

CIGI QUALITA MOSIM 2023

Exploration des processus pour l'évaluation des pertes de cadence

MATHILDE GUENDON¹, CAMELIA DADOUCHI^{1,2}, BRUNO AGARD^{1,2}

¹ LABORATOIRE EN INTELLIGENCE DES DONNEES
DEPARTEMENT DE MATHÉMATIQUES ET GENIE INDUSTRIEL,

ÉCOLE POLYTECHNIQUE DE MONTREAL, CP 6079, SUCCURSALE CENTRE-VILLE, MONTREAL, QUEBEC, CANADA
MATHILDE.GUENDON@POLYMTL.CA, CAMELIA.DADOUCHI@POLYMTL.CA, BRUNO.AGARD@POLYMTL.CA

² CENTRE INTERUNIVERSITAIRE DE RECHERCHE SUR LES RESEAUX D'ENTREPRISE, LA LOGISTIQUE ET LE TRANSPORT
(CIRRELT)

Résumé – À l'heure de l'Industrie 4.0, une multitude de données est enregistrée tout au long des processus de production. Ces données sont riches d'information et peuvent servir à l'acquisition de nouvelles connaissances qui permettront, en particulier, de s'attaquer aux sources des pertes en industrie. Afin de tirer profit de certaines de ces données, ce travail propose une méthodologie de valorisation de données industrielles visant l'analyse des pertes de cadence de machines d'assemblage semi-automatiques par l'évaluation des séquences. La méthodologie est inspirée de la démarche CRISP-DM (Cross-Industry Standard Process for Data Mining) et propose une visualisation des données et une étude des processus permettant aux gestionnaires de production d'identifier les meilleures pratiques et d'assurer le transfert de connaissance entre les différentes ressources des lignes de production. La méthodologie d'évaluation des pertes de cadence proposée se décline en trois phases: (1) Compréhension du cas et préparation des données, (2) Visualisation et analyse exploratoire des données et (3) Exploration de processus.

Abstract – In the age of Industry 4.0, a multitude of data is recorded throughout the production processes. This data is rich in information and can be used to acquire new knowledge that will allow, in particular, to address the sources of losses in industry. In order to take advantage of some of these data, this work proposes a methodology of industrial data valorisation aiming at the analysis of cadence of semi-automatic assembly machines by the evaluation of the sequences. The methodology is inspired by the CRISP-DM (Cross-Industry Standard Process for Data Mining) approach and proposes a data visualization and a process study allowing production managers to identify the best practices and to ensure the transfer of knowledge between the different resources of the production lines. The proposed rate loss assessment methodology consists of three phases: (1) Case understanding and data preparation, (2) Data visualization and exploratory analysis, and (3) Process mining.

Mots clés – Industrie 4.0, Performance, Pertes de cadence, Exploration de processus, Intelligence Artificielle.

Keywords – Industry 4.0, Performance, Downtime, Process Mining, Artificial Intelligence.

1 INTRODUCTION

Dans le contexte manufacturier actuel, les méthodes de production ont beaucoup évolué, notamment en ce qui concerne le suivi de la production. Les développements autour de l'industrie 4.0 permettent maintenant la surveillance des équipements en temps réel. Malgré tout, les pertes de cadence sur les lignes de production persistent pour différentes raisons et nécessitent des interventions supplémentaires pour rétablir la production. L'analyse de l'impact de ces interventions supplémentaires s'impose afin d'identifier les effets de ces dernières sur la productivité. La diversité des produits, des composantes et des caractéristiques ajoutent de la complexité quant à l'évaluation de la productivité.

La littérature propose plusieurs outils d'aide à la décision basés sur les données [Bhogal et Garb, 2020]. Cependant, leur exploitation dans le domaine manufacturier et plus précisément

pour les pertes de cadence en production reste encore limitée au regard des diverses avenues possibles [Souza et al., 2021].

Dans le présent article, une méthodologie de valorisation de données industrielles pour l'évaluation des pertes de cadence est proposée. Cette méthodologie repose principalement sur des outils d'analyse exploratoire ainsi que des outils d'exploration de processus. Cet article se décompose de la façon suivante : la section 2 présente l'état de l'art, dans lequel nous définirons la valorisation de données et les pertes de cadence étudiées dans ce travail. L'état de l'art sera aussi l'occasion d'adresser les différentes recherches s'intéressant à la valorisation de données pour l'étude des pertes de cadence afin d'inscrire ce travail dans leur continuité. La section 3 présentera la méthodologie développée afin d'évaluer les pertes de cadence, par l'identification de bonnes pratiques de production, dans un contexte donné. La section 4 mettra en œuvre la méthodologie proposée sur un cas d'étude. Finalement, la section 5 présentera les résultats, les limites perçues et quelques perspectives.

2 ÉTAT DE L'ART

2.1 Définitions

Afin de gérer leur production, les usines surveillent de près de nombreux indicateurs. Les indicateurs les plus fréquemment utilisés dans le milieu manufacturier sont ceux regroupés sous le concept de taux de rendement global (« Overall Equipment Effectiveness » en anglais), qui évaluent la productivité. Ils sont issus de la théorie des six grandes pertes de Nakajima. Les pertes identifiées concernent la disponibilité lorsqu'il s'agit de pannes machines et de changements de séries. Elles touchent à l'efficacité lorsque ce sont des micro-arrêts et des cadences ralenties. Enfin, elles impactent la qualité lorsqu'il y a des défauts de qualité et des défauts causés lors de la mise en fonctionnement des machines [Soltanali et al., 2021].

Dans cette revue de littérature, les pertes de cadence feront référence aux micro-arrêts, aux baisses de cadence ainsi qu'aux arrêts non-planifiés ne nécessitant pas de maintenance.

Selon l'Office Québécois de la Langue Française [2019], la valorisation de données fait référence au « processus de collecte, de traitement et d'analyse de données, permettant l'utilisation optimale de celles-ci dans la poursuite d'un objectif donné ». Lorsqu'il s'agit de données industrielles, l'enjeu principal est d'être capable d'en extraire de la connaissance. Comme le soulignent Dogan et Briant [2021], « La majorité des problèmes industriels sont riches en données, mais pauvres en savoir ». La valorisation de données est alors un atout majeur pour soutenir les différentes prises de décision et rester compétitif.

2.2 Valorisation de données pour l'étude des pertes de cadence

La valorisation de données industrielles concernant les pertes de cadence a été traitée par différentes techniques, notamment en utilisant la **détection d'anomalies**. Plusieurs chercheurs ont utilisé l'analyse de séries temporelles par des techniques telles que la méthode des hypergraphes de contraintes [Liang et al., 2021] ou les réseaux de neurones [Souza et al., 2021]. D'autres se concentrent plutôt sur la détection de **comportements menant à des anomalies de production** : Bhogal et Garg [2020] proposent des approches de processus-mining, Bouché et Zanni-Merk [2010] utilisent des approches de classification tel que les arbres de décision avec les algorithmes Random Forest, d'autres utilisent des méthodes Extreme Gradient Bosting [Klaeger et al., 2019] notamment pour pouvoir remonter à la cause des anomalies détectées [Souza et al., 2021]. D'autres approches comme l'**extraction de modèles** et l'**analyse des causes racines** ont aussi été développées. Certains travaux vont au-delà de la détection d'anomalie, en effet Bhogal et Garg [2020] analysent les performances de production et les différents flux afin d'identifier des motifs de pannes et de générer de la connaissance pour réduire les pertes de cadence engendrées par les pannes et le processus de réparation. Bouche et Zanni-Merk [2010] créent un arbre permettant la visualisation de modèles de pannes et des relations séquentielles menant à elles les plus probables. Dagnino [2019] réalise de la fouille de motifs séquentiels et analyse des alarmes horodatées au sein de systèmes de contrôles de processus, afin de détecter les modes de défaillance de valves de contrôles régulées en pression. **Des méthodologies différentes pour appréhender les pertes de cadences en production sont aussi proposées**. Pabolu et al. [2022] s'intéressent à la prédiction du temps d'opération le plus confortable pour un opérateur sur ligne. Ils traitent alors les données concernant les conditions de travail grâce à des outils d'apprentissage supervisé. Hrcka et al. [2017] réalisent de la fouille de données textuelles à partir de formulaires de micro-arrêts remplis par des opérateurs. Chen et

al. [2019] poursuivent la détection d'anomalies avec une méthode statistique de régression afin de choisir le meilleur ensemble de paramètres menant à une occurrence non défectueuse et à un temps de soudure optimal.

La valorisation de données concernant les pertes de cadence sur ligne de production est donc majoritairement répartie entre la **détection d'anomalies** et l'**identification de modèles de pannes**, bien que d'autres approches moins fréquentes existent.

2.3 Analyse critique

Les résultats présentés ci-dessus sont issus de l'utilisation de divers outils permettant de faire de la détection d'anomalies pour traiter le problème de réduction des pertes de cadence. Cependant, peu d'intérêt est porté à ce qui vient en amont de la panne. Çiflikli et al. [2010] soulèvent la possibilité d'identifier à la fois des comportements normaux et des points aberrants qui ne correspondent à aucun modèle préalablement déterminé, puisque les données ne sont pas labellisées.

[Haasbroek et al., 2018] souligne le manque d'applications de ces méthodes à des cas concrets avec une quantité importante de données. De plus, lorsque des cas pratiques sont étudiés, la connaissance métier n'est pas forcément prise en compte [Alfeo et al., 2020] et lorsqu'elle l'est, elle ne l'est que partiellement [Haasbroek et al., 2018]. Cependant, l'utilisation des connaissances des experts du domaine peut permettre de donner une direction aux recherches effectuées et d'obtenir de meilleurs résultats Alfeo et al. [2020], même si parfois ces connaissances peuvent biaiser les méthodes développées et les résultats obtenus [Trunzer et al., 2017]. Les anomalies peuvent alors être détectées à partir de caractéristiques sélectionnées grâce aux connaissances métiers, puis classées dans des modes prédéfinis. D'après Cerquitelli et al. [2020], la valorisation de données de production porte réellement ses fruits lorsqu'elle est à l'origine de nouvelles actions. Mais, Liang et al. [2020] évoquent la difficulté à quantifier l'impact de chaque anomalie détectée, ce qui peut poser un problème dans la priorisation des actions à mener. D'autres travaux ne considèrent pas directement les pertes de cadence mais des problématiques relatives telles que maintenance prédictive [Souza et al., 2021], le signalement d'anomalies [Dagnigo 2019] ou la réparation de pannes [Bhogal et Garg, 2020].

Enfin, certains travaux étudient les pertes de cadence avec une approche simplifiée vis-à-vis de la réalité de la production. Klaeger et al. [2019] ne distinguent pas les arrêts planifiés des arrêts non planifiés. Bhogal et Garg [2020] exploitent des données collectées manuellement en faible quantité et jugent la performance de production en ne considérant que les durées des activités de production. Cette approche de la performance se retrouve dans le travail de Bouché et Zanni-Merk [2020]. La seule mesure de la performance est celle du nombre de bonnes pièces produites en un certain temps. Quant à eux, Hrcka et al. [2017] présentent une démarche d'identification des conditions dans lesquels se produit une panne par traitement de données textuelles. Cette démarche permet de réaliser une première identification des pannes, mais aucune proposition n'est faite pour l'éviter ou la corriger.

Suite à cet état de l'art, trois manquements peuvent être identifiés. Premièrement, le manque d'intérêt porté aux processus et aux pratiques qui en découlent pour expliquer la cause des comportements. Ensuite, certains travaux s'appuient sur une faible quantité de données et n'exploitent que partiellement les connaissances des experts du domaine. Finalement, la visualisation des données est très peu sollicitée bien qu'elle soit un outil essentiel pour l'interprétabilité des résultats. En effet, la visualisation des données peut présenter

une grande quantité de données dans un format simple et facile à comprendre afin de communiquer l'information de manière claire et efficace [Embarak, 2018] or dans la majorité des travaux, les résultats sont des modèles ou des méthodologies qui plus difficile à expliquer et donc à exploiter dans les entreprises. Être capable de produire des visualisations rapprochant la compréhension de tous faciliterait l'accessibilité d'outils de détection aux PME.

Ce sont ces trois points qui seront au cœur de la méthodologie développée.

3 METHODOLOGIE PROPOSEE

La méthodologie proposée a pour objectif général d'améliorer la productivité de lignes de production. Elle propose d'évaluer les pertes de cadence à partir de l'analyse de données industrielles, notamment En proposant une nouvelle approche d'analyse de données séquentielles dans le contexte manufacturier de manière à identifier les bonnes pratiques.

Une bonne pratique sera définie par le meilleur comportement identifié face à une machine afin d'éviter les pertes de cadence dans une situation particulière.

La méthodologie proposée s'inspire de CRISP-DM [IBM, 2021] qui reste aujourd'hui la méthodologie de valorisation de données la plus utilisée. La méthodologie proposée dans cet article s'en distingue par son adaptation à l'analyses des pertes de cadences. Cette méthodologie se décline en 3 étapes : (1) la compréhension du cas et la préparation des données, afin de bien saisir les enjeux et les ressources à disposition pour la suite de l'étude ; (2) la visualisation et l'analyse exploratoire des données, afin de faire un état des lieux des lignes de production étudiées, aussi bien en mesurant leur productivité qu'en observant les différents événements qui s'y produisent ; enfin (3) la découverte de processus, qui sera l'occasion de comprendre les comportements de production afin d'identifier les bonnes pratiques.

3.1 Compréhension du cas et préparation des données

Lors de cette première phase, il est question de recueillir toutes les informations relatives au contexte manufacturier.

Les objectifs sont clarifiés afin de bien délimiter le périmètre de l'étude. Cette phase de compréhension est alimentée de discussions avec les experts du domaine afin d'avoir une meilleure prise en compte des conditions de production et comprendre les interactions entre la section étudiée et le reste du système de production. Il faut tout d'abord extraire les données pertinentes. Ici, ce sont les données de production et des événements machine qui sont intéressantes. Les données doivent pouvoir rendre compte, de façon détaillée, des différents produits réalisés, ainsi que différents événements qui se produisent lors du processus étudié. Une analyse des corrélations est réalisée ainsi qu'une identification de données manquantes, doublons et points aberrants.

Après ces premières étapes, les données sont préparées et formatées pour l'étude. Les données de production doivent permettre l'identification unique de chacun des produits réalisés en incluant des caractéristiques comme le type de produit, sa composition, ainsi que les ressources allouées pour sa réalisation. Concernant les données événementielles, il faut être capable d'identifier chaque événement, son type, et son occurrence lors du processus. Il faut parfois enrichir les données avec des informations complémentaires, afin de matérialiser des informations tacites du domaine et non présentes dans les données extraites. On parle alors d'enrichissement des données, aussi appelé *Feature Engineering*. Certaines informations peuvent être réparties dans différentes tables, il peut alors être

nécessaire de les regrouper dans une seule et même table par des opérations de jointure. Si les données manquantes sont présentes en faible proportion, elles peuvent être supprimées. Il en est de même pour les points aberrants et les doublons.

3.2 Visualisation et analyse exploratoire des données

Afin d'identifier les pratiques menant aux meilleures performances de production, il est nécessaire de mesurer les performances actuelles des lignes dans différents contextes de productions. Pour cela, il faut une métrique permettant la mesure de la productivité en tenant compte des différents facteurs pouvant l'affecter. Dans l'optique de l'identification de bonnes pratiques, et afin de distinguer les conditions non contrôlables des conditions contrôlables, nous proposons le concept de « profil de production ». On appellera alors « profil de production » un groupement d'informations qui caractérisent un opérateur sur une ligne de production, comme son expérience, la machine sur laquelle il travaille ou encore les types de produit qu'il réalise. Le profil de production doit rendre possible la distinction entre différents niveaux de productivité pour des profils de production similaires. La création de profils permettra alors de mettre en parallèle des profils comparables et d'évaluer leur niveau de productivité avant de s'intéresser à leurs pratiques.

3.2.1 Identification des facteurs influents sur la production
Afin de préciser les profils de production, il est important de comprendre quels sont les facteurs qui régissent la production. S'agit-il d'une production multiproduit ? Quelles sont les ressources à disposition ? Quelles sont les périodes de travail ? Existe-t-il une saisonnalité ? Il faut ensuite de pouvoir quantifier l'impact de ces facteurs. Pour cela, le choix d'une métrique permettant de juger la performance de la production s'impose. Elle correspond souvent à la métrique utilisée par l'entreprise pour rendre compte de sa productivité. Elle peut varier selon l'industrie à laquelle cette méthodologie s'applique. Elle est alors évaluée pour chacun des profils de production ayant une expérience similaire. Dépendamment de la métrique choisie, elle peut, à ce stade, être très générale. L'objectif de cette étape est de comprendre quels sont les éléments qui caractérisent la production et ainsi de venir affiner par la suite la métrique choisie. Afin de continuer le développement de cette méthodologie, la métrique sélectionnée est par exemple : le nombre de produits réalisés par unité de temps.

3.2.2 Statistiques de production

À la lumière des facteurs identifiés, différentes statistiques sont déterminées au moyen d'outils d'analyse exploratoire, afin de préciser la métrique sélectionnée et d'évaluer l'impact des facteurs externes non métrisable sur la productivité.

Par exemple, pour le nombre de produits réalisés par unité de temps, dans un premier temps, il faut établir les temps moyens de production pour une référence particulière considérant que les produits peuvent avoir des niveaux de complexité variables ou des ordres de grandeurs différents. Il est alors possible de préciser un profil de production par les produits qu'il réalise.

3.2.3 Statistiques d'événements

De la même façon, les événements sont maintenant étudiés. Un événement est alors un état du processus de production qui diffère de la production normale. Cette section propose de résumer les événements rencontrés lors du processus de production étudié, pour chacun des profils de production. L'étude statistique de ces événements est aussi l'occasion de comprendre leur impact sur la cadence de production : pour un profil donné, quelle est la proportion d'événements qui mènent

à un ralentissement de la cadence ? à un arrêt planifié ou non planifié ?

Les historiques de production doivent permettre l'identification de profils de production ayant des similarités, par exemple ceux pour lesquels on retrouve les mêmes types de produits. Pour chacun des profils identifiés comme semblables, les statistiques d'événements sont évaluées. Ces nouvelles informations ont pour but de préciser un profil au regard des événements rencontrés au cours du processus de production. Cela permettra de comparer des comportements relatifs à des profils similaires (section 3.3). Cette approche comportementale est alors rendue possible par l'exploration de processus.

3.3 Exploration de processus

Selon Van der Aalst [2016], l'exploration de processus permet de valoriser des données événementielles afin de comprendre des comportements. La visualisation et l'analyse exploratoire ont permis de faire un état des lieux des pertes de cadence et d'introduire les profils de production. L'exploration de processus tente alors d'offrir une explication à ces pertes de cadence d'un point de vue comportemental.

3.3.1 Présentation du processus étudié

Selon Van der Aalst [2016], un processus est caractérisé par : des cas, des concepts et des dates. Dans cette méthodologie destinée à l'industrie manufacturière, un processus correspondra aux différents événements rencontrés par plusieurs produits, à différents moments sur une ligne de production.

3.3.2 Modélisation sous forme de graphes dirigés

L'exploration de processus permet de modéliser les processus sous différentes formes, à l'aide de différents outils. Ce sont les graphes dirigés qui sont choisis ici. Ils représentent un bon outil pour la représentation visuelle des relations directes pondérées entre des événements, favorisant ainsi l'interprétabilité des résultats [Dupuis et al., 2021]. Afin de pouvoir tracer des graphes dirigés, la table de données doit être formatée avec trois attributs essentiels : l'identifiant unique de chaque produit (le cas), l'identifiant de l'événement que ce produit rencontre (le concept) ainsi qu'un horodatage (la date). Le tableau 1 montre un processus de production par lequel les produits A, B et C passent par différentes étapes 1, 2 et 3.

Tableau 1 Quelques lignes d'un exemple de table de données événementielles

Produit	Horodatage	Événement
A	01/01/2022 8 :10	1
A	01/01/2022 8 :11	2
A	01/01/2022 8 :12	3
B	01/01/2022 8 :15	1
C	01/01/2022 8 :35	1
C	01/01/2022 8 :36	2
...

À partir d'une telle table formatée, il est possible de tracer des graphes dirigés tels que représentés sur la figure 1.

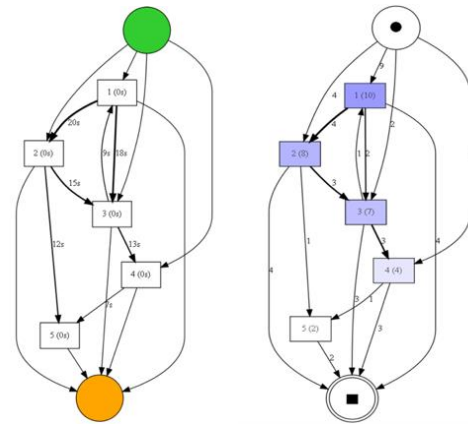


Figure 1 Graphes dirigés avec temps moyen de transition (gauche) et occurrences des événements (droite)

Le point vert et le point noir représentent le début du processus étudié. Le point orange et le carré noir représentent la fin du processus. Entre ces deux étapes, des étapes intermédiaires apparaissent sous forme de rectangles. Ce sont les différents états du processus. Sur le graphe de gauche, le temps moyen pour passer d'un état à un autre est inscrit sur les flèches de transitions. Sur le graphe de droite, le nombre d'occurrences des transitions est inscrit sur les flèches et le nombre d'occurrences d'un état donné est inscrit dans la case de cet état.

Dans les sections suivantes, le potentiel des graphes dirigés sera exploité dans le but d'identifier les transitions dites « critiques », à savoir les transitions qui impactent la cadence de production. Les transitions les plus fréquentes ainsi que les transitions les plus chronophages seront étudiées.

3.3.3 Filtrage des graphes

À cause du volume important de données d'événement à traiter, l'interprétation des graphes dirigés peut s'avérer complexe. Afin de se concentrer sur les transitions affectant la cadence de production, les graphes sont filtrés en amont suivant le principe de Pareto. Ainsi, les transitions qui représentent 80% des occurrences sont conservées, ainsi que les transitions qui occupent 80% du temps total.

3.3.4 Étude des transitions fréquentes

L'analyse des transitions fréquentes a pour objectif, dans un premier temps, de mettre la lumière sur des transitions entre deux états du processus qui pourraient être révélatrices d'un comportement anormal. Par la suite, et grâce à la création des profils de production, il est possible de comparer les graphes dirigés pour des profils de production similaires. Une fois les transitions filtrées, il s'agit de tracer le graphe affichant, d'une part les occurrences de ces transitions, puis celui affichant les durées moyennes. Il en découle alors une possibilité d'identifier comportements de production différents. Dans l'industrie manufacturière, la majorité des comportements sont soumis à des normes, ainsi, les opérateurs sont guidés quant aux actions à prendre selon les diverses situations, afin de garantir d'une part, une certaine cadence de production et d'autre part, d'assurer leur propre sécurité. Les graphes dirigés peuvent être analysés par les experts du domaine à la lumière de ces normes et de leurs connaissances pour distinguer des bonnes pratiques.

3.3.5 Étude des transitions chronophages

Il s'agit ici de ne pas oublier des transitions importantes. Les transitions à considérer sont celles dont la durée est élevée. De la même façon que précédemment, il s'agit ici de comparer les graphes dirigés affichant les durées moyennes des transitions les plus grandes pour chacun des profils sélectionnés. Après avoir identifié des écarts de durées de transitions importants, il sera

possible déceler les pratiques correspondantes. L'étude des transitions critiques permet d'offrir un support visuel aux experts du domaine. Les graphes sont des outils qui permettent de cibler des transitions anormales et ainsi de guider les experts dans la priorisation des pratiques à améliorer. L'identification des bonnes pratiques peut alors être faite en combinant l'étude des graphes dirigés à des observations sur le terrain permettant une meilleure compréhension des profils de production et ainsi des actions pratiques améliorant la cadence. Les meilleures pratiques pourront être transmises par la formation des opérateurs.

4 CAS D'ETUDE

4.1 Compréhension et préparation des données

4.1.1 Compréhension du cas d'études

Une usine québécoise, faisant partie d'une multinationale manufacturière de l'industrie automobile, a modifié la composition de ses lignes de production, notamment en remplaçant une majorité de ses machines d'assemblage par des machines semi-automatiques. Grâce à la surveillance de ces machines, une grande quantité de données est disponible, offrant un grand potentiel d'exploitation. Cette usine montre la volonté d'augmenter sa cadence de 10% car elle est actuellement entravée par différentes pertes. L'usine évalue sa production à partir du nombre de produits réalisés par jour.

Diverses contraintes ajoutant de la complexité à l'étude sont à prendre en compte. La machine étudiée est divisée en trois parties, fonctionnant en parallèle. Ainsi, deux produits peuvent être à deux étapes différentes de leur assemblage sur la même machine et au même moment. Ces machines semi-automatiques nécessitent une intervention des opérateurs. Enfin, la production de cette usine est multiproduits, dont la taille, la composition et d'autres caractéristiques varient.

Par l'application de la méthodologie de valorisation de données pour l'évaluation des pertes de cadence, l'objectif est alors d'améliorer la productivité de l'usine en mesurant sa productivité actuelle et en identifiant les pratiques à privilégier.

4.1.2 Compréhension des données

Les journaux d'événements, de production et diverses tables de définitions permettent de récolter les données nécessaires. L'exploration se fait par la mise en parallèle de différents attributs. Par exemple, la figure 2 met en parallèle le type de produit (Matière) et les opérateurs (Main d'œuvre).

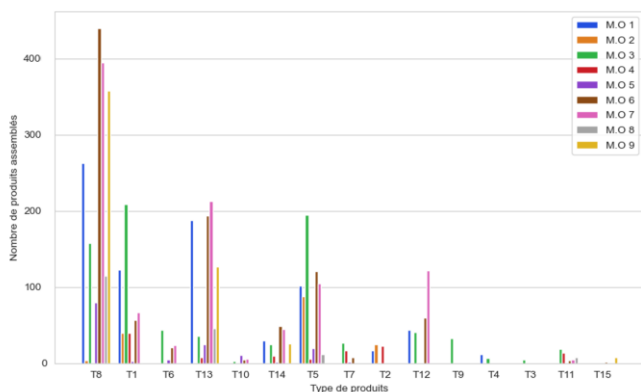


Figure 2 Distribution du nombre de produits assemblés en fonction du type de produit, pour différents opérateurs, sur une période de 3 mois

Ce graphe donne beaucoup d'informations sur la réalité de la production, par exemple tous les opérateurs d'une même

machine ne semblent pas assembler les mêmes types de produits et certains types de produits sont plus fréquents que d'autres. Les analyses ont montré différents points de complexité à prendre en compte pour évaluer la productivité et comprendre les facteurs qui peuvent l'impacter. Cette phase d'exploration sera approfondie par la suite afin d'être plus proche de la réalité de la production.

4.1.3 Préparation des données

Dans le cadre de l'évaluation des pertes de cadence pour l'usine partenaire, trois tables sont étudiées : la table du journal de production (Table 2), la table du journal des événements (Table 3) et la table de définition des codes événements (Table 4).

Tableau 2 Format du journal de production

Identifiant unique	Utilisateur	Types de produit	Machine	Date	Durée (s)
12345	ABC	T1	M1	01/01/2022 17:38:00	52

Tableau 3 Format du journal d'événements

Code	Début	Fin	Machine
123	01/01/2022 17:39:00	01/01/2022 17:42:00	M1

Tableau 4 Format de la table de définition des codes événements

Code	Catégorie	Définition
123	C1	xxx

La construction et l'intégration des données permet de rassembler les informations relatives aux événements dans une même table triée chronologiquement. La connaissance des experts du domaine permet de préciser différents événements, notamment pour savoir si des événements sont susceptibles de mener à un arrêt ou à un ralentissement, et s'ils apparaissent alors que la machine est sous la supervision d'un opérateur, et non lors d'opérations de maintenance par exemple.

4.2 Visualisation des données et analyse exploratoire

Dans ce cas d'étude, la métrique choisie pour mesurer la productivité est le nombre de produits réalisés par unité de temps. L'échantillon de données étudiées couvre 3 mois de production sur une machine donnée. Le nombre de produits réalisés par unité de temps est évalué pour chacun des opérateurs de cette machine.

4.2.1 Identification des facteurs influant la production

Lors de compréhension des données (4.1.2), différents attributs ont été mis en parallèle et semblent avoir un impact sur la production. C'est le cas de la main d'œuvre.

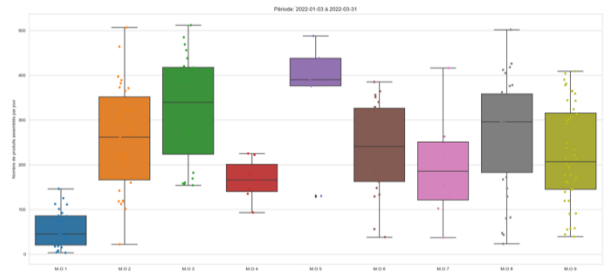


Figure 3 Distribution du nombre de produits assemblés par jour et par opérateur

La figure 3 représente le nombre de produits assemblés par jour pour chacun des opérateurs de la machine étudiée, sur la période de 3 mois. Sur une même machine, les performances sont différentes d'un opérateur à l'autre. Cependant, ce graphique ne tient pas compte de toute la complexité d'une production semi-

automatique et multiproduits, d'autres informations sont nécessaires. Il s'agira donc de recenser un maximum d'informations permettant de préciser les différentes façons de produire, au moyen de « profils de production », permettant de connaître l'expérience de l'opérateur, les types de produit sur lesquels il travaille, son quart de travail, etc.

4.2.2 Statistiques de production

La métrique de productivité utilisée par le partenaire est le nombre de produits réalisés par unité de temps. Ici, elle sera définie par la somme des temps d'opération sur la période étudiée. Cela permet de prendre en compte les différents événements qui se produisent sur la ligne de production. Pour la suite de cette étude, différents profils sont sélectionnés avec les experts du domaine. Ces profils ont une expérience similaire, ils ont assemblé un nombre important de produits et ont réalisé un nombre similaire d'heures de production sur 3 mois. Dans un premier temps, l'intérêt est porté sur les produits. Pour ces trois profils, les distributions des types de produits qu'ils assemblent sont représentées sur la figure 4.

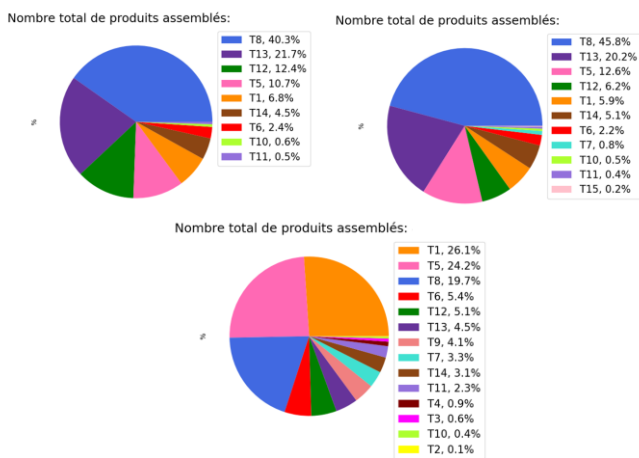


Figure 4 Proportion des types de produits assemblés pour 3 profils

Deux profils semblent avoir une distribution similaire et seront détaillés par la suite. Pour ces deux profils, la distribution des temps d'opération pour chaque type de produits est représentée sur la figure 5.

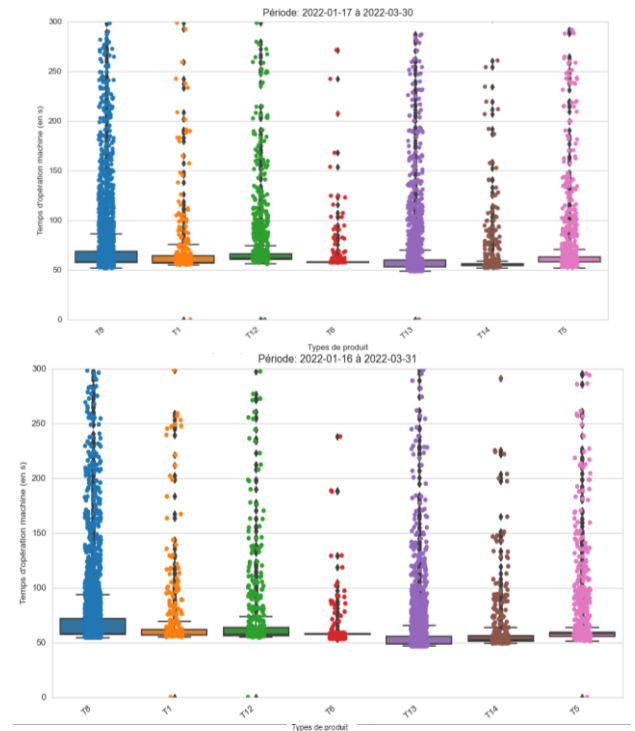


Figure 5 Distribution des temps d'opération par type de produits assemblés pour les profils 1 (haut) et 2 (bas)

La variabilité des temps d'opérations apparaît très importante. Les valeurs élevées s'expliquent par la présence de changement de série qui nécessitent un temps d'opération supérieur sur certains produits afin de paramétrer correctement la machine. Mais des temps supérieurs sont aussi causés par des événements venant perturber la production. Les médianes témoignent d'une cadence de production proche des valeurs attendues.

En se penchant sur la productivité des profils de production, selon les différents types de produits, des écarts significatifs peuvent tout de même apparaître, notamment pour le type de produits T13. Dès lors qu'une différence de productivité est pointée, l'analyse continue avec l'analyse des événements venant diminuer la cadence de production.

4.2.3 Statistiques d'événements

De la même façon que précédemment, il est ici question de préciser les profils de production en étudiant les événements rencontrés lors de la production. Dans un premier temps, des statistiques classiques sont mesurées : nombre d'événements, durés, arrêts, ralentissements, etc. Ainsi, le profil 2 affiche moins d'événements durant la période étudiée, et la distribution des événements est étudiée pour chacun des profils sélectionnés sur la période (Figure 6).

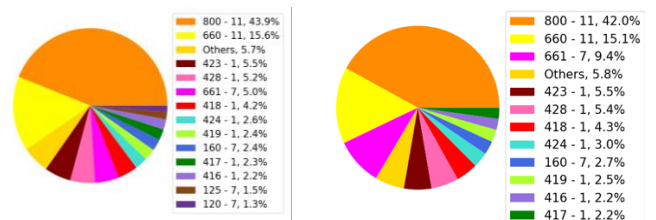


Figure 6 Proportion des événements rencontrés sur la période considérée, pour les profils 1 et 2

Un premier écart apparaît dans la fréquence des événements, notamment pour le code 661 relatif à une correction suite à l'apparition d'une anomalie de matériel non conforme. Cela consiste en une première étape dans l'identification des bonnes pratiques.

4.3 Exploration des processus

4.3.1 Présentation du processus étudié

Dans le cadre de cette étude, le processus correspond aux différents états par lesquels passe un produit lors de son assemblage. Les états sont ici matérialisés par les différents codes événements. À partir de la jointure des deux tables de production et d'événements, seules les lignes relatives aux événements sont conservées. Cette nouvelle table diffère de la table des événements initiale car un événement ou une suite d'événement peuvent dès lors être associés à un produit et un opérateur en particulier.

4.3.2 Modélisation sous forme de graphes dirigés

Les processus de production sont maintenant représentés par des graphes dirigés. Pour chacun des deux profils sélectionnés, les trois attributs essentiels sont l'identificateur unique du produit, l'horodatage ainsi que les événements rencontrés (représenté par le code d'identification de l'événement).

4.3.3 Filtrage des graphes

Afin de conserver des graphes lisibles et d'après le principe de Pareto, les graphes dirigés sont filtrés de façon à ne conserver que les transitions qui ont un impact significatif sur la durée totale et les occurrences totales des transitions (Figure 7).

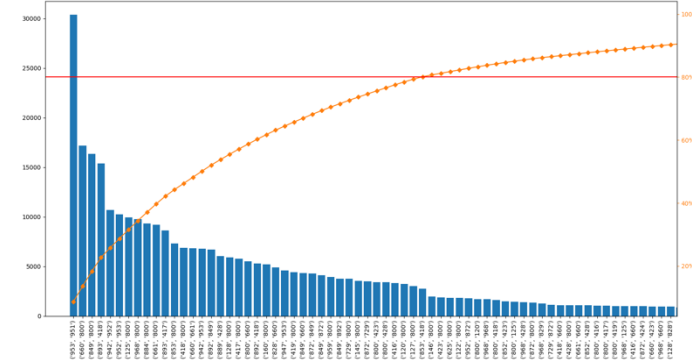


Figure 7 Graphe de Pareto des durées des transitions Profil 1

4.3.4 Études des transitions fréquentes

Dès lors que les transitions les moins fréquentes sont filtrées, ce sont les durées moyennes des transitions restantes qui sont étudiées, pour les deux profils sélectionnés.

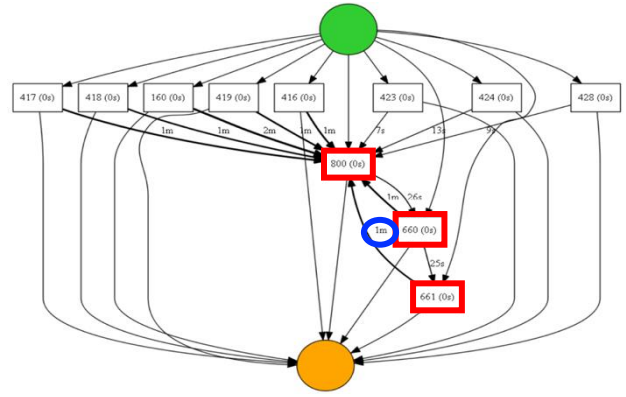
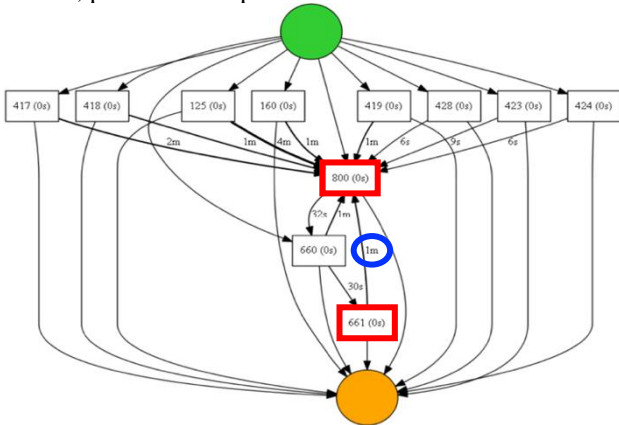
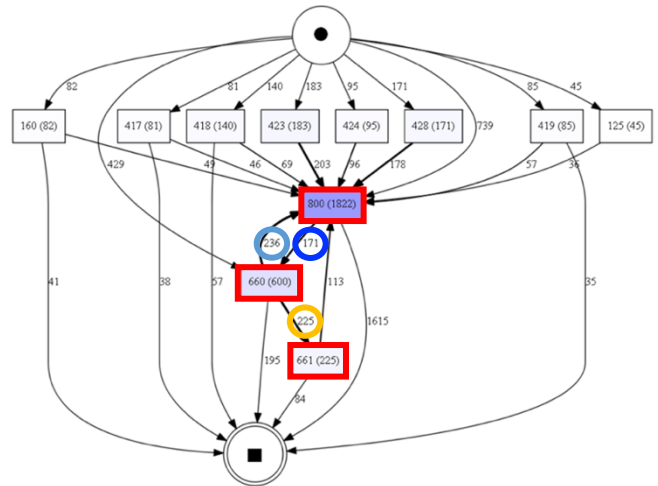


Figure 8 DFGs filtrés selon les transitions les plus fréquentes, pondérés des durées moyennes, pour les profils 1 et 2

D'après la figure 8, certaines transitions dépassent de manière importante les valeurs attendues. La connaissance métier indique notamment qu'un changement de composantes (codes 4xx) peut mener à un arrêt de la machine s'il n'est pas complété dans une certaine période avant le retour en production (code 800). Pour les deux profils étudiés, un écart apparaît pour le code 417 ce qui peut provenir d'une différence de comportement quant aux changements de composantes.

Concernant les occurrences de ces transitions fréquentes, les graphes sont représentés sur la figure 9.



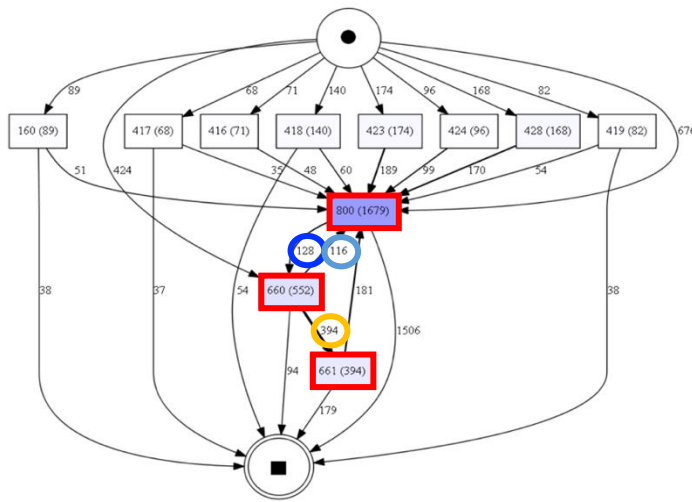


Figure 9 DFGs filtrés selon les transitions les plus fréquentes, pondérés des occurrences pour les profils 1 et 2

Certains codes ne sont pas présents pour les deux profils et une différence notable apparaît pour l'occurrence des transitions 800/660, 660/661 et 661/800 (Figure 9). Cette boucle correspond à un arrêt de la machine pour détection d'un matériel non conforme (660) et nécessite l'intervention de l'opérateur pour corriger cette non-conformité (661) avant le retour en production (800). Ici, il est par exemple possible de distinguer les deux profils : l'un semble mettre plus de temps à corriger la non-conformité (30s en moyenne contre 25s), ainsi qu'à en corriger seulement 225 non-conformités, tandis que le second profil en corrige 394 durant la période étudiée.

4.3.5 Études des transitions chronophages

Cette partie consiste à mettre la lumière sur d'autres types de transitions critiques, à savoir les transitions qui n'apparaissent pas forcément de façon répétitive mais qui, lorsqu'elles apparaissent, prennent un temps trop important.



Figure 10 Agrandissement du DFG filtré selon les transitions les plus chronophages, pondéré des durées moyennes pour le profil 1

De nouveaux codes, apparaissent (Figure 10). La même approche que précédemment peut être utilisée pour analyser des sous-parties du processus de production. Par exemple, les codes entourés sur la figure 10 font référence à des temps d'ajustement des paramètres lors d'un changement de série. Leur analyse pourrait mener à une meilleure compréhension des actions prises lors d'un changement de série et à l'identification des bonnes pratiques liées à cette étape de production.

Les graphes dirigés permettent ainsi aux ingénieurs sur le terrain de cibler des transitions peu fréquentes mais chronophages et ainsi d'observer de mieux comprendre ce qui peut mener à ces écarts.

5 CONCLUSION

À l'aide d'une démarche permettant l'identification des bonnes pratiques de production, l'objectif de la méthodologie présentée est d'évaluer les pertes de cadence dans un contexte de production semi-automatique. Inspirée de la méthodologie CRISP-DM, ce travail s'appuie sur des outils d'analyse exploratoire de données et de découverte de processus.

Cette méthodologie se distingue de celle proposée dans la littérature par sa capacité à représenter de manière visuelle, simple et claire les différents comportements face à des étapes ciblées d'un processus de production. En effet, la visualisation par DFG permet de comprendre les séquences menant à des niveaux de productivité différents. Elle présente donc un outil visuel, que les ingénieurs de production peuvent utiliser pour observer et améliorer les pratiques de production de manière simple et instinctive. Son applicabilité a été démontrée grâce au cas d'étude d'une usine québécoise.

La méthodologie développée présente cependant quelques limites. En effet, une plus grande précision aurait pu être accordée à la prise en compte de la complexité des machines de production. L'utilisation de filtres sur les graphes dirigés doit être faite avec soin et peut mener à des résultats erronés si les filtres sont mal ajustés. Notons aussi que seule la productivité a été considérée dans ce cas d'étude, d'autres critères de performance pourraient être explorés.

6 REMERCIEMENTS

Les auteurs souhaitent exprimer leur reconnaissance au partenaire industriel de ce travail, pour l'accessibilité aux informations et données nécessaires, leur confiance et leur implication dans ce projet.

Nous remercions aussi MITACS, projet IT25894, pour le financement de ce projet.

7 REFERENCES

- Alfeo, A. L., Cimino, M. G., Manco, G., Ritacco, E., & Vaglini, G. (2020). *Using an autoencoder in the design of an anomaly detector for smart manufacturing*. *Pattern Recognition Letters*, 136, 272-278.
- Bhogal, R., & Garg, A. (2020). *Anomaly detection and fault prediction of breakdown to repair process using mining techniques*. In 2020 International Conference on Intelligent Engineering and Management (ICIEM) (pp. 240-245). IEEE.
- Bouche, P., & Zanni-Merk, C. (2011). *Improving the performance of production lines with an expert system using a stochastic approach*. *Simulation*, 87(5), 363-383.
- Cerquitelli, T., Pagliari, D. J., Calimera, A., Bottaccioli, L., Patti, E., Acquaviva, A., & Poncino, M. (2021). *Manufacturing as a data-driven practice: methodologies, technologies, and tools*. *Proceedings of the IEEE*, 109(4), 399-422.
- Chen, Y. C., Chien, C. F., Chen, Y. Y., & Wang, C. Y. (2019). *A Data Mining Approach for Optimizing Manufacturing Parameters of Wire Bonding Process in IC Packaging Industry and Empirical Study*. In 2019 IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE) (pp. 53-57). IEEE.
- Çiflikli, C., & Kahya-Özyirmidokuz, E. (2010). *Implementing a data mining solution for enhancing carpet manufacturing productivity*. *Knowledge-Based Systems*, 23(8), 783-788.

7. Dagnino, A. (2019). *Data mining methods to analyze alarm logs in IoT process control systems*. In 2019 IEEE 15th international conference on automation science and engineering (CASE) (pp. 323-330). IEEE.
8. Dogan, A., & Birant, D. (2021). *Machine learning and data mining in manufacturing*. Expert Systems with Applications, 166, 114060.
9. Dupuis, A., Dadouchi, C., & Agard, B. (2022). *Predicting crop rotations using process mining techniques and Markov principals*. Computers and Electronics in Agriculture, 194, art. 106686.
10. Haasbroek, A., Strydom, J. J., McCoy, J. T., & Auret, L. (2018). *Fault Diagnosis for an Industrial High Pressure Leaching Process with a Monitoring Dashboard*. IFAC-PapersOnLine, 51(21), 117-122.
11. Hrcka, L., Simoncicova, V., Tadanai, O., Tanuska, P., & Vazan, P. (2017). *Using text mining methods for analysis of production data in automotive industry*. In Artificial Intelligence Trends in Intelligent Systems: Proceedings of the 6th Computer Science On-line Conference 2017 (CSOC2017), Vol 1 6 (pp. 393-403). Springer International Publishing.
12. IBM. (2021). *Big Data Analytics*. Récupéré sur <https://www.ibm.com/analytics/big-data-analytics>
13. Klaeger, T., Schult, A., & Oehm, L. (2019). *Using anomaly detection to support classification of fast running packaging processes*. In 2019 IEEE 17th International Conference on Industrial Informatics (INDIN) (Vol. 1, pp. 343-348). IEEE.
14. Liang, Z., Wang, H., Ding, X., & Mu, T. (2021). *Industrial time series determinative anomaly detection based on constraint hypergraph*. Knowledge-Based Systems, 233, art. 107548.
15. Office Québécois de la Langue Française. (2019). *Fiche terminologique - Valorisation de données*. Récupéré sur https://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26557454
16. Pabolu, V. K. R., Shrivastava, D., & Kulkarni, M. S. (2022). *A Dynamic System to Predict an Assembly Line Worker's Comfortable Work-Duration Time by Using the Machine Learning Technique*. Procedia CIRP, 106, 270-275.
17. Soltanali, H., Khojastehpour, M., & Farinha, J. T. (2021). *Measuring the production performance indicators for food processing industry*. Measurement, 173, art. 108394.
18. Souza, M. L. H., da Costa, C. A., de Oliveira Ramos, G., & da Rosa Righi, R. (2021). *A feature identification method to explain anomalies in condition monitoring*. Computers in Industry, 133, art. 103528.
19. Trunzer, E., Weiß, I., Folmer, J., Schrüfer, C., Vogel-Heuser, B., Erben, S. & Vermum, C. (2017). *Failure mode classification for control valves for supporting data-driven fault detection*. In 2017 IEEE international conference on industrial engineering and engineering management (IEEM) (pp. 2346-2350). IEEE.
20. Van Der Aalst (2016). *Process Mining: Data Science in Action*. Springer.
21. Embarak O. (2018). *The importance of data visualization in business intelligence*. *Data Analysis and Visualization Using Python: Analyze Data to Create Visualizations for BI Systems*, 85-124.