcentage de rétraction. En contrepartie, il est difficile de déterminer quelle proportion de la performance de prédiction perdue est attribuable à ces facteurs ainsi qu'aux facteurs relatifs à la qualité de l'eau. De la même façon, il se pourrait que certaines sources de variabilité ne soient pas captées dans les données utilisées. À titre d'exemple, grâce à des efforts déployés en parallèle au projet de recherche, l'entreprise a été en mesure de diminuer substantiellement son taux de non-conformité de la couleur grâce à la standardisation des méthodes de travail. Cette amélioration supporte le fait qu'une certaine partie de la variabilité du procédé était imputable aux méthodes de travail, une source de variabilité qui n'avait pas été prise en compte dans l'ensemble de variables.



2. Étendre la modélisation prédictive aux autres styles. Plus précisément, le modèle développé dans le cadre de ce projet de recherche ne concerne que les styles de tissu « 2334 » et « 2335 » et c'est pourquoi une avenue intéressante aurait été de faire le même exercice de modélisation avec les autres styles fabriqués par Duvaltex. Cependant, l'objectif n'est pas de concevoir un seul modèle pour tous les styles, mais bien d'obtenir un certain nombre de modèles pouvant être appliqués aux différents styles de l'entreprise (par exemple, on peut penser à un modèle par style).
3. Ajouter des données pour l'étape de la modélisation. Tel que mentionné dans la discussion sur le modèle prédictif retenu, une faiblesse du modèle est le peu de données utilisées pour entraîner le modèle. Le fait d'avoir plus de données ferait en sorte que les facteurs identifiés lors de la modélisation seraient plus représentatifs de la tendance générale, ce qui améliorerait probablement la capacité de prédiction du modèle. De plus, avoir peu de données augmente les risques d'avoir une partition entraînement/validation trop hétérogène (par exemple, avoir trop de données d'une couleur présentant un comportement différent des autres couleurs dans l'ensemble de données de validation).
4. Une autre piste intéressante aurait été de refaire l'exercice de modélisation avec d'autres procédés ou processus de l'entreprise. Puisque le cadre méthodologique développé dans le contexte de ce projet de recherche a été conçu pour être générique et applicable dans différentes situations, il y aurait beaucoup de valeur en termes d'apprentissage de la méthode et des processus d'appliquer le cadre à d'autres processus de l'entreprise.
5. Tester davantage de méthodes de modélisation prédictive. Plusieurs autres méthodes issues des techniques de forage de données (« *data mining* ») auraient peut-être donné de meilleurs résultats, mais auraient nécessité davantage de données et de temps. Dans un contexte où la maximisation de la performance du modèle prédictif est primordiale, l'exploration de davantage de méthodes pourrait s'avérer nécessaire.
6. Enfin, une autre avenue qui n'a pas été explorée lors du formatage et du nettoyage des données consiste à utiliser des méthodes d'échelonnage des données (« *scaling* »). Plus précisément, ces méthodes permettent en quelque sorte de standardiser l'ordre de grandeur des données. À titre d'exemple, une certaine variable dont les valeurs se situent entre 100 et 1000 n'aura pas le même effet qu'une variable qui varie, quant à elle, de -1 à 1. Bien qu'aucune méthode de ce type n'a été appliquée dans cette modélisation prédictive, il est généralement recommandé de le faire afin d'obtenir de meilleurs résultats (Kantardzic, 2011).

### 4.3. Étape 3 : Modélisation prescriptive

Tel qu'illustré dans la Figure 4, la prochaine étape du cadre méthodologique consiste à l'analyse prescriptive. Le principe de cette étape est d'utiliser les résultats du modèle prédictif afin de proposer des actions permettant de réduire les impacts de la variabilité du procédé ou tout simplement de réduire la variabilité des sources de variabilité identifiées dans l'étape de l'analyse de la situation initiale. Qu'il s'agisse de proposer directement des solutions ou de concevoir un système permettant de proposer des solutions d'amélioration, l'objectif est d'instrumenter les résultats du modèle prédictif.

Dans le cas d'application avec l'entreprise Duvaltex, les contraintes de temps liées au projet de recherche ont fait en sorte qu'il était impossible de développer un système prescriptif. La décision de proposer des pistes d'amélioration basées sur la modélisation prédictive a plutôt été prise. Le Tableau 8 résume quelques options considérées pour la suite du projet de recherche. Une description plus détaillée des options suivra ensuite.

Tableau 8: Différentes options possibles pour la poursuite du projet Duvaltex

#	Option	Description	Retombées	Effort requis estimé
1	Étude approfondie des facteurs les plus importants	Cette option consiste à explorer et à étudier plus en profondeur chacun des facteurs identifiés dans le tableau 2 et à tenter d'identifier la cause d'origine qui amène de la variabilité au niveau du procédé et si possible d'éliminer ou de réduire cette cause racine. Ultimement, il s'agit de travailler à réduire la variabilité en amont du procédé de certains facteurs lorsque possible afin de faciliter la compréhension et la gestion de la variabilité intrinsèque.	Permettrait de mieux quantifier l'impact des sources de variabilité incontrôlable.	++
2	Modifier le procédé ou les produits	En fonction des prédicteurs les plus importants, il serait possible d'apporter des améliorations au procédé et aux produits.	Améliorer la qualité des produits et réduire la variabilité.	++
3	Développer une interface d'aide à la décision à partir du modèle prédictif	Cette option consiste à développer une interface utilisant des données en temps réel pour supporter la prise de décision au niveau du procédé.	Conception d'un outil permettant de supporter le travail des opérateurs.	+++
4	Développer un outil de simulation pour tester le résultat du procédé de teinture	Il s'agit d'utiliser conjointement la simulation et le modèle prédictif afin de tester différents scénarios du procédé de teinture.	Permettrait de mieux concevoir les nouvelles recettes et même d'améliorer les recettes existantes.	+

Il est à noter que pour chacune des options, le temps requis est estimé de façon très approximative selon la nature de l'option. De plus, aucune analyse de faisabilité ou analyse économique n'a été conduite.

## Option 1

Cette première option consiste à utiliser les résultats du modèle prédictif, plus précisément les facteurs identifiés qui ont le plus d'impact sur la variabilité du procédé, afin d'orienter des initiatives de diminution de la variabilité des sources de variabilité contrôlables. Il est évident qu'il est impossible d'éliminer totalement la variabilité, mais l'exercice de modélisation effectuée dans ce projet et particulièrement les facteurs identifiés lors de la régression permettent d'orienter ces initiatives.

## Option 2

Pour ce qui est de cette seconde option, il s'agit de l'opportunité d'utiliser les résultats de la modélisation prédictive pour améliorer les procédés et les produits actuels ainsi que les futurs produits. À titre d'exemple, on remarque dans le Tableau 1 que le niveau d'eau moyen dans la teindeuse est un facteur ayant un impact important sur la probabilité d'obtenir une non-conformité de la couleur (« *off-shade* »). Ce faisant, il pourrait être pertinent d'étudier plus en profondeur dans quelle mesure ce facteur affecte la couleur et d'ainsi tenter d'identifier une plage de valeurs qui minimisent les risques, plage qui sera prise en compte lors de la conception des futurs produits.

## Option 3

Dans le cadre de cette troisième option, il serait possible de développer une interface d'aide à la décision conçue à partir du modèle prédictif présenté précédemment. Cette interface pourrait proposer, par exemple, des actions correctives pour augmenter la probabilité de réussite d'une teinture. Cependant, contrairement aux deux options précédentes, la réalisation de ce projet nécessite l'accomplissement de certains projets intermédiaires :

1. Tout d'abord, il serait nécessaire d'améliorer la récolte des données afin de permettre d'utiliser des données en temps réel ou à une fréquence plus élevée ;
2. Il serait fortement conseillé de parfaire le modèle prédictif proposé précédemment en appliquant les pistes d'améliorations discutées dans la Section « Amélioration du modèle prédictif », et ce, afin de maximiser l'utilité et la précision d'une telle interface ;
3. Dans le contexte d'une solution permettant de proposer des actions correctives, il serait nécessaire de bien connaître et de définir les corrélations entre les sources de variabilité. Plus précisément, si un tel système détecte une dérive du procédé augmentant la probabilité d'une non-conformité de la couleur (« *off-shade* »), ce même système propose alors des ajustements ou des modifications au niveau des sources de variabilité sur lesquelles l'entreprise a un certain contrôle. De plus, afin que cette solution prescriptive demeure pertinente, le modèle prédictif et l'interface devront être mis à jour selon une fréquence prédéterminée.

## Option 4

Dans un même ordre d'idées, cette quatrième option consiste à utiliser le modèle prédictif proposé afin de développer un outil de simulation qui permettra de tester des recettes de teinture et possiblement d'améliorer les recettes actuelles. Au même titre que l'option 3 présentée ci-haut, il serait conseillé d'améliorer le modèle prédictif avant de poursuivre avec cette option.

### **4.4. Étape 4 : Implantation, contrôle et suivi**

Dans le contexte de cette étude de cas, cette dernière étape d'implantation, de contrôle et de suivi n'a pas été incluse dans la portée du projet défini initialement.

## **Chapitre 5 Mise en œuvre en contexte industriel – Cas Fromagerie L’Ancêtre**

Pour ce qui est de la deuxième étude de cas effectuée dans le contexte d’une preuve de concept de la première étape du cadre méthodologique, une collaboration a été mise sur pied avec la Fromagerie L’Ancêtre, entreprise fabriquant du fromage biologique située à Bécancour. La mission de l’entreprise s’énonce comme suit : « La Fromagerie L’Ancêtre produit une importante variété de beurres et de fromages biologiques savoureux de qualité supérieure depuis 1992. La certification biologique de nos produits nous permet de participer au développement durable, de respecter l’environnement et d’offrir des aliments biologiques que l’on prend le temps de déguster avec plaisir. Notre large gamme de produits vous fait vivre le plaisir de manger au quotidien. Elle agrmente vos repas santé, est votre meilleure alliée pour la collation et accompagne merveilleusement votre petit déjeuner sans gluten » (L’Ancêtre, 2018). En addition à cela, elle préconise six valeurs : le développement durable, l’innovation et le développement de produits biologiques, l’excellence de ses pratiques, l’intégrité, la passion et le plaisir de manger en famille (L’Ancêtre, 2018).

Les opérations de l’entreprise sont séparées en deux établissements. Le premier établissement, en plus de servir de comptoir de vente de ses produits, est chargé des opérations de réception et de stockage du lait, de pasteurisation ou de thermisation, de coagulation, de finition, de mise en moule, de mise en boîte et de refroidissement du fromage. Pour ce qui est du deuxième établissement, il est chargé de stocker le fromage en boîte afin de le faire vieillir, d’effectuer la découpe, la mise en sachet ainsi que la mise en boîte.

La problématique étudiée dans le cas présent porte sur la variation de la densité et des dimensions des blocs de fromage. Après les étapes de thermisation, de coagulation, de finition, de mise en moule et de mise en boîte, le fromage est entreposé sous forme de blocs d’environ 19 kilogrammes. Cette variation engendre des pertes lors de la découpe finale du fromage.

Le procédé de découpe utilise un équipement de découpe à l’ultrason. En premier lieu, le bloc est coupé en deux plus petits blocs. Par la suite, ces blocs sont recoupés en cinq, puis enfin en neuf blocs d’approximativement 200 grammes. L’entreprise a un certain standard à respecter au niveau du pourcentage de blocs en dessous de 200 grammes et les blocs largement supérieurs à 200 grammes sont recoupés manuellement. Les retailles et les blocs de fromage non conformes sont retirés; une partie peut être utilisée pour faire du fromage râpé tandis que le reste doit être vendu à perte.

La problématique de variabilité est causée par le fait que les blocs initiaux de fromage de 19 kilogrammes ont souvent des dimensions et une densité variables d’un bloc à l’autre. Cependant, l’équipement de découpe, quant à lui, utilise des valeurs constantes de distance pour effectuer sa découpe, peu importe les dimensions du bloc.

C'est ce phénomène qui fait en sorte que les blocs finaux de la série de découpes varient autour de 200 grammes. Ce sont habituellement les premiers ou les derniers blocs qui ne respectent pas les standards de poids.

## 5.1. Étape 1 : Analyse de la situation initiale

À première vue, il apparaît que les caractéristiques du lait utilisé pour la production du fromage, certaines caractéristiques du fromage avant la mise en moule de même que les caractéristiques dimensionnelles des blocs de fromage avant la découpe sont des pistes d'exploration prometteuses pour la diminution des pertes lors du procédé de découpe :

- Pour ce qui est des caractéristiques du lait utilisé dans le procédé, il semble que seules certaines soient mesurées, mais que les données ne sont pas utilisées pour identifier des corrélations avec la performance du procédé. À titre d'exemple, l'entreprise est consciente de l'importance du taux de gras et du taux de protéine du lait de même que de leur impact sur la densité du fromage. Cependant, bien que l'entreprise mesure et conserve ces données afin de dégager des tendances, elle n'utilise pas ces données pour prédire la performance du procédé et ajuster les paramètres du procédé en conséquence. Bien que d'autres facteurs entrent en jeu pour déterminer la densité finale du fromage, le taux de gras et le taux de protéine doivent tout de même en expliquer une partie.
- Une autre piste d'amélioration concerne le processus de mise en moule du fromage. Tout d'abord, un gain relativement simple à obtenir passerait par la modification de l'étagère de moules afin de s'assurer que ceux-ci ne soient jamais mal alignés (les moules alignés non convenablement donnent une forme non standard et génèrent des pertes supplémentaires). Également, dans l'optique d'être en mesure de prédire la densité du fromage, mesurer l'humidité avant la mise en moule permettrait de corrélérer cette donnée avec la masse finale des blocs de fromage (plus le fromage est humide, plus il perdra de lactosérum lors du pressage et plus il sera léger).
- La dernière piste importante à explorer consiste à utiliser les dimensions individuelles de chaque bloc de fromage (longueur, hauteur, largeur) afin d'ajuster les paramètres de découpe, et ce, dans le but d'optimiser les patrons de découpe du fromage (tableau 9).

Tableau 9: Résumé de la découpe des blocs de fromage

État 1	Découpe	État 2	Découpe	État 3	Découpe	État 4
≈18.5 kg	→	≈9.25 kg	→	≈1.85 kg	→	≈0.2055 kg
Découpe en 2			Découpe en 5		Découpe en 9	
Découpe faite sur la base de mesures fixes						

Par ailleurs, des changements ont été apportés dernièrement au procédé de découpe afin que la découpe en cinq longes de fromage soit faite à partir d'un pourcentage de la longueur totale. Les longes de 1,85 kg (en moyenne) ainsi produites sont un peu plus uniformes, ce qui diminue les pertes.

Afin de bien représenter la compréhension de la problématique, un diagramme des processus a été bâti (Figure 12). Ce diagramme présente, en plus des étapes du procédé de production du fromage, les flux d'information et de données pour chacune de ces étapes.

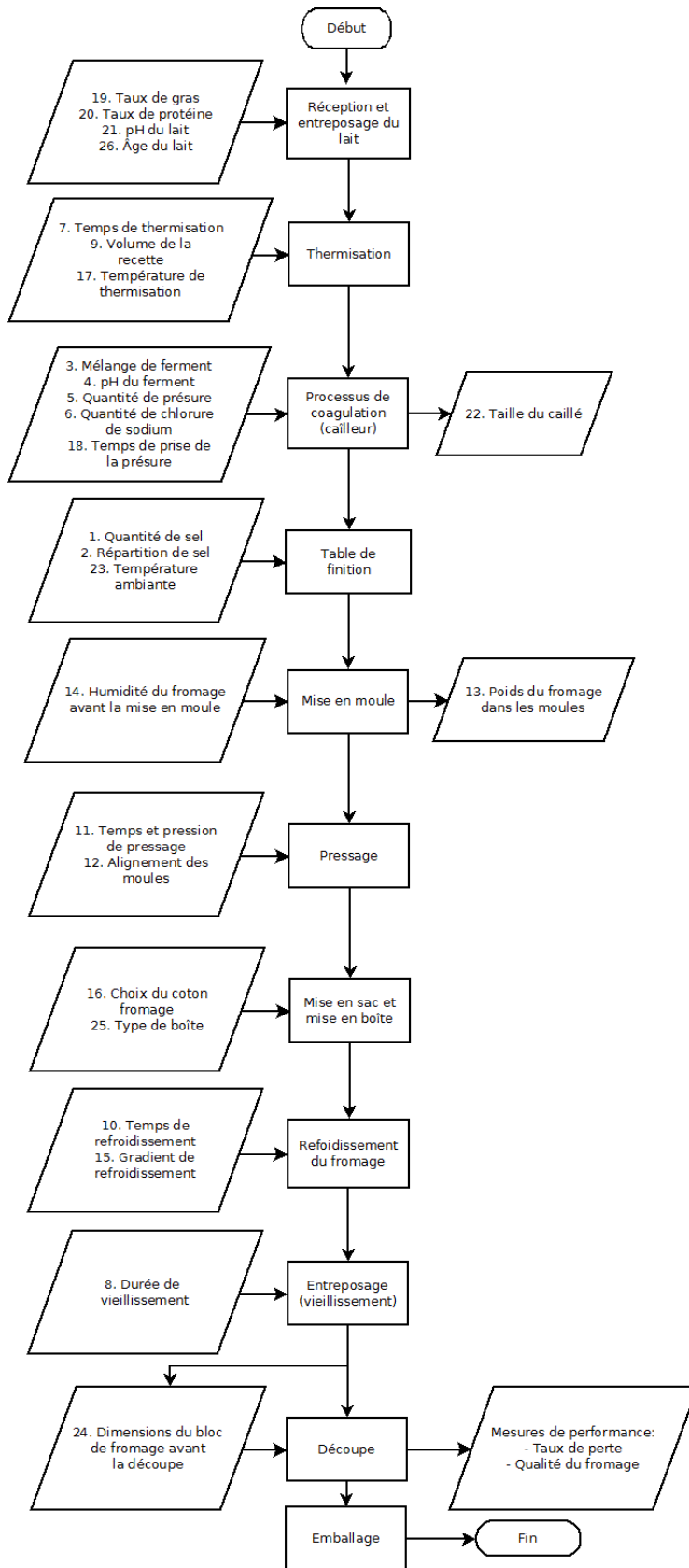


Figure 13: Diagramme des processus avec flux d'information



Dans ce diagramme, les rectangles représentent les étapes du processus tandis que les parallélogrammes représentent les entrées ou sorties d'information ou de données. L'idée est de représenter les étapes critiques du procédé qui nécessitent une quantité importante de paramètres à contrôler.

Le diagramme de la Figure 13 représente les relations entre les différentes causes indirectes et le problème de perte lors du procédé de découpe. Au fil des rencontres et des discussions qui ont eu lieu dans le cadre de ce cas d'étude, il est apparu pertinent de créer une forme de représentation pour souligner les relations entre les différentes causes indirectes et le problème de perte de fromage lors du procédé de découpe et c'est ainsi qu'il a semblé pertinent d'utiliser une carte conceptuelle. Selon cette carte conceptuelle, on remarque que les deux causes sous-jacentes principales au problème de perte lors de la découpe sont la friabilité<sup>11</sup> variable du fromage ainsi que les dimensions variables des blocs de fromage.

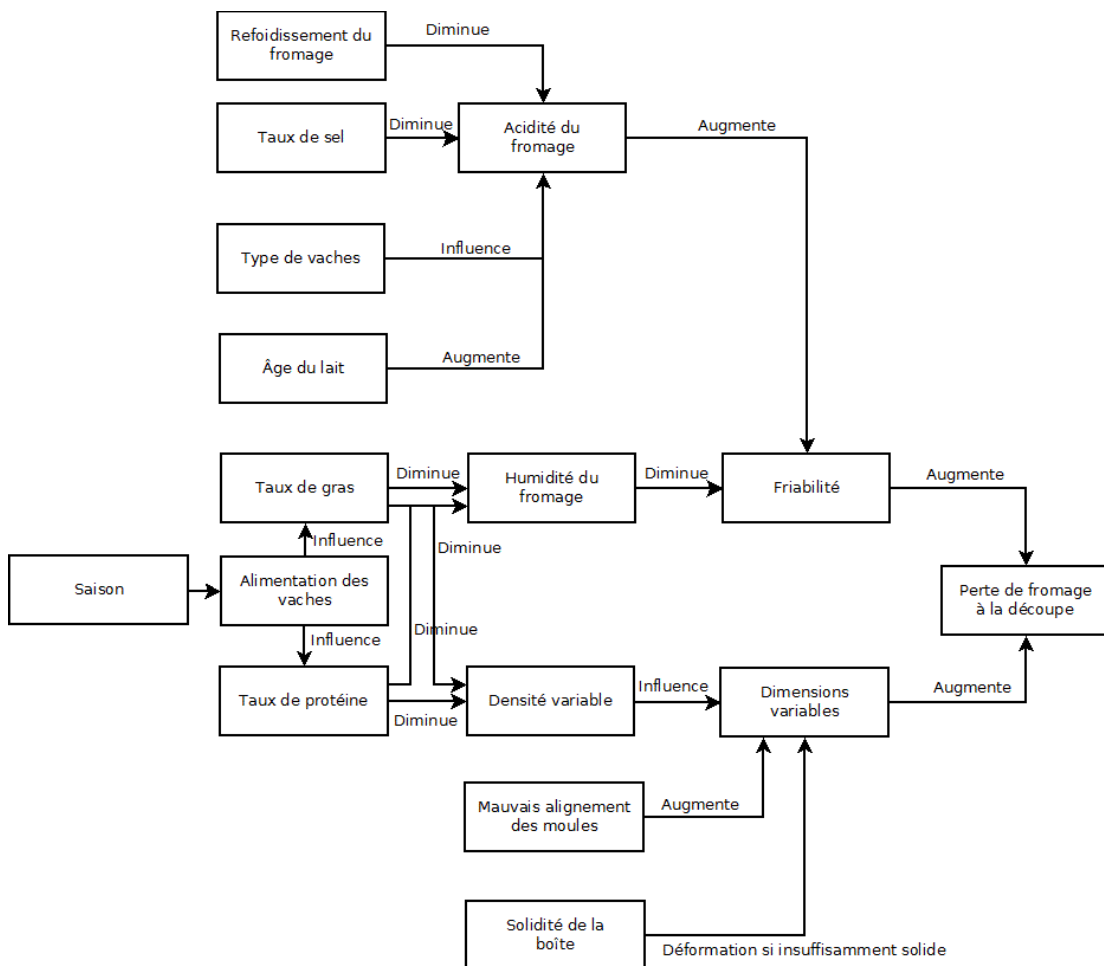


Figure 14: Carte conceptuelle représentant les causes indirectes de la perte de fromage du procédé de découpe

<sup>11</sup> Quelque chose de friable se définit comme quelque chose « qui peut aisément être réduit en petits fragments, en poudre, en poussière » selon le Larousse en ligne (<http://larousse.fr/dictionnaires/francais/friable/35225?q=Friable#35194>).

À l'instar de la première étude de cas Duvaltex, la mise au point d'un diagramme d'Ishikawa permet de présenter sous une autre perspective les sources de variabilité identifiées au cours des discussions avec le partenaire industriel. Cependant, la problématique représentée dans le diagramme suivant (Figure 15) n'est pas le taux de perte du procédé de découpage, mais bien la variation de la densité ou des dimensions des blocs de fromage puisqu'il s'agit des causes indirectes les plus importantes des pertes occasionnées lors de la découpe. De surcroît, il est important de considérer encore une fois que cette représentation ne traite aucunement de l'importance relative de ces sources de variabilité.

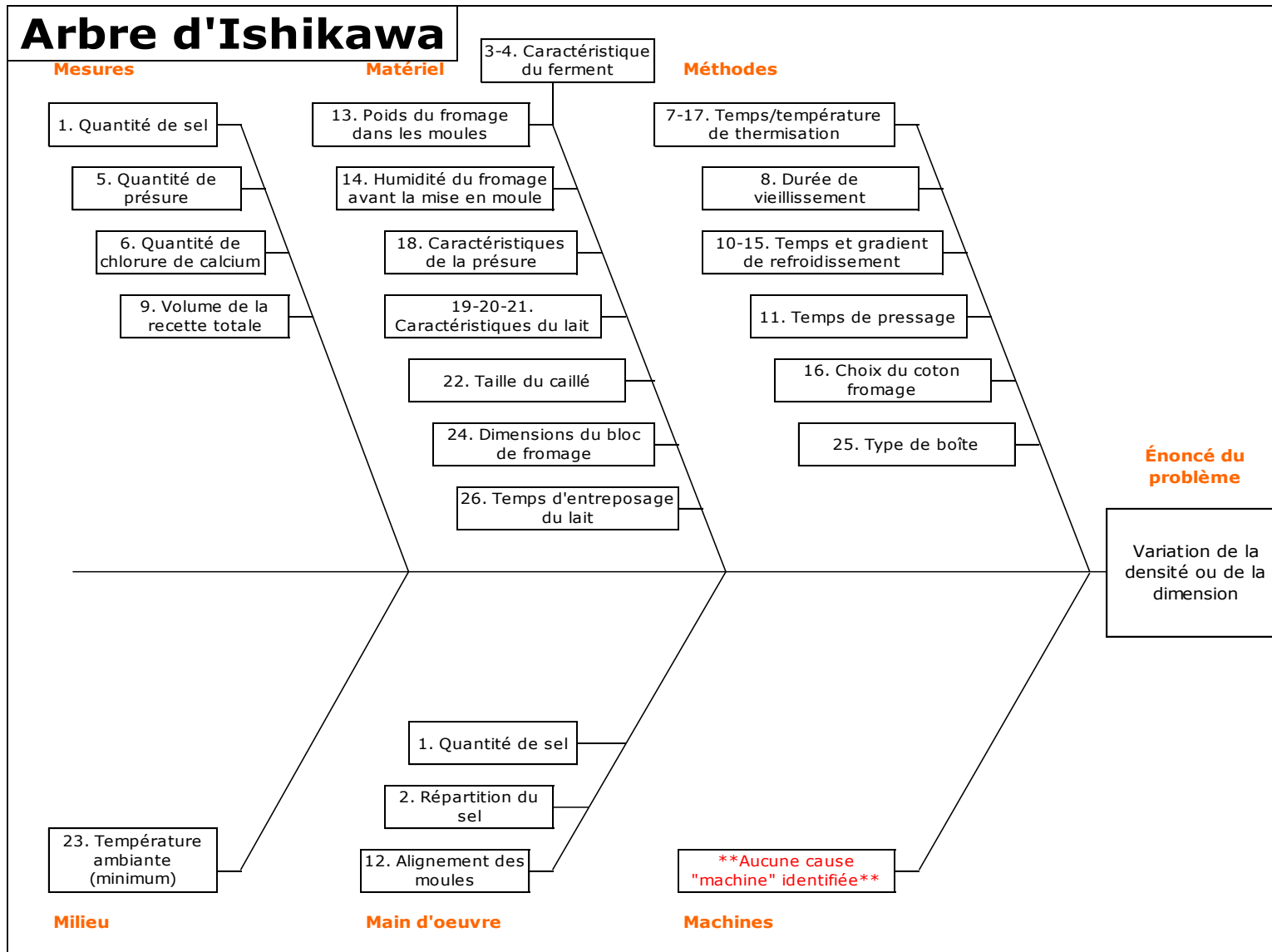


Figure 15: Représentation des sources de variabilité dans un diagramme de causes et effet pour le cas L'Ancêtre

Le Tableau 10 présente la liste des sources de variabilité identifiées lors des visites et des rencontres avec la Fromagerie L'Ancêtre. Sans dire qu'elle est complète, elle contient tout de même les principales sources identifiées par le personnel de l'entreprise qui peuvent affecter les pertes de fromage lors du procédé de découpe. Pour chacune de ces sources de variabilité, une importance relative est estimée par rapport à la variation perçue lors de la découpe et l'échelle suivante sera utilisée : très faible, faible, moyenne, forte et très forte.

Il est à noter qu'il s'agit d'une évaluation préliminaire qualitative, l'objectif étant d'abord d'identifier des pistes à prioriser. Cette évaluation est évidemment basée sur les discussions et indications du personnel technique et du personnel de gestion de l'entreprise.

Tableau 10: Présentation des différentes sources de variabilité, d'une courte description de même que de certaines précisions à savoir si elles sont ou peuvent être mesurées et contrôlées.

#	Source de variabilité	Description	Mesurée (oui ou non)	Contrôlée (oui/non) et comment	Importance relative estimée p/r à la variation	Actions à poser	
						Pour mesurer	Pour contrôler
1	Quantité de sel	Il s'agit de la quantité de sel ajoutée au fromage à l'étape de la finition. On peut rapporter ce taux en g/100g par exemple. L'introduction du sel dans la recette est reconnue comme pouvant arrêter l'acidification du fromage.	Oui	Oui, selon le volume de la recette, les opérateurs mettent le sel à la main à partir d'une chaudière (variation possible).	Faible	Une amélioration possible au niveau de l'incorporation du sel serait d'utiliser un système automatisé de distribution qui permettrait de mettre une quantité constante et précise de sel en fonction du volume total de la recette, du pH du lait et des autres paramètres de production.	Une amélioration possible au niveau de l'incorporation du sel serait d'utiliser un système automatisé de distribution qui permettrait de mettre une quantité constante et précise de sel en fonction du volume total de la recette, du pH du lait et des autres paramètres de production.
2	Répartition du sel	En plus de la quantité totale de sel mise dans le mélange lors de la finition du fromage, on doit considérer la répartition de cette quantité de sel dans le fromage. La quantité totale de sel est ainsi divisée en trois chaudières qui sont incorporées à intervalle spécifique pour permettre une meilleure incorporation. Le sert est, en plus du goût, utilisé pour stopper la progression du pH; une mauvaise répartition du sel peut donc affecter l'acidité, la texture et la saveur du fromage.	Non	Non	Faible	Une procédure d'échantillonnage multiple devrait être mise en place afin d'obtenir une mesure de la répartition du sel.	Une meilleure intégration du sel de même qu'un meilleur procédé de brassage (la problématique semble ambiguë à savoir si la répartition du sel provient du brassage ou de la méthode d'insertion du sel). Différentes options existent pour améliorer la répartition du sel via son incorporation; on peut penser à l'automatisation du processus ou bien l'amélioration de la façon de faire en se basant sur les données récoltées à l'aide d'un échantillonnage multiple.
3	Mélange de ferments	Il s'agit de la quantité de chacun des ferments du mélange de ferments.	Oui, la réception des ferments est accompagnée d'un certificat d'analyse des ferments. Cependant, les spécifications des ferments ne sont pas validées.	Oui, indirectement. Le contrôle consiste à la sélection du mélange auprès du fournisseur selon certaines tolérances.	Faible	S.O.	Une amélioration serait de mettre en place un processus de validation lors de la réception des ferments afin de confirmer les spécifications demandées, selon la faisabilité de cette validation.
4	pH du ferment	Le mélange de ferments présente une mesure d'acidité, donnée sur l'échelle pH.	Oui, la réception des ferments est accompagnée d'un certificat d'analyse des ferments. Cependant, les spécifications des ferments ne sont pas validées.	Oui, indirectement. Le contrôle consiste à la sélection du mélange auprès du fournisseur selon certaines tolérances.	Faible	S.O.	Une amélioration serait de mettre en place un processus de validation lors de la réception des ferments afin de confirmer les spécifications demandées, selon la faisabilité de cette validation.
5	Quantité de présure	Il s'agit de la quantité de présure utilisée dans le procédé de coagulation (« caillage »).	Oui	Oui, la quantité mise dans le procédé est mesurée	Faible	S.O.	S.O.

#	Source de variabilité	Description	Mesurée (oui ou non)	Contrôlée (oui/non) et comment	Importance relative estimée p/r à la variation	Actions à poser	
						Pour mesurer	Pour contrôler
				manuellement à l'aide d'un cylindre gradué.			
6	Quantité de chlorure de calcium	Il s'agit de la quantité de chlorure de calcium ajoutée au fromage.	Oui	Oui <sup>12</sup>	Très faible	S.O.	S.O.
7	Temps de thermisation	Il s'agit du temps pendant lequel le lait est thermisé (la thermisation est le procédé de chauffage du lait pour éliminer une partie des bactéries, similaire à la pasteurisation, mais moins chaud).	Oui	Oui, avec des contrôleurs électroniques.	Faible	S.O.	S.O.
8	Durée de vieillissement	Il s'agit de la durée que le fromage en question est vieilli, si vieillissement il y a.	Oui	Oui, il s'agit en fait d'une variable de contrôle du procédé.	Moyenne	La durée de vieillissement n'est pas difficile à mesurer, mais son impact indirect sur les pertes reste à préciser. La mise au point d'un historique de production ou d'un plan d'expérience permettrait de développer la compréhension de l'effet du vieillissement sur la friabilité du fromage en fonction du type de fromage.	La durée de vieillissement est déjà contrôlée puisqu'il s'agit d'une variable de contrôle.
9	Volume total de la recette	Il s'agit du volume total de lait utilisé pour faire le fromage.	Oui	Oui, à l'aide de débitmètres.	Faible	S.O.	S.O.
10	Temps de refroidissement	Il s'agit de la durée pendant laquelle le fromage en moule est refroidi. Combinée au gradient de refroidissement, cette source de variabilité a un impact sur l'acidité finale du fromage ainsi que sur sa texture et sa saveur.	Oui	Oui, l'entreprise décide la durée pendant laquelle les blocs de fromage en moule sont refroidis. De plus, le refroidissement est contrôlé indirectement via la méthode de palettisation en cheminée qui permet une meilleure circulation d'air autour des caisses de fromage.	Faible	S.O.	S.O.
11	Temps et pression de pressage	Il s'agit du temps pendant lequel le fromage est soumis à une pression dans les moules ainsi que la mesure de cette même pression. Ces deux paramètres sont spécifiques au type de fromage produit.	Oui	Oui, il s'agit d'une variable de contrôle.	Faible	S.O.	S.O.
12	Alignement des moules	Il s'agit de l'alignement des moules dans la presse de moules. Un alignement non adéquat va faire en sorte que les blocs de fromage auront une forme non adéquate et	Oui (observable)	Oui, mais il s'agit d'une opération manuelle et des erreurs peuvent survenir.	Forte	L'entreprise devrait noter chaque fois que des moules sont mal alignés et conserver l'information pour des analyses futures.	L'entreprise devrait modifier les étagères de moules pour faire en sorte qu'il soit impossible de mal aligner les moules (modifier les rails

<sup>12</sup> La façon dont le chlorure de calcium est contrôlé est à confirmer avec la fromagerie.

#	Source de variabilité	Description	Mesurée (oui ou non)	Contrôlée (oui/non) et comment	Importance relative estimée p/r à la variation	Actions à poser	
						Pour mesurer	Pour contrôler
		augmentera les pertes. Le phénomène semble provenir des couverts de moules qui sont munis de légères bosses, probablement pour faciliter l'évacuation du liquide lors de la presse.					de guidage ou les couverts des moules).
13	Poids du fromage dans les moules	Il s'agit de la quantité de fromage en termes de poids contenu dans chaque moule.	Oui, il s'agit d'une variable de contrôle selon la sorte de fromage. De plus, le poids après démoulage est mesuré et enregistré.	Oui, il s'agit d'une variable de contrôle. Cependant, la mesure du poids après démoulage est conservée, mais n'est actuellement pas utilisée pour un contrôle.	Faible	S.O.	Une meilleure connaissance et compréhension du procédé de pressage permettrait un meilleur contrôle de la quantité finale de fromage dans les moules (prendre en compte la quantité de liquide et la densité du fromage, par exemple).
14	Humidité du fromage avant la mise en moule	Il s'agit de la quantité de lactosérum contenue dans le fromage avant le procédé de pressage.	Oui, mais seulement mesurée en aval.	Non	Moyenne	Investiguer pour être capable de mesurer le taux d'humidité avant le pressage et prédire le poids du fromage.	Selon les mesures d'humidité, il sera possible d'investiguer si une corrélation existe entre l'humidité avant pressage et la densité finale du fromage.
15	Gradient de refroidissement	Il s'agit du gradient auquel le fromage est refroidi, plus précisément le nombre de degrés perdus par unité de temps (degré Celsius/heure par exemple). Le gradient a un effet sur l'acidité et la friabilité du fromage ainsi que sur sa texture et sa saveur.	Non	Oui, le refroidissement est contrôlé indirectement via la méthode de palettisation en cheminée qui permet une meilleure circulation d'air autour des caisses de fromage.	Faible	Pour être capable de mesurer le gradient de refroidissement du fromage, l'entreprise doit mesurer périodiquement la température du fromage; étudier la possibilité d'utiliser un thermocouple sans fil pour mesurer automatiquement le gradient.	L'utilisation d'un thermocouple sans fil relié au système de ventilation pourrait permettre d'avoir un contrôle adapté du gradient de refroidissement. Sinon, une méthode plus simple serait d'ajuster le refroidissement selon des mesures prises manuellement.
16	Choix du « coton fromage »	Il s'agit d'une pellicule en plastique ou en tissu utilisée dans la boîte pour absorber l'humidité du fromage. Dépendamment de la porosité de la pellicule, il peut y avoir des effets différents sur le fromage.	Oui	Oui, l'entreprise a le choix du « coton fromage ».	Forte <sup>13</sup>	S.O.	Une amélioration serait d'effectuer un plan d'expérience pour comparer les différents « cotons fromage » avec les différents produits pour déterminer quels types de « cotons fromage » minimisent les pertes lors du démoulage.
17	Température de thermisation	Il s'agit de la température à laquelle le lait est chauffé dans le procédé de thermisation.	Oui	Oui, et cette température est normalisée.	Faible	S.O.	S.O.
18	Temps de prise de la présure	Il s'agit du temps que nécessite la présure pour effectuer son action de coagulation	Oui, ces données sont enregistrées	Non	Faible	S.O.	Les données sur les temps de prise pourraient permettre de dégager les

<sup>13</sup> Il est à noter que l'importance relative de cette source de variabilité a été évaluée à moyenne puisque cette importance relative est estimée par rapport à la variation de la densité, des dimensions et de la friabilité du fromage, qui à leur tour causent des pertes à la découpe. Donc les pertes générées par le fait que le fromage colle parfois aux « cotons fromage » est un problème à part; en revanche, ce problème modifie les dimensions finales des blocs de fromage entraînant ainsi des pertes à la découpe expliquant ainsi l'importance relative estimée à moyenne.

#	Source de variabilité	Description	Mesurée (oui ou non)	Contrôlée (oui/non) et comment	Importance relative estimée p/r à la variation	Actions à poser	
						Pour mesurer	Pour contrôler
		dans le procédé de « caillage ». Ce temps peut varier selon la composition du lait (protéines et minéraux) ainsi que la quantité de présure utilisée.	dans une base de données.				relations existantes entre les caractéristiques du lait est la réaction de la présure.
19	Taux de gras	Il s'agit de la quantité de matières grasses contenue dans le lait utilisé pour la fabrication du fromage. Le taux de gras a un impact sur la friabilité du fromage de même que sur sa densité finale.	Oui, les données sont enregistrées dans une base de données.	Non, l'entreprise reçoit un mélange de lait de leur fournisseur et est contrainte d'utiliser ce lait pour leurs produits.	Forte	S.O.	Pour des raisons de normes et certifications biologiques ainsi que pour des raisons de stratégie d'affaires, l'entreprise ne peut utiliser d'additifs pour corriger le taux de gras. Une autre solution serait cependant d'utiliser les mesures récoltées pour mieux comprendre et évaluer l'impact du taux de gras sur la perte finale de fromage.
20	Taux de protéine	Il s'agit de la quantité de protéine contenue dans le lait utilisé pour la fabrication du fromage. Le taux de protéine a un impact sur la friabilité du fromage de même que sur sa densité.	Oui, les données sont enregistrées dans une base de données.	Non, l'entreprise reçoit un mélange de lait de leur fournisseur et est contrainte d'utiliser ce lait pour leurs produits.	Forte	S.O.	Pour des raisons de normes et certifications biologiques ainsi que pour des raisons de stratégie d'affaires, l'entreprise ne peut utiliser d'additifs pour corriger le taux de protéine. Une autre solution serait cependant d'utiliser les mesures récoltées pour mieux comprendre et évaluer l'impact du taux de protéine sur la perte finale de fromage.
21	pH du lait	Il s'agit du taux d'acidité du lait utilisé dans la recette de fabrication de fromage.	Oui, des mesures du pH sont faites à la réception et durant la fabrication et ces données sont enregistrées dans une base de données,	Non, l'entreprise n'effectue aucun contrôle ou correction du taux d'acidité du lait reçu pour la production. Cependant, si la mesure du pH ne respecte pas les normes du CQ, le lait peut être refusé à la réception.	Moyenne	S.O.	Une méthode pour mieux contrôler les effets de l'acidité serait d'utiliser les données du pH récoltées pour identifier une corrélation entre le taux d'acidité et les pertes générées à la fin du procédé.
22	Grosseur du caillé	Il s'agit de la taille des morceaux de caillé, l'état intermédiaire du lait avant de devenir complètement du fromage. Cette taille du « caillé » varie en fonction des couteaux utilisés ainsi que de la vitesse de rotation de ces couteaux et peut avoir un impact sur l'humidité et la texture du fromage.	Non	Non, l'entreprise ne contrôle pas cette source de variabilité.	Moyen	Pour être en mesure de mesurer quantitativement la grosseur du caillé, il serait nécessaire d'étudier les différentes méthodes qui existent sur le marché (utilisation de vision artificielle ou mesure manuelle peut-être).	Pour être en mesure de mieux contrôler cette source de variabilité, l'entreprise pourrait identifier la corrélation entre les paramètres du procédé de coagulation (quantité de présure, quantité totale de lait, couteaux utilisés, vitesse de rotation des couteaux, etc.) et la grosseur du caillé.



#	Source de variabilité	Description	Mesurée (oui ou non)	Contrôlée (oui/non) et comment	Importance relative estimée p/r à la variation	Actions à poser	
						Pour mesurer	Pour contrôler
23	Température ambiante	Il s'agit de la température ambiante dans l'usine lors du procédé de finition (table de finition). La température peut avoir un effet sur le taux d'acidité du fromage.	Non	Non, l'entreprise ne contrôle pas la température ambiante.	Faible	Mesurer <sup>14</sup> la température ambiante est très facile; probablement que l'entreprise a déjà sur place un thermomètre. Il s'agirait seulement de la mesurer de temps à autre et de conserver les données pour ensuite tenter de la corrélérer avec l'évolution du pH durant le procédé de finition.	Un système de climatisation dans un milieu manufacturier est très complexe et dispendieux lorsque ce n'est pas nécessaire.
24	Dimensions du bloc de fromage avant la découpe	Il s'agit des dimensions du bloc de fromage suite à sa sortie de la boîte et avant la découpe.	Oui	Non, l'entreprise essaie de garder les dimensions constantes, mais ne contrôle pas directement la dimension des blocs.	Forte	Une mesure sommaire des dimensions pourrait être prise manuellement puis conservée. Il semble cependant que certains capteurs de l'équipement de découpe fournissent déjà certaines mesures du bloc; cependant, prendre les mesures extérieures du bloc ne règle pas les problèmes de surface courbe, en cas de déformation du bloc. Une solution serait l'utilisation d'un scanner tridimensionnel qui serait en mesure d'indiquer l'ensemble des dimensions à l'équipement de découpe.	Obtenir les mesures de dimensions du bloc permettrait de mettre au point un programme d'optimisation afin de minimiser les pertes en fonction de ces dimensions.
25	Type de boîte	Il s'agit du type de boîte utilisé pour entreposer les blocs de fromage. Selon le type et l'épaisseur de carton utilisé, le refroidissement du fromage est affecté.	Oui	Oui, mais l'entreprise tente différents types de boîtes pour évaluer son impact sur le refroidissement.	Moyenne	S.O.	L'entreprise est déjà en train d'étudier l'impact des différents types de boîtes sur le refroidissement du fromage. Un exercice qui pourrait permettre de mieux structurer cette étude serait la mise au point d'un plan d'expérience pour tester les différents types de boîtes expérimentés jusqu'à maintenant.
26	Temps d'entreposage du lait	Il s'agit du temps entre la réception du lait et son utilisation pour la fabrication d'un lot de fromage.	Oui, cette donnée est enregistrée dans une base de données (date de réception du lait et âge du lait).	Non, mais un meilleur ordonnancement permettrait peut-être de minimiser ces délais (hypothèse) en ayant une meilleure coordination entre la réception des lots de lait et la mise en production de ces derniers.	Faible	S.O.	L'utilisation des données sur le temps d'entreposage et le taux d'acidité du lait et du fromage à la toute fin permettrait d'identifier une corrélation, si elle existe.

<sup>14</sup> La mesure systématique de la température ambiante pourrait permettre ultimement de mesurer son impact sur les pertes finales via le taux d'acidité final.

On remarque ainsi que les points soulevés au début de cette section, soit le besoin d'améliorer les mesures et le contrôle du lait, la mise en place d'une structure de contrôle du procédé de mise en moule et l'utilisation des dimensions des blocs de fromage afin d'optimiser la découpe, sont confirmés par l'importance relative des sources de variabilité 14, 19, 20, 21 et 24 du tableau 10.

La dernière perspective intéressante est de classer les sources de variabilité dans une matrice présentant le caractère contrôlable des sources en fonction de leurs impacts (connus ou inconnus) et de leur observabilité. Cette forme de matrice permet de déceler quelles sources de variabilité risquent d'être plus problématiques que les autres, par exemple, celles dont l'impact est inconnu. Il est à noter que dans cette matrice, lorsque l'on fait référence au caractère contrôlable d'une source de variabilité, il s'agit de l'état actuel des choses et non de la faisabilité technique de ce même contrôle.

Identification des sources de variabilités	Contrôlable		Non Contrôlable	
	Valeur observable	Valeur non observable	Valeur observable	Valeur non observable
<b>Observabilité</b>				
<b>Impact</b>				
<b>Impact(s) connu(s)</b>	1. Quantité de sel	18. Temps de prise de la présure	19. Taux de gras	
	2. Répartition du sel		20. Taux de protéine	
	3. Mélange de ferment		21. pH du lait	
	4. pH du ferment		22. Grosseur du caillé	
	5. Quantité de présure		23. Température ambiante	
	6. Quantité de chlorure de calcium		24. Dimensions du bloc de fromage avant la découpe	
	7. Temps de thermisation			
	8. Durée du vieillissement			
	9. Volume de la recette totale			
	10. Temps de refroidissement			
	11. Temps de pressage			
	12. Alignement des moules			
	13. Poids du fromage dans les moules			
	14. Humidité du fromage avant la mise en moule			
	15. Gradient de refroidissement			
	16. Choix du coton fromage			
	17. Température de thermisation			
<b>Impact(s) inconnu(s)</b>	25. Type de boîte			26. Temps d'entreposage du lait

Figure 16: Tableau représentant l'observabilité, la contrôlabilité ainsi que la connaissance des impacts de chaque source de variabilité du projet L'Ancêtre.

Tout comme pour le cas de Duvaltex, les discussions, analyses et représentations ont permis de dégager des recommandations d'amélioration qui ont été soumises à l'entreprise Fromagerie L'Ancêtre. Ces recommandations sont basées uniquement sur des conclusions d'un processus d'investigation hautement qualitatif et pour cette raison, une rétroaction de la part de l'entreprise a été nécessaire afin de confirmer la validité et la faisabilité de ces recommandations.

### **Recommandation n° 1 : Mesure et utilisation des caractéristiques du lait**

Comme il a été mentionné précédemment, il apparaît que les caractéristiques du lait utilisé pour la production du fromage restent relativement problématiques considérant qu'il est impossible pour l'organisation d'obtenir une constance au niveau de ces caractéristiques. Pour cette raison, la première recommandation serait l'utilisation des données des caractéristiques du lait à la réception afin de concevoir des modèles statistiques.

L'objectif ultime est d'utiliser les données sur les caractéristiques du lait afin de dégager des tendances et des corrélations statistiques entre celles-ci et les caractéristiques finales du fromage (par exemple, la densité du fromage, l'humidité avant le pressage, la friabilité du fromage, etc.). Ces corrélations statistiques permettraient de mieux comprendre l'impact des caractéristiques du lait sur la performance du procédé.

### **Recommandation n° 2 : Optimisation de la découpe**

À l'instar des dernières modifications faites sur le procédé de découpe, une seconde recommandation serait de mettre en place un processus permettant d'incorporer les dimensions et la densité des blocs de fromage dans le procédé de découpe afin d'optimiser les portions et de minimiser les pertes.

Considérant la variation importante au niveau des dimensions des blocs de fromage avant la découpe, l'utilisation d'un programme mathématique d'optimisation des patrons de coupe incorporé à la programmation de l'équipement permettrait de réduire considérablement les pertes. Plutôt que d'utiliser les données de production en amont afin de mieux contrôler ces dimensions, une solution alternative consiste à adapter le procédé de découpe aux variations dimensionnelles des blocs de fromage.

De plus, la façon de prendre les mesures dimensionnelles importe peu pour le moment : elles peuvent être prises manuellement, mais dans une perspective de transition vers l'industrie 4.0, une prise de mesures automatique serait plus appropriée.

### **Recommandation n° 3 : Amélioration du contrôle du procédé de mise en moule**

Dans l'optique d'obtenir des dimensions plus constantes des blocs de fromage, il semble que le procédé de mise en moule pourrait gagner d'un meilleur contrôle des paramètres. Tout d'abord, il serait possible de récolter et de stocker les données sur l'humidité du fromage avec le pressage afin d'être en mesure de prédire

la masse de chacun des blocs après pressage (peut-être que d'autres paramètres devront être pris en compte afin d'y parvenir).

Un autre gain possible concerne l'alignement des moules. Un alignement inadéquat des moules semble être un problème récurrent qui cause des déformations de certains blocs de fromage. Or, il s'agirait de modifier l'étagère de moule afin d'y intégrer un dispositif « anti-erreur » (« *poka-yoke* ») qui empêcherait que les moules soient mal alignés ou d'améliorer la maintenance des goupilles actuellement en place pour éviter le glissement des moules.

## **5.2. Étapes 2 à 4 : Modélisation prédictive, modélisation prescriptive et implantation, contrôle et suivi**

Contrairement à l'étude de cas Duvaltex, ces étapes du cadre méthodologique n'ont pas été amorcées dû à une certaine incompatibilité au niveau de la problématique. En effet, le fait qu'une solution quasiment triviale existait pour régler la problématique (la modification ou l'amélioration de l'équipement de découpe aurait permis d'une certaine façon de minimiser les pertes) faisait en sorte que le cas n'aurait pas été suffisamment représentatif pour l'application des autres étapes du cadre méthodologique.

## Chapitre 6 Discussion et limites de la recherche

À la lumière des résultats obtenus dans les deux cas d'étude, il semble que le cadre méthodologique générique développé dans cette recherche atteigne les objectifs préalablement définis. Bien que des méthodes visant à mieux gérer la variabilité des processus existent dans la littérature, ces méthodes ont souvent une lacune au niveau de la compréhension de l'origine de cette variabilité et c'est pourquoi un cadre méthodologique permettant d'identifier des sources de variabilité tout en permettant de minimiser l'impact de la variabilité sur la performance des procédés a été développé. Le cadre en question de même que les résultats de son application dans l'industrie textile ont d'ailleurs fait l'objet d'une publication scientifique lors d'une conférence internationale (voir Annexe 6).

Bien que l'application des étapes 1 et 2 du cadre dans le contexte d'une preuve de concept ait présenté des résultats très intéressants pour les entreprises collaboratrices tout en apportant des connaissances sur leurs procédés, cet exercice possède tout de même certaines limites. Tout d'abord, une constatation faite lors du démarrage de la modélisation prédictive est que l'analyse qualitative effectuée lors de l'étape « Analyse de la situation initiale » semble un peu trop en surface pour que la transition vers la seconde étape du cadre se fasse parfaitement. Ce manque de profondeur est dû principalement à la nature qualitative de cette analyse, mais aussi, dans certains cas, à l'incertitude liée aux sources mêmes de la variabilité (des données sont nécessaires pour confirmer qu'une source de variabilité est bel et bien une source). Une autre limite est liée à la preuve concept effectuée avec Duvaltex mais qui constitue aussi une opportunité pour la suite du projet réside au niveau de la contrainte de temps imposée par le projet de recherche en soi qui a contribué à restreindre le nombre de méthodes testées, à n'utiliser un historique de production que de quelques mois et à ne modéliser que deux styles de tissus en particulier. En contrepartie, même s'il ne s'agissait que d'une preuve de concept, l'application du cadre dans les deux cas d'études a permis d'apporter aux entreprises beaucoup de connaissances sur leur procédé.

Dans le même ordre d'idées, l'application du cadre peut devenir coûteuse en termes de temps, et ce, en fonction de la complexité de la problématique abordée. De plus, dans certains contextes, l'application du cadre peut nécessiter certaines connaissances dans le domaine de l'analyse des données (par exemple pour identifier les bonnes méthodes de modélisation) ainsi que dans le domaine d'application ciblé. Il est évident que certaines connaissances sont acquises lors de l'analyse de la situation initiale, mais selon la complexité du procédé industriel ciblé, cela pourrait s'avérer insuffisant pour développer une compréhension adéquate pour les étapes subséquentes du cadre. En revanche, si l'on compare avec d'autres méthodologies similaires telles que la méthodologie « *CRISP-DM* » mentionnée dans le Chapitre 1, le besoin en termes de connaissances et de compétences en analyse de données est moindre, ce qui fait en sorte que le cadre

méthodologique générique semble être plus approprié pour l'application dans des secteurs industriels que de telles méthodologies destinées à des problématiques d'affaires.

En comparaison avec la méthodologie « *CRISP-DM* » (voir chapitre 1, section Méthodes), probablement l'une des méthodologies qui s'apparente le plus avec le cadre méthodologique développé dans le contexte de ce projet de recherche, le cadre méthodologique générique développé ici cherche à mettre davantage l'accent sur l'apprentissage et la découverte des connaissances en ayant une approche d'analyse de la situation initiale plus approfondie et plus structurée. De surcroît, le cadre méthodologique a été conçu spécifiquement pour aborder les problématiques de variabilité intrinsèque aux procédés industriels, contrairement à la méthodologie « *CRISM-DM* » qui ne cible pas un type de problématique en particulier et qui cherche plutôt à modéliser un large ensemble de données.

Sur une autre note, il est important de considérer l'apprentissage relié à l'application du cadre méthodologique. Il est probable que l'efficacité de l'application du cadre suive la courbe d'apprentissage et devienne ainsi de plus en plus efficace, et ce, considérant aussi l'appropriation du cadre par l'utilisateur (Zangwill and Kantor, 1998). Dans ce sens, plusieurs outils sont proposés dans le chapitre 3; ces outils ont été testés et appliqués dans le contexte des cas d'étude industriels, mais ils pourraient tout de même être améliorés ou modifiés selon les situations. Un autre élément relié à l'apprentissage des utilisateurs concerne la connaissance des utilisateurs des nouvelles méthodes de modélisation. La recherche effectuée dans le domaine de la modélisation prédictive est en constante évolution et il peut s'avérer pertinent pour l'utilisateur du cadre méthodologique de consulter les plus récents avancements dans ce domaine.

Une autre limite du cadre méthodologique est sa dépendance à la qualité et la quantité des données fournies pour les étapes de modélisation. En effet, tout comme plusieurs formes de modélisation, une mauvaise qualité des données ou une quantité insuffisante de données peuvent faire en sorte que le modèle prédictif ait des performances décevantes. Cette limite souligne l'importance de la première étape du cadre, l'analyse de la situation initiale, où l'identification des mesures de performance et l'identification des sources de variabilité ont à être effectuées avec précaution et minutie.

Enfin, une dernière limite de la démarche se situe au niveau de la preuve de concept : bien que l'application des deux premières étapes du cadre ait atteint les objectifs définis initialement, la preuve de concept n'a pas pu s'étendre aux étapes 3 et 4 d'un point de vue pratique.

# Conclusion

Ce mémoire présente un cadre méthodologique générique permettant d'améliorer la gestion des sources de variabilité et de leurs impacts sur les procédés industriels. Suite à une étude de la littérature, ce cadre a été développé pour ensuite être appliqué à deux reprises dans un contexte industriel, application faite en premier lieu sur un procédé de teinture d'une entreprise de textile et ensuite sur un procédé de découpe de fromage d'une entreprise alimentaire, et ce, à titre de preuves de concept. Les conclusions de ces applications ont par la suite permis de réviser le cadre pour qu'il reflète davantage la réalité industrielle.

L'application du cadre méthodologique pour une problématique retrouvée dans l'industrie du textile a permis de développer un modèle prédictif présentant des précisions de prédiction variant de 92 à 96%, des fractions de faux négatifs variant de 27 à 61% ainsi que des fractions de faux positifs variant de 0 à 6%, et ce, en se basant sur des sources de variabilité identifiées grâce au cadre méthodologique. Cette modélisation prédictive a par la suite permis d'identifier et de proposer des pistes d'amélioration au procédé industriel ciblé par l'application du cadre.

Par contre, ce cadre méthodologique générique présente quelques limites. La preuve de concept effectuée sous la forme d'application en contexte industriel n'a été faite que partiellement. La dépendance du cadre à la qualité et la quantité de données de modélisation de même que la nécessité d'avoir une certaine base de connaissances du procédé industriel ciblé et des méthodes de modélisation sont d'autres limites de ce dernier.

En termes d'avenues de recherche d'intérêt pour ce cadre méthodologique, une application complète du cadre méthodologique en entreprise, incluant une implantation complète d'un modèle prescriptif, pourrait faire l'objet d'un projet de recherche futur. Il serait ainsi pertinent d'évaluer si l'identification initiale des sources de variabilité est suffisamment adaptée aux réalités industrielles pour se rendre jusqu'à une implantation complète et fonctionnelle.

Dans un même ordre d'idées, une limite énoncée précédemment concerne la qualité et la quantité des données. Un ajout particulièrement intéressant au cadre méthodologique serait d'y ajouter une méthodologie de diagnostic permettant d'évaluer l'adéquation de la qualité et de la quantité des données pour l'application du cadre en question, plus précisément pour les étapes de modélisation. À l'heure actuelle, le cadre méthodologique ne prévoit qu'une étape pour décliner les sources de variabilité en variables de modélisation et en données tangibles à récolter, mais il ne prévoit pas de forme d'évaluation de la qualité et de la quantité de ces données. Il pourrait s'agir d'une étape préalable à la modélisation qui permettrait de garantir une meilleure performance lors de cette modélisation. Enfin, dans le contexte d'une démarche de transition



numérique dans les milieux industriels, il serait pertinent de situer l'application du cadre dans une démarche plus globale afin de maximiser les résultats bénéfiques. Bien que la numérisation des procédés n'a été que très peu mentionnée jusqu'à maintenant, l'application du cadre permet d'obtenir des modèles et d'acquérir des connaissances qui sont préalables à un processus de numérisation.

# Références

- Abrahamson, E., 1996. Management fashion. *Academy of management review* 21, 254–285.
- Berger, R., 2014. Industry 4.0: The new industrial revolution-How Europe will succeed. *Roland Berger strategy consultants, Maart*.
- Bertin, E., Crespi, N., Magedanz, T., 2013. *Evolution of Telecommunication Services*.
- Boell, S.K., Cecez-Kecmanovic, D., 2010. Literature reviews and the hermeneutic circle. *Australian Academic & Research Libraries* 41, 129–144.
- Choudhary, A.K., Harding, J.A., Tiwari, M.K., 2009. Data mining in manufacturing: a review based on the kind of knowledge. *Journal of Intelligent Manufacturing* 20, 501.
- Cook, D.F., Shannon, R.E., 1992. A predictive neural network modelling system for manufacturing process parameters. *The International Journal Of Production Research* 30, 1537–1550.
- De Mast, J., Lokkerbol, J., 2012. An analysis of the Six Sigma DMAIC method from the perspective of problem solving. *International Journal of Production Economics* 139, 604–614.
- Commission d'enrichissement de la langue française, 2018. Fiche terminologique de l'internet des objets. URL : [http://www.granddictionnaire.com/ficheOqlf.aspx?Id\\_Fiche=26544585](http://www.granddictionnaire.com/ficheOqlf.aspx?Id_Fiche=26544585) (consulté le 19/04/19).
- Duvaltex, 2018. Site internet de Duvaltex - *Mission, vision et valeurs*. URL : <http://www.duvaltex.com/fr/entreprise/#histoire> (consulté le 25/07/18).
- Ghods, A., 2006. Dimensionality reduction a short tutorial. *Department of Statistics and Actuarial Science, Univ. of Waterloo, Ontario, Canada* 37, 38.
- Gilchrist, A., 2016. Industry 4.0: the industrial internet of things. Apress.
- Guh, R.-S., 2005. A hybrid learning-based model for on-line detection and analysis of control chart patterns. *Computers & Industrial Engineering* 49, 35–62.
- Hines, W.W., Montgomery, D.C., Goldsman, D.M., Borror, C.M., 2008. *Probability and statistics in engineering*. John Wiley & Sons.
- Hofmann, E., Rüs, M., 2017. Industry 4.0 and the current status as well as future prospects on logistics. *Computers in Industry* 89, 23–34.
- Jaccard, J., Turr, R., 2003. Interaction effects in multiple regression. Sage.
- Jain, S., Narayanan, A., Lee, Y.-T.T., 2018. Comparison of Data Analytics Approaches Using Simulation, in: *Proceedings of the 2018 Winter Simulation Conference*. 1084–1095.
- Kagermann, H., Helbig, J., Hellinger, A., Wahlster, W., 2013. *Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Securing the future of German manufacturing industry; final report of the Industrie 4.0 Working Group*. Forschungsunion.

- Kantardzic, M., 2011. *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons.
- Kleijnen, J. P. 2017. Simulation Optimization through Regression or Kriging Metamodels. In *High-Performance Simulation-Based Optimization*. Springer, Cham, 115-135.
- Konrad, B., Lieber, D., Deuse, J., 2013. Striving for Zero Defect Production: Intelligent Manufacturing Control Through Data Mining in Continuous Rolling Mill Processes, in: *Robust Manufacturing Control, Lecture Notes in Production Engineering*. Springer, Berlin, Heidelberg, 215–229.
- Kusiak, A., 2017. Smart manufacturing. *International Journal of Production Research* 1–10.
- Fromagerie L'Ancêtre, 2018. Site internet de la Fromagerie L'Ancêtre - *Mission et valeurs*. URL : <http://fromagerieancetre.com/fr/a-propos/mission-et-valeurs/> (consulté le 30/07/18).
- Larman, C., 2004. *Applying UML and patterns: an introduction to object oriented analysis and design and iterative development, 3rd ed.* Pearson Education.
- Larousse, 2018a. *Le Larousse en ligne, définition de « variable »*. URL : <https://www.larousse.fr/dictionnaires/francais/variable/81108> (consulté le 25/07/18).
- Larousse, 2018b. *Le Larousse en ligne, définition de « incertain, incertaine »*. URL : [https://www.larousse.fr/dictionnaires/francais/incertain\\_incertaine/42220](https://www.larousse.fr/dictionnaires/francais/incertain_incertaine/42220) (consulté le 25/07/18).
- Lei, Q., Yu, H., Wu, M., She, J., 2016. Modeling of complex industrial process based on active semi-supervised clustering. *Engineering Applications of Artificial Intelligence* 56, 131–141.
- Liao, Y., Deschamps, F., Loures, E. de F.R., Ramos, L.F.P., 2017. Past, present and future of Industry 4.0 - a systematic literature review and research agenda proposal. *International Journal of Production Research* 55, 3609–3629.
- Lieber, D., Konrad, B., Deuse, J., Stolpe, M., Morik, K., 2012. Sustainable Interlinked Manufacturing Processes through Real-Time Quality Prediction, in: *Leveraging Technology for a Sustainable World*. Springer, Berlin, Heidelberg, pp. 393–398.
- Lieber, D., Stolpe, M., Konrad, B., Deuse, J., Morik, K., 2013. Quality Prediction in Interlinked Manufacturing Processes based on Supervised & Unsupervised Machine Learning. *Procedia CIRP, Forty Sixth CIRP Conference on Manufacturing Systems 2013* 7, 193–198.
- Lock, D., 2007. *Project Management (9e ed.)* Gower Publishing.
- Lu, X., Liu, C., Huang, M., 2015. Online Probabilistic Extreme Learning Machine for Distribution Modeling of Complex Batch Forging Processes. *IEEE Transactions on Industrial Informatics* 11, 1277–1286.
- Lu, Z.J., Xiang, Q., Wu, Y. m, Gu, J., 2015. Application of support vector machine and genetic algorithm optimization for quality prediction within complex industrial process, in: *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*. 98–103.
- Metz, C.E., 1978. Basic principles of ROC analysis. *Semin Nucl Med* 8, 283–98.
- Ministère de l'Économie, de la Science et de l'Innovation, 2017. *Industrie 4.0 : les défis de la quatrième révolution industrielle*. URL : <https://www.economie.gouv.qc.ca/bibliotheques/outils/gestion-dune->

- entreprise/industrie-40/industrie-40-les-defis-de-la-quatrieme-revolution-industrielle/ (consulté 17/04/18).
- Moeuf, A., Pellerin, R., Lamouri, S., Tamayo-Giraldo, S., Barbaray, R., 2017. The industrial management of SMEs in the era of Industry 4.0. *International Journal of Production Research* 1–19.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Journal of Clinical Epidemiology* 62, 1006–1012.
- Morin, M., Paradis, F., Rolland, A., Wery, J., Gaudreault, J., Laviolette, F., 2015. Machine Learning-based Metamodels for Sawing Simulation, in: *Proceedings of the 2015 Winter Simulation Conference, WSC '15*. IEEE Press, 2160–2171.
- Nisbet, R., Elder, J., Miner, G., 2009. *Handbook of statistical analysis and data mining applications*. Academic Press.
- Noyel, M., 2015. *Thèse de doctorat : Contrôle intégré du pilotage d'atelier et de la qualité des produits : application à la société ACTA mobilier*. Université de Lorraine.
- Pyzdek, T., 2003. *The Six Sigma Handbook: The Complete Guide for Greenbelts, Blackbelts, and Managers at All Levels, Revised and Expanded Edition*.
- Office québécois de la langue française, 2017a. *Fiche terminologique de l'infonuagique*. URL : [http://www.granddictionnaire.com/ficheOqlf.aspx?Id\\_Fiche=26544585](http://www.granddictionnaire.com/ficheOqlf.aspx?Id_Fiche=26544585) (consulté le 19/04/19).
- Office québécois de la langue française, 2017b. *Fiche terminologique de la fabrication additive*. URL : [http://www.granddictionnaire.com/ficheOqlf.aspx?Id\\_Fiche=26544533](http://www.granddictionnaire.com/ficheOqlf.aspx?Id_Fiche=26544533) (consulté le 19/04/19).
- Office québécois de la langue française, 2017c. *Fiche terminologique des mégadonnées*. URL : [http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id\\_Fiche=26507313](http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26507313) (consulté le 28/04/19).
- Office québécois de la langue française, 2017d. *Fiche terminologique du forage de données*. URL : [http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id\\_Fiche=8375476](http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8375476) (consulté le 28/04/19).
- Office québécois de la langue française, 2017e. *Fiche terminologique de l'intelligence artificielle*. URL [http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id\\_Fiche=8385376](http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8385376) (consulté le 28/04/19).
- Salim, R., Johansson, J., 2016. The Influence of Raw Material on the Wood Product Manufacturing. *Procedia CIRP, Factories of the Future in the digital environment - Proceedings of the 49th CIRP Conference on Manufacturing Systems* 57, 764–768.
- Shapiro, A., Dentcheva, D., Ruszczyński, A., 2009. *Lectures on stochastic programming: modeling and theory*. SIAM.
- Tufféry, S., 2011. *Data mining and statistics for decision making*. Wiley Chichester.
- Van der Vorst, J.G., Beulens, A.J., 2002. Identifying sources of uncertainty to generate supply chain redesign strategies. *International Journal of Physical Distribution & Logistics Management* 32, 409–430.
- Zangwill, W.I., Kantor, P.B., 1998. Toward a Theory of Continuous Improvement and the Learning Curve. *Management Science* 44, 910–920.

Zhong, R.Y., Xu, X., Klotz, E., Newman, S.T., 2017. Intelligent Manufacturing in the Context of Industry 4.0: A Review. *Engineering* 3, 616–630.

Zhou, K., Liu, T., Zhou, L., 2015. Industry 4.0: Towards future industrial opportunities and challenges, in: *Fuzzy Systems and Knowledge Discovery (FSKD), 2015 12th International Conference on*. 2147–2152.

## Annexe 1 Équations des données de modélisation

Colonne	Description de la donnée	Description du calcul
Durée approximative du plateau de chauffage (min)	Il s'agit d'une durée moyenne calculée du plateau de chauffage dans le procédé de teinture.	<p>Cette donnée est calculée à l'aide des données des fichiers de format « .dbf » qui sont en fait des séries temporelles du procédé de teinture. Considérant que ces séries temporelles représentent 1 donnée par 2 minutes, une colonne ajoutée présente une valeur de « 2 » si la température obtenue dans la teindeuse est à moins de 1 °C de la température maximale du procédé :</p> $f(T) = \begin{cases} 2, & \text{Max}(T) - T < 1 \\ 0, & \text{Max}(T) - T \geq 1 \end{cases}$ <p>où T est la température du procédé à un temps t et où Max(T) est la température maximale obtenue durant le procédé de teinture.</p>
Débit moyen de tissu dans la teindeuse (kg/min)	Il s'agit de la quantité de tissu qui passe dans la buse Venturi par unité de temps (en kilogrammes par minutes).	Le débit moyen est calculé en multipliant la masse linéique du tissu (en kg/verges) par la vitesse moyenne de rotation du tissu dans la teindeuse (en verges/minutes).
Taux de rétractation de la longueur dans la teindeuse (%)	Il s'agit du différentiel de longueur d'un lot (dans le calcul, on présuppose que la longueur des pièces de tissu dans les différents tubes d'une	Cette variable a été calculée à l'aide des séries temporelles du procédé de teinture (les fichiers de format « .dbf »). Pour y parvenir, la première étape consistait à calculer une longueur estimée calculée en multipliant le temps de cycle (en secondes) par la vitesse (en verges/seconde). Ensuite, une moyenne mobile de même qu'une dérivée ont été

	<p>teindeuse est similaire) entre la longueur en début de procédé et la longueur en fin de procédé.</p>	<p>calculées pour chaque donnée temporelle. À l'aide de ces deux dernières données et de plusieurs autres conditions, un indicateur de stabilité du procédé est calculé pour chaque donnée temporelle. En voici la liste :</p> <p>La dérivée de la ligne ainsi que les 4 dérivées précédentes doivent être inférieures à 1 (une dérivée proche de 0 indique une valeur constante);</p> <p>La longueur calculée de la ligne de même que les 4 longueurs calculées précédentes doivent être différentes de 0;</p> <p>Il doit s'agir du même lot pour la ligne utilisée et la ligne précédente;</p> <p>Le temps de cycle de la donnée temporelle doit être supérieur à 30 secondes;</p> <p>La longueur calculée doit être inférieure à 400 (une valeur supérieure à 400 peut indiquer que le capteur magnétique a sauté un tour).</p> <p>Enfin, une fois l'indicateur de stabilité calculé pour chaque ligne de la série de données temporelles, il a suffi de déterminer, pour chaque lot, la longueur maximale et minimale du lot de tissu lorsque le procédé de teinture était stable. La différence de longueur divisée par sa longueur initiale donne ainsi le taux de rétraction dans la teindeuse :</p> $\text{Taux de rétraction (\%)} = \frac{[\text{Max}(\text{Longueur calculée}   \text{Indicateur stabilité} = \text{stable}) - \text{min}(\text{Longueur calculée}   \text{indicateur stabilité} = \text{stable})]}{\text{Max}(\text{Longueur calculée}   \text{Indicateur stabilité} = \text{stable})}$
--	---	---

## Annexe 2 Liste des variables du cas d'étude Duvaltex

Sources de variabilité	Impact estimé par rapport à la variabilité	Nombre de variables	Variables	Contrôlable	Moment où la variable est disponible (avant, pendant, après)	Explications
pH de l'eau	Forte	1	pH de la teindeuse	Oui	Avant	Mesuré une seule fois avant le démarrage du procédé de teinture.
Composition du mélange d'additifs chimiques	Faible	0	-	-	-	-
Quantité de chaque additif	Forte	10	TRIDEFOAM AGD TRICOBUFF AC-50 CONC SERA LUBE M-CF PE-CLEAR CONC TRICOSPERSER SN1**voir 3028**2% ROQUEST EA EGANAL PS LIQ. DORAFast AKS SOUDE CAUSTIQUE 50%LIQUIDE TRITEX TD (BAGS)	Oui	Avant	-
Couleur des colorants	Faible	0	-	-	-	-
Quantité de colorant	Moyenne	23	DIANIX RUBINE CC DIANIX BLACK CCR DIANIX BLEU AM SLR 200% DIANIX BLEU CC DIANIX BLEU E-R 150% DIANIX NAVY CC DIANIX RED CC DIANIX RED AM-SLR DIANIX YELLOW AM-SLR DIANIX YELLOW BROWN CC DIANIX YELLOW CC DISPERSE SCARLET Y DISPERSE VIOLET 3RL FORON BRILL. BLUE S-BG LUMACRON BLUE ERD LUMACRON YELLOW ERD TRICOBLANC RGF 250% TRICOSPERSER SN TRISÉTILE RED ACE TRISÉTILE BLUE CM TRISÉTILE BR. RED FB 200% TRISÉTILE TURQUOISE GL 200% TRISÉTILE YELLOW 3GLE 400%	Oui	Avant	-
Type de colorant	Très faible	0	-	-	-	-
Patron de tissu	Faible	0	-	-	-	Le style a été utilisé
Type de tissu	Forte	0	-	Oui	Avant	pour effectuer un tri



						dès le début de l'analyse (la modélisation n'a été faite que sur les styles « 2334 » et « 2335 »).
Nombre de tubes de la teindeuse	Faible	0	-	-	-	-
Profil de chauffage et refroidissement	Faible	2	Température du plateau de chauffage (degrés C)	Oui	Avant	-
			Durée approximative du plateau de chauffage (min)	Oui	Avant	Il s'agit d'une variable calculée. Voir Annexe 2.
Vitesse de rotation du tissu dans la teindeuse	Faible à moyenne	3	Temps de cycle moyen (sec)	Oui	Après	-
			Vitesse moyenne (vgs/min)	Oui	Après	-
			Débit moyen de tissu dans la teindeuse (kg/min)	Non	Après	Il s'agit d'une variable calculée. Voir Annexe 2.
Différentiel de pression	Faible	2	Pression différentielle moyenne (BAR)	Oui	Après	-
			Pression différentielle maximale (BAR)	Oui	Après	-
Température du séchoir	Faible	0	-	-	-	-
Vitesse de roulement du tissu dans le séchoir	Faible	0	-	-	-	-
Tension du tissu dans le séchoir	Moyenne	0	-	-	-	Considérant le contrôle en temps réel de ce paramètre, le fait qu'il s'agit d'un autre procédé et les projets déjà en cours en lien avec le procédé de séchage, cette source de variabilité a été omise délibérément.
Choix du séchoir	Très faible	0	-	-	-	-
Température ambiante	Très faible	0	-	-	-	-
Présence de polluants dans l'eau	Faible à moyenne	1	Turbidité résiduelle (UTN)	Non	Avant	-
Présence de rouille dans l'eau	Moyenne	0	-	-	-	Ces données n'étaient pas disponibles ou étaient insuffisantes (trop de données manquantes) pour les étapes de modélisation.
Taux d'alun	Faible	0	-	-	-	
Taux de polymère	Moyenne	0	-	-	-	
Taux de microsable	Très faible	0	-	-	-	
Granularité du colorant	Faible	0	-	-	-	
Répartition du colorant	Très faible	0	-	-	-	
Nombre de recyclages du tissu	Moyen	0	-	-	-	
Contamination du polymère du tissu	Moyenne à forte	0	-	-	-	
Caractéristiques du colorant	Moyen	69	Mesure L, A et B de chaque colorant (voir « Quantité de colorant »)	Non	Avant	-
Caractéristiques du tissu	AJOUTÉE	7	Masse linéique	Non	Avant	Autres sources de variabilité non identifiées au départ du projet lors de l'analyse de la situation initiale. Elles se sont ajoutées lors de la modélisation. Le pourcentage de rétraction de la longueur dans la teindeuse est une variable calculée. Voir Annexe 2.
			Poids du lot	Oui	Avant	
			Longueur initiale demandée pour le lot	Oui	Avant	
			Largeur	Oui	Avant	
			Présence d'un nœud	Non	Pendant	
			Pourcentage de rétraction de la longueur dans la teindeuse	Non	Après	
Quantité d'eau	AJOUTÉE	2	Niveau d'eau moyen (pouces)	Oui	Après	
			Niveau d'eau maximal (pouces)	Oui	Après	

## Annexe 3 Détails de la régression par la méthode des moindres carrés

Les équations de la méthode sont tirées de Hines et *al.* (2008) pp.410-411.

Tout d'abord, le modèle de régression multiple peut être défini par l'équation suivante :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \epsilon_i \quad (1)$$

Si l'on définit l'erreur par la variable  $L$ , on obtient :

$$L = \sum_{i=1}^n \epsilon_i^2 \quad (2)$$

$$L = \sum_{i=1}^n (Y_i - \beta_0 - \sum_{j=1}^k \beta_j X_{ij})^2 \quad (3)$$

Afin d'être en mesure de déterminer les coefficients  $\beta$ , on peut minimiser  $L$  par rapport aux différents  $\beta_k$ . Les estimateurs  $\xi_k$  des coefficients  $\beta_k$  doivent donc satisfaire à ces deux équations :

$$\left. \frac{\partial L}{\partial \beta_0} \right|_{\xi_0, \xi_1, \dots, \xi_k} = -2 \sum_{i=1}^n (Y_i - \xi_0 - \sum_{j=1}^k \xi_j X_{ij}) = 0 \quad (4)$$

Et

$$\left. \frac{\partial L}{\partial \beta_j} \right|_{\xi_0, \xi_1, \dots, \xi_k} = -2 \sum_{i=1}^n (Y_i - \xi_0 - \sum_{j=1}^k \xi_j X_{ij}) X_{ij} = 0 \quad j=1,2,\dots,k \quad (5)$$

La simplification des équations 4 et 5 permet d'obtenir les « équations normales » de la méthode des moindres carrés :

$$\begin{aligned} n\xi_0 + \xi_1 \sum_{i=1}^n X_{i1} + \xi_2 \sum_{i=1}^n X_{i2} + \dots + \xi_k \sum_{i=1}^n X_{ik} &= \sum_{i=1}^n Y_i \\ \xi_0 + \xi_1 \sum_{i=1}^n X_{i1}^2 + \xi_2 \sum_{i=1}^n X_{i1}X_{i2} + \dots + \xi_k \sum_{i=1}^n X_{i1}X_{ik} &= \sum_{i=1}^n X_{i1}Y_i \\ \dots & \dots \\ \xi_0 + \xi_1 \sum_{i=1}^n X_{ik}X_{i1} + \xi_2 \sum_{i=1}^n X_{ik}X_{i2} + \dots + \xi_k \sum_{i=1}^n X_{ik}X_{ik} &= \sum_{i=1}^n X_{ik}Y_i \end{aligned} \quad (6)$$

Avec l'utilisation de la notation matricielle pour la résolution de ces « équations normales », on obtient donc une équation où l'on cherche à minimiser  $L$  :

$$L = \sum_{i=1}^n \epsilon_i^2 = \epsilon' \epsilon = (Y - X\beta)'(Y - X\beta) \quad (7)$$

Où

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & \dots & X_{1k} \\ 1 & X_{21} & \dots & X_{2k} \\ \dots & \dots & \dots & \dots \\ 1 & X_{n1} & \dots & X_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix} \quad \text{et} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{bmatrix}$$

Avec quelques manipulations de l'équation 7, on cherche ainsi à satisfaire l'équation suivante :

$$\left. \frac{\partial L}{\partial \beta} \right|_{\xi} = -2X'Y + 2X'X\xi = \mathbf{0} \quad (8)$$

Et enfin :

$$\xi = (X'X)^{-1}X'Y \quad (9)$$

Quoiqu'il ne soit pas précisé dans la documentation du logiciel JMP comment les estimateurs de la régression multiple par la méthode des moindres carrés sont déterminés, il s'agit ici d'un exemple analytique démontrant comment calculer ces derniers.

## Annexe 4 Détails de la régression nominale logistique

Les équations de la méthode sont tirées de Tufféry (2011) pp.437-449.

Dans un premier temps, cette méthode tente de modéliser une variable  $Y$  en fonction d'un ensemble de régresseurs  $x_i$  qu'il est possible de mettre dans une matrice  $X$  :

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_p \end{bmatrix} \quad (10)$$

De cette façon, on tente de définir l'espérance de  $Y$  en fonction des valeurs de  $X$  :

$$E(Y|X = x) = Prob(Y = 1|X = x) \quad (11)$$

S'il s'agissait d'une régression linéaire, nous pourrions utiliser un hyperplan afin d'estimer  $Y$  en fonction de  $X$  :

$$E(Y|X = x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (12)$$

Cependant, dans certaines situations où cette forme linéaire n'est pas adaptée, il est possible d'utiliser une fonction de transfert et sa fonction inverse, la fonction de liaison. Dans le cas d'étude présentée dans ce mémoire, la fonction de transfert utilisée est la fonction logit :

$$\text{Fonction de transfert} = \frac{e^t}{1+e^t} \quad (13)$$

$$\text{Fonction de liaison} = \log\left(\frac{\mu}{1-\mu}\right) \quad (14)$$

Avec la fonction de transfert, on peut modéliser la probabilité de  $Y$  par :

$$Prob(Y = 1|X = x) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}} = \pi(x) \quad (15)$$

Contrairement à la régression par la méthode des moindres carrés, les coefficients de la régression nominale logistique (ou dans ce cas-ci, binaire logistique) sont estimés en maximisant la fonction de vraisemblance. La fonction de vraisemblance consiste à définir un estimateur  $\beta$  (pour une distribution discrète) suivant l'équation :

$$L(\beta, x^1, x^2, \dots, x^n) = Prob_{\beta}(X = x^1) * Prob_{\beta}(X = x^2) * \dots * Prob_{\beta}(X = x^n) \quad (16)$$

En revanche, il n'existe pas de méthode analytique pour déterminer le maximum de  $L$  et c'est pourquoi des méthodes numériques itératives comme les algorithmes Newton-Raphson ou Fisher (Tufféry, 2011) sont utilisées.

# Annexe 5 Détails des modèles prédictifs du cas d'étude Duvaltex

## Modèle 1

Prédiction de la conformité du lot, soit  $Y =$

-2.71193358  
+ 0.0249324195 • Température du plateau de chauffage (degrés C)  
+ 0.0029677683 • Temps de cycle moyen (sec)  
+ 0.0019473655 • Durée approximative du plateau de chauffage (min)  
+ -0.007469297 • Niveau d'eau moyen (pouces)  
+ -0.462573215 • Pression différentielle moyenne (BAR)  
+ 0.0881597616 • TRIDEFOAM AGD  
+ -0.032781123 • TRICOBUFF AC-50 CONC  
+ 0.0021116813 • SERA LUBE M-CF  
+ -0.058294792 • PE-CLEAR CONC  
+ 0.0099504978 • ROQUEST EA  
+ -0.012425039 • DORAFast AKS  
+ -0.10987927 • SOUDE CAUSTIQUE 50%LIQUIDE  
+ 0.0001318305 • Poids de la pièce (kg)  
+ 0.0007659894 • Longueur initiale demandée pour le lot (vgs)  
+ 1.2656235183 • Pourcentage de rétraction longueur dans la teindeuse (%)  
+ -0.100170236 • Turbidité résiduel (UTN)  
+ 0.1467911348 • DIANIX BLACK CCR  
+ 2.6131696912 • DIANIX BLEU CC  
+ -0.100336867 • LUMACRON BLUE ERD  
+ -0.087051339 • DIANIX BLACK CCR [L]  
+ 4.9458842916 • DIANIX BLACK CCR [A]  
+ -0.155352581 • LUMACRON BLUE ERD [A]  
+ -0.393383681 • DIANIX BLACK CCR [B]  
+ 0.045900187 • DIANIX BLEU CC [B]

Où la présence d'un « *off-shade* » (non-conformité) est déterminée par l'équation suivante :

$$Si \begin{cases} Prédiction Y \geq 0.5 \rightarrow Off - shade = 0 \\ Prédiction Y < 0.5 \rightarrow Off - shade = 1 \end{cases}$$

## Modèle 2

Équation de la fonction de liaison de la régression nominale logistique, soit  $\text{Lin}[0] =$

11.308268162  
+ -1.471426031 • pH teindeuse  
+ -0.063347076 • Durée approximative du plateau de chauffage (min)  
+ 0.1185610226 • Niveau d'eau moyen (pouces)  
+ 0.0412679507 • Débit moyen de tissu dans la teindeuse (kg/min)  
+ -3.580338091 • TRIDEFOAM AGD  
+ 1.259476718 • TRICOBUFF AC-50 CONC  
+ -0.945900582 • EGANAL PS LIQ.  
+ 0.4641593014 • DORAFast AKS  
+ -0.005175967 • Poids de la pièce (kg)  
+ -0.009316382 • Longueur initiale demandée pour le lot (vgs)  
+ -46.59618464 • Pourcentage de rétraction longueur dans la teindeuse (%)  
+ 1.4564152523 • Turbidité résiduel (UTN)  
+ -19.92565931 • DIANIX RED CC  
+ -0.832628458 • LUMACRON BLUE ERD  
+ -9.777561022 • LUMACRON YELLOW ERD  
+ 14.16586989 • TRISETILE RED ACE  
+ 0.7228333425 • DIANIX BLACK CCR [L]  
+ 0.0411781793 • DIANIX RUBINE CC [A]  
+ -52.35713848 • DIANIX BLACK CCR [A]  
+ 0.235880973 • DIANIX BLEU CC [A]  
+ 12.124634452 • LUMACRON BLUE ERD [A]  
+ 4.8891380098 • DIANIX BLACK CCR [B]  
+ 0.7630187207 • DIANIX RED CC [B]



Probabilité d'obtenir une non-conformité de la couleur (« *off-shade* »), soit Prob[0] =

$$\frac{1}{\left(1 + \text{Exp} \left(-\text{Lin}[0]\right)\right)}$$

## Modèle 3

Prédiction de la conformité du lot (deux parties), soit  
 $Y =$

$$\begin{aligned}
 & -1.453703336 \\
 & + \text{Match}(\text{Merge}(0=\text{NON}, 1=\text{OUI})? \begin{cases} "0" \Rightarrow -0.021063511 \\ "1" \Rightarrow 0.0210635106 \\ \text{else} \Rightarrow . \end{cases} ) \\
 & + 0.0019936929 \cdot \text{pH teindeuse} \\
 & + 0.0175991785 \cdot \text{Température du plateau de chauffage (degrés C)} \\
 & + 0.0023458569 \cdot \text{Durée approximative du plateau de chauffage (min)} \\
 & + -0.007267251 \cdot \text{Niveau d'eau moyen (pouces)} \\
 & + -0.002241063 \cdot \text{Niveau d'eau maximal (pouces)} \\
 & + -0.148587593 \cdot \text{Pression différentielle moyenne (BAR)} \\
 & + -3.696386839 \cdot \text{Masse linéique (kg/vgs)} \\
 & + -0.000290663 \cdot \text{Débit moyen de tissu dans la teindeuse (kg/min)} \\
 & + -0.042865292 \cdot \text{TRICOBUFF AC-50 CONC} \\
 & + -0.002514265 \cdot \text{SERA LUBE M-CF} \\
 & + 0.0097482999 \cdot \text{ROQUEST EA} \\
 & + 0.0804334744 \cdot \text{EGANAL PS LIQ.} \\
 & + -0.019699454 \cdot \text{SOUDE CAUSTIQUE 50\%LIQUIDE} \\
 & + 0.0001180054 \cdot \text{Poids de la pièce (kg)} \\
 & + 0.0008044277 \cdot \text{Longueur initiale demandée pour le lot (vgs)} \\
 & + 3.0317787604 \cdot \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} \\
 & + -0.10817188 \cdot \text{Turbidité résiduel (UTN)} \\
 & + 0.0542945961 \cdot \text{DIANIX BLACK CCR} \\
 & + 2.4124591331 \cdot \text{DIANIX BLEU CC} \\
 & + 0.8291545098 \cdot \text{DISPERSE VIOLET 3RL} \\
 & + -0.034211557 \cdot \text{TRISETILE BLUE CM} \\
 & + 0.0259532044 \cdot \text{TRISETILE TURQUOISE GL 200\%} \\
 & + 0.0175130732 \cdot \text{Largeur (vgs)} \\
 & + -0.07168645 \cdot \text{DIANIX BLACK CCR [L]} \\
 & + 3.4853361784 \cdot \text{DIANIX BLACK CCR [A]} \\
 & + 0.2588491327 \cdot \text{DIANIX BLEU E-R 150\% [A]} \\
 & + -0.000935092 \cdot \text{DIANIX RED CC [A]} \\
 & + 0.0065019686 \cdot \text{DIANIX YELLOW BROWN CC [A]} \\
 & + -0.050143523 \cdot \text{DIANIX YELLOW CC [A]} \\
 & + -0.126971523 \cdot \text{LUMACRON BLUE ERD [A]} \\
 & + -0.374709808 \cdot \text{DIANIX BLACK CCR [B]} \\
 & + 0.04326607 \cdot \text{DIANIX BLEU CC [B]} \\
 & + \left( \text{pH teindeuse} - 4.7931322259 \right) \\
 & + \left( \left( \text{Durée approximative du plateau de chauffage (min)} - 43.368770764 \right) \cdot 0.0094035617 \right) \\
 & + \left( \text{pH teindeuse} - 4.7931322259 \right) \\
 & + \left( \left( \text{Niveau d'eau moyen (pouces)} - 11.710687361 \right) \cdot -0.021888669 \right) \\
 & + \left( \text{Niveau d'eau maximal (pouces)} - 17.396465382 \right) \\
 & + \left( \left( \text{Pression différentielle moyenne (BAR)} - 0.5257599878 \right) \cdot -0.124566629 \right) \\
 & + \left( \text{Durée approximative du plateau de chauffage (min)} - 43.368770764 \right) \\
 & + \left( \left( \text{Masse linéique (kg/vgs)} - 0.4335773874 \right) \cdot 0.1014762642 \right) \\
 & + \left( \text{Niveau d'eau moyen (pouces)} - 11.710687361 \right) \\
 & + \left( \left( \text{Masse linéique (kg/vgs)} - 0.4335773874 \right) \cdot 0.4831980628 \right)
 \end{aligned}$$

$$\begin{aligned}
 & + \left( \text{Pression différentielle moyenne (BAR)} - 0.5257599878 \right) \\
 & + \left( \left( \text{Masse linéique (kg/vgs)} - 0.4335773874 \right) \cdot 33.219542043 \right) \\
 & + \left( \text{Débit moyen de tissu dans la teindeuse (kg/min)} - 74.844009105 \right) \\
 & + \left( \left( \text{Débit moyen de tissu dans la teindeuse (kg/min)} - 74.844009105 \right) \cdot 0.000987062 \right) \\
 & + \left( \text{pH teindeuse} - 4.7931322259 \right) \cdot \left( \left( \text{Poids de la pièce (kg)} - 538.18845515 \right) \cdot -0.00041835 \right) \\
 & + \left( \text{Pression différentielle moyenne (BAR)} - 0.5257599878 \right) \\
 & + \left( \left( \text{Poids de la pièce (kg)} - 538.18845515 \right) \cdot -0.002171163 \right) \\
 & + \left( \text{Niveau d'eau moyen (pouces)} - 11.710687361 \right) \\
 & + \left( \left( \text{Longueur initiale demandée pour le lot (vgs)} - 348.28502091 \right) \cdot 0.0005677124 \right) \\
 & + \left( \text{Débit moyen de tissu dans la teindeuse (kg/min)} - 74.844009105 \right) \\
 & + \left( \left( \text{Longueur initiale demandée pour le lot (vgs)} - 348.28502091 \right) \cdot -0.000077835 \right) \\
 & + \left( \text{pH teindeuse} - 4.7931322259 \right) \\
 & + \left( \left( \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \right) \cdot 3.0242865322 \right) \\
 & + \left( \text{Durée approximative du plateau de chauffage (min)} - 43.368770764 \right) \\
 & + \left( \left( \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \right) \cdot -0.04318609 \right) \\
 & + \left( \text{Niveau d'eau moyen (pouces)} - 11.710687361 \right) \\
 & + \left( \left( \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \right) \cdot 0.407501701 \right) \\
 & + \left( \text{Niveau d'eau maximal (pouces)} - 17.396465382 \right) \\
 & + \left( \left( \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \right) \cdot -0.37346354 \right) \\
 & + \left( \text{Masse linéique (kg/vgs)} - 0.4335773874 \right) \\
 & + \left( \left( \left( \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \right) \right) \cdot 58.397762347 \right) \\
 & + \left( \text{Poids de la pièce (kg)} - 538.18845515 \right) \\
 & + \left( \left( \left( \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \right) \right) \cdot -0.002257618 \right) \\
 & + \left( \text{Longueur initiale demandée pour le lot (vgs)} - 348.28502091 \right) \\
 & + \left( \left( \left( \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \right) \right) \cdot -0.048146805 \right) \\
 & + \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \\
 & + \left( \left( \left( \left( \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \right) \right) \right) \cdot 8.265158215 \right) \\
 & + \left( \text{Durée approximative du plateau de chauffage (min)} - 43.368770764 \right) \\
 & + \left( \left( \text{Turbidité résiduel (UTN)} - 0.4702104097 \right) \cdot -0.012779953 \right) \\
 & + \left( \text{Longueur initiale demandée pour le lot (vgs)} - 348.28502091 \right) \\
 & + \left( \left( \text{Turbidité résiduel (UTN)} - 0.4702104097 \right) \cdot 0.0036307397 \right) \\
 & + \left( \text{Masse linéique (kg/vgs)} - 0.4335773874 \right) \\
 & + \left( \left( \text{Largeur (vgs)} - 62.110770584 \right) \cdot 0.3677069779 \right) \\
 & + \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \\
 & + \left( \left( \text{Largeur (vgs)} - 62.110770584 \right) \cdot -0.341343446 \right)
 \end{aligned}$$

Où la présence d'un « *off-shade* » (non-conformité) est déterminée par l'équation suivante :

$$S_i \begin{cases} \text{Prédiction } Y \geq 0.5 \rightarrow \text{Off - shade} = 0 \\ \text{Prédiction } Y < 0.5 \rightarrow \text{Off - shade} = 1 \end{cases}$$

## Modèle 4

Équation de la fonction de liaison de la régression

nominale logistiques, soit  $\text{Lin}[0] =$

2421.9439641

+ Match(Nœud (0=NON, 1=OUI)?  $\begin{pmatrix} \text{"0"} \Rightarrow 226.90287245 \\ \text{"1"} \Rightarrow -226.9028724 \\ \text{else} \Rightarrow . \end{pmatrix}$ )

+ -12.76101649 • Température du plateau de chauffage (degrés C)

+ -10.4511919 • Durée approximative du plateau de chauffage (min)

+ -10.74393498 • Niveau d'eau moyen (pouces)

+ -0.409092426 • Niveau d'eau maximal (pouces)

+ 763.19787164 • Pression différentielle moyenne (BAR)

+ -246.04908 • Pression différentielle maximale (BAR)

+ -1531.181787 • Masse linéique (kg/vgs)

+ -1.868702528 • Vitesse moyenne (vgs/min)

+ 4.464468896 • Débit moyen de tissu dans la teindeuse (kg/min)

+ 1.1668698429 • SERA LUBE M-CF

+ -0.088234905 • Poids de la pièce (kg)

+ -0.15010127 • Longueur initiale demandée pour le lot (vgs)

+ -600.6064857 • Pourcentage de rétraction longueur dans la teindeuse (%)

+ 12.667859643 • Turbidité résiduel (UTN)

+ -330.9036816 • DIANIX RED CC

+ -52.68053057 • DIANIX YELLOW CC

+ 2.9930869139 • LUMACRON BLUE ERD

+ 33.368616488 • TRISITILE RED ACE

+ 30.701154932 • TRISITILE BLUE CM

+ 0.9179811165 • Largeur (vgs)

+ 2.5707796434 • DIANIX BLACK CCR [L]

+ -0.899814115 • DIANIX YELLOW BROWN CC [L]

+ 300.93785718 • LUMACRON BLUE ERD [A]

+ 15.442073011 • DIANIX RED CC [B]

$\left( \text{Température du plateau de chauffage (degrés C)} - 129.48744608 \right)$

+  $\cdot \left( \left( \text{Température du plateau de chauffage (degrés C)} - 129.48744608 \right) \cdot 19.306057365 \right)$

$\left( \text{Niveau d'eau moyen (pouces)} - 11.710687361 \right)$

+  $\cdot \left( \left( \text{Niveau d'eau moyen (pouces)} - 11.710687361 \right) \cdot -4.730909873 \right)$

$\left( \text{Durée approximative du plateau de chauffage (min)} - 43.368770764 \right)$

+  $\cdot \left( \left( \text{Niveau d'eau maximal (pouces)} - 17.396465382 \right) \cdot -0.856465407 \right)$

$\left( \text{Niveau d'eau moyen (pouces)} - 11.710687361 \right)$

+  $\cdot \left( \left( \text{Niveau d'eau maximal (pouces)} - 17.396465382 \right) \cdot 3.5198004315 \right)$

$\left( \text{Niveau d'eau maximal (pouces)} - 17.396465382 \right)$

+  $\cdot \left( \left( \text{Pression différentielle moyenne (BAR)} - 0.5257599878 \right) \cdot 175.83099392 \right)$

$\left( \text{Durée approximative du plateau de chauffage (min)} - 43.368770764 \right)$

+  $\cdot \left( \left( \text{Pression différentielle maximale (BAR)} - 0.6462077076 \right) \cdot -72.1956726 \right)$

$\left( \text{Durée approximative du plateau de chauffage (min)} - 43.368770764 \right)$

+  $\cdot \left( \left( \text{Masse linéique (kg/vgs)} - 0.4335773874 \right) \cdot 20.211292394 \right)$

$\left( \text{Niveau d'eau moyen (pouces)} - 11.710687361 \right)$

+  $\cdot \left( \left( \text{Masse linéique (kg/vgs)} - 0.4335773874 \right) \cdot -295.2177807 \right)$

$\left( \text{Température du plateau de chauffage (degrés C)} - 129.48744608 \right)$

+  $\cdot \left( \left( \text{Vitesse moyenne (vgs/min)} - 172.74387753 \right) \cdot -1.109476076 \right)$

$\left( \text{Pression différentielle maximale (BAR)} - 0.6462077076 \right)$

+  $\cdot \left( \left( \text{Débit moyen de tissu dans la teindeuse (kg/min)} - 74.844009105 \right) \cdot -34.87616731 \right)$

$\left( \text{Pression différentielle maximale (BAR)} - 0.6462077076 \right)$

+  $\cdot \left( \left( \text{Poids de la pièce (kg)} - 538.18845515 \right) \cdot 0.8513958631 \right)$

$\left( \text{Niveau d'eau moyen (pouces)} - 11.710687361 \right)$

+  $\cdot \left( \left( \text{Longueur initiale demandée pour le lot (vgs)} - 348.28502091 \right) \cdot -0.090604232 \right)$

$\left( \text{Vitesse moyenne (vgs/min)} - 172.74387753 \right)$

+  $\cdot \left( \left( \text{Longueur initiale demandée pour le lot (vgs)} - 348.28502091 \right) \cdot -0.028555703 \right)$

$\left( \text{Durée approximative du plateau de chauffage (min)} - 43.368770764 \right)$

+  $\cdot \left( \left( \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \right) \cdot 98.983970831 \right)$

$\left( \text{Niveau d'eau moyen (pouces)} - 11.710687361 \right)$

+  $\cdot \left( \left( \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \right) \cdot -471.8123156 \right)$

$\left( \text{Niveau d'eau maximal (pouces)} - 17.396465382 \right)$

+  $\cdot \left( \left( \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \right) \cdot 246.65609404 \right)$

$\left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right)$

+  $\cdot \left( \left( \left( \text{Pourcentage de rétraction longueur dans la teindeuse (\%)} - 0.1389017825 \right) \right) \cdot 1715.8880406 \right)$

$\left( \text{Durée approximative du plateau de chauffage (min)} - 43.368770764 \right)$

+  $\cdot \left( \left( \text{Turbidité résiduel (UTN)} - 0.4702104097 \right) \cdot 15.82340806 \right)$

$\left( \text{Durée approximative du plateau de chauffage (min)} - 43.368770764 \right)$

+  $\cdot \left( \left( \text{Largeur (vgs)} - 62.110770584 \right) \cdot -0.61215362 \right)$

Probabilité d'obtenir une non-conformité de la couleur (« *off-shade* »), soit Prob[0] =

$$\frac{1}{\left(1 + \text{Exp} \left(-\text{Lin}[0]\right)\right)}$$

# Annexe 6 Article de conférence (Manufacturing Modelling, Management and Control – 9th MIM 2019 Conference)<sup>15</sup>

## A data-driven framework to deal with intrinsic variability of industrial processes: An application in the textile industry

Patrice Lajoie, Jonathan Gaudreault, Nadia Lehoux, Maha Ben Ali

*CRISI Research Consortium for Industry 4.0 Systems Engineering,  
Université Laval, Québec City, Québec, CANADA (e-mail: patrice.lajoie.1@ulaval.ca,  
jonathan.gaudreault@ift.ulaval.ca, nadia.lehoux@gmc.ulaval.ca, maha.ben-ali.1@ulaval.ca)*

---

**Abstract:** In many industries (e.g. natural resources processing, food processing, etc.), variation is intrinsic to the process. Data captured by advanced technologies and sensors can drive in-depth reflection to avoid and resolve potential issues related to process variability. In this study, we offer practical steps as a guidance for production engineers to deal with the intrinsic variability of industrial processes. We propose a methodological framework in three steps, which includes the definition and the characterization of the industrial process, predictive modeling and prescriptive analytics. In addition, we illustrate the framework application with a real case study from the textile industry. Copyright © 2019 IFAC.

**Keywords:** Probabilistic & statistical models in industrial plant control, Industrial and applied mathematics for production, Industry 4.0, intrinsic variability, data analytics

---

### 1. INTRODUCTION

In industry, processes variability affects system performance and is considered a major cause of defect (Yeh and *al.*, 2012; Pinarbasi and *al.*, 2016). Detection of variability sources (related to raw material, the process itself or external factors) is substantial to improve performance and to monitor process variability.

Literature usually focuses on final-product quality issues and the problem is addressed using “problem-solving” approaches, having better process control in mind. Six Sigma’s DMAIC is a well-known approach regarding to this.

However, in many industries, variation is intrinsic to the process, instead of being the symptom of sub-optimal process control. A good example of this is wood lumber production; trees are highly variable, which makes it difficult to determine the basket of products resulting for a given log (Morin *et al.*, 2015). Another example is the food processing industry. Recipes must be adapted according to many factors (e.g. characteristics of ingredients such as humidity, size of flakes, etc). Therefore, the goal is really

to “control the process very well”, but to “understand and define a bunch of alternative processes”.

For Industry 4.0 factories, advanced technologies and sensors provide large amounts of different types of data. Analytics is needed in order to uncover hidden patterns, correlations and other useful information.

The objective of this paper is to propose a standard framework/method allowing production engineers to systematically identify the sources of intrinsic variability, to measure/document their effect, in order to know them and define alternative production processes/practices. The framework was developed while working with different companies related to the CRISI Research Consortium for Industry 4.0 Systems Engineering. To support the presentation of the framework, we will describe an actual case study that was carried out in a textile dyeing mill.

The remainder of this paper is organized as follows. Section 2 gives an overview of related work and presents how predictive tools and data-mining techniques can be used for classic variability reduction. Section 3 presents the proposed framework to deal with intrinsic variability

---

<sup>15</sup> Il ne s’agit pas du format final de l’article mais le contenu est représentatif.

and illustrates how it can be applied using the case study. Section 4 presents the conclusion.

## 2. RELATED WORK

### 2.1 Variability reduction approaches

Traditional approaches for variability reduction can be grouped into five generic categories (MacKay and *al.*, 1997): (1) introducing or tightening output inspection, (2) introducing or improving feedback control, (3) reducing variation in process inputs, (4) introducing or improving feedforward control and (5) desensitizing the process to input variation.

Various statistical process control tools are presented in literature to monitor the variability of the transformed quality characteristics, including control charts (Chiu and *al.*, 2007; Simmons Ivy and *al.*, 2009; Djauhari and *al.*, 2016) to report out-of-control states. However, these methods have less been used in multivariate control (Markovic and *al.*, 2017; Fan and *al.*, 2017) since they are increasingly complex to manage in multivariate environments, when data volumes continue to grow exponentially.

In the quality management field, the DMAIC (Define, Measure, Analyze, Improve and Control) process has been widely used as an improvement cycle, particularly for Six Sigma projects (see Smetkowska and *al.*, 2018 for a case study). However, De Mast and Lokkerbol (2012) have highlighted its limitation specifically from a problem-solving perspective: the DMAIC process appears unsuited for industrial environments in which human dynamics, subjective perceptions, and personal values are important.

### 2.2 Predictive tools and data-mining techniques for variability control/reduction

Manufacturers are now aiming to avoid potential issues using predictive approaches. Lee and *al.* (2013) define Predictive Manufacturing System as an emerging field to deal with issues on process variability. The idea is to ensure that data that are constantly being generated in a factory can be transformed, through predictive analytics, into information needed to make decisions. This is crucial for intelligent manufacturing processes (Jeong and *al.*, 2007), in which concepts as automatic data collection tools, real-time data analysis and decentralized decision making are fundamental for process control and improvement, contrary to conventional tools used to reduce variability (Schall, 2012), which require involving different operations teams.

In order to analyze large amounts of data, data mining techniques are usually used (Tahmasebian and *al.*, 2017). There are plenty studies in literature that compare different predictive models/approaches in specific industrial

settings since it is most of the time difficult to know which model will best fit the data (Gorman & Toman, 1966; Nisbet and *al.*, 2009). Jain and *al.* (2018) for example, compare neural networks and Gaussian process regression to estimate cycle times for incoming orders of a steel company. Logistic regression is widely used to detect quality events for manufacturing systems (Escobar and Morales-Menendez, 2018), where the defect is formulated as a binary value. However, there are no universal answers, which means models comparison should be the core of the production engineer's methodology.

Among examples, CRISP-DM (Cross Industry Standard Process for Data Mining) is a methodology defined by Wirth and Hipp (2000) that includes business understanding, data understanding, data preparation, modeling and evaluation. However, a generic framework is needed to support the engineering process considering its specificities and managers can face challenges to implement the CRISP-DM (Kampker and *al.*, 2018).

## 3. PROPOSED FRAMEWORK

In this section, we propose a generic framework to understand processes with intrinsic variability.

As shown in Fig. 1, the framework is composed of three different phases that will be described in the next subsections. In order to illustrate how it can be applied in a real industrial context, we will in parallel describe how the methodology was applied to analyze a textile dyeing process.

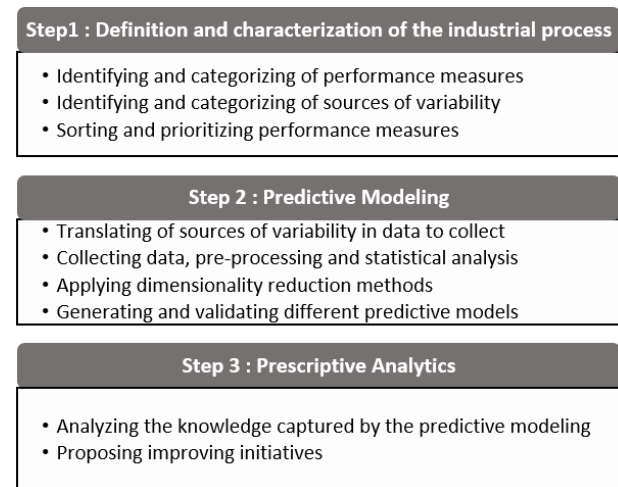


Fig. 1. Methodological framework in three steps.

### 3.1 Definition and characterization of the industrial process

Process owners first need to identify the variables, or performance measures (e.g.: fabric color compliance to

standard), for which they want to control and predict the variability. We call them  $Y=y_1..y_i$ .

Second, we ask them to identify the potential sources of variability ( $X=x_1..x_i$ ) that can affect Y. We have proposed four categories that make their identification easier: (1) characteristics of raw material, (2) the operational parameters, (3) environmental parameters and (4) the random factors (see Fig. 2 **Erreur ! Source du renvoi introuvable.**). These categories cover most case of variability, and presenting them during a brainstorming session makes easier identifying the factors. In consulting practice, Fig. 2 can be displayed on a wall and process experts fill the wall using post-it papers.

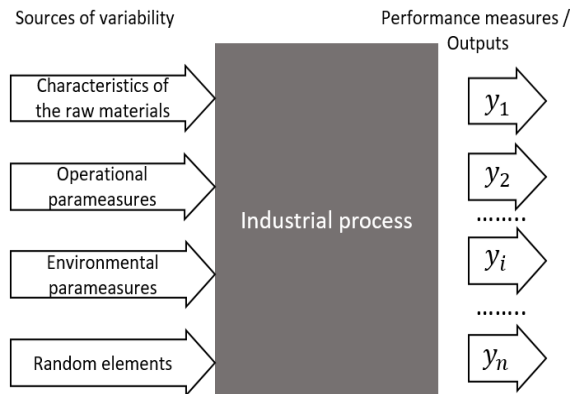


Fig. 2. Template used for the characterization of the process.

In order to capture experts' comprehension of the process, we then provide a standard form represented by Table 1. It can be used to better characterize the sources of variability and their relations to Y. For each  $y_i$ , a copy of the table has to be filled with the experts. It tells which of the previous varying factor has an impact over the  $y_i$ , specifies if its variability is known to be controllable (e.g: quantity for each dye) or uncontrollable (e.g: amount of residual contaminant in the water) and if we know how much it can affect  $y_i$ .

Finally, for each  $x_i$  one must identify datapoints that can be measured in the mill (see Table 1).

For our industrial case study, managers asked us to focus on the fabric color compliance to standard as the Y to manage. Experts identified 27 factors affecting textile fabric color compliance: 17 were identified as controllable and 10 as uncontrollable. A total of 120 corresponding datapoints were available. Some are provided by sensors monitoring the dyeing process, some are related to quality control, whereas others are simply parameters of driving the process.

### 3.2 Predictive modeling

The second phase is dedicated to the design and validation of a predictive model linking X and Y, with the goal of helping engineer to understand the correlation between the different  $x_j$  and  $y_i$ .

Of course, we do not only want a model that best fits the data, we want a model that will perform correctly for the prediction of additional data points not seen while training the model. For this reason, it is common practice to randomly split dataset into a training set and a testing set, in order to assess the goodness-of-fit and the predictive capacity. Different models are trained with the training set and then compared according to how they perform on the testing set (Jain and *al.*, 2018). Since the dataset is split randomly, the experiments are carried out several times and the results are reported using confidence intervals.

For our industrial process, we collected data for 602 production runs. We then experimented two types of models to predict the fabric color compliance to standard (the output Y). First, a nominal logistic regression (NLR) model, where Y is considered as a Boolean value (the color conforms to standard or not). Second, a standard least square regression (SLSR) model, where Y is considered as a continuous numeric value (telling "how much" the color conforms to the standard).

We first trained the models considering only simple (main) effects (that is, each  $x_i$  is a factor for the model). Then, we added crossed and quadratic effects (i.e. the product of each pair  $x_i*x_j$  is considered as an additional factor, as well as  $x_i*x_i$ ).

Table 2 presents how well the models, built using the training set, perform when evaluated on the testing set.

We evaluated the prediction accuracy of each model in terms of: (1) Overall accuracy (correctly predicting color compliance), (2) False negative fraction (compliance is predicted but textile actually failed quality control), and (3) False positive fraction (non-compliance is predicted but textile actually passed quality control). Overall Accuracy (OA), False Negative Fraction (FNF) and False Positive Fraction (FPF) are expressed by equations (Metz, 1978):

$$OA = \frac{\text{Number of correct predictions}}{\text{Total number of observations}} \quad (1)$$

$$FNF = \frac{\text{Number of false negative predictions}}{\text{Number of actually positive predictions}} \quad (2)$$

$$FPF = \frac{\text{Number of false positive predictions}}{\text{Number of actually negative predictions}} \quad (3)$$

The best model would have a high overall accuracy (OA), a low false negative rate (FNF) and a low false positive (FPF).

The first three models in Table 2 have a similar performance regarding the overall accuracy (OA). They also perform better than model 4 regarding the false



positive rate (FPF). However, model 4 is the best model regarding the false negative rate (FNF).

The false negative fraction (FNF) is the most important indicator, as we really want to avoid predicting that a textile will meet quality standard if it is not the case. As a matter of fact, non-compliances represent approximately 8% of all the cases available in the historical data. It is imperative for the predictive model to be able to detect most of them. With this detail in mind, model 4 was selected as the best model.

Considering an Overall Accuracy of 92.47% and a False Negative Fraction of 27.37%, this model might be viewed as excellent for such a complex multivariate industrial environment. Furthermore, it allows process owners to have insight as to which factors affect the most the process output and to act upon these, which will be more detailed in Section 3.3.

### 3.3 Prescriptive Analytics

This third step of the generic framework is oriented toward using the knowledge captured by the predictive model.

Due to the huge number of factors, it is a complex task for the process owners to determine precisely the causes of color non-compliance. Identifying the most significant datapoints based on the predictive model, they can propose improving initiatives, and even equipment or process modifications (similar to what one would do in a problem-solving situation). As an example, Table 3 presents the 10 datapoints (subfactors) having the highest  $p$ -values and the corresponding factor from Table 1.

Table 1. Characterization of the sources of variability (factors identified by experts) and the number of corresponding datapoints (subfactors that can actually be measured in the mill)

	Controllable		Uncontrollable	
	Observable	Unobservable	Observable	Unobservable
<i>Known impacts</i>	None	None	None	None
<i>Unknown impacts</i>	<ul style="list-style-type: none"> <li>- Water pH (1)</li> <li>- Amount of water used (2)</li> <li>- Chemical additives quantity (10)</li> <li>- Quantity of each dye (23)</li> <li>- Dyes characteristic (0)</li> <li>- Color of each dye (69)</li> <li>- Dye type (0)</li> <li>- Fabric characteristics (7)</li> <li>- Number of dying tube (0)</li> <li>- Heating gradient (0)</li> <li>- Duration of heating peak (1)</li> <li>- Temperature of heating peak (1)</li> <li>- Cooling gradient (0)</li> <li>- Fabric rotation speed (3)</li> <li>- Differential pressure (2)</li> <li>- Fabric rolling speed (0)</li> <li>- Dryer type (0)</li> </ul>	None	<ul style="list-style-type: none"> <li>- Ambient temperature (0)</li> <li>- Amount of residual contaminants (1)</li> <li>- Presence of rust in the water (0)</li> <li>- Amount of residual alum in the water (0)</li> <li>- Amount of residual polymer in the water (0)</li> <li>- Amount of residual micro sand (0)</li> </ul>	<ul style="list-style-type: none"> <li>- Dye granularity (0)</li> <li>- Dye dispersion (0)</li> <li>- Number of times the fabric has been recycled (0)</li> <li>- Contamination of the polymer used (0)</li> </ul>

Table 2. Models prediction performance (prediction color compliance) with 95 % confidence intervals

	<b>Model</b>	<b>Effects</b>	<b>Overall accuracy (OA)</b>	<b>False Negative Fraction (FNF)</b>	<b>False Positive Fraction (FPF)</b>
1	Standard least square regression (SLSR)	Simples	94.87% ± 1.45%	61.05% ± 15.08%	0.36% ± 1.51%
2	Nominal logistic regression (NLR)	Simples	94.87% ± 1.48%	51.58% ± 15.95%	1.17% ± 1.75%
3	Standard least square regression (SLSR)	Simples, crossed and quadratics	95.62% ± 1.05%	50.53% ± 12.32%	0.45% ± 1.43%
4	Nominal logistic regression (NLR)	Simples, crossed and quadratics	92.47% ± 2.14%	27.37% ± 16.44%	5.83% ± 2.49%

**Table 3. Datapoints (subfactors) with highest p-values and the corresponding factor from Table 1**

Datapoints (Factors)	p-value
Average water level (amount of water used)	0.046
Maximal differential pressure * Fabric flow rate (quantity of water used and fabric rotation speed)	0.050
Fabric weight (fabric characteristics)	0.072
Black dye lightness (dyes characteristics)	0.083
Quantity of blue dye Y (quantity of each dye)	0.090
Retraction rate (fabric characteristics)	0.097
Retraction rate * Retraction rate (fabric characteristics)	0.104
Average water level * Maximal water level (amount of water used)	0.105
Average differential pressure (differential pressure)	0.105
Average water level * Average water level (amount of water used)	0.106

Furthermore, most of the model value comes from the following. It allows defining new production processes. For a new production batch, it is possible to measure in the plant the values of the factors  $x_i$  identified in Table 1 as both *uncontrollable* and *observable*. Then, the ideal value for the  $x_i$  that are *controllable* can be derived from the predictive model. There is a drawback related to this approach: a predictive model is intended to be good in an *interpolation* context (that is, for the range of values it was trained/tested for) and we need to avoid using it in an *extrapolation* context. Therefore, the model recommendation should not be used directly as a prescription, but (first) to validate process engineer's intuition, and (then) to define new documented/tested processes/recipes.

By documenting how Key Performance Indicators (KPI) of the plant evolve over a long period, it is possible to monitor the efficiency of the improvement process.

#### 4. CONCLUSION

This paper addresses the need of industrials to understand processes with intrinsic variability. We introduce a framework/method allowing production engineers to systematically identify the sources of intrinsic variability and to measure/document their effects in order to define alternative production processes/practices. In

addition, we provide a case study showing how this method can be applied for a textile dyeing mill. Using predictive modeling and prescriptive analytics, the framework has allowed developing process knowledge and identifying improvement initiatives. Finally, it is worth recall that the generic framework can be applied in many other industrial contexts and sectors facing intrinsic process variability.

#### ACKNOWLEDGEMENTS

Funding for this research was provided by CRIQ and MITACS. The authors would like to thank Corinne Chabot and Jean Wéry (CRIQ), Sandee Noonan, Mathieu Lachance and Mathieu Brousseau (Duvalex).

#### REFERENCES

- Chiu, J.-E., Juang, T.-Y., Kuo, T.-I., Koo, T.-Y. (2007). Monitoring the Average and Variability in Multivariate Process. *J. of the Chinese Institute of Industrial Engineers*, volume 24, 252–258.
- De Mast, J., Lokkerbol, J. (2012). An analysis of the Six Sigma DMAIC method from the perspective of problem solving. *Int. J. of Production Economics*, volume 139, 604–614.
- Djahhari, M.A., Sagadavan, R., Li, L.S. (2016). Monitoring multivariate process variability when subgroup size is small. *Quality Engineering*, volume 28, 429–440.
- Escobar, C.A., Morales-Menendez, R. (2018). Machine learning techniques for quality control in high conformance manufacturing environment. *Advances in Mechanical Engineering*, volume 10, 1-16.
- Fan, J., Shu, L., Zhao, H., Yeung, H. (2017). Monitoring multivariate process variability via eigenvalues. *Computers & Industrial Engineering*, volume 113, 269–281.
- Gorman, J.W., Toman, R. (1966). Selection of variables for fitting equations to data. *Technometrics*, volume 8, 27–51.
- Jain, S., Narayanan, A.N., Lee, Y.-T.T. (2018). Comparison of data analytics approaches using simulation. *WinterSim 2018*, 1084-1095.
- Jeong, M.K., Lu, J.-C., Zhou, W., Ghosh, S.K. (2007). Data-reduction method for spatial data using a structured wavelet model. *Int. J. of Production Research*, volume 45, 2295–2311.
- Kampker, A., Heimes, H., Bühner, U., Lienemann, C., Krotil, S. (2018). Enabling Data Analytics in Large

- Scale Manufacturing. *Procedia Manufacturing*, volume 24, 120–127.
- Lee, J., Lapira, E., Yang, S., Kao, A. (2013). Predictive manufacturing system-Trends of next-generation production systems. *IFAC 2018*, volume 46, 150–156.
- MacKay, R.J., Steiner, S.H. (1997). Strategies for variability reduction. *Quality Engineering*, volume 10, 125–136.
- Markovic, S., Kerc, J., Horvat, M. (2017). The interprocess NIR sampling as an alternative approach to multivariate statistical process control for identifying sources of product-quality variability. *Drug development and industrial pharmacy*, volume 43, 379–389.
- Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in nuclear medicine*, volume 8, 283–298.
- Morin, M., Paradis, F., Rolland, A., Wery, J., Gaudreault, J. & Laviolette, F. (2015). Machine Learning-based Metamodels for Sawing Simulation. *WinterSim 2015*, 2160–2171.
- Nisbet, R., Elder, J., Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. Academic Press.
- Pinarbasi, M., Yüzükirmizi, M., Toklu, B. (2016). Variability modelling and balancing of stochastic assembly lines. *Int. J. of Production Research*, volume 54, 5761–5782.
- Schall, S.O. (2012). Variability reduction: a statistical engineering approach to engage operations teams in process improvement. *Quality Engineering*, volume 24, 264–279.
- Simmons Ivy, J., Black Nembhard, H., Baran, K. (2009). Quantifying the impact of variability and noise on patient outcomes in breast cancer decision making. *Quality Engineering*, volume 21, 319–334.
- Smetkowska, M., Mrugalska, B. (2018). Using Six Sigma DMAIC to Improve the Quality of the Production Process: A Case Study. *Procedia-Social and Behavioral Sciences*, volume 238, 590–596.
- Tahmasebian, S., Ghazisaeedi, M., Langarizadeh, M., Mokhtaran, M., Mahdavi-Mazdeh, M., Javadian, P. (2017). Applying data mining techniques to determine important parameters in chronic kidney disease and the relations of these parameters to each other. *J. of renal injury prevention*, volume 6, 83-87.
- Wirth, R., Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining, 4th Int. Conference on the Practical Applications of Knowledge Discovery and Data Mining, 29–39.
- Yeh, A.B., Li, B., Wang, K. (2012). Monitoring multivariate process variability with individual observations via penalised likelihood estimation. *Int. J. of Production Research*, volume 50, 6624–6638.

