

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

**ÉTUDE DES FACTEURS AFFECTANT LA FIABILITÉ DES
TRANSFORMATEURS DE PUISSANCE**

MÉMOIRE PRÉSENTÉ

COMME EXIGENCE PARTIELLE DE LA

MAÎTRISE EN INGÉNIERIE, CONCENTRATION GÉNIE INDUSTRIEL

PAR

MATHIEU PAYETTE

MARS 2020

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

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

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

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

Ce mémoire a été dirigé par :

| | |
|--|---------------------------------------|
| Georges Abdul-Nour, Professeur | Université du Québec à Trois-Rivières |
| Prénom et nom, directeur de recherche, grade | Rattachement institutionnel |

Jury d'évaluation de l'essai ou du mémoire :

| | |
|--------------------------------|---------------------------------------|
| Georges Abdul-Nour, Professeur | Université du Québec à Trois-Rivières |
| Prénom et nom, grade | Rattachement institutionnel |

| | |
|-------------------------------|---------------------------------------|
| François Gauthier, Professeur | Université du Québec à Trois-Rivières |
| Prénom et nom, grade | Rattachement institutionnel |

| | |
|-----------------------------------|--------------------------------|
| Souheil-Antoine Tahan, Professeur | École de Technologie Supérieur |
| Prénom et nom, grade | Rattachement institutionnel |

REMERCIEMENTS

Tout d'abord, je souhaiterais remercier mon directeur de recherche, le professeur Georges Abdounour, sans qui ce projet n'aurait pas été possible. Il a su me conseiller et me diriger tout au long de ce projet enrichissant, je me sens particulièrement chanceux d'avoir pu travailler sous sa direction.

Je tiens aussi à remercier les chercheurs de l'Institut de recherche en électricité du Québec (IREQ), en particulier Olivier Blancke, qui m'a supervisé et épaulé dans l'avancement de ce projet, ainsi que Alain Côté, chercheur et chargé du projet PRIAD. J'aimerais aussi souligner le travail de tous les collaborateurs du projet PRIAD, qui ont rendu possible cette recherche, notamment Dragan Komljenovic et Jean-François Boudreau, chercheur à l'IREQ.

Je remercie également mon père, Normand Payette, pour ses conseils et sa disponibilité tout au long de mon parcours. C'est à lui que je dois ma passion pour le génie industriel, passion qu'il m'a transmise depuis mon plus jeune âge. J'aimerais aussi souligner tous les encouragements et le soutien dont m'ont témoigné ma mère, Céline, et mes frères, Étienne et Olivier. Merci de me pousser à me dépasser, et surtout, merci d'avoir toujours été présent.

Enfin, j'aimerais remercier ma femme, Marie-Michèle. Tu es à mes côtés depuis le début, avant même notre périple universitaire. Tu m'as tellement appris, toujours disponible pour me conseiller et m'épauler. Je t'aime, merci du fond du cœur!

SOMMAIRE

La modélisation statistique des défaillances auxquelles sont soumis les équipements est une activité essentielle au processus de gestion des actifs du réseau de transport d'électricité. Ces modèles affectent les stratégies et la planification organisationnelles de même que la prise de décision en gestion des actifs. L'étude de la fiabilité est d'autant plus complexe, considérant l'hétérogénéité de la population d'équipements à traiter. Effectivement, ces actifs possèdent différentes caractéristiques spécifiques, diverses contraintes d'opérations, de même qu'un historique d'entretien particulier. En conséquence, les méthodes appliquées doivent être adaptées à ces contraintes; ce projet présente l'application de modèles spécifiques aux systèmes réparables, de même que des modèles considérant l'effet de covariables sur le taux de défaillance. Un modèle de Processus de Poisson Non-Homogène (NHPP) est appliqué aux données de maintenance, afin d'estimer la fonction d'intensité générique de la population d'équipements. À cet effet, une fonction de puissance est estimée par la méthode non paramétrique de la « *Mean Cumulative Function* », prenant en compte les censures dans les données. Par la suite, des modèles d'Intensité proportionnelle (modèles de Cox) sont appliqués aux données, afin d'estimer l'effet des covariables d'intérêts sur la fiabilité. L'application de ces modèles permettra d'analyser les équipements de transport d'électricité d'Hydro-Québec, en particulier les transformateurs de puissance. Les résultats de l'étude démontrent la pertinence de ces méthodes dans la gestion des actifs. En exprimant la fiabilité en fonction du temps et de cofacteurs, les décideurs auront accès à de l'information plus complète et plus spécifique pour prendre des décisions. De plus, ce projet permet de démontrer l'importance de la qualité des données dans la gestion des actifs, compte tenu de la sensibilité des modèles statistiques utilisés en fiabilité.

Table des matières

| | |
|--|-----|
| REMERCIEMENTS | v |
| SOMMAIRE | vi |
| 0.0 Liste des symboles et abréviations..... | xv |
| 0.1 Lexique des termes en fiabilité | xvi |
| INTRODUCTION | 1 |
| CHAPITRE 1 | 2 |
| MISE EN CONTEXTE..... | 2 |
| 1.1 Problématique | 3 |
| 1.2 Questions de recherche | 3 |
| 1.3 Objectif de recherche | 4 |
| 1.4 Objectifs secondaires | 4 |
| 1.5 Considérations et limites de l'étude..... | 5 |
| 1.6 Concepts de base..... | 6 |
| CHAPITRE 2 | 10 |
| REVUE DE LITTÉRATURE | 10 |
| 2.0 Gestion des actifs | 10 |
| 2.1 Concepts de base en fiabilité et maintenance | 17 |
| 2.2 Modèles statistiques en fiabilité..... | 18 |
| 2.3 Système non réparable | 18 |
| 2.4 Système réparable | 19 |
| 2.5 Modèle HPP | 22 |
| 2.6 Estimateur du maximum de vraisemblance : | 23 |

| | | |
|---|---|----|
| 2.7 | Analyse de survie..... | 24 |
| 2.8 | Qualité et nettoyage des données..... | 33 |
| CHAPITRE 3 | | 37 |
| MÉTHODOLOGIE..... | | 37 |
| 3.0 | Analyse préliminaire et préparation des données | 37 |
| 3.1 | Modélisation du taux de défaillance | 38 |
| 3.2 | Modélisation et régression avancées..... | 41 |
| CHAPITRE 4 | | 44 |
| Analyse préliminaire | | 44 |
| 4.0 | Processus de saisie de données | 44 |
| 4.1 | Structure des données | 46 |
| 4.2 | Analyse exploratoire des données..... | 48 |
| 4.3 | Audit des données..... | 49 |
| CHAPITRE 5 | | 56 |
| Modélisation du taux de défaillance | | 56 |
| 5.0 | Données préliminaires | 56 |
| 5.1 | Résultats..... | 58 |
| 5.2 | Discussions et Recommandations..... | 60 |
| CHAPITRE 6 | | 62 |
| Analyse de survie | | 62 |
| 6.1 | Préparation des données..... | 62 |
| 6.2 | Analyse Préliminaire..... | 67 |
| 6.3 | Résultats..... | 71 |
| CONCLUSION | | 81 |

Recommandations82

LISTE DE RÉFÉRENCES84

Liste des tableaux

| | |
|---|----|
| Tableau 1.1: Sigles | xv |
| Tableau 2.1: Notation mathématique | 21 |
| Tableau 2.2: Description des équations du modèle HPP | 22 |
| Tableau 2.3: Équations du modèle NHPP..... | 23 |
| Tableau 2.4: Exemple de données (processus de comptage) | 26 |
| Tableau 2.5: Description des variables de l'exemple | 26 |
| Tableau 2.6: Exemple de données (format strates) | 27 |
| Tableau 2.7: Fonction d'intensité des modèles de survie | 29 |
| Tableau 2.8: Description des variables des modèles de survie | 30 |
| Tableau 2.9: Résumé des articles | 31 |
| Tableau 2.10: Résumé des articles (suite)..... | 32 |
| Tableau 4.1: Récapitulatif du type de documents et de leur émission | 46 |
| Tableau 4.2: Règles d'affaires des événements de maintenance | 53 |
| Tableau 4.3: Résultats de l'application des règles d'exclusion..... | 54 |
| Tableau 5.1: Échantillon des données utilisées..... | 57 |
| Tableau 5.2: Rappel des équations de la loi de puissance..... | 59 |
| Tableau 6.1: Description des variables de base du modèle..... | 62 |
| Tableau 6.2: Descriptions des variables d'indicateur d'état..... | 63 |
| Tableau 6.3: Description des variables d'équipements | 64 |
| Tableau 6.4: Description des variables d'avis | 64 |
| Tableau 6.5: Description des variables d'ordre | 64 |
| Tableau 6.6: Description des variables hybrides..... | 65 |
| Tableau 6.7: Sortie du modèle préliminaire PWP-TT | 68 |
| Tableau 6.8: Sortie du modèle préliminaire PWP-TT (suite) | 69 |
| Tableau 6.9: Extraction de champs textes pour les Ordres de défaillances anormales.... | 71 |
| Tableau 6.10: Résultats de l'application du modèle Andersen-Gill | 72 |
| Tableau 6.11: Résumé de l'impact des variables significatives pour la période terminant en 2018 | 73 |

| | |
|--|----|
| Tableau 6.12: Résumé des coefficient et intervalle de confiance pour le modèle AG | 74 |
| Tableau 6.13: Résumé de l'impact des variables significatives pour le modèle PWP | 77 |
| Tableau 6.14: Comparaison entre les deux modèles | 77 |
| Tableau 6.15: Résumé des coefficient et intervalle de confiance pour le modèle PWP .. | 78 |
| Tableau 6.16: Paramètre de la fonction d'intensité | 80 |

Liste des figures

| | |
|---|----|
| Figure 1.1: Schéma d'un transformateur de puissance (Reliability, 2012) | 7 |
| Figure 2.1: Champs d'application de la GDA (Amadi-Echendu et al., 2011)..... | 10 |
| Figure 2.2: Portée de la gestion des Actifs (IAM, 2015) | 12 |
| Figure 2.3: Activités de Gestion des Actifs (IAM, 2015)..... | 13 |
| Figure 2.4: Évaluation de la fiabilité d'un transformateur de puissance (Rexhepi, 2017) | 16 |
| Figure 2.5 Exemple de courbe baignoire (King & Jewett, 2010) | 19 |
| Figure 2.6: Politiques de remplacement (Leemis, 2009) | 20 |
| Figure 2.7: Processus ponctuel pour les systèmes réparables (Rausand & Høyland, 2003) | 21 |
| Figure 2.8: Approches pour l'évaluation et le contrôle de la qualité des données (Berti-Équille, 2018)..... | 35 |
| Figure 2.9: Cercle vertueux pour la gestion de la qualité des données (Berti-Équille, 2018) | 35 |
| Figure 3.1: Exemple de données | 40 |
| Figure 3.2: Exemple du modèle Andersen-Gill | 42 |
| Figure 3.3: Exemple du modèle conditionnel | 42 |
| Figure 3.4 : Exemple du modèle marginal | 42 |
| Figure 3.5: Méthode d'application des modèles de fiabilités | 43 |
| Figure 4.1: Processus de maintenance et de saisies de l'information | 45 |
| Figure 4.2 : Chronologie des activités de maintenance..... | 49 |
| Figure 4.3: Structure simplifiée des tables de la BD maintenance | 49 |
| Figure 4.4: Pourcentage des champs remplis par caractéristiques | 51 |
| Figure 4.5: Nombre d'ordre émis en fonction du temps..... | 52 |
| Figure 5.1: Nombre attendu de défaillances cumulées en fonction de l'âge..... | 58 |
| Figure 5.2: Taux de réparation en fonction de l'âge des équipements..... | 59 |
| Figure 5.3 : Nombre de défaillances cumulées en fonction de l'âge avec IC | 60 |
| Figure 6.1: Illustration de la variable Ratio_Ordre_Avis..... | 66 |

| | |
|---|----|
| Figure 6.2: Illustration de la variable Ratio_1 | 66 |
| Figure 6.3: Illustration de la variable Ratio_2 | 67 |
| Figure 6.4: Graphique de la fonction de risque cumulative | 70 |
| Figure 6.5: Test statistique du modèle AG..... | 76 |
| Figure 6.6: Graphique du risque cumulatif pour le modèle AG | 76 |
| Figure 6.7: Test statistique du modèle PWP | 78 |
| Figure 6.8: Graphique de la fonction de risque cumulative pour le modèle PWP..... | 79 |

Liste des équations

| | |
|---------------------|----|
| Équation 2.1 | 22 |
| Équation 2.2 | 22 |
| Équation 2.3 | 22 |
| Équation 2.4 | 22 |
| Équation 2.5 | 22 |
| Éq. 2.6 | 23 |
| Éq. 2.7 | 23 |
| Éq. 2.8 | 23 |
| Éq. 2.9 | 23 |
| Éq. 2.10 | 23 |
| Équation 2.11 | 23 |
| Équation 2.12 | 24 |
| Équation 4.1: | 53 |
| Équation 5.1 | 59 |
| Équation 5.2 | 59 |
| Équation 5.3 | 59 |
| Équation 5.4 | 59 |
| Équation 5.5 | 59 |
| Équation 5.6 | 59 |

0.0 Liste des symboles et abréviations

Tableau 1.1: Sigles

| Sigle | Sigle anglais | Signification | Traduction (anglais) |
|-------|---------------|--|---|
| HQ | - | Hydro-Québec | - |
| HQT | - | Hydro-Québec TransÉnergie | - |
| PRIAD | - | Projet de Robustesse, d'Intégration et d'Aide à la Décision pour la gestion d'actifs d'Hydro-Québec TransÉnergie | - |
| IREQ | - | Institut de recherche en électricité du Québec. | - |
| - | MTTF | Temps moyen de fonctionnement avant la défaillance | « mean time to failure » |
| - | MTBF | Temps moyen de bon fonctionnement | « mean time between failure » |
| - | ROCOF | Taux d'occurrence des événements de défaillances | « rate of occurrence of failures » |
| MBF | RCM | Maintenance basée sur la fiabilité | « Reliability centered maintenance » |
| AMDEC | FMECA /FMEA | Analyse des modes de défaillance, de leurs effets et de leur criticité | « failure modes, effects and criticality analysis » |
| GDA | AM | Gestion des actifs | Asset management |
| BD | DB | Base de données relationnelle | « DataBase » |
| HPP | HPP | Processus de Poisson Homogène | « Homogeneous Poisson Process » |
| NHPP | NHPP | Processus de Poisson Non-Homogène | « Non-Homogeneous Poisson Process » |
| AG | AG | Modèle Andersen-Gill | - |
| PWP | PWP | Modèle Prentice-Williams-Peterson | - |
| WLW | WLW | Modèle Wei-Lin-Weissfeld | - |

0.1 Lexique des termes en fiabilité

Système:

Ensemble déterminé de composants interconnectés ou en interaction.

Composant:

Partie élémentaire d'un système.

Défaillance:

Arrêt de la capacité d'un composant ou d'un système à accomplir une fonction requise (Rausand & Høyland, 2003).

Disponibilité:

Capacité d'un bien à être en état d'accomplir une fonction requise, c'est-à-dire qu'il n'est ni en réparation, ni en panne et dans des conditions données et à un instant donné (Rausand & Høyland, 2003).

Fiabilité:

Aptitude d'un composant ou d'un système à accomplir sa fonction requise, dans des conditions données et pendant un intervalle de temps connu (Rausand & Høyland, 2003).

Mécanisme de défaillance:

Processus physique ou chimique qui entraîne l'endommagement progressif d'un système résultant en une perte progressive de la fonctionnalité et menant à une défaillance (Lazzaroni, Cristaldi, Peretto, Rinaldi, & Catelani, 2011).

Mode de défaillance:

Effets/manifestations physiques résultant d'une défaillance (Lazzaroni et al., 2011).

INTRODUCTION

La complexification des équipements et des systèmes industriels requiert une expertise très poussée, notamment dans le domaine de l'ingénierie et de la fiabilité. De surcroît, certaines organisations ont à gérer des milliers d'équipements distincts qui génèrent une importante quantité et diversité de données de maintenance. Ces facteurs rendent la tâche difficile pour les gestionnaires, puisque les outils de gestion classique sont souvent désuets. En effet, ces outils sont plutôt orientés vers la gestion des opérations (planification, ordonnancement, etc.) et sont souvent moins robustes pour la modélisation et les prédictions complexes. En revanche, le développement fulgurant de nouvelles techniques permet désormais le traitement des mégadonnées, et de considérer beaucoup plus de paramètres dans la prise de décision. À cet effet, ce mémoire proposera une démarche de modélisation de la fiabilité des systèmes réparables, afin d'améliorer la précision des informations nécessaires aux processus décisionnels en gestion des actifs (GDA). Les méthodes identifiées seront ensuite appliquées au cas des transformateurs de puissances du système de transport d'électricité d'Hydro-Québec.

Pour commencer, le chapitre 1 permettra de mettre en contexte le projet au sein de l'organisation d'Hydro-Québec. La problématique et les questions de recherches qui en découlent seront présentées, ainsi que les objectifs principaux et secondaires. Une brève définition des concepts permettra d'assurer une bonne compréhension tout au long des chapitres suivants. Ensuite, le chapitre 2 établira la connaissance et les avancements dans le domaine. La revue de littérature portera sur les concepts de base en fiabilité et en maintenance et en gestion des actifs. Puis les enjeux concernant la qualité des données et les modèles mathématiques et statistiques en fiabilité seront décrits. Le chapitre 3 présentera la méthodologie utilisée dans la réalisation de ce travail. En outre, la méthode d'acquisition de données, la méthodologie pour la préparation et le formatage des données, les techniques d'analyse utilisées, ainsi que les actions désirées qui en découlent seront établies. Les chapitre 4, 5 et 6 serviront à présenter les résultats, l'analyse de ces derniers, puis ils seront discutés et critiqués. Le dernier chapitre servira à la conclusion du mémoire, ainsi qu'à la formulation de recommandations.

CHAPITRE 1

MISE EN CONTEXTE

Pour commencer, il importe de mentionner que ce projet de recherche est effectué en partenariat avec Hydro-Québec, par le biais de la Chaire de recherche Hydro-Québec en gestion des actifs, à l'Université du Québec à Trois-Rivières. Celui-ci fait partie d'un plus grand projet d'innovation au sein d'Hydro-Québec, soit le Programme de Robustesse, d'Intégration et d'Aide à la Décision pour la gestion d'actifs d'Hydro-Québec TransÉnergie (PRIAD). Le projet PRIAD est piloté par une équipe de chercheurs de l'IREQ, l'institut de recherche d'Hydro-Québec.

Pour ce qui est d'Hydro-Québec (HQ), il s'agit d'une société d'État québécoise dont l'unique actionnaire est le gouvernement du Québec. HQ est responsable de la production, du transport et de la distribution de l'électricité de la province canadienne. L'organisation d'Hydro-Québec a été scindée en plusieurs divisions fonctionnelles, afin de respecter la réglementation de la « *Federal Energy Regulatory Commission* ». Ainsi, pour exporter l'électricité à l'extérieur du Québec, les divisions de production (Hydro-Québec Production), de distribution (Hydro-Québec Distribution) et de transport (Hydro-Québec TransÉnergie) ont été créées. Hydro-Québec TransÉnergie (HQT) est la division d'HQ qui est responsable du transport de l'électricité à travers la province. Cette division est sous la réglementation de la Régie de l'énergie du Québec, qui détermine notamment les tarifs du transport de l'énergie. Ainsi, l'institut de recherche d'HQ (IREQ) est mandaté pour mener les projets d'innovations tout en conservant cette division fonctionnelle.

Quant au projet PRIAD, il a été initié afin de répondre aux besoins de l'équipe de Planification et gestion des actifs (PGA) et de la direction principale Planification, expertise et soutien opérationnel (DESO) qui a identifié différentes problématiques, notamment sur la qualité des données ainsi que sur les modèles et les systèmes informatiques utilisés dans leurs modèles de GDA. L'équipe de l'IREQ a comme mandat de regrouper et de traiter les données pertinentes à la GDA de TransÉnergie tout en améliorant les différents modèles de comportement des actifs. Dans ce contexte, la problématique et les objectifs du mémoire seront orientés vers le besoin du client.

1.1 Problématique

Comme mentionné, l'équipe qui est chargée de la gestion des actifs fait face à différents enjeux, notamment en ce qui a trait à la qualité des données disponibles pour les analyses de fiabilité. Les modèles de comportements des actifs, ainsi que les systèmes informatiques en place ne répondent plus aux besoins de l'équipe PGA, qui requiert le développement de nouveaux outils plus adaptés. Qui plus est, il existe peu d'étude sur l'application de modèle de fiabilité pour les systèmes réparables permettant de quantifier l'impact de variables d'intérêts et qui sont appliquées par des contextes réels (Shwu-Tzy, Landers, & Rhoads, 2006).

1.2 Questions de recherche

Le contexte opérationnel particulier d'Hydro-Québec TransÉnergie rend la GDA très complexe, d'autant plus qu'essentiel. Ce contexte favorise l'émergence de plusieurs questionnements :

- Comment modéliser le comportement d'équipements de transport d'électricité, dans un contexte pratique, avec des données comportant des biais et des incertitudes?
- Quels sont les modèles qui décrivent adéquatement la fiabilité d'équipements réparables?
- Comment représenter mathématiquement le comportement en fiabilité d'une famille d'actifs ayant des caractéristiques et des environnements d'opérations différents (famille d'équipements hétérogènes)?
- Comment quantifier l'impact de certaines variables sur le taux de défaillance d'une famille d'équipements?
- Comment peut-on traiter les biais et les incertitudes au sein des données à analyser?
- Doit-on classifier les équipements, de façon à avoir des modèles plus performants, en développant des familles plus homogènes?

1.3 Objectif de recherche

Découlant de ces questions, ce projet vise l'application de nouveaux modèles de comportements des actifs du réseau de transport d'électricité, dans le but d'améliorer la compréhension des facteurs affectant la fiabilité. Le modèle et la méthodologie seront appliqués dans une étude de cas, qui se focalisera sur les transformateurs de puissance. Un transformateur sera considéré comme étant un système réparable, c'est-à-dire un système qui peut être réparé après un événement de défaillance.

1.4 Objectifs secondaires

Outre l'application de nouveaux modèles de comportements, cette étude traitera aussi la qualité des données. Effectivement, la performance d'un modèle mathématique est intrinsèquement liée à la qualité des données dont il découle. Pour ce faire, des stratégies seront identifiées afin d'optimiser la qualité des données à l'étude.

De plus, des méthodes analytiques seront appliquées afin d'explorer le jeu de donnée («*data mining*»), pour en faire sortir le maximum de connaissances. Des méthodes de statistiques seront présentées à cet effet. Ensuite, des modèles de survie seront appliqués pour quantifier les facteurs influençant la fiabilité d'une même famille d'équipements. Les résultats permettront de déterminer l'applicabilité de cette méthode dans le contexte de HQT.

1.5 Considérations et limites de l'étude

Selon les experts consultés au début du projet, l'étude des caractéristiques de fiabilité des équipements représente un grand défi en raison de la qualité et de la diversité des données opérationnelles. Les erreurs peuvent venir des changements de systèmes au fil des années, ou bien des changements de méthodes quant à l'acquisition de ces données. Le premier constat est que la qualité des données doit être améliorée pour effectuer une étude de fiabilité. Comme mentionné plus haut, les transformateurs sont considérés comme des systèmes réparables. Cela signifie que les défaillances sont traitées par équipement, et non par composantes de l'équipement, et que, ce système peut être réparé et ramené à un état de fonctionnement.

Cette étude de cas vise à présenter une démarche d'analyse de la fiabilité des équipements réparables de réseau électrique. Elle se concentrera en particulier sur les transformateurs de puissance. Cela exclut donc les transformateurs de distribution, les transformateurs de mesures, ou tout autre type d'équipements. Différentes techniques de modélisation seront présentées et appliquées, mais le but de l'étude n'est pas le développement ou l'amélioration de ces techniques.

1.6 Concepts de base

Transport de l'électricité :

Pour acheminer l'électricité de la centrale hydroélectrique jusqu'au client, l'énergie doit parcourir des centaines, voire des milliers de kilomètres. À cet effet, Hydro-Québec a développé et mis en place un réseau permettant de limiter les pertes énergétiques lors du transport de l'électricité (Hydro-Québec, 2019c). L'entreprise québécoise transporte l'électricité par des lignes électriques haute tension (735 kV) qui sont supportées par différents types de pylônes. Le réseau commence donc à la source, où l'énergie est produite, puis est relié à des postes qui servent à moduler la tension, avant d'être acheminé vers le client.

Poste électrique :

Un poste électrique est une partie du réseau qui regroupe en un même lieu des appareils de mesures électriques, des transformateurs de tension, des transformateurs de courant et des transformateurs de puissance. Ils sont aussi munis d'équipements de protection (parafoudre, disjoncteur, *etc.*), de sectionneurs et de plusieurs autres équipements. Un poste sert non seulement à gérer la tension du courant, mais aussi à minimiser les impacts reliés à la perte de courant d'un tronçon, en minimisant les distances des lignes (Hydro-Québec, 2019d).

Transformateur:

Les transformateurs sont des appareils qui permettent d'élever ou d'abaisser soit la tension (différence de potentiel en volt), le courant (intensité de courant en ampère) ou la puissance (kilo volts) électrique dans un réseau.

Dans un premier temps, un transformateur de courant est un « *transformateur de mesure dans lequel le courant secondaire est, dans les conditions normales d'emploi, pratiquement proportionnel au courant primaire et déphasé par rapport à celui-ci d'un angle voisin de zéro, pour un sens approprié des connexions* », c'est-à-dire qu'il s'agit d'un appareil qui abaisse l'intensité de courant, afin d'alimenter différents appareils de mesures et de compteurs (Hydro-Québec, 2019a; internationale, 2002).

Ensuite, un transformateur de tension est un « *transformateur de mesure dans lequel la tension secondaire est, dans les conditions normales d'emploi, pratiquement proportionnel à la tension primaire et déphasée par rapport à celle-ci d'un angle voisin de zéro, pour un sens approprié des connexions* », c'est-à-dire qu'il s'agit d'un appareil qui abaisse la tension du courant (le voltage), afin d'alimenter différents appareils de mesures et de compteurs (Hydro-Québec, 2019a; internationale, 2003). Finalement, un transformateur de puissance est un :

Appareil statique à deux enroulements ou plus qui, par induction électromagnétique, transforme un système de tension et courant alternatif en un autre système de tension et de courant de valeurs généralement différentes à la même fréquence dans le but de transmettre de la puissance électrique. (Hydro-Québec, 2019b)

En d'autres termes, il permet d'augmenter ou d'abaisser la tension dans le réseau électrique (Hydro-Québec, 2019b; internationale, 2011). C'est d'ailleurs cet équipement qui sera à l'étude dans ce mémoire et qui est illustré ci-dessous.

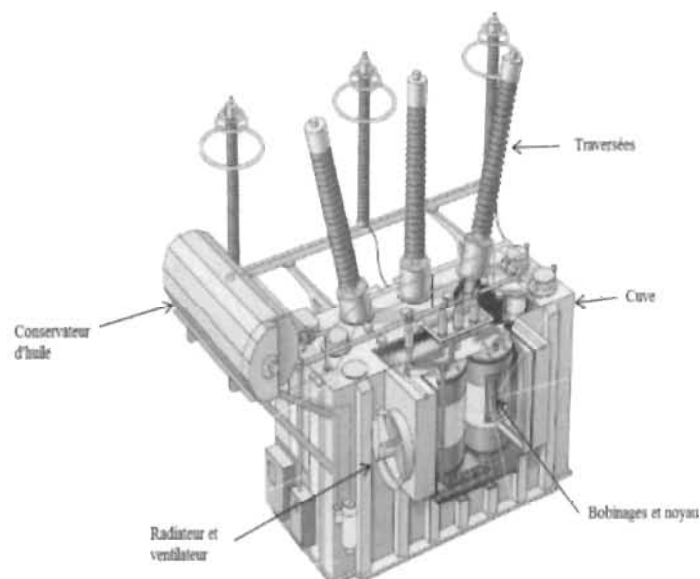


Figure 1.1: Schéma d'un transformateur de puissance (Reliability, 2012)

Fiabilité :

La fiabilité se définit comme étant la probabilité qu'un système remplisse ses fonctions d'usage, durant une certaine période de temps et selon certaines conditions environnementales. Dans ce contexte, une défaillance est un événement qui rend le système inapte à remplir ses fonctions. La défaillance peut causer un arrêt total du système, mais dans certains cas, elle ne fait qu'empêcher le système de fonctionner correctement (Wildi & Sybille, 2005). En addition, la fiabilité peut être exprimée soit en fonction du temps ou en fonction du nombre de cycles d'utilisation, selon les particularités du système à l'étude.

Maintenance :

La maintenance est l'ensemble des activités exécutées dans le but de maintenir ou de rétablir les fonctions d'usage d'un système. Dans cette étude de cas, les différents types de maintenance seront définis comme suit :

Maintenance corrective :

La maintenance corrective survient lorsqu'une défaillance cause la panne du système, signifiant un arrêt total des fonctions du système.

Maintenance préventive :

La maintenance préventive est un type de maintenance qui survient avant l'arrivée d'une défaillance. Elle peut être effectuée à intervalles constants ou selon certaines conditions. Le but est donc de diminuer la probabilité qu'une défaillance ait lieu. Elle regroupe entre autres la maintenance systématique, conditionnelle et prévisionnelle.

La maintenance systématique est une maintenance qui est effectuée en fonction de l'utilisation de l'équipement, c'est-à-dire selon le nombre de cycles ou selon le temps d'utilisation. Les conditions (cycles ou temps) sont prédéterminées à l'avance.

La maintenance conditionnelle survient lorsque le système démontre des signes d'usures et que cette usure dépasse un seuil préalablement établi. Ces signes sont souvent précurseurs d'événements de défaillance, c'est pourquoi il convient de faire la maintenance. Ces éléments d'usures peuvent être révélés lors d'une inspection de routine, mais peuvent aussi être détectés par des capteurs.

La maintenance prévisionnelle, parfois traduite par maintenance prédictive, est une forme de maintenance conditionnelle. Bien qu'une maintenance soit déclenchée lorsque l'usure d'une composante dépasse un certain seuil, la maintenance prévisionnelle, elle, repose sur le monitoring en continu des équipements. Cela permet de constater l'évolution des dégradations et de prévoir une éventuelle défaillance (Levitt, 2011).

Mégadonnées :

Mégadonnées, plus communément appelé « *Big Data* », est le terme usuel pour désigner un ensemble de données très volumineux, dans lequel il est difficile d'extraire des connaissances par les capacités humaines, et même par les techniques informatiques classiques.

Ce chapitre a permis de décrire le contexte du projet, en plus de définir la problématique et les objectifs spécifiques de recherche. De plus, il aura servi à introduire certains concepts de base, facilitant la compréhension du mémoire. Le chapitre suivant présente la revue de littérature, qui recense les contributions scientifiques en lien avec le projet.

CHAPITRE 2

REVUE DE LITTÉRATURE

Le chapitre précédent a permis de définir les objectifs et les limites en lien avec ce projet. La revue de littérature permettra d'établir les connaissances scientifiques existantes en lien avec le domaine de recherche, en plus de présenter les différents avancements.

2.0 Gestion des actifs

Concept de base en GDA :

Tout d'abord, un actif est défini comme étant un objet auquel une entité légale lui attribue une valeur. Il existe deux types d'actifs : l'actif d'ingénierie et l'actif financier. L'actif d'ingénierie est celui dont il sera question durant ce texte (Amadi-Echendu, Brown, Willett, & Mathew, 2011). L'actif financier a une valeur établie de façon contractuelle entre les entités légales. Il s'agit, par exemple, d'actions boursières, de droit de brevet. L'actif d'ingénierie a une valeur qui existe, indépendamment des contrats établis entre différents partis. L'actif d'ingénierie peut être un équipement, de l'inventaire, un bâtiment, *etc.* La valeur de l'actif d'ingénierie peut être monétaire, ou selon ses aptitudes, par exemple la capacité de production d'une machine par période de temps (Amadi-Echendu et al., 2011). La relation entre les différentes sphères de la GDA et l'implication de la fiabilité et de la maintenance sont démontrées à la figure suivante :

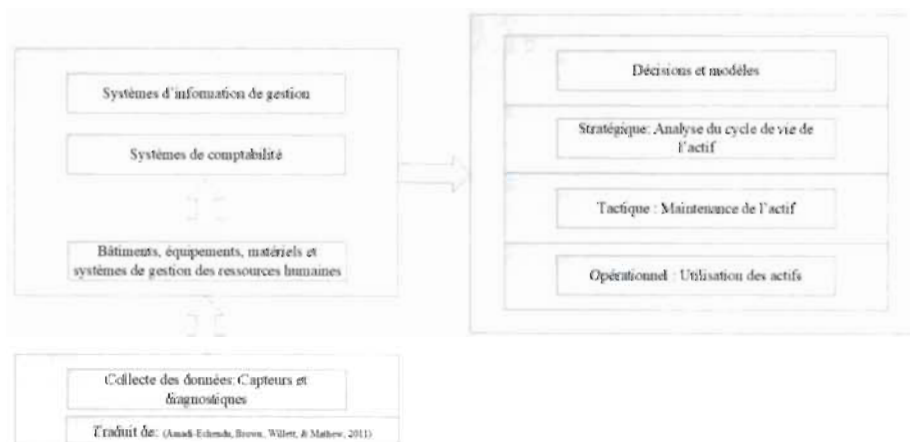


Figure 2.1: Champs d'application de la GDA (Amadi-Echendu et al., 2011)

Pour continuer, la série de norme ISO 55000, publiée en 2014, a permis d'accroître l'acceptation et la reconnaissance de la discipline de la gestion des actifs (IAM, 2015). La norme ISO 55000 décrit la gestion des actifs comme étant l'ensemble des activités

coordonnées d'une organisation pour réaliser la valeur d'un actif. L'actif se définit comme un bien, une chose ou une entité qui a de la valeur, potentielle ou actuelle, pour une organisation. La réalisation de l'actif implique la balance des coûts, des risques, des opportunités et des bénéfices de performance [Traduction libre]. (International Organisation for Standardization, 2015). À cet effet, chaque organisation doit établir ce que représente la valeur, selon le contexte et les objectifs qui lui sont propres. La norme ISO établit aussi les principes fondamentaux de la gestion des actifs :

- Valeur :
 - La fonction de l'actif est de générer de la valeur pour une organisation et la GDA a pour objectif la réalisation de cette valeur (International Organisation for Standardization, 2015).
- Positionnement :
 - La GDA s'aligne aux objectifs d'entreprise (positionnement). Ces objectifs se concrétisent par les décisions, la planification et les activités financières et techniques de l'entreprise (International Organisation for Standardization, 2015). Les objectifs doivent se transposer dans les politiques et les stratégies de gestion (IAM, 2015).
- Leadership :
 - Le leadership et la culture d'entreprise sont des déterminants essentiels à la réalisation de la valeur des actifs. Un engagement à tous les niveaux est essentiel à l'atteinte des objectifs de GDA (International Organisation for Standardization, 2015).
- Garantie :
 - La GDA doit garantir que les actifs rempliront leurs fonctions, et ce, de façon régulière et durable (IAM, 2015; International Organisation for Standardization, 2015).

À partir de ces principes et standards, la « *Institute of Asset Management* » (IAM) a élaboré un modèle de gestion des actifs. Celui-ci est présenté à la Figure 2.2, et la Figure 2.3 détaille les différentes activités qui ont été assignées pour chaque thème de la GDA.

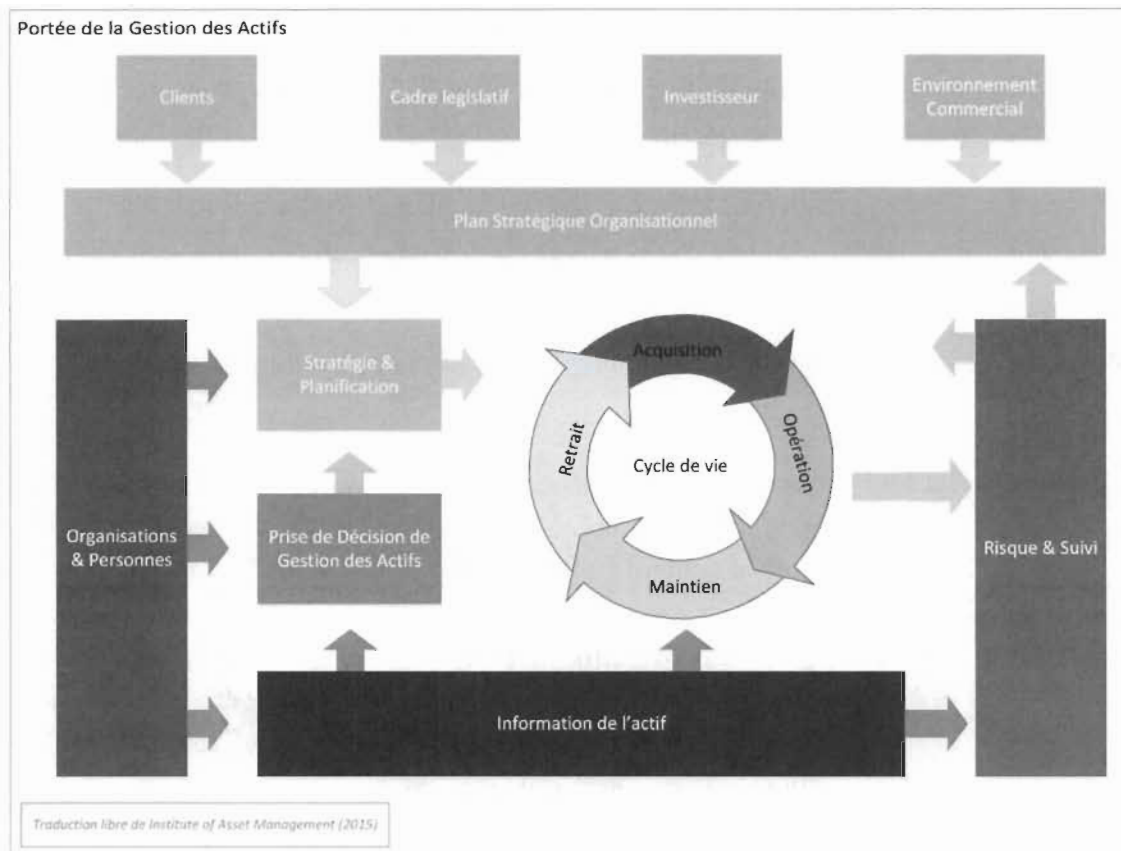


Figure 2.2: Portée de la gestion des Actifs (IAM, 2015)

Chacun des sujets est défini comme suit :

- Stratégie et planification :
 - Regroupe les activités qui favorisent la mise en œuvre du positionnement de l'entreprise sur les besoins (IAM, 2015).
- Prise de décisions de gestion des actifs :
 - Regroupe les activités de prise de décision sur l'acquisition et la création, de même que les opérations et la maintenance de l'actif (IAM, 2015).
- Livraison du cycle de vie :
 - Regroupe les activités d'exécution et de contrôle des plans et des stratégies de l'entreprise définis préalablement (IAM, 2015).
- Information de l'actif :
 - Regroupe les activités relatives à la mise en place des structures informatiques pour le traitement des données, de même que l'identification de l'information et de la qualité de données requise (IAM, 2015).
- Organisations et personnes :
 - Regroupe les activités de mise en place et/ou de restructuration organisationnelle favorisant la gestion des actifs. La GDA peut mener à des changements importants, notamment dans les rôles et responsabilités (IAM,

2015). Dans cette sphère d'activité, les principes de leadership énoncés plus haut sont fondamentaux.

- Risque et suivi :
 - Regroupe les activités d'identifications et de gestion des risques inhérents aux actifs. Cela inclut aussi les activités de suivi, qui assure entre autres l'atteinte des objectifs d'entreprise, de même que l'amélioration continue de ces processus.

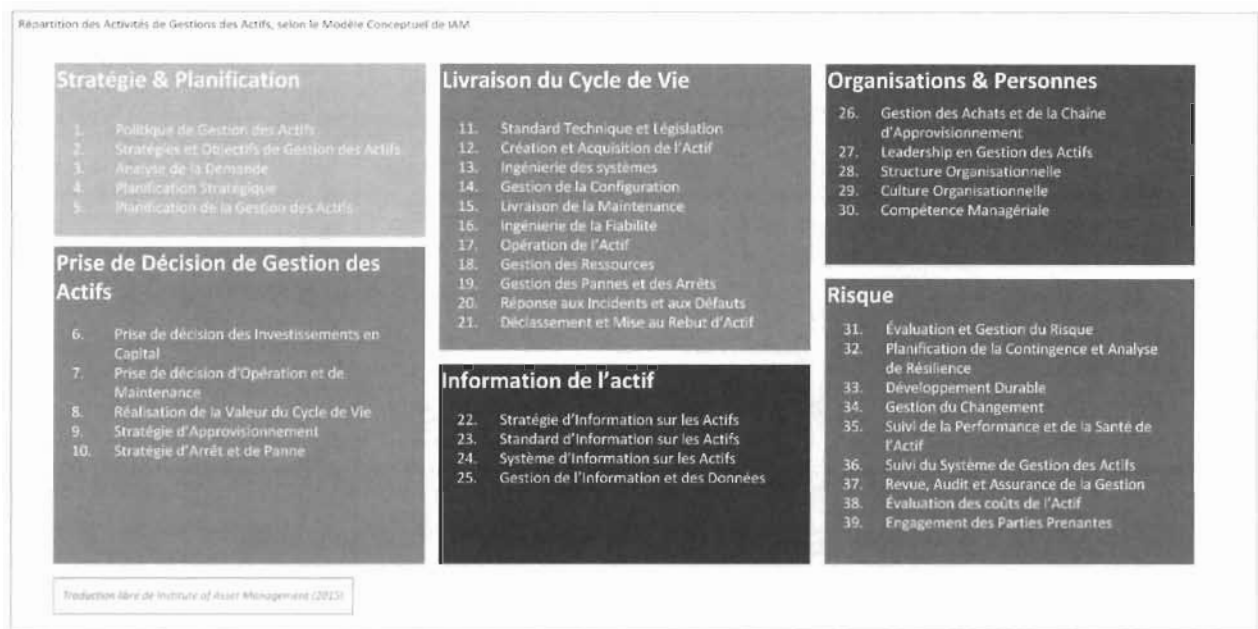


Figure 2.3: Activités de Gestion des Actifs (IAM, 2015)

En ce qui concerne ce projet, il permet de travailler sur plusieurs de ces activités de GDA, principalement l'ingénierie de la fiabilité. Un des objectifs est de déterminer les facteurs affectant la fiabilité des équipements. À cet effet, l'effet des politiques de maintenance sera étudié (planification de la GDA et opération de l'actif). Qui plus est, les modèles qui sont appliqués serviront d'intrant à PGA, qui œuvre notamment dans la stratégie et la planification, de même que la prise de décision en gestion des actifs. Finalement, ce projet touche aussi aux activités informationnelles de l'actif.

Davis (2007) propose une définition intégrée des concepts clés en GDA. Selon lui, la GDA est un processus d'amélioration continue visant l'amélioration de la fiabilité, de la disponibilité, de la sécurité et de la longévité des actifs. La GDA ne se limite pas à la maintenance, mais doit mettre en œuvre toutes les fonctions d'une organisation. Les piliers

en gestion des actifs reposent principalement sur la connaissance des équipements, de leur contexte opérationnel, de leur design, *etc.* Elle repose évidemment sur les activités de maintenance, mais doit être considérée dans une vue de performance globale au sein de l'organisation (Davis, 2007).

À cet effet, les transformateurs de puissances sont des actifs critiques du réseau de transport d'électricité. Ce projet permettra d'appliquer de nouvelles techniques de modélisation et, en conséquence, augmenter la compréhension du risque de défaillance des transformateurs dans le réseau électrique. Cela permettra de mettre en lien les caractéristiques techniques, les conditions d'utilisations, ainsi que les politiques de maintenance avec la probabilité de défaillance des équipements dans le but d'améliorer les performances des actifs.

Application de la GDA

Les articles suivants présentent les applications de la GDA, en particulier celles portant sur les réseaux électriques et aux transformateurs de puissance.

Pour commencer, Carneiro (2013) présente les détails d'une méthode de priorisation des équipements à risque, centrée sur la maintenance prédictive et appliquée aux transformateurs de puissance. L'étude se concentre sur la simulation de différents scénarios de maintenance, en utilisant les données historiques des équipements, l'analyse des types et des modes de défaillances et les caractéristiques techniques des équipements. Le but est d'optimiser en parallèle les coûts ainsi que l'effet de la maintenance sur la fiabilité des équipements, pour finalement produire un plan de maintenance qui tient compte des caractéristiques du système (Carneiro, 2013).

Singh, Sinha et Joshi (2016) présentent une revue des différentes démarches d'application en maintenance basée sur la fiabilité (MBF) comme stratégie de GDA, dans le contexte d'un poste électrique. Tout d'abord, les étapes de la MBF sont définies :

1. Identification des actifs critiques :
 - a. Analyser le comportement des composantes et leurs interactions envers les différents systèmes, lorsqu'une panne se produit.
2. Analyse des modes de défaillance, de leurs effets et de leur criticité (AMDEC) :
 - a. Examiner les modes de défaillance, ainsi que les causes, la fréquence, les effets, la gravité et les différentes méthodes de détection qui en découlent.

3. Développement des actions de maintenance basées sur la fiabilité :
 - a. Développer des tâches pour diminuer les risques associés à chaque mode de défaillance identifié.
4. Comparaison des tâches de maintenance :
 - a. Comparer les nouvelles tâches développées avec celles existantes, et voir comment les intégrer dans une nouvelle planification.

Ensuite, ils présentent différentes méthodes analytiques, comme les estimateurs linéaires (MTBF, MTTF, *etc.*) ainsi que des approches statistiques de base. Ils recensent aussi les causes potentielles de problèmes d'équipements qui sont détectables par les données (Singh, Sinha, & Joshi, 2016):

- Problèmes de design d'équipements :
 - Classement des défaillances par types d'équipements, pour retrouver des récurrences dans les composantes impliquées.
- Problèmes reliés aux matériaux :
 - Analyse des données démontre une usure prématurée de certains équipements.
- Problèmes de design du système :
 - Analyse des défaillances récurrentes d'un équipement démontre qu'il a mal été sélectionné dans la composition du système.
- Problème dans les procédures de maintenance :
 - Comparatif les données d'un poste à un autre, selon les différentes équipes de maintenance.
- Problèmes de procédure d'opération :
 - Analyse des données démontre une usure prématurée de certains équipements.

Finalement, ils présentent différentes méthodes de détection d'anomalies, ainsi qu'une méthode de calcul pour l'optimisation des coûts-bénéfices.

Velasquez-Contreras, Sanz-Bobi et Galceran Arellano (2011) présentent les éléments d'un modèle de GDA pour la distribution d'électricité. L'implémentation d'un système intelligent pour la détection et le diagnostic des défaillances ainsi que l'estimation du taux de panne sont montrés en exemple. Comme première étape, ils ont développé une plateforme informatique pour l'intégration des données des différents systèmes informatiques de l'entreprise. Cette plateforme, en plus d'inclure les données de maintenance, regroupe les conditions d'opération, les historiques d'opérations et les spécifications de chaque équipement. Ensuite, ils ont développé un module de détection d'anomalies; des capteurs enregistrent les données en temps réel qui sont traitées par un

réseau de neurones. Le réseau de neurones est entraîné pour détecter l'apparition de modes de défaillance. Un module de diagnostic permet d'évaluer les conditions obtenues sur l'état du système et de confirmer s'il y a apparition de modes de défaillances. Ces deux modules enregistrent les données, qui sont transférées dans un module d'optimisation. Un autre module est utilisé afin d'estimer le taux de défaillance des équipements; un modèle de Markov caché a été utilisé pour déterminer les probabilités de transitions du système. Une estimation des taux de défaillance permet de calculer le MTTF, ainsi que différents indicateurs d'état du système. Tous les modules développés servent d'intrants à un optimiseur, qui prend en considération l'ensemble de ces variables pour proposer des plans de maintenance optimisée (Velasquez-Contreras, Sanz-Bobi, & Galceran Arellano, 2011).

Rexhepi (2017) présente les résultats d'analyses statistiques des transformateurs de puissances du système de transport d'énergie au Kosovo. L'étude recense les types de défaillances, les analyses de fiabilité et les politiques de maintenances adoptées. Les données obtenues démontrent que 61% des pannes sont causées par un court-circuit externe du système de distribution et 7% par une trop grande intensité du courant dans le système. Pour chaque composante impliquée, il calcule la probabilité de défaillance, selon le type de panne entraînée, démontré à la Figure 2.4. L'étude confirme l'importance de surveiller (monitoring) les conditions d'opération des équipements, dans le but de mieux gérer les transformateurs de puissance (Rexhepi, 2017).

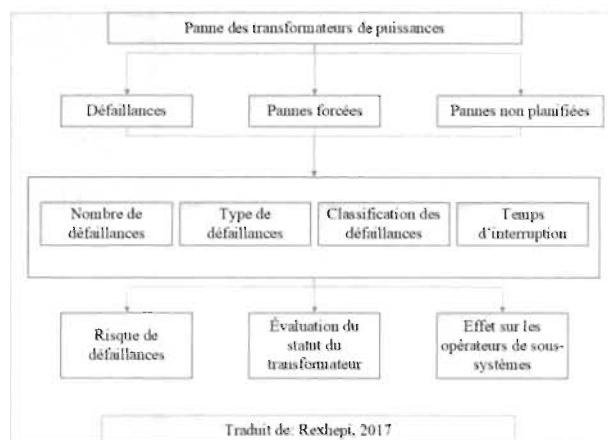


Figure 2.4: Évaluation de la fiabilité d'un transformateur de puissance (Rexhepi, 2017)

Amadi-Echendu et Mafutsana (2016) présentent les développements, autant dans la technologie des transformateurs de puissance, que dans la gestion de ce type d'actif. En effet, ils proposent une revue bibliographique des tendances de 1970 à 2014. Ils exposent

les différents changements technologiques qui ont permis d'améliorer les matériaux et la structure des composantes, notamment par l'évolution des techniques de modélisation par ordinateur. La publication fait aussi état des avancements dans le monitoring des équipements à l'aide de capteurs, ce qui permet une meilleure application de la MBF et un meilleur suivi de la santé des équipements (Amadi-Echendu & Mafutsana, 2016).

2.1 Concepts de base en fiabilité et maintenance

Le but des activités de maintenance est d'assurer et de maintenir la performance d'un produit ou système. La maintenance se distingue de la production de biens dans le fait que la qualité ou la conformité d'un bien est fonction de la performance de celui-ci, selon des spécifications préétablies (tolérance, % défectueux, *etc.*). En production, le bien est conforme s'il respecte les standards et qu'il est en mesure d'effectuer l'action pour laquelle il a été conçu. Lorsqu'il est question de maintenance, la fiabilité est la capacité de ce bien à effectuer ses actions d'usage selon un intervalle de temps donné. Ainsi, la fiabilité peut être exprimée comme étant la probabilité qu'un bien soit en mesure de performer ses fonctions d'usages dans un intervalle de temps et dans des conditions préétablies (Lazzaroni et al., 2011).

La fiabilité s'exprime sous la forme de modèle mathématique, ou sous forme d'estimateur statistique, comme le temps moyen avant la défaillance (MTTF) ou le temps moyen de bon fonctionnement (MTBF). Le but de l'étude de fiabilité est de prévenir et/ou de réduire la fréquence des défaillances en appliquant l'ingénierie et les diverses techniques statistiques. C'est aussi d'identifier et d'éliminer, si possible, les mécanismes qui causent des défaillances, en dépit des efforts de prévention. Enfin, les fiabilistes doivent estimer la fiabilité de nouveaux composants (design) et analyser les données (O'Connor & Kleyner, 2011), notamment à l'aide de tests accélérés de durée de vie.

En ce qui a trait à la maintenance, il s'agit des actions visant la mise en œuvre des objectifs de production d'une entreprise. Elle inclut l'ensemble des activités de planification, de gestion et d'exécution de l'entretien des équipements, dans le but d'atteindre ces objectifs d'entreprise. En fiabilité, on distingue deux types de systèmes, les systèmes réparables et les systèmes non réparables.

2.2 Modèles statistiques en fiabilité

Il existe plusieurs façons de représenter mathématiquement la fiabilité, soit avec des modèles non paramétriques et des modèles paramétriques.

Modèles non paramétriques

Les analyses non paramétriques permettent de prédire la fiabilité d'un système, sans l'utilisation de distribution statistique. La construction d'histogramme est largement employée pour étudier le comportement des systèmes, de même que l'analyse de la moyenne et de la variance d'un échantillon. Ces méthodes permettent d'estimer la fonction de densité de probabilité, sans avoir recours à une loi de distribution spécifique (Lewis, 1996).

Modèles paramétriques

De leur côté, les analyses paramétriques permettent de choisir la distribution de probabilité appropriée et d'en évaluer les paramètres, à partir de données d'entretien des équipements. Plusieurs facteurs viennent influencer le choix d'une distribution, que ce soit selon l'ajustement de la courbe ou par le contexte opérationnel des phénomènes à l'étude (Lewis, 1996).

2.3 Système non réparable

Lorsqu'il est question d'un système non réparable, cela signifie qu'une défaillance entraîne la fin de la vie du système. Il s'agit alors de calculer la probabilité de survie, ou taux de risque, afin de déterminer l'occurrence de cet événement. Le temps moyen de fonctionnement avant la défaillance est utilisé afin d'exprimer la durée de vie du système en question (O'Connor & Kleyner, 2011).

Le taux de défaillance d'un système est caractérisé selon trois états, soit décroissant en début de vie, constant durant la vie utile ou en croissance lors de la fin de vie. Ces comportements sont généralement présentés sous la forme de la courbe baignoire.

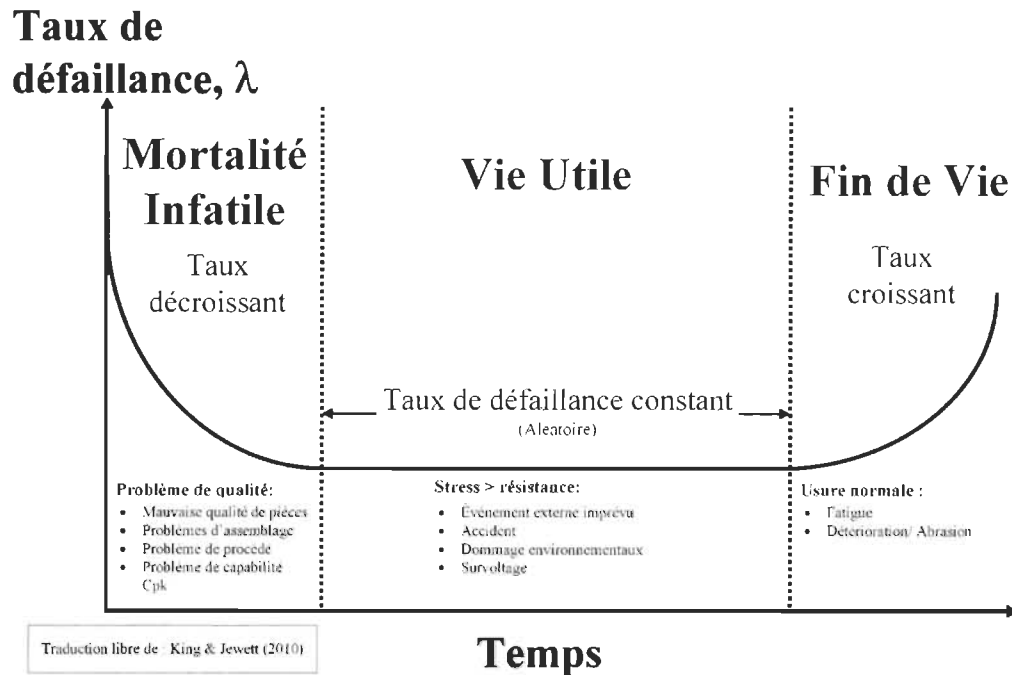


Figure 2.5 Exemple de courbe baignoire (King & Jewett, 2010)

Les défaillances en début de vie sont souvent associées à des défauts dus à la production, à l'assemblage, *etc.* Différents contrôles sont utilisés afin d'empêcher ces produits de se rendre au client. Lorsqu'il s'agit de défaillance durant la vie utile, il est question de défaillance aléatoire, c'est-à-dire qu'un niveau de stress inhabituel sur le système dépasse sa capacité et cause la défaillance. Finalement, lorsque le nombre de défaillances est croissant (en fin de vie), elles sont généralement associées au vieillissement et à la dégradation physique des composantes (King & Jewett, 2010).

2.4 Système réparable

Dans le cas d'un système réparable, il se distingue par le fait que lorsqu'un événement de défaillance survient, le système peut être ramené dans un état de fonctionnement. La fiabilité est alors exprimée comme le taux d'occurrence des événements de défaillances (ROCOF ou taux de réparation), ou comme le temps moyen de bon fonctionnement (MTBF).

Le taux de défaillance peut varier avec le temps, ou demeurer constant. Un taux constant, comme pour les systèmes non réparables, est associé à des défaillances aléatoires causées par des facteurs externes. Dans d'autres cas, le taux de défaillance peut être

décroissant, si une réparation vient améliorer l'état de santé d'un équipement. Finalement, l'usure globale du système causera progressivement un taux de défaillance en croissance.

Le choix d'un modèle mathématique dépend grandement du type de décision à prendre, en plus du type de système à l'étude. Dans la représentation des systèmes réparables, il y a trois cas de figure auxquels les modèles de probabilités permettent de répondre. La première situation est lorsqu'il faut remplacer un élément non réparable d'un système. Ce type de problème, nommé modèle de remplacement, permet entre autres de déterminer le nombre de pièces de remplacement en inventaire, afin de minimiser les coûts de stockage et de calculer la probabilité de pénurie de stock. Cela permet aussi de choisir différente politique de remplacement :

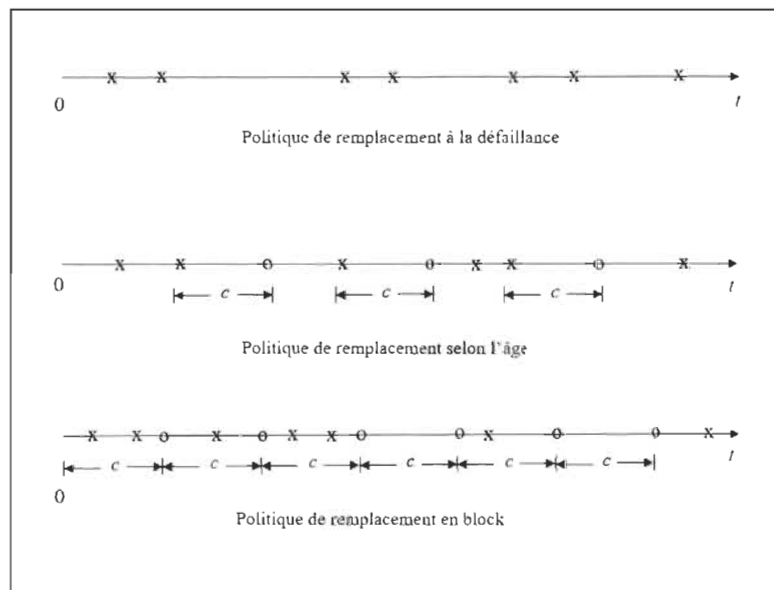


Figure 2.6: Politiques de remplacement (Leemis, 2009)

La Figure 2.6, tiré de Leemis (2009), permet d'identifier trois politiques de remplacement. Les défaillances sont illustrées par un « x » et les remplacements par un « o ». La première politique stipule qu'un remplacement de pièce est effectué seulement lorsque celle-ci subit une défaillance. Dans la seconde politique, un remplacement de la pièce est effectué lorsqu'elle atteint un âge prédéfini « c ». La troisième politique consiste à remplacer la composante à chaque intervalle de temps « c », peu importe l'âge de la pièce (Leemis, 2009). Pour continuer, la seconde problématique se résout à l'aide des modèles de maintenance. Ces modèles sont utilisés lorsqu'un système doit subir de la maintenance

préventive, ainsi que de la maintenance corrective. Le but est donc de déterminer à quel moment il faut effectuer la maintenance préventive, par exemple, à quelle fréquence changée l'huile d'un moteur à combustion. Finalement, il y a les modèles de réparation, qui décrivent la situation où l'action de maintenance s'effectue seulement dans le cas où le système défaille (Leemis, 2009).

Afin de pouvoir répondre à ces situations, il faut utiliser des modèles de probabilités pour déterminer l'occurrence des défaillances. Différents modèles ont été développés, se distinguant selon plusieurs hypothèses. Si le temps de maintenance est négligeable par rapport au temps de fonctionnement par exemple, le temps de défaillance est modélisé par un processus ponctuel. Dans le cas contraire, certains utilisent notamment les modèles de Markov (Leemis, 2009). Pour cette étude, il est raisonnable de faire l'hypothèse que les temps de réparations sont négligeables par rapport au temps de fonctionnement. Ainsi, les probabilités peuvent être déterminées entre autres par les modèles HPP et NHPP, qui seront décrits plus bas.

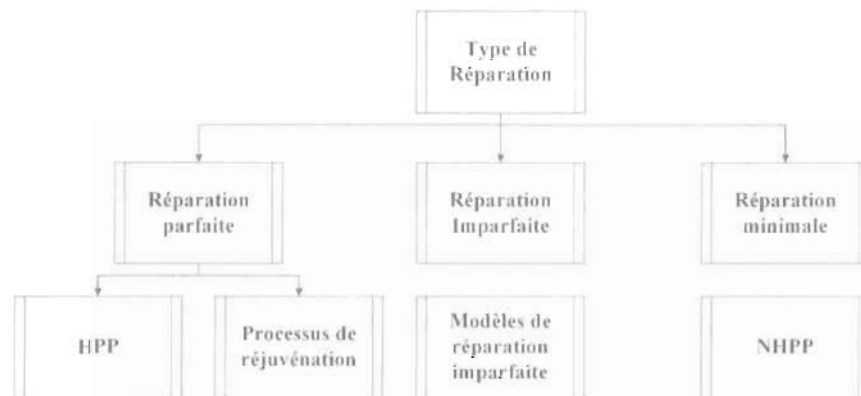


Figure 2.7: Processus ponctuel pour les systèmes réparables (Rausand & Hoyland, 2003)

Afin d'assurer une cohérence tout au long du texte, la notation suivante sera appliquée au modèle HPP et NHPP :

Tableau 2.1: Notation mathématique

| Variable | Type de variable | Signification |
|--------------|-------------------------------------|-------------------------------------|
| k | Paramètre de forme | $k^{\text{ième}}$ défaillance |
| λ | Paramètre d'échelle | Taux d'occurrence des défaillances |
| $\lambda(t)$ | Fonction d'intensité | Taux de défaillance à l'instant t |
| t | Paramètre de positionnement | Temps |
| e | Constante de Neper ($\ln(e) = 1$) | |

2.5 Modèle HPP

Le modèle HPP (« *Homogeneous Poisson Process* ») est une forme simple et populaire pour les calculs probabilistes des systèmes réparables. Par définition, ce modèle considère un temps entre les défaillances constant, c'est-à-dire qu'il suit une distribution exponentielle de λ et que ces événements sont indépendants et identiquement distribués (Leemis, 2009). On peut donc décrire le modèle HPP par les équations de la distribution exponentielle :

Tableau 2.2: Description des équations du modèle HPP

| # | Fonction | Équation | Nom | Description |
|--------------|----------|--------------------------|---|--|
| Équation 2.1 | $f(t)$ | $\lambda e^{-\lambda t}$ | Fonction de densité de probabilité | |
| Équation 2.2 | $S(t)$ | $e^{-\lambda t}$ | Fonction de survie | Probabilité que le système fonctionne au temps T |
| Équation 2.3 | $F(t)$ | $1 - e^{-\lambda t}$ | Fonction de densité de probabilité cumulative | Temps à la prochaine défaillance $S(t) = 1 - F(t)$ |
| Équation 2.4 | $h(t)$ | λ | Fonction d'intensité | Taux de défaillance |
| Équation 2.5 | $H(t)$ | λt | Fonction d'intensité cumulative | |

Étant donné que le taux de réparation est constant, ce modèle est le plus souvent utilisé lorsque le système est dans sa période de vie utile (voir Figure 2.5). De plus, cela implique qu'une réparation ramène le système dans son état original « *as good as new* », c'est pourquoi on parle de maintenance parfaite (Technology, 2003).

Modèle NHPP

Contrairement à HPP, le modèle NHPP « *Non-Homogeneous Poisson Process* » ne considère pas une maintenance parfaite. La fonction d'intensité ($h(t) = \lambda(t)$) n'est pas nécessairement constante, elle peut évoluer en fonction de l'âge des équipements et on y réfère comme étant un processus de loi de puissance. En d'autres termes, les réparations antérieures n'affecteront pas les performances futures du système (Technology, 2003). De plus, le modèle HPP est un cas particulier de NHPP. Effectivement, ce modèle offre beaucoup de flexibilité, en plus de permettre de modéliser des systèmes en détérioration ou en régénération. Aussi, la fonction d'intensité de la loi de puissance possède la même forme

que la fonction de risque d'une distribution Weibull, présenté au tableau suivant (Leemis, 2009; Technology, 2003).

Tableau 2.3: Équations du modèle NHPP

| # | Fonction | Équation | Nom | Description |
|----------|----------------------------------|---|---|--|
| Éq. 2.6 | $f(t)$ | $\left(\frac{k}{\lambda}\right) \left(\frac{t}{\lambda}\right)^{k-1}$ | Fonction de densité de probabilité | |
| Éq. 2.7 | $S(t)$ | $e^{-\left(\frac{t}{\lambda}\right)^k}$ | Fonction de survie | Probabilité que le système fonctionne au temps t |
| Éq. 2.8 | $F(t)$ | $1 - e^{-\left(\frac{t}{\lambda}\right)^k}$ | Fonction de densité de probabilité cumulative | Temps à la prochaine défaillance $S(t) = 1 - F(t)$ |
| Éq. 2.9 | $\lambda(t) = \frac{f(t)}{S(t)}$ | $\left(\frac{k}{\lambda}\right) \left(\frac{t}{\lambda}\right)^{k-1}$ | Fonction d'intensité | taux de défaillance |
| Éq. 2.10 | $H(t)$ | $\left(\frac{t}{\lambda}\right)^k$ | Fonction d'intensité cumulative | |

L'équation de la fonction d'intensité de défaillance peut aussi s'écrire sous la forme :

$$\text{Équation 2.11} \quad h(t) = k\lambda t^{k-1}$$

En observant l'Équation 2.11, lorsque $k = 1$:

$$h(t) = 1 * \lambda^1 t^{1-1} = \lambda$$

Comme démontré ci-haut, la fonction d'intensité HPP est la même que pour NHPP lorsque $k = 1$.

Dans le cas de NHPP, le système est considéré « *as bad as old* » après sa réparation. Cela signifie que le système, plutôt que de revenir aux conditions initiales, est remis dans son état précédant la défaillance (Gámiz, Kulasekera, Limnios, & Lindqvist, 2011).

2.6 Estimateur du maximum de vraisemblance :

En statistique, il existe deux principales méthodes afin d'estimer les paramètres d'une distribution; la méthode des moindres carrés et l'estimation du maximum de vraisemblance. L'estimation des paramètres des modèles utilisés en fiabilité se fait généralement par l'estimation du maximum de vraisemblance (Myung, 2003).

Le but cette méthode est de déterminer les paramètres de la fonction de densité de probabilité susceptible de produire les données observées, considérant le modèle d'intérêt (Myung, 2003). Pour ce faire, il faut maximiser la fonction de vraisemblance, qui correspond au produit des fonctions de densités de probabilités :

Équation 2.12

$$L(t, \theta) = \prod_{i=1}^n f(t_i, \theta)$$

Où θ représente le vecteur de paramètres inconnus. De façon pratique, maximiser la fonction de vraisemblance revient à trouver les valeurs estimées de θ , en résolvant les dérivées de la fonction, lorsque celle-ci est égale à zéro. En pratique, il est plus facile de résoudre ces équations en appliquant le logarithme à la fonction $L(t, \theta)$, c'est pourquoi il est souvent question de fonction de log-vraisemblance, ou « *log likelihood* » (Leemis, 2009; Myung, 2003).

2.7 Analyse de survie

Les modèles présentés plus haut permettent d'obtenir de l'information essentielle en GDA. Ces modèles permettent d'estimer le taux de défaillance observé des différents actifs analysés. En revanche, ceux-ci ne prennent en compte que les événements de défaillances, sans considérer l'influence d'autres facteurs sur le système. À cet égard, les modèles de survie permettent de déterminer une relation entre ces événements de défaillances et certaines variables (Gámiz et al., 2011; Gorjian, Ma, Mittinty, Yarlalagadda, & Sun, 2010). Les modèles les plus connus sont le modèle de risque proportionnel (modèle de Cox), le modèle additif d'Aalen et le modèle de vie accélérée. Encore une fois, ces approches ne permettent que de traiter une seule défaillance. Plusieurs chercheurs ont donc tenté de développer des extensions capables de traiter le cas d'événements récurrents. Les variations du modèle de Cox les plus communes sont le modèle Andersen-Gill, Wei-Lin & Wiessfield ainsi que Prentice-Williams-Peterson (Therneau & Grambsch, 2000).

Extensions du modèle de Cox

Cette section a été élaborée à partir de l'ouvrage de Therneau et Grambsch (2000), qui présente les extensions du modèle de Cox, ainsi que le tutoriel de Amorim et Cai (2014), qui présentent un cas d'application des mêmes modèles. Ceux-ci sont catégorisés comme des événements récurrents ordonnés, puisque la séquence des événements affecte le résultat du modèle. Ces extensions du modèle de Cox doivent leur popularité au fait que plusieurs outils permettent facilement leur implémentation. Effectivement, une fois que les données ont été formatées correctement, l'application du modèle est presque similaire au modèle de Cox. Il ne faut que remplacer l'estimation de la variance par une estimation corrigée, qui tient compte des possibles corrélations (Therneau & Grambsch, 2000).

Correction de la variance

Pour le modèle de Cox standard, la variance est estimée en considérant que chaque observation (événements de défaillance) est indépendante. Cependant, lorsqu'un équipement subit plusieurs défaillances, ces observations ne peuvent plus être considérées indépendantes. Ainsi, la variance est estimée par une méthode de rééchantillonnage (Jackknife), afin de prendre en compte qu'un équipement subit des défaillances de façon récurrente (Therneau & Grambsch, 2000). De façon pratique, il suffit de considérer le numéro d'identification de l'équipement comme l'identifiant d'un groupe d'observation.

Structure des données

Comme mentionnée plus haut, ces extensions du modèle de Cox ont une implémentation presque similaire au modèle de base, ce qui facilite grandement leur application. La première différence réside dans la façon de formater les données. Cette section permettra de démontrer deux façons, suivie d'un exemple tiré du tutoriel de (Amorim & Cai, 2014).
Méthode de type processus de comptage :

- Chaque équipement est représenté par une série de défaillances (observations)
 - Chaque événement de défaillance k d'un équipement i représente une observation

Tableau 2.4: Exemple de données (processus de comptage)

| id | tstart | tstop | status | event | trt | numtum | size |
|----|--------|-------|--------|-------|-----|--------|------|
| 1 | 0 | 1 | 0 | 1 | 1 | 1 | 3 |
| 2 | 0 | 4 | 0 | 1 | 1 | 2 | 1 |
| 3 | 0 | 7 | 0 | 1 | 1 | 1 | 1 |
| 4 | 0 | 10 | 0 | 1 | 1 | 5 | 1 |
| 5 | 0 | 6 | 1 | 1 | 1 | 4 | 1 |
| 5 | 6 | 10 | 0 | 2 | 1 | 4 | 1 |
| 6 | 0 | 14 | 0 | 1 | 1 | 1 | 1 |
| 7 | 0 | 18 | 0 | 1 | 1 | 1 | 1 |
| 8 | 0 | 5 | 1 | 1 | 1 | 1 | 3 |
| 8 | 5 | 18 | 0 | 2 | 1 | 1 | 3 |
| 9 | 0 | 12 | 1 | 1 | 1 | 1 | 1 |
| 9 | 12 | 16 | 1 | 2 | 1 | 1 | 1 |
| 9 | 16 | 18 | 0 | 3 | 1 | 1 | 1 |

Tableau 2.5: Description des variables de l'exemple

| | |
|--------|---|
| id | Identifiant du patient |
| tstart | Temps du début de l'intervalle d'observation |
| tstop | Temps de fin de l'intervalle d'observation |
| status | status = 1 si l'événement est récurrent et 0 s'il est censuré |
| event | le $k^{\text{ième}}$ événement |
| trt | Covariables |
| numtum | |
| size | |

Cet exemple illustre un essai clinique sur différents patients, ce jeu de données est largement utilisé dans la littérature et est connu sous le nom de « *bladder dataset* ». Les observations du patient 5, 8 et 9 mettent en lumière que les observations sont soit l'entrée dans l'étude ou bien un événement d'intérêt. Dans ce cas-ci, seulement trois patients ont observé l'événement d'intérêt. Les autres données ont donc une censure à droite (status = 0), de même que pour la dernière observation de ces individus (5,8,9) (Amorim & Cai, 2014; Therneau & Grambsch, 2000). Dans ce jeu de données, la colonne « *event* », présente une incrémentation à chaque événement, soit une nouvelle strate. Il peut donc être utilisé pour le modèle AG, ainsi que le modèle PWP.

Méthode par strate :

- Le nombre maximum d'événements de défaillances vécus par les équipements représentent le nombre de strates.
 - Pour chaque équipement i , il y a une observation par strates k .

Tableau 2.6: Exemple de données (format strates)

| id | time | status | event | trt | numtum | size |
|----|------|--------|-------|-----|--------|------|
| 1 | 1 | 0 | 1 | 1 | 1 | 3 |
| 1 | 1 | 0 | 2 | 1 | 1 | 3 |
| 1 | 1 | 0 | 3 | 1 | 1 | 3 |
| 1 | 1 | 0 | 4 | 1 | 1 | 3 |
| | | | | | | |
| 5 | 6 | 1 | 1 | 1 | 4 | 1 |
| 5 | 10 | 0 | 2 | 1 | 4 | 1 |
| 5 | 10 | 0 | 3 | 1 | 4 | 1 |
| 5 | 10 | 0 | 4 | 1 | 4 | 1 |
| | | | | | | |
| 9 | 12 | 1 | 1 | 1 | 1 | 1 |
| 9 | 16 | 1 | 2 | 1 | 1 | 1 |
| 9 | 18 | 0 | 3 | 1 | 1 | 1 |
| 9 | 18 | 0 | 4 | 1 | 1 | 1 |

Cet exemple de données stratifiées ne représente que les données des patients 1, 5 et 9. Comme le nombre d'événements maximum est de quatre (incluant l'entrée dans l'étude), chaque patient a quatre observations. Les strates sont identifiées par la variable « *event* ». À titre comparatif, le jeu de donnée complet compte 190 observations dans le premier format, tandis qu'il en contient 340 dans l'autre.

Un autre facteur important de la structure des données est l'échelle de temps utilisé. Le premier cas illustre un intervalle de temps entre deux événements. Dans le second cas, le temps est calculé depuis le début de l'étude pour chaque observation. Dans la littérature, on désigne souvent ce modèle comme « *gap time model* ». De plus, les modèles peuvent aussi contenir des cofacteurs variant dans le temps, c'est-à-dire que leurs valeurs peuvent changer tout au long de l'étude. Finalement, le jeu de donnée peut être stratifié, donc subdivisé en sous-population homogène. De plus, chaque observation est assignée à une

seule strate. À partir de ces concepts, les extensions du modèle de Cox se décrivent comme suit :

➤ Modèle Andersen-Gill, ou modèle d'incrément indépendant :

Hypothèses :

- Les événements de défaillance d'un équipement sont mutuellement indépendants
- Les événements de défaillances ne changent pas la fonction de référence (strates).

Particularités :

- Données formatées selon un processus de comptage
- Pas de stratification des données
- Échelles de temps possible :
 - Temps d'intervalle entre les événements
- Permet d'estimer l'effet global des variables d'intérêt.

➤ Modèle Wei-Lin & Weissfeld, ou modèle marginal :

Hypothèses :

- Les événements de défaillances changent la fonction de référence
- Traite le problème comme étant un problème de risque concurrent non ordonné.

Particularités :

- Le nombre maximum de défaillances vécues représente un nombre de strates
 - Chaque observation d'un équipement est placée dans la strate correspondant à sa k ème défaillance
- Stratification des données
 - Chaque strate k a sa propre fonction de référence $\lambda_{0k}(t)$
 - Données formatées par stratification
- Échelle de temps possible :
 - Temps depuis l'entrée dans l'étude

➤ Modèle Prentice-Williams-Peterson (PWP), ou modèle conditionnel :

Hypothèse :

- Un équipement ne peut être à risque de subir sa $k^{ième}$ défaillance s'il n'a pas subi sa $k - 1^{ième}$.

Particularités :

- Données formatées selon un processus de comptage
- Chaque événement est assigné à une différente couche
- Échelles de temps possible :
 - Temps depuis l'entrée dans l'étude
 - Temps d'intervalle entre les événements
- La fonction de référence varie selon la strate

PWP-TT :

- Modèle PWP avec une structure de donnée de type « *total time* »
 - Même échelle de temps que le modèle AG
 - Évalue l'effet de variables d'intérêt pour un événement k depuis le début de la période d'étude.

PWP-GT :

- Modèle PWP avec une structure de donnée de type « *gap time* »
 - Assume un processus de renouvellement « *renewal process* »
 - Utile lorsque :
 - Les défaillances ne sont pas fréquentes
 - L'hypothèse du renouvellement est adéquate
 - Évalue l'effet de variables d'intérêt pour un événement k depuis l'événement précédent.

Tableau 2.7: Fonction d'intensité des modèles de survie

| Modèles de survie | | |
|---------------------------------------|-----------------------|---|
| Nom du modèle | Particularité | Fonction d'intensité/ de risque |
| Modèle à risque proportionnel | - | $\lambda_i(t) = \lambda_0(t)e^{\beta X_i(t)}$ |
| Modèle d'Andersen-Gill | Événements récurrents | $\lambda_i(t) = \lambda_0(t)e^{\beta X_i(t)}$ |
| Prentice-Williams-Peterson total time | Événements récurrents | $\lambda_{ik}(t) = \lambda_{0k}(t)e^{\beta_k X_i(t)}$ |
| Prentice-Williams-Peterson gap time | Événements récurrents | $\lambda_{ik}(t) = \lambda_{0k}(t - t_{k-1})e^{\beta_k X_i(t)}$ |
| Méthode de Wei-Lin-Weissfeld | Événements récurrents | $\lambda_{ik}(t) = \lambda_{0k}(t)e^{\beta_k X_i(t)}$ |
| Modèles à fragilité partagés | Événements récurrents | $\lambda_{ik}(t) = \lambda_{0k}(t) Z_i e^{\beta_k X_i(t)}$ |

Tableau 2.8: Description des variables des modèles de survie

| Variable | Signification |
|----------------|--|
| i | Équipement/ individus i de la population à l'étude |
| t | Temps |
| $\lambda_i(t)$ | Risque/intensité de défaillance de l'équipement i au temps t |
| $\lambda_0(t)$ | Fonction de risque/intensité de référence |
| β | Vecteur des coefficients (effet de la variable d'intérêt) |
| X_i | Matrice des covariables de l'équipement i |
| $L(\beta)$ | Fonction de vraisemblance |

Applications et avancées

Gorjian *et al.* (2010) proposent une synthèse des différents modèles de survie, ainsi que les forces et faiblesses de chacun. Selon lui, la force du modèle de Cox est qu'il n'est pas nécessaire de spécifier une distribution de référence. De plus, les coefficients peuvent être estimés en utilisant une fonction de vraisemblance partielle, ce qui permet une estimation efficace de l'effet des covariables. En revanche, le modèle de base permet seulement de traiter les systèmes non réparables, puisqu'il considère les défaillances létales (Gorjian *et al.*, 2010). De son côté, le modèle de risque additif permet de déterminer les variables d'intérêts d'un système réparable, lorsque celui-ci subit une maintenance imparfaite. Cependant, l'auteur souligne que l'hypothèse d'additivité du risque peut mener à une estimation du risque négatif, et mène donc à des résultats peu réalistes (Gorjian *et al.*, 2010). Les modèles d'intensité proportionnelle, aussi produite par Cox, sont une famille d'extensions permettant de traiter les systèmes réparables. Comme le modèle de base, il s'agit de modèles simples d'utilisation et très répandus. Les coefficients sont calculés à partir de l'estimation du maximum de vraisemblance (Gorjian *et al.*, 2010). D'autre part, Shwu-Tzy, Landers et Rhoads (2006) comparent différentes variantes des modèles d'intensité proportionnelle. Ils s'intéressent à leur application à l'ingénierie. Toutefois, puisque ces méthodes sont peu utilisées en ingénierie, ils considèrent aussi les approches du domaine médical. En outre, ils testent des variantes semi-paramétriques, et des variantes paramétriques (suivant un Processus de Poisson Non-Homogène). Les modèles à l'étude sont le modèle AG, PWP et WLW. Le tableau suivant présente un réseau des contributions évaluées par Shwu-Tzy, Landers et Rhoads (2006) :

Tableau 2.9: Résumé des articles

| Auteurs | Articles | But de l'étude | Systèmes à l'étude | Modèles | Variables | Conclusion |
|--|--|--|--|---------|---|---|
| T. L. Landers & H. E. Soroudi | Robustness of a semi-parametric proportional intensity model | Déterminer les conditions d'utilisation du modèle PWP, pour ensuite utiliser les résultats pour prédire le temps entre les événements. | Données simulées à partir d'un modèle NHPP suivant une loi de puissance | PWP | Taux d'occurrence des défaillances | L'application du modèle est appropriée lorsque le taux d'occurrence des défaillances se situe entre 1 et 3 et que la taille de l'échantillon est supérieure à 60 observations |
| W. M. Qureshi, T. L. Landers, and E. E. Gbur | Robustness evaluation of a semi-parametric proportional intensity model | | | | Grosseur de l'échantillon | |
| T. L. Landers, S. T. Jiang, and J. R. Peek | Semi-parametric PWP model robustness for log-linear increasing rates of occurrence of failures | Étudier la robustesse du modèle, dans le cas d'un processus NHPP | Données simulées à partir d'un modèle NHPP suivant une fonction d'intensité log linéaire | PWP | Taux d'occurrence des défaillances Grosseur de l'échantillon | L'application du modèle est appropriée lorsque le taux d'occurrence des défaillances se situe entre 0 et 3 et que la taille de l'échantillon est supérieure à 60 observations |
| S. Jiang, T. L. Landers, and T. R. Rhoads | Semi-parametric proportional intensity models robustness for right-censored recurrent failure data | Étudier la robustesse des modèles, dans le cas d'un processus NHPP avec des données censurées à droite | Données simulées à partir d'un modèle NHPP suivant une loi de puissance | PWP | Taux d'occurrence des défaillances Grosseur de l'échantillon | Pour les petits échantillons, PWP performe bien lorsque la majorité (80%) des équipements présentent des censures, ainsi que pour de grands échantillons hautement censurés. |
| | | | | AG | | |
| S. Jiang, T. L. Landers, and T. R. Rhoads | Proportional intensity models robustness with overhaul intervals | Évaluer la perturbation des modèles lorsque le temps d'indisponibilité du système est grand. | | PWP | Temps de réparation | PWP réussit à bien performer lorsque la durée d'interruption est inférieure à la moitié du temps de fonctionnement précédent. |
| | | | | AG | Grosseur de l'échantillon | AG performe assez bien, peu importe le temps de réparation, si l'échantillon est petit |

Tableau 2.10: Résumé des articles (suite)

| Auteurs | Articles | But de l'étude | Systèmes à l'étude | Modèles | Variables | Conclusion |
|---|--|--|---|-----------------------|---|--|
| J. I. Ansell and M. J. Phillips | Practical aspects of modeling of repairable systems data using proportional hazards models | Déterminer l'impact de covariable sur les défaillances d'un système de photocopieur | Données de défaillance d'un système de photocopieur | PHM | Taux de défaillance | L'impact des variables de stress et de températures s'est avéré significatif. |
| | | | | | Covariables | Aussi, l'estimation du taux de défaillance par une méthode paramétrique permet d'avoir une estimation plus précise. |
| J. Ansell, T. Archibald, J. Dagpunar, L. Thomas, P. Abell, and D. Duncalf | Assessing the maintenance in a process using a semi-parametric approach | Analyse des actions non planifiées de maintenance et de réparation sur le processus de filtration | Donnée d'un système de traitement d'eau municipal : Filtre à graviter | PWP modifié | Débit maximum de l'eau | Les covariables ont un impact significatif sur la fiabilité du système |
| | | | | | Débit moyen de l'eau | |
| | | | | | Processus (performance, adéquation, etc.) | |
| J. Ansell, T. Archibald, J. Dagpunar, L. Thomas, P. Abell, and D. Duncalf | Analysing maintenance data to gain insight into systems performance | Identifier les covariables de maintenance qui ont un effet sur la fiabilité du système de filtration d'eau | Donnée d'un système de traitement d'eau municipal : Filtre à graviter | Modèle de Crow | Taux d'occurrences des défaillances | La remise à neuf des équipements a un impact significatif sur le taux de défaillance |
| | | | | Modèle de Cox & Lewis | Effet des remises à neuf | |
| D. Kumar and U. Westberg | Proportional hazards modeling of time-dependent covariates using linear regression: A case study | Étudier les facteurs qui causent la défaillance des câbles d'alimentation des équipements | Système de chargeur "LHD" | PHM | Causes mécaniques | Le type de câble et la présence de nouveau joint de soudure a un impact significatif sur le taux de défaillance. De plus, les données démontrent que la variable du type de câble est dépendante du temps. |
| | | | | | Causes thermiques | |

Finalement, Percy et Alkali (2006) proposent une variante du modèle d'intensité proportionnelle, qui prend en compte l'effet de la maintenance sur le taux de réparation des équipements. Le modèle est ensuite appliqué à un jeu de données publié par l'industrie pétrolière. En outre, leur modèle démontre une bonne performance dans la prise en compte des variables d'intérêts (Percy & Alkali, 2006).

2.8 Qualité et nettoyage des données

Comme mentionné plus haut, la qualité des données est un enjeu important du projet. En ce sens, la plupart des projets qui nécessitent l'utilisation de données d'opération saisies manuellement font face au même problème. Bien que ce mémoire ne traite pas directement cette problématique, le contexte appliqué du projet force à adopter de bonnes pratiques dans le nettoyage des données. Cette section décrit différentes techniques qui ont été développées.

Dans un premier temps, Ho et ses collaborateurs (2015) proposent une étude de cas, appliquer dans la GDA des équipements du domaine minier, afin de recenser différents problèmes issus des données. Ils soulignent que les données enregistrées sont souvent erronées, manquant de détails et de précision, et par conséquent peu propice à aider la prise de décision (Ho, Hodkiewicz, Pun, Petchey, & Li, 2015). L'attention se porte sur différents champs des bases de données relationnelles (BD) et l'étude révèle des faiblesses au niveau de :

- L'assignation des tâches de maintenance au bon sous-système
- La cohérence dans l'assignation des codes pour décrire le type de travail
 - Type de maintenance (conditionnelle, préventif, correctif, *etc.*)
- La cohérence dans l'assignation des codes pour décrire les composants à maintenir
- L'utilisation des champs texte pour la description des tâches
 - Mauvaise utilisation
 - Champs non standards
- L'utilisation réelle des équipements
 - Incapacité de déduire l'usage réel à partir des données

Ensuite, Hodkiewicz et Ho (2016) proposent de développer une méthode efficace pour corriger les données, en identifiant les problèmes de qualité à partir des bons de travail de maintenance pour pouvoir faire des analyses de fiabilité. En effet, une méthode de nettoyage de donnée basée sur des règles est développée et appliquée au cas des

équipements miniers, sur un jeu de données réelles. Celle-ci permet de démontrer qu'il est possible de corriger les erreurs d'une base de données de maintenance et de les rendre aptes au traitement statistique. Le nettoyage se fait en partie par la récupération d'informations essentielles qui sont consignées dans les champs de texte libre.

Prasad *et al.* (2011) développent un ensemble d'outils pour personnaliser les règles de nettoyages de données, ainsi que les dictionnaires. La principale contribution est de développer un outil d'organisation des règles de nettoyage en élaborant une structure arborescente. La méthodologie débute par l'examen des données, où sont recensées les différentes erreurs. Ensuite, à partir de cette analyse, des règles et des dictionnaires de synonymes et de variantes sont élaborés. Le client peut alors décider d'appliquer les règles sur ces données, puis faire des analyses (Prasad *et al.*, 2011).

Quant à Murphy (2019), il étudie la qualité des données, du point de vue de l'utilisateur. Plutôt que de corriger les données, il analyse l'importance des facteurs humains relatifs à la qualité des données saisies dans un système informatique. La variance dans la qualité des données est influencée grandement par l'utilisateur qui la saisit. Cela démontre la nécessité de développer une meilleure gestion des groupes de travailleurs qui les produisent. En se basant sur la théorie des comportements planifiés, il évalue l'impact de l'attitude, de la perception, et des intentions envers le processus de collecte de données. Il prend aussi en considération l'effet de la pression du temps et de la rétroaction du personnel-cadre auprès des opérateurs. En outre, il démontre l'importance de l'implication et de la valorisation du travail des opérateurs envers la saisie des données, dans le but d'en améliorer la qualité (Murphy, 2009).

Berti-Équille (2018) présente les différentes approches, ainsi qu'une méthodologie afin de quantifier la qualité des données (diagnostique) et de gérer cette qualité (correction). L'article se divise en trois parties : la gestion et la mesure de la qualité des données, et la correction de ces données (Berti-Équille, 2018). La Figure 2.8 résume les approches de gestion de la qualité des données, tandis que la Figure 2.9 résume une démarche de gestion de la qualité et de la correction des données. De plus, l'auteure détermine plusieurs métriques afin de mesurer la qualité, selon les différents niveaux d'analyse des données.

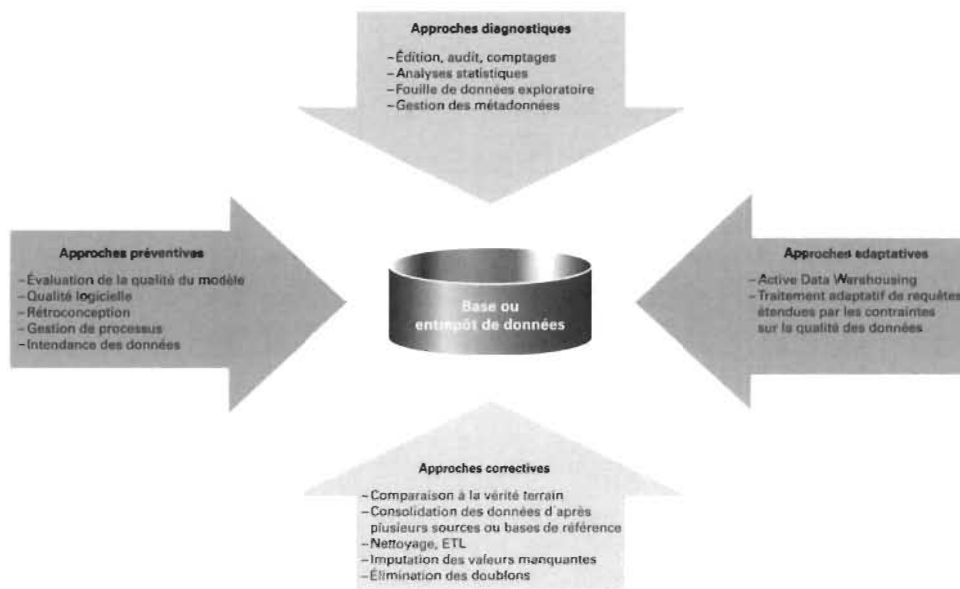


Figure 2.8: Approches pour l'évaluation et le contrôle de la qualité des données (Berti-Équille, 2018)

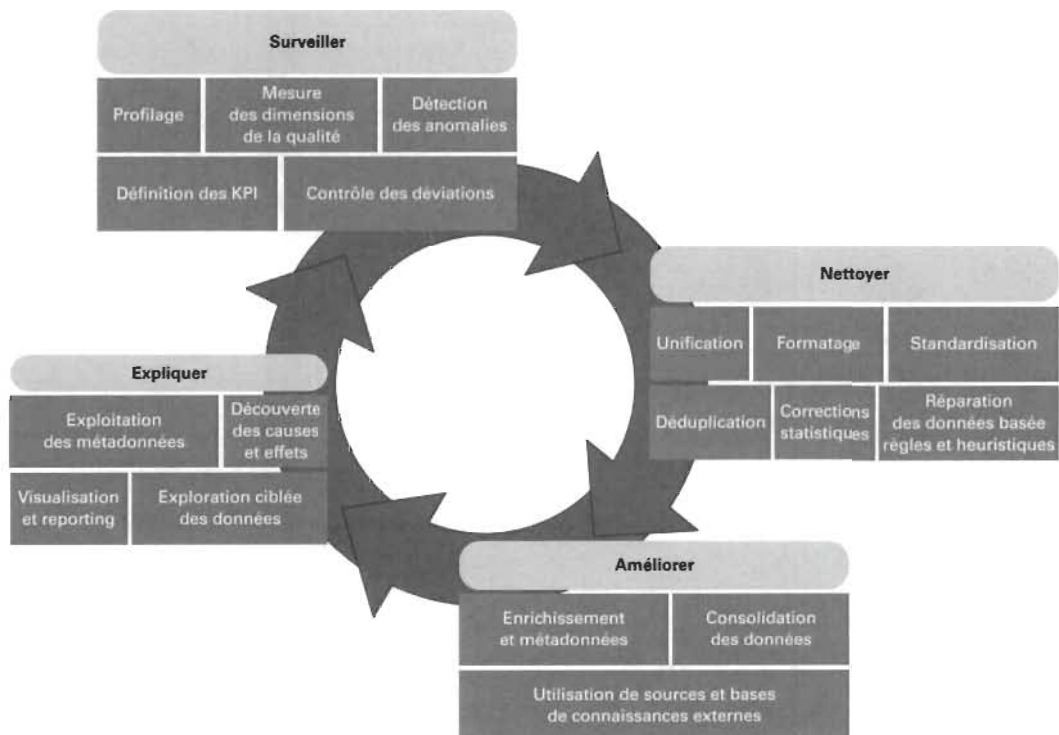


Figure 2.9: Cercle vertueux pour la gestion de la qualité des données (Berti-Équille, 2018)

Finalement Hodkiewicz et Montgomery (2014) élaborent une approche afin de déterminer les données nécessaires à chaque type de décisions, dans un contexte de GDA. À cet effet, ils ont listé différents champs d'une BD qui correspondrait à des types de décisions à prendre pour la maintenance. De cette façon, il est possible d'évaluer la capacité d'une entreprise à effectuer la GDA, dans diverses sphères décisionnelles. Cela permet aussi de dresser le portrait d'une entreprise et de donner des indications sur les informations manquantes à ajouter dans leur système, s'ils veulent prendre certaines décisions stratégiques.

En terminant, la revue de littérature a permis d'établir les connaissances scientifiques en lien avec les sujets abordés. Les concepts de la gestion des actifs et des modélisations statistiques en fiabilité ont été introduits, de même que les avancées scientifiques du domaine. Le chapitre suivant présente la méthodologie, qui permettra de répondre aux questions et aux objectifs de recherche.

CHAPITRE 3

MÉTHODOLOGIE

Ce chapitre présente la méthodologie mise en application à chaque étape du projet. Le développement de celle-ci est basé sur les questions et les objectifs de recherche, de même que sur la revue de littérature. L'objectif principal de cette recherche est d'améliorer la compréhension des facteurs affectant la fiabilité des actifs d'un réseau de transport d'électricité. Les moyens employés à cette fin sont l'application de modèles fiabilistes à une étude de cas portant sur les transformateurs de puissances. Ces moyens sont détaillés dans ce chapitre.

Dans un premier temps, les méthodes d'acquisitions et d'analyses préliminaires sont décrites au point 3.0. En second, les méthodes de modélisations du taux de défaillance sont expliquées au point 3.1. Finalement, les modèles de régressions utilisés sont exposés au point 3.2. L'application de ces méthodes, ainsi que les résultats et discussions sont présentés respectivement au chapitres 4, 5 et 6.

3.0 Analyse préliminaire et préparation des données

Analyse exploratoire des données

Pour commencer, il importe de reconnaître l'une des contraintes majeures d'un tel projet; la qualité des données. L'utilisation de données réelles, souvent saisies de façon manuelle, nécessite un traitement préliminaire rigoureux (filtre, nettoyage, etc.). L'analyse exploratoire permettra d'extraire un maximum de connaissances des données de base. D'autre part, cela contribuera à augmenter la qualité du jeu de données, en performant les traitements adéquats à la préparation de ces données.

Des méthodes de visualisation (histogrammes, nuages de points, diagrammes de quartiles) de même que des statistiques descriptives des données (moyenne, médiane, variance, *etc.*) permettront d'enrichir les analyses préliminaires. Ces méthodes faciliteront la détection de données aberrantes, d'erreurs de saisies, de données manquantes et parfois même de déceler certaines tendances.

Nettoyage des données

Une fois que des anomalies ont été détectées dans les données, il est impératif de mettre en place des stratégies afin de pallier celles-ci. L'objectif vise à augmenter la qualité des données afin de diminuer les biais et les incertitudes inhérentes à chaque type d'erreurs. Ainsi, la création de règles d'exclusion de données favorisera l'élimination des données aberrantes, des doublons, des entrées vides, *etc.* Ensuite, un audit de la qualité des données permettra d'en apprécier la complétude, la précision (moyenne, écart type, mode, *etc.*) ainsi que la compréhension.

Préparation des données

Pour poursuivre, les données doivent être préparées afin qu'elles puissent être traitées, selon les méthodes de modélisation choisies. Une fois nettoyées, ces données doivent être analysées pour en connaître l'utilité. Il faut sélectionner les variables qui enrichiront les modèles, en tenant compte de la complétude et des exigences de base des modèles. Certaines variables devront être écartées, puisqu'elles alourdissent le modèle sans apporter d'information pertinente. D'autres seront exclues en raison de la qualité des données. Aussi, certaines variables seront combinées afin d'augmenter la performance de traitement sans diminuer la qualité du modèle. Certaines de ces variables devront être modifiées pour que l'algorithme puisse les traiter, notamment les variables qualitatives. En outre, il est important que les variables nominales et ordinales soient traitées de façon distincte, afin d'obtenir des résultats cohérents.

Ensuite, les données seront filtrées, afin de cibler une population précise et de réduire la quantité de données à traiter. Dans ce cas-ci, les filtres serviront à cibler la famille d'équipement des transformateurs et à observer une fenêtre d'étude sur quelques années seulement.

3.1 Modélisation du taux de défaillance

Une fois que les analyses préliminaires ont été faites et que les données ont été nettoyées et filtrées, il est possible de calculer un taux de défaillance de façon plus rigoureuse. Comme présentée dans les chapitres précédant, selon le cas à l'étude, le calcul du taux de défaillance peut être fait selon plusieurs méthodes statistiques. Pour ce projet, la fiabilité des transformateurs de puissances sera étudiée. En cas de défaillance, ce type d'équipement est soit remplacé ou réparé. De ce fait, il est impératif de ne considérer que

les modèles adaptés aux systèmes réparables. À partir de la littérature, deux méthodes de calcul ont été relevées, soit le modèle HPP et le modèle NHPP qui sera appliqué pour le projet. L'une de ces méthodes considère un taux de réparation optimiste comme constant à travers la vie de l'équipement et l'autre pessimiste avec un taux de défaillance pouvant évoluer au cours de la vie de l'équipement; le calcul du taux de défaillance permettra de supporter le processus décisionnel en gestion des actifs. Pour déterminer la fonction d'intensité du modèle NHPP, la méthode développée par le Dr Wayne B. Nelson sera appliquée (Nelson, 2003). Cette méthode non paramétrique est basée sur la « *mean cumulative function* » (MCF).

Tout d'abord, le type de donnée traité comporte certaines spécificités; bien que certains équipements soient en service depuis plusieurs années, les données de maintenance sont récentes. L'analyse est réalisée sur un échantillon dans un intervalle de temps récent, avec un historique de donnée tronqué. Certains équipements ont donc des événements censurés, c'est-à-dire qu'un équipement peut avoir subi une défaillance avant ou après la période d'observation, sans en avoir connu durant la période étudiée. La méthode est décrite selon les instructions tirées de (Nelson, 2003) :

1. Ordonner chaque observation en ordre croissant selon l'âge lors de l'événement
 - a. Âge lors de la défaillance
 - b. Âge de censure : lors de la mise en service ou du retrait durant la période échantillon
2. Déterminer le nombre d'unités qui entrent l'échantillon à l'âge t_i
 - a. $+N_i$: Nombre d'équipements qui entrent l'échantillon à l'âge t_i (censure à gauche)
 - b. $-N_i$: Nombre d'équipements qui sortent l'échantillon à l'âge t_i (censure à droite)
3. Calculer le nombre d'équipements qui sont à risque à l'âge t_i
 - a. Pour chaque observation, le nombre d'équipements à risque de subir une défaillance
 - i. $r_i = r_{i-1} + N_i$
4. Calculer l'incrément du nombre moyen de récurrences par unité à risque
 - a. Correspond à $m_i = \frac{1}{r_i}$
 - b. Pour un âge censuré, $m_i = 0$
5. Calculer la somme cumulative des incréments (MCF)
 - a. $M_i^* = m_i + M_{i-1}^*$
 - b. Pour un âge censuré, $M_i^* = 0$
6. Tracer le graphique des M_i^* en fonction de l'âge t_i

7. Tracer la courbe d'ajustement, afin de déterminer les paramètres d'une loi de puissance

La Figure 3.1 est un exemple de données structuré selon cette méthode :

| 1. Âge t_i | 2. Entrée N_i | 3. Nombre à risque r_i | 4. Incrémental % $m_i = 100/r_i$ | 5. Cumulative % $M_i^* = M_{i-1}^* + m_i$ |
|-----------------|--------------------|-----------------------------|-------------------------------------|--|
| 0 | +149 | 149 | | |
| 0 | +195 | 344 | | |
| 0.17 | | 344 | 0.29 | 0.29 |
| 0.17 | | 344 | 0.29 | 0.58 |
| 1 | +458 | 802 | | |
| 1.34 | | 802 | 0.12 | 0.71 |
| 2.17 | | 802 | 0.12 | 0.83 |
| 2.58 | | 802 | 0.12 | 0.96 |
| 2.59 | +164 | 966 | | |
| 3.3 | | 966 | 0.10 | 1.06 |
| 3.65 | | 966 | 0.10 | 1.16 |
| 4.14 | | 966 | 0.10 | 1.27 |
| 4.14 | -195 | 771 | | |
| 4.45 | +356 | 1127 | | |
| 4.47 | | 1127 | 0.09 | 1.35 |
| 4.47 | | 1127 | 0.09 | 1.44 |
| 4.62 | | 1127 | 0.09 | 1.53 |
| 4.62 | | 1127 | 0.09 | 1.62 |
| 4.65 | | 1127 | 0.09 | 1.71 |
| 4.79 | | 1127 | 0.09 | 1.80 |
| 5.09 | -149 | 978 | | |

Figure 3.1: Exemple de données

1. « Âge t_i »: Âge à l'événement, censuré ou non
2. « Entrée N_i » :
 - a. + Lorsque des équipements entrent dans l'étude durant la période observée (ex. : mise en service)
 - b. - Lorsque des équipements sortent de l'étude durant la période observée (ex. : retrait d'équipement, défaillance létale, etc.)
3. « Nombre à risque » r_i : Nombre d'équipement à risque d'avoir une défaillance, à l'âge t_i
4. « Incrémental % m_i »: Correspond à l'incrément du nombre moyen de récurrences par unité à risque. Dans ce cas-ci, l'auteur a préféré utiliser $100/r_i$ afin de l'exprimer en pourcentage

5. « *Cumulative % M_i^** » : Correspond au cumulatif du nombre moyen de récurrences par unité à risque

La courbe d'ajustement sur les données permet donc de déterminer les paramètres de la loi de puissance, donc du taux de défaillance. La méthode des moindres carrés permet de produire une courbe en utilisant les paramètres de la loi minimisant la somme des écarts au carré entre les points et cette courbe.

3.2 Modélisation et régression avancées

Pour terminer, des modèles de régression seront appliqués. La revue de littérature a permis d'établir les méthodes existantes pour modéliser les covariables. Ces méthodes seront donc appliquées aux cas des transformateurs, afin de déterminer la ou les méthodes les plus appropriées pour décrire les variables affectant le taux de défaillance. Du même coup, les méthodes qui seront retenues serviront aussi à améliorer la connaissance des équipements et des facteurs qui en affecte la fiabilité. Tout comme les modèles de calcul du taux de défaillance, les modèles de régressions qui seront testés sont adaptés aux systèmes réparables.

Les modèles retenus sont des extensions du modèle de Cox, soit les modèles Prentice-Williams-Peterson (PWP) et le modèle Anderson-Gill. Selon l'étude menée par Shwu-Tzy, Landers et Rhoads (2006), les modèles PWP offrent une bonne performance et sont généralement ceux qui sont le plus utilisés en fiabilité. De plus, l'hypothèse du modèle PWP est que le risque d'occurrence d'un événement est influencé par les événements qui l'ont précédé (Amorim & Cai, 2014). Cette hypothèse semble appropriée au contexte de la fiabilité des transformateurs. Comme le modèle AG et le modèle PWP-TT utilisent la même structure de données, il sera facile d'implémenter les deux modèles et de les comparer.

Une fois que ceux-ci auront été appliqués, des tests statistiques permettront de déterminer si les modèles développés sont statistiquement valides et ainsi discriminer certaines méthodes. Les modèles valides seront alors analysés, afin de déterminer si certaines variables ont un effet significatif sur le taux de défaillance observé.

Les méthodes d'application ont été développées à partir de l'ouvrage de Therneau et Grambsch (2000), ainsi que du tutoriel d'Amorim et Cai (2014). Le code des modèles de survie a été développé sur la plateforme Databricks, dans le langage de programmation R. Pour ce faire, la librairie « *survival* » a été utilisée. Cette librairie est la référence dans l'application des modèles de survie sur R et a été développée par Therneau et Lumley (2014). Voici un exemple de la formulation qui sera utilisée :

```
library (survival)|
modeAG = coxph(Surv(tstart,tstop,status) ~ var1 + var2+ ... + vark,
method="breslow", robust=TRUE, data = example1)
summary(model.1)
```

Figure 3.2: Exemple du modèle Andersen-Gill

- library (survival)
 - Importation de la librairie
- coxph(Surv(tstart,tstop,status)
 - Déclaration du modèle de survie Cox
 - Les variables sont déclarées comme un processus de comptage
- ~ var1 + var2 + ... + vark
 - Déclaration des champs de covariables
- method="breslow"
 - Cette méthode permet une estimation précise de la fonction de survie
- robust=TRUE
 - Estimation de la variance par la méthode « *Jackknife* »
- data = example1
 - Déclaration du jeu de données

```
modelPWTT=coxph(Surv(tstart, tstop, status)~var1+var2+...+vark+cluster(id)+s
trata(event), method="breslow", data=example1)
summary(modelPWTT)
```

Figure 3.3: Exemple du modèle conditionnel

- cluster(id)
 - Estimation de la variance par la méthode « *grouped Jackknife* »
- strata(event)
 - Stratification sur la variable du nombre d'événements

```
modelWLW = coxph(Surv(time,status) ~ var1 + var2+ ... + vark + cluster(id),
method="breslow", data = example2)
summary(modelWLW)|
```

Figure 3.4 : Exemple du modèle marginal

- coxph(Surv(time,status)
 - Déclaration des données par strate
- data = example2

- Jeu de donnée en format par strate

Ces trois exemples permettent de comprendre comment déclarer les deux types de jeu de données. De plus, les méthodes d'approximation de la variance sont démontrées, de même que la stratification des variables. L'un des cas de figure possibles est que le jeu de données présente des variables qualitatives. Dans ce cas, la variable *var1* est déclarée avec le terme *factor(var1)*.

Finalement, la Figure 3.5 résume les étapes qui seront appliquées au projet, et met en relief que ce projet ne se veut pas une réponse absolue aux facteurs qui affectent la fiabilité. Le projet permet plutôt de mettre en place une boucle d'amélioration des modèles statistiques en application chez HQ et ouvre la discussion sur les variables d'influences. En outre, ces modèles sont voués à évoluer; ils permettent de mettre à jour les connaissances, mais surtout de soulever les informations à recueillir dans le futur.

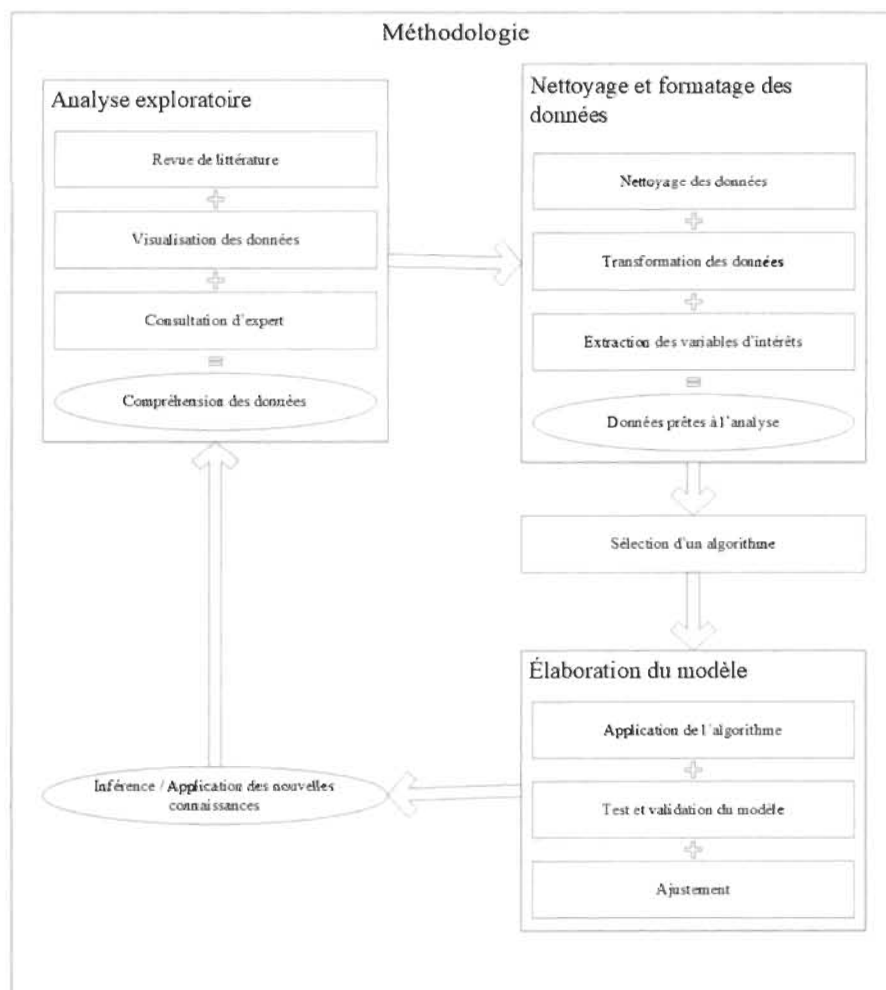


Figure 3.5: Méthode d'application des modèles de fiabilité

CHAPITRE 4

ANALYSE PRÉLIMINAIRE

Pour commencer, il est primordial d'analyser les données brutes, afin d'assurer la qualité, la faisabilité et la validité des modèles qui seront appliqués. Pour ce projet, l'approche de diagnostic est employée, plus précisément l'audit, le comptage, la fouille exploratoire et l'analyse statistique des données. Par la suite, une approche corrective est appliquée pour diminuer l'incertitude sur les données. L'élimination des doublons et un nettoyage sont ensuite mis en œuvre. Étant donné le fonctionnement de l'IREQ par rapport aux autres divisions d'Hydro-Québec, les champs d'action pour assurer la qualité des données sont limités à ces deux approches. Toutefois, les analyses permettront de faire des recommandations aux donneurs d'ordres, qui pourront initier la mise en place de mesures préventives et/ou adaptatives.

4.0 Processus de saisie de données

Il est impératif de comprendre le processus de saisie de donnée de maintenance chez Hydro-Québec. La Figure 4.1 illustre ce processus, ainsi que chaque étape menant à la saisie de données. Celle-ci permet d'illustrer la fonction de chaque type de documents, ainsi que la codification des types d'intervention. Dans un premier temps, il existe deux types de documents utilisés pour la maintenance; les avis et les ordres de maintenance.

L'avis est émis par un opérateur/inspecteur et est transmis au chef de maintenance, lors de la détection d'anomalie ou de défaillance. Le chef de maintenance doit alors décider de la priorité de l'action à apporter. Le chef de maintenance émet alors un ordre qui est transmis aux opérateurs, pour qu'une maintenance soit effectuée. Les ordres d'inspections systématiques sont émis dans le but de détecter l'apparition des signes de dégradation des équipements. Si l'appareil inspecté présente un problème, un avis est transmis au chef. Si la dégradation dépasse les conditions d'opération préétablie, il émettra un ordre de maintenance conditionnelle, et un opérateur effectuera les réparations. Lorsqu'un événement de défaillance est confirmé, il y aura un avis de maintenance corrective, ainsi qu'un ordre de maintenance corrective.

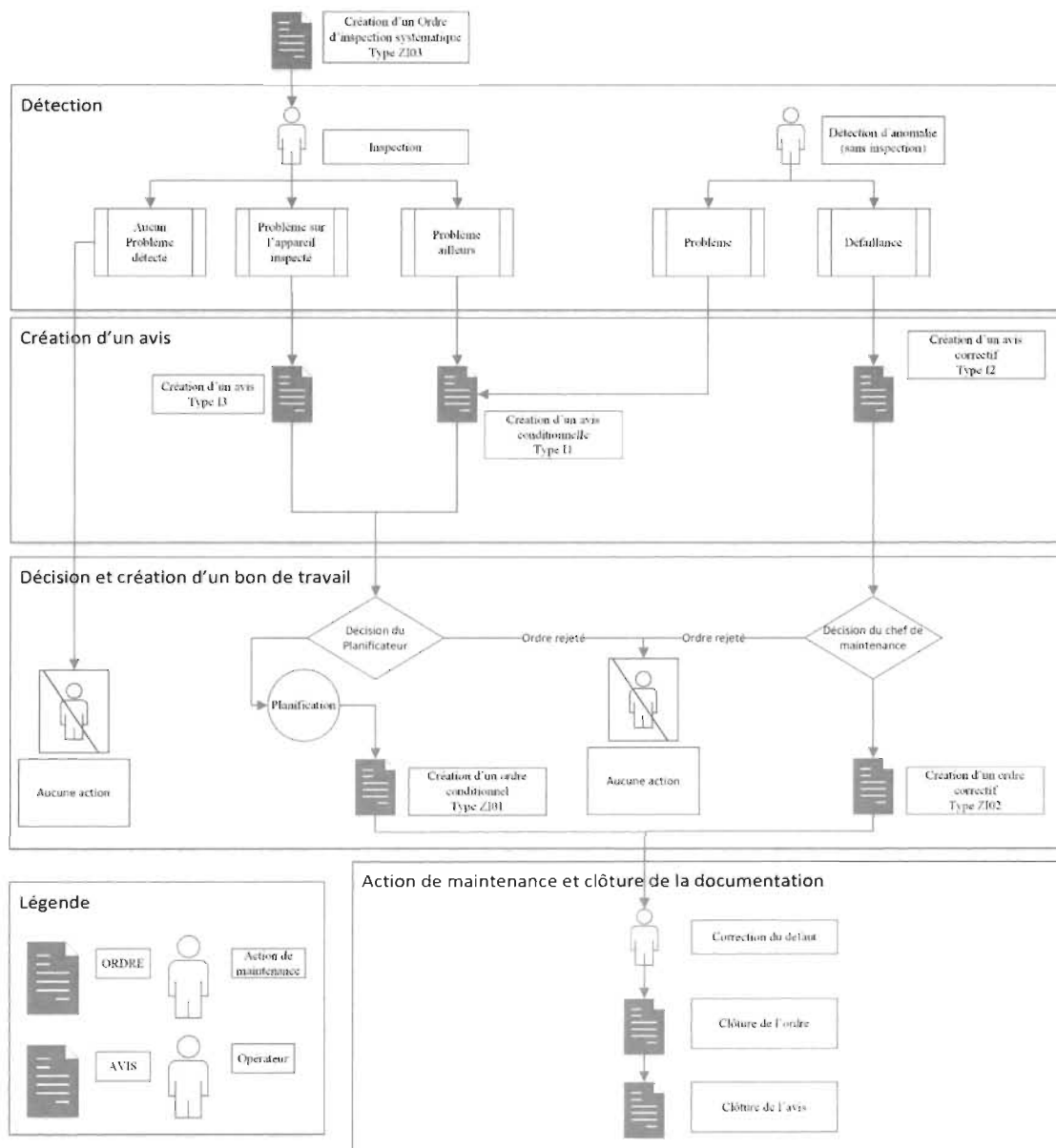


Figure 4.1: Processus de maintenance et de saisies de l'information

Tableau 4.1 : Récapitulatif du type de documents et de leur émission

| TYPE | | | DÉCLENCHÉ PAR | PROBLÈME DÉTECTÉ | ÉMIS PAR |
|-------|-----------|--------------|--|--|---------------------|
| AVIS | Préventif | Conditionnel | Détection d'anomalie (sans ordre) | Problème constaté sur un équipement | Opérateur |
| AVIS | Préventif | Conditionnel | Ordre d'inspection systématique | Problème constaté sur un autre équipement que celui qui doit être inspecté | Inspecteur |
| AVIS | Préventif | Systématique | Ordre d'inspection systématique | Problème sur l'appareil inspecté | Inspecteur |
| AVIS | Correctif | Correctif | Détection d'anomalie (sans ordre) | Défaillance | Opérateur |
| ORDRE | Préventif | Conditionnel | Avis conditionnel ou Avis systématique | Problème constaté sur un équipement et validé par le chef de maintenance | Chef de maintenance |
| ORDRE | Préventif | Systématique | Planification de maintenance | Aucun | Ordonnanceur |
| ORDRE | Correctif | Correctif | Avis correctif | Défaillance constatée et validée par le chef de maintenance | Chef de maintenance |

D'un point de vue pratique, cela signifie qu'un événement de défaillance sera signalé par un avis correctif, auxquels un ordre correctif est rattaché. En contrepartie, la méthode n'a pas toujours été employée de cette manière, et la codification des activités de maintenances a évolué au fil du temps. Ces facteurs influenceront la façon dont les données seront analysées ultérieurement.

4.1 Structure des données

La section précédente a permis d'établir la façon dont fonctionne la saisie des données d'opération qui alimentent les modèles de fiabilité. Cette section permet de décrire comment ces processus se formalisent de façon informatique. Pour cette étude, HQ a fourni un accès à une partie de la base de données (BD) de maintenance dans le but de produire

les études de fiabilité. Une base de données relationnelle est dotée de plusieurs tables. Ces tables contiennent des champs (variables/attributs), ainsi que l'information permettant de faire la relation d'une table à l'autre (clés). Chaque ligne d'une table représente un enregistrement d'information distinct. Pour traduire les informations récoltées lors des tâches de maintenance, il y a trois tables qui sont principalement utilisées. La table des avis de maintenance, la table des ordres, ainsi que la table des équipements.

Table des avis

La table des avis permet de lister toutes les anomalies qui ont été détectées sur un équipement par les opérateurs et les exploitants. Chaque avis est unique; un identifiant lui est assigné lors de son émission. L'opérateur y assigne aussi le type d'avis, tel que décrit précédemment. Qui plus est, chaque avis émis doit être relié à un équipement, par son numéro d'identification. En plus de ces champs, un avis contient d'autres informations comme le numéro d'ordre associé, la date d'émission et la date de clôture de l'avis, ainsi que des champs texte permettant d'identifier les composantes impliquées ainsi que les causes potentielles des anomalies.

Table des ordres

La table des ordres permet de lister toutes les actions de maintenance qui ont été effectuées sur un équipement. Chaque ordre est unique; un identifiant unique lui est assigné lors de son émission par le chef de maintenance. Celui-ci lui assigne le type d'ordre respectif et, comme pour les avis, il doit être associé à un numéro d'identification d'équipement. En plus de ces champs, un ordre contient d'autres informations comme le numéro d'ordre associé, les dates d'actions planifiées et réalisées, ainsi que des champs texte permettant d'identifier les opérations effectuées, la localisation de l'équipement, *etc.*

Tables des équipements

La table d'équipements, ou table d'inventaire permet de lister chaque équipement en service à TransÉnergie. Chaque équipement possède un identifiant unique. La table inventaire permet aussi d'associer l'équipement à une famille tel que les transformateurs de puissance, les sectionneurs, les inductances, *etc.* On y retrouve diverses informations comme :

- Date de mise en service

- Date de remise à neuf
- Date de remplacement
- Fabricant
- Tension
- Localisation
- Année de fabrication

4.2 Analyse exploratoire des données

Premiers constats

L'analyse du processus de saisie ainsi que la structure des données permettent de faire une première analyse. Dans une étude de fiabilité, l'événement d'intérêt est la défaillance. Dans le système informatique, ces défaillances sont enregistrées comme un ordre de maintenance corrective. Comme ces ordres sont rattachés à un équipement, il est possible de connaître l'historique des défaillances des équipements, en plus de tous leurs entretiens (ordre systématique et conditionnel). Les ordres sont aussi rattachés à un avis, il est donc possible de retracer cet historique. Cependant, les données réelles démontrent que les ordres ne sont pas toujours rattachés à des avis.

Qui plus est, il est important de considérer qu'il s'agit d'un jeu de données réelles et que certaines données sont enregistrées par un opérateur (tables d'avis et d'ordre). Certains champs sont remplis automatiquement, par exemple la date d'émission et l'identifiant de l'avis/ordre. D'autres champs sont entrés manuellement, mais le risque d'erreur de saisies est peu probable, par exemple, l'identifiant de l'équipement. Les experts estiment qu'il peut y avoir plus d'incertitude sur le type d'avis émis ou le type d'ordre. De plus, le processus peut engendrer des écarts de temps quant aux activités de maintenance. La Figure 4.2 permet d'illustrer ces écarts :

- L'inspection ne se réalise pas nécessairement à la date prévue
 - Il y a donc un retard d'inspection
- Il y a un délai entre la découverte d'une anomalie et la correction du problème
 - Délai entre la découverte de l'anomalie et la création de l'avis
 - Délai de traitement de l'avis par le chef de maintenance et la création de l'ordre
 - Délai entre la création de l'ordre et la tâche de maintenance
 - Délai entre la tâche de maintenance et la clôture de l'ordre

De plus, cela ne prend pas en compte que l'anomalie est probablement arrivée avant la date prévue de l'inspection.

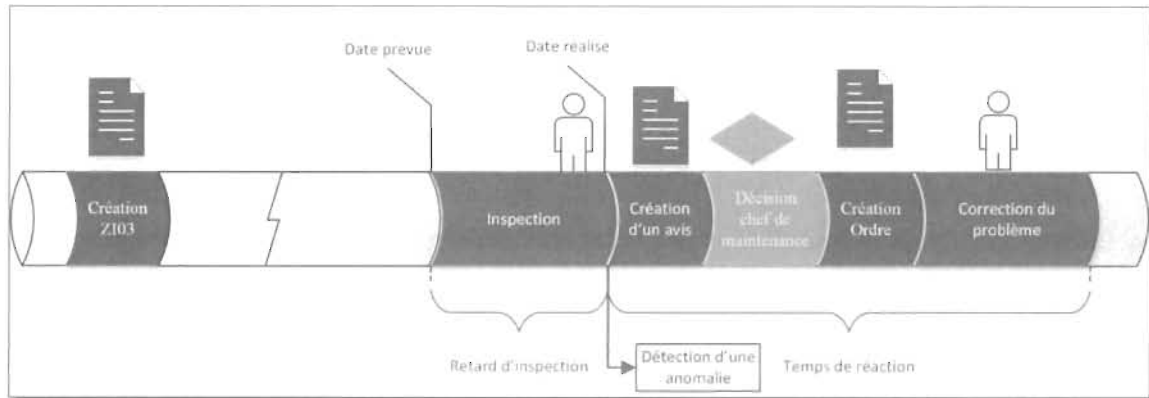


Figure 4.2 : Chronologie des activités de maintenance

4.3 Audit des données

Périmètre de l'audit et analyse des tables

Comme mentionné plus haut, les données qui serviront à l'analyse de fiabilité sont les tables d'avis, d'inventaire et d'équipement.

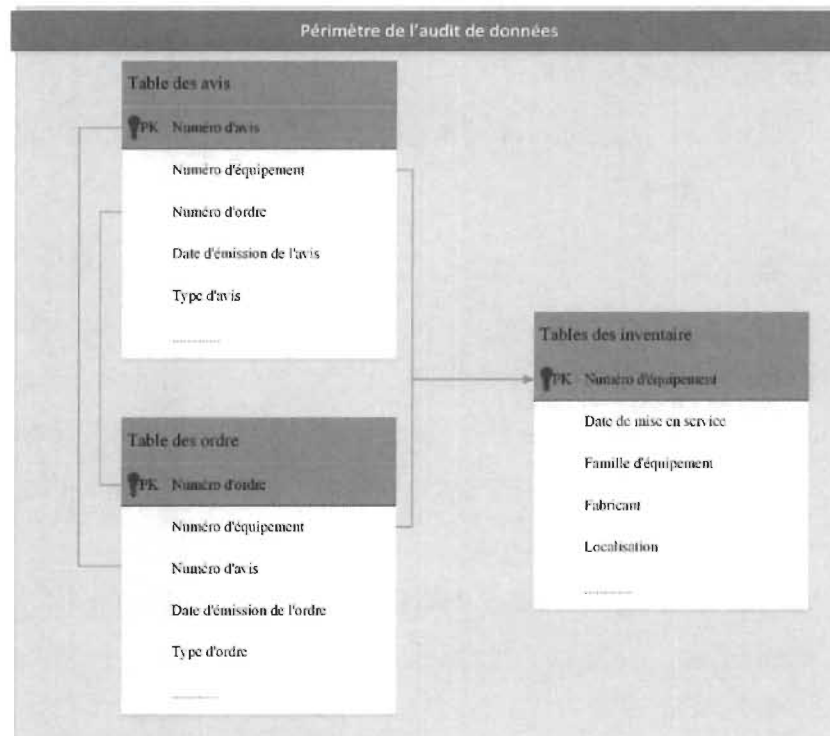


Figure 4.3: Structure simplifiée des tables de la BD maintenance

Comme l'illustre la Figure 4.3, la clé primaire de chaque table est un numéro d'identifiant unique, indiqué par la clé et les initiales PK (« *primary key* »). Les numéros

d'identifiant se retrouvent dans les autres tables à titre de clé étrangère, afin d'assurer la relation. Effectivement, on retrouve le numéro d'équipement comme attribut de la table des ordres, par exemple.

Contraintes sur les données

La première analyse portera sur la modélisation du taux de défaillance avec la méthode NHPP. Les données doivent donc être mises en forme de façon à ce que chaque entrée représente une défaillance unique. Il faut donc s'assurer que chaque ordre correctif est unique et, dans la mesure du possible, éliminer les entrées qui ne sont pas réellement des défaillances réparables. De plus, les dates de mises en service, de même que les dates d'interventions doivent être présentes et valides. La seconde partie porte sur l'analyse de survie, afin de déterminer quelles covariables ont une influence sur le taux de défaillance des équipements. Les contraintes du modèle NHPP pour la qualité des données s'appliquent aussi à ces modèles. Par contre, pour qu'une variable soit intégrée au modèle, il doit y avoir suffisamment d'entrées (complétude). Effectivement, si une variable a trop de données manquantes, cela diminuera de façon considérable le jeu de données et nuira au résultat. Enfin, les dates consignées dans les tables devront respecter une chronologie adéquate. Un équipement ne peut subir de défaillance avant sa mise en service, par exemple.

Complétude des données

Il faut auditer les données, afin de déceler les entrées qui peuvent être nuisibles à l'analyse. À cet effet, les données d'inventaires, d'ordre et d'avis de maintenance ont été croisées, de façon à ce que chaque entrée corresponde à un événement de maintenance. Chaque événement de maintenance correspond donc à une combinaison unique entre le code d'équipement, le numéro d'ordre et le numéro d'avis (s'il y en a un). Pour chaque champ, le nombre de valeurs nulles ou vides a été décompté à la Figure 4.4.

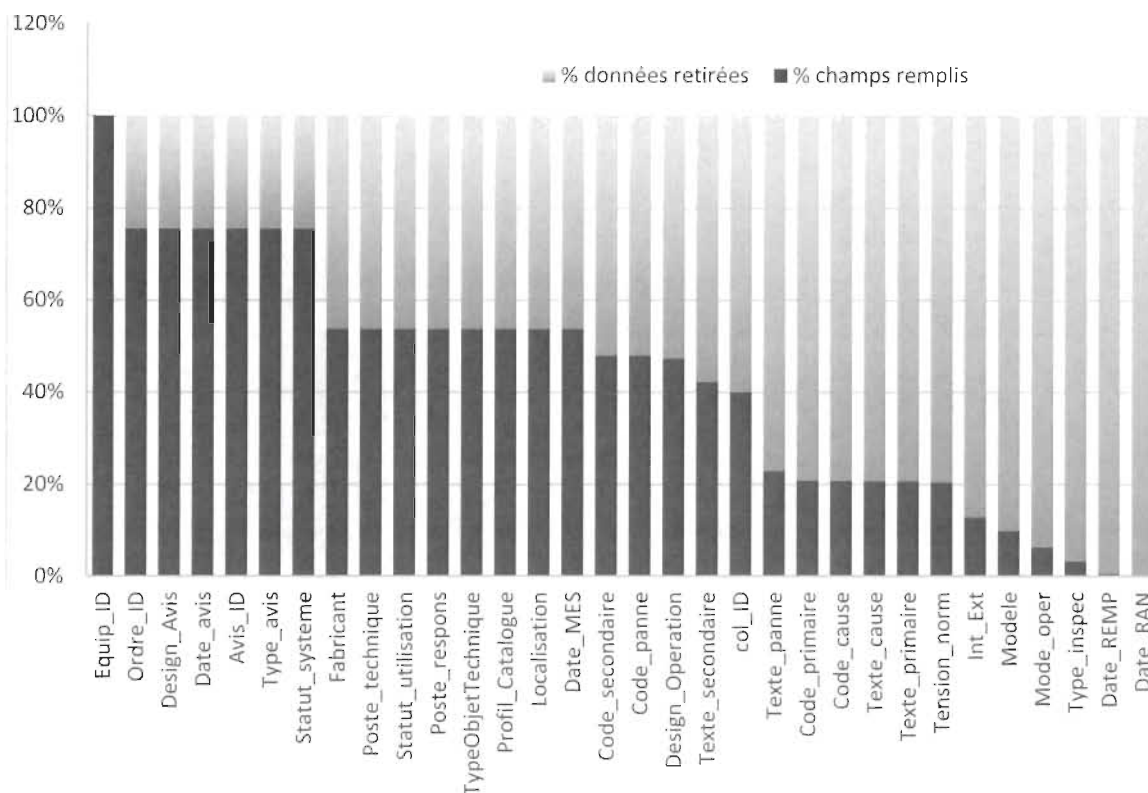


Figure 4.4: Pourcentage des champs remplis par caractéristiques

Cette figure démontre qu'il y a plusieurs champs qui ne sont presque jamais remplis. Comme les données à saisir n'ont pas toujours été les mêmes, ces observations ne sont pas surprenantes. Cependant, les données qui permettent d'identifier les événements de défaillances semblent généralement bien remplies, soit les numéros d'équipements, d'ordres et d'avis, ainsi que les dates d'émission. Ce sont principalement ces données qui serviront à modéliser le taux de défaillance. Comme mentionné précédemment, certains équipements d'Hydro-Québec sont en service depuis plusieurs décennies. Cependant, les politiques de maintenance ainsi que les systèmes d'information ont pu changer au cours de ces années. La Figure 4.5 présente l'évolution du nombre d'ordres émis au fil des années.

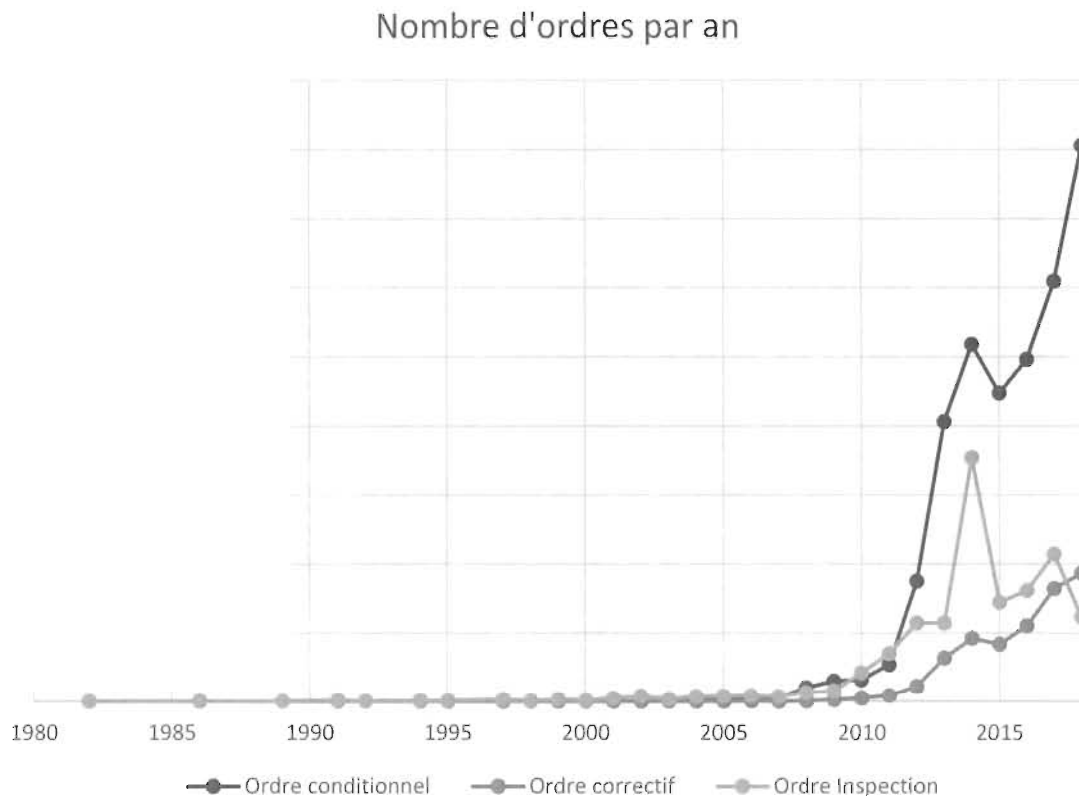


Figure 4.5: Nombre d'ordre émis en fonction du temps

En observant la Figure 4.5, il semble que peu d'ordres aient été enregistrés avant 2010, ce qui soulève plusieurs questions. En outre, les données débutent en 1982 et s'échelonnent sur une période de plus de trente ans. Selon les experts de TransÉnergie, plusieurs changements de système ont eu lieu; les plus récents en 2008 et en 2013. Cette absence de données peut être expliquée par plusieurs hypothèses :

- Changement de codification des ordres et avis
 - Et/ou changement du processus de saisie
- Augmentation des signalements à l'implantation d'un nouveau logiciel (formation)
- Perte de données dans les changements de systèmes
- Absence de documentation (1982-2000+)

À partir de tous ces facteurs, des règles d'exclusion ont été élaborées, afin de pouvoir assurer la qualité des analyses subséquentes. Le Tableau 4.2 décrit les différentes règles qui ont été appliquées pour contrôler la qualité des données.

Tableau 4.2: Règles d'affaires des événements de maintenance

| Table | Règles d'affaires (Règles de rejet) | Raison |
|------------|--|---|
| | Aucune règle | |
| Inventaire | Avis sans numéro d'équipement | Tout avis/ordre doit être lié à un équipement |
| Inventaire | Équipement présent dans la table d'avis, mais qui n'est pas dans la table d'inventaire | Si l'équipement n'est pas dans l'inventaire, il ne peut être lié à son historique complet |
| Inventaire | Équipement sans date de mise en service | Impossible de connaître l'âge de l'équipement |
| Inventaire | Mise en service en 1901-01-01 exclue | Valeur automatique, donc pas une date réelle |
| Inventaire | Date de remise à neuf incohérente | Date de remise à neuf avant la mise en service |
| Inventaire | Date de remplacement incohérente | Date de remplacement avant la mise en service |
| Avis | Date avis nulle | Date avis nulle |
| Avis | Date d'avis incohérente | Date d'avis avant la mise en service |
| Avis | Avis sans ordre | Tout avis doit être lié à un ordre; assume qu'aucune maintenance n'a eu lieu |
| Avis | Avis Rejeté | L'avis a été rejeté par le chef de maintenance |
| Avis | Avis Doubleton | Entrée en double |

Ces règles permettront donc de ne conserver que les entrées qui sont réellement des événements de maintenance. De plus, cela assure que les dates qui sont entrées dans le système sont valides et suivent la chronologie des événements. À partir de l'outil DataBricks, ces règles ont pu être écrites avec une syntaxe similaire à une requête SQL, soit avec le module `pyspark.sql`. L'Équation 4.1 permet d'illustrer la structure d'écriture de chacune des règles :

Équation 4.1:

```
dataFrame = dataFrame.filter(col("Date_MES").isNotNull())
```

Ces règles permettent de créer un « *DataFrame* » qui pourra être utilisé pour les analyses de fiabilités. Les résultats de l'application de ces filtres, pour tous les équipements de TransÉnergie, sont présentés au tableau suivant.

Tableau 4.3: Résultats de l'application des règles d'exclusion

| # | Règle d'affaires | % d'équipement restant | % d'avis conditionnel restant | % d'avis correctif restant | % d'avis inspection restant |
|----|--|------------------------|-------------------------------|----------------------------|-----------------------------|
| 0 | Tous équipements HQT dans les tables d'inventaire, d'avis et d'ordre | 100.0% | 100.0% | 100.0% | 100.0% |
| 1 | Avis sans numéro d'équipement | 100.0% | 91.0% | 87.0% | 98.4% |
| 2 | Équipement présent dans la table d'avis, mais qui n'est pas dans la table d'inventaire | 69.0% | 38.9% | 36.0% | 39.8% |
| 3 | Équipement sans date de mise en service | 69.0% | 38.9% | 36.0% | 39.8% |
| 4 | Mise en service en 1901-01-01 exclue | 69.0% | 38.9% | 36.0% | 39.8% |
| 5 | Date de remise à neuf incohérente | 69.0% | 38.9% | 36.0% | 39.8% |
| 6 | Date de remplacement incohérente | 69.0% | 38.9% | 36.0% | 39.8% |
| 7 | Date avis nulle | 69.0% | 38.9% | 36.0% | 39.8% |
| 8 | Date d'avis incohérente | 69.0% | 38.3% | 34.7% | 39.8% |
| 9 | Avis sans ordre | 69.0% | 23.7% | 33.4% | 22.7% |
| 10 | Avis Rejeté | 69.0% | 23.7% | 33.4% | 22.7% |
| 11 | Avis Doublon | 69.0% | 23.7% | 33.4% | 22.7% |

Ce tableau démontre l'importance d'appliquer des règles, afin de nettoyer les données. Les deux premières règles permettent d'assurer que chaque avis est lié à un équipement; pour conduire l'analyse de fiabilité, il est primordial que les défaillances soient associées correctement à un équipement. De plus, ceux-ci doivent être disponibles dans la table d'inventaire, afin d'avoir toutes les informations qui y sont reliées. Les règles 3, 5, 6, 7 et 8 permettent d'assurer que les avis ont des dates enregistrées, et que ces dates

sont cohérentes avec la chronologie des événements. En plus d'appliquer ces règles génériques, il a été nécessaire d'appliquer des filtres, dans le but de réduire la population à l'étude. La Figure 4.5 démontre le besoin de filtrer sur un intervalle de temps, pour que les données de défaillances soient bien identifiées. Ainsi, la période de 2013 à 2018 sera étudiée dans le but de produire les analyses de fiabilité.

Enfin, le jeu de données initial contient toutes les familles équipements, il faut donc le filtrer pour n'avoir que les transformateurs de puissances. Ces règles et ces filtres sont génériques. Les modèles qui seront présentés exigent de nouvelles mises en forme pour le jeu de données, ainsi que différentes spécificités. Des filtres supplémentaires seront employés en vue d'obtenir de meilleurs résultats. Ces filtres seront présentés dans les deux prochains chapitres, qui concernent respectivement l'application de la modélisation du taux de défaillance et des modèles de survie.

CHAPITRE 5

MODÉLISATION DU TAUX DE DÉFAILLANCE

L'objectif de ce chapitre est de présenter la démarche détaillée, de même que les résultats de l'analyse **de fiabilité** pour les transformateurs de puissances. À cet effet, les étapes pour y arriver seront présentées dans les sections suivantes :

- Extraction et préparation des données pour l'étude;
- Présentation des résultats et analyse;
- Recommandations.

5.0 Données préliminaires

Exigences

Les données ont été filtrées dans l'intention de ne conserver que les entrées correspondant à des ordres de types correctifs (défaillances fonctionnelles). De cette façon, chaque ligne correspond à un événement de défaillance unique et est rattachée à un seul équipement. Comme défini, la période observée s'étend de 2013 à 2018, puisque c'est dans cet intervalle de temps que les données semblent les mieux consignées. Notez que le dernier changement de système date de 2013, ce qui laisse aussi croire que les données précédentes n'ont probablement pas toutes été transférées dans le système actuel.

Préparation

Ensuite, le chapitre 3 a permis de déterminer les exigences pour les données, en ce qui a trait à l'application de la méthode de la MCF :

1. Ordonner chaque observation en ordre croissant selon l'âge lors de l'événement
2. Déterminer le nombre d'unités qui entrent l'échantillon à l'âge t_i
3. Calculer le nombre d'équipements qui sont à risque à l'âge t_i
4. Calculer l'incrément du nombre moyen de récurrences par unité à risque
5. Calculer la somme cumulative des incréments (MCF)
6. Tracer le graphique des M_i^* en fonction de l'âge t_i
7. Tracer la courbe d'ajustement, afin de déterminer les paramètres d'une loi de puissance

Le Tableau 5.1 présente un échantillon des données utilisées, afin de montrer comment la méthode a été appliquée au jeu de donnée d'Hydro-Québec.

5.1 Résultats

Les données calculées à partir de la méthode MCF permettent de produire le graphique de la Figure 5.1.

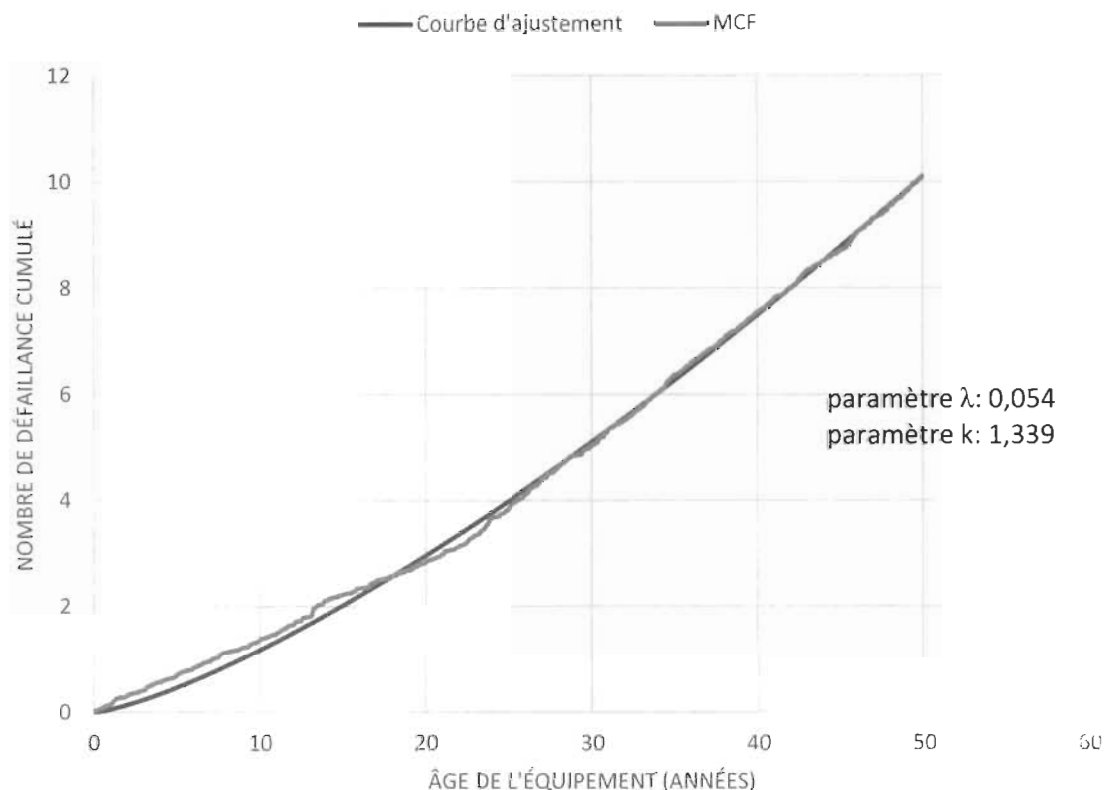


Figure 5.1: Nombre attendu de défaillances cumulées en fonction de l'âge

Ce graphique illustre la fonction MCF qui représente le nombre de défaillances cumulées en fonction de l'âge des équipements. Ce faisant, cela correspond à la fonction $H(t)$, lorsque les données suivent un processus de loi de puissance. L'ajustement de la courbe, par la méthode des moindres carrés, permet d'obtenir les paramètres k et λ des équations de la fonction d'intensité $h(t)$ et de la fonction cumulative $H(t)$, telle que présentée au Tableau 5.2. Pour ce faire, la fonction « `scipy.optimize.curve_fit` » de la librairie « `SciPy` » a été appliquée à l'Équation 5.5.

Tableau 5.2: Rappel des équations de la loi de puissance

| # | Fonction | Équation |
|--------------|----------------------------------|------------------------------------|
| Équation 5.1 | $f(t)$ | $k\lambda t^{k-1}e^{-\lambda t^k}$ |
| Équation 5.2 | $S(t)$ | $e^{-\lambda t^k}$ |
| Équation 5.3 | $F(t)$ | $1 - e^{-\lambda t^k}$ |
| Équation 5.4 | $\lambda(t) = \frac{f(t)}{S(t)}$ | $k\lambda t^{k-1}$ |
| Équation 5.5 | $H(t)$ | λt^k |

Le taux de défaillance instantanée peut donc être obtenu par l'équation suivante :

Équation 5.6 $\lambda(t) = 1,338 * 0,054 * t^{0.339}$

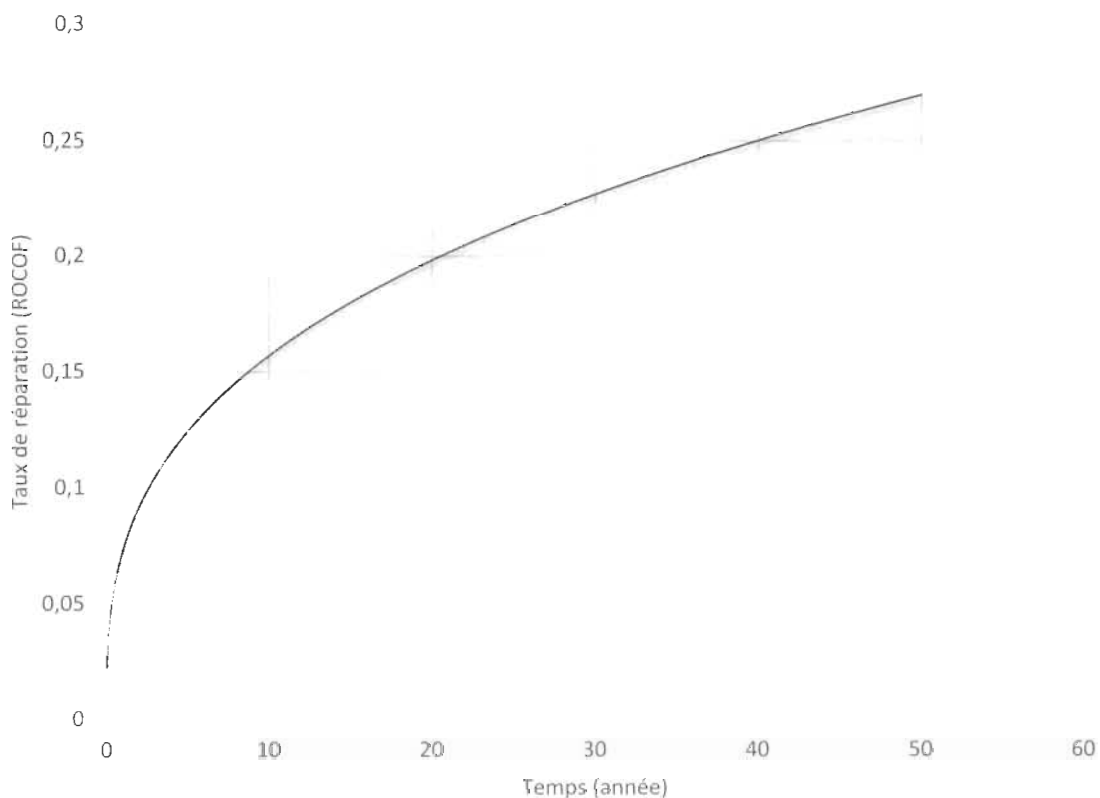


Figure 5.2: Taux de réparation en fonction de l'âge des équipements

La Figure 5.2 illustre graphiquement le taux de réparation des équipements, en fonction de leur âge. L'hypothèse d'un modèle HPP est que le taux de réparation est constant. En observant ce graphique, il est évident qu'une telle hypothèse mènerait à des estimations

éloignées du nombre de défaillances que subira chaque équipement. Le modèle NHPP propose un taux qui varie selon l'âge des équipements, ce qui semble approprié dans ce cas-ci. Effectivement, il est évident que le taux de réparation augmente en fonction du temps à la Figure 5.2. De plus, vers dix ans, la vitesse d'augmentation du taux semble varier de façon linéaire. À la Figure 5.3, la courbe de la fonction cumulative est représenté avec l'intervalle de confiance (IC), pour l'estimation à chaque pas de temps. Cet intervalle représente la borne supérieur et inférieur de l'estimation à un niveau de confiance de 95%. La figure illustre que les bornes de l'IC croissent avec le temps, ce qui indique que l'erreur d'estimation est plus grande pour les équipements plus âgé. Par contre, les bornes restent bien groupées autour de la fonction, ce qui indique que l'estimation est précise.

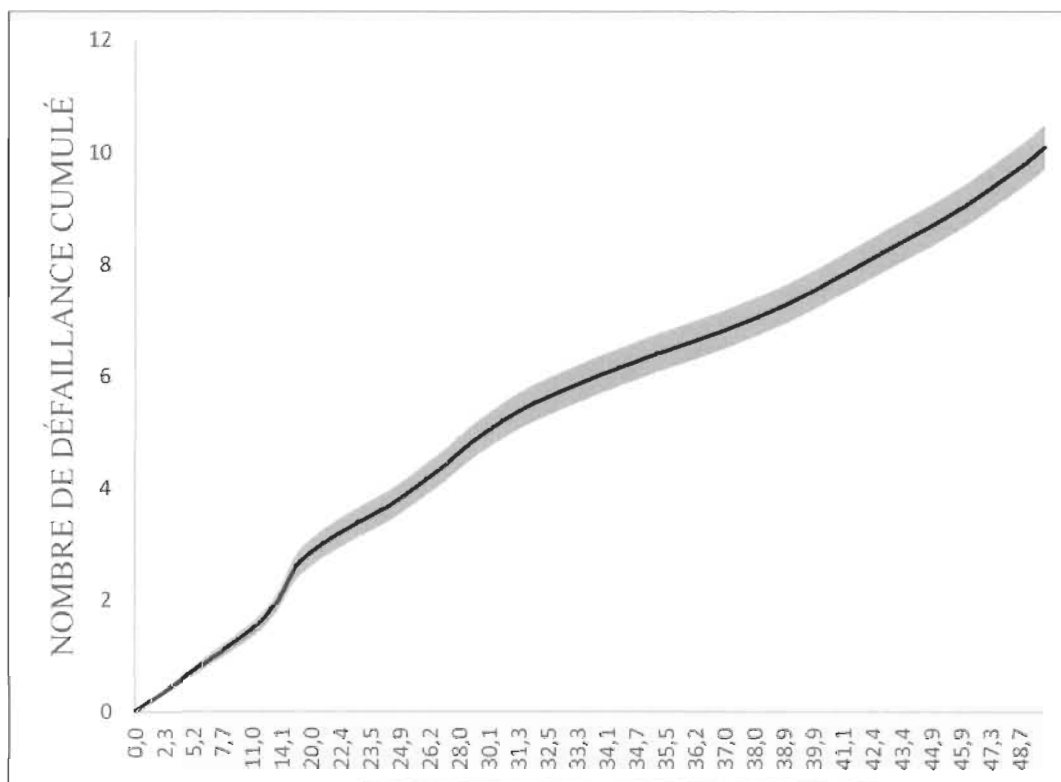


Figure 5.3 : Nombre de défaillances cumulées en fonction de l'âge avec IC

5.2 Discussions et Recommandations

À partir des résultats obtenus, les responsables de la gestion des actifs seront en mesure d'obtenir des estimations plus précises du taux de défaillance de leurs équipements. Effectivement, les équations permettent de déterminer rapidement un taux de défaillance en fonction de l'âge des équipements, ce qui n'était pas le cas auparavant. De plus, cette

approche tient compte des censures, ce qui est une avancée importante compte tenu de la structure des données présentées par la Figure 4.5. Cela permettra de prendre des décisions plus appropriées, en tenant compte du nombre d'équipements en service pour chaque tranche d'âge, plutôt que de considérer un taux constant. De plus, les valeurs obtenues permettront de réaliser des simulations, afin de déterminer des politiques d'entretien qui maximisent les objectifs d'entreprise.

CHAPITRE 6

ANALYSE DE SURVIE

L'objectif de ce chapitre est de présenter la démarche détaillée, ainsi que les résultats de l'analyse de survie des transformateurs de puissances. À cet effet, les étapes pour y arriver sont présentées dans les sections suivantes :

- Préparation des données pour l'étude
- Analyse préliminaire
- Résultats et analyse des modèles
- Recommandations

6.1 Préparation des données

Les données qui seront utilisées seront essentiellement les mêmes que celles appliquées au modèle NHPP. Effectivement, les mêmes filtres et règles ont été employés, mais le format des données est différent. Encore une fois, chaque ligne correspond à un événement de défaillance unique et est rattachée à un seul équipement. La période observée s'étend également de 2013 à 2018. Comme présenté au chapitre 3, les données pour les modèles PWP-TT et AG doivent être exprimées sous la forme d'un processus de comptage. Ensuite, il a été nécessaire de sélectionner les variables à analyser pour les modèles de survie, et les transformer pour qu'elles soient traitables par l'algorithme. Les Tableau 6.1 au Tableau 6.6 résument les variables considérées.

Tableau 6.1: Description des variables de base du modèle

| Variabes | Description | Type de variable | % bonne donnée |
|--------------|---|------------------|----------------|
| tstart | Temps du début de l'intervalle d'observation | Modèle | 100 |
| status | status = 1 si l'événement est récurrent et 0 s'il est censuré | Modèle | 100 |
| tstop | Temps de fin de l'intervalle d'observation | Modèle | 100 |
| Equip_ID | Identifiant de l'équipement | Modèle | 100 |
| numEvenement | le k_ième événement | Modèle | 100 |

Tableau 6.1 présente les variables de base du modèle, tel que présenté dans la méthodologie.

Tableau 6.2: Descriptions des variables d'indicateur d'état

| Variabes anonymes | Description | Type de variable | % bonne donnée |
|-------------------|---------------------------------------|------------------|----------------|
| Cote 1 | Indicateurs de l'état de l'équipement | Cote | 100 |
| Cote 2 | | Cote | 100 |
| Cote 3 | | Cote | 100 |
| Cote 4 | | Cote | 100 |
| Cote 5 | | Cote | 100 |
| Cote 6 | | Cote | 100 |
| Cote 7 | | Cote | 100 |
| Cote 8 | | Cote | 100 |

Le Tableau 6.2 présente les variables indiquant l'état des équipements selon une cote. Ces champs permettent aux opérateurs terrain de mettre une cote de « santé » de l'équipement, selon l'état qui est constaté lors d'une maintenance. Ces variables sont ordinales, c'est-à-dire qu'elles sont qualitatives, mais qu'il existe une échelle de valeurs qui permet de les ordonner.

Quant au Tableau 6.3, il présente les variables qui sont relatives aux caractéristiques de l'équipement. Les variables CPC_Bin, CARACT_TENSION, Classe d'actif et Fabricant, sont des variables qualitatives et devront donc être traitées telles que décrites dans la méthodologie. Aussi, certaines variables comportent des données manquantes (94% de complétude). Cependant, l'information que ces variables pourraient apporter au modèle fait en sorte qu'elles seront tout de même incluses, et donc une partie des entrées seront rejetées.

Tableau 6.3: Description des variables d'équipements

| Variables | Description | Type de variable | % bonne donnée |
|------------------------|--|------------------|----------------|
| Fabricant | Nom du fabricant de l'équipement | Équipement | 100 |
| CPC_Bin | Présence d'un CPC = 1 | Équipement | 100 |
| CARACT_TENSION | Tension maximum de l'équipement | Équipement | 100 |
| Classe d'actif | Si l'équipement a une inductance shunt | Équipement | 100 |
| Nombre de phases | Triphasé ou Monophasé | Équipement | 100 |
| Latitude | Latitude | Équipement | 94.13 |
| Longitude | Longitude | Équipement | 94.13 |
| EQUI_AnneeConstruction | Année de fabrication de l'équipement | Équipement | 98.94 |

Tableau 6.4: Description des variables d'avis

| Variables | Description | Type de variable | % bonne donnée |
|---------------------------|-----------------------------------|------------------|----------------|
| Nombre_Avis | Nombre total d'avis | Avis | 100 |
| Nombre_Avis_Conditionnels | Nombre total d'avis conditionnels | Avis | 100 |
| Nombre_Avis_Correctifs | Nombre total d'avis correctif | Avis | 100 |
| Nombre_Avis_Inspection | Nombre total d'avis systématique | Avis | 100 |

Tableau 6.5: Description des variables d'ordre

| Variables | Description | Type de variable | % bonne donnée |
|----------------------------|-------------------------------------|------------------|----------------|
| Nombre_Ordre_Correctif | Nombre total d'avis correctif | Ordre | 100 |
| Nombre_Ordre_Inspection | Nombre total d'ordres conditionnels | Ordre | 100 |
| Nombre_Ordre_Conditionnels | Nombre total d'ordres correctif | Ordre | 100 |

Les Tableau 6.4 et Tableau 6.5 présentent les variables sur les ordres et les avis. Il s'agit en fait du décompte de nombre d'ordre et d'avis émis par type.

Tableau 6.6: Description des variables hybrides

| Variables | Description | Type de variable | % bonne donnée |
|-------------------------------|--|------------------|----------------|
| Ratio_Ordre_Avis | Rapport du nombre d'ordres systématiques par rapport au nombre total d'avis | Hybride | 100 |
| Nombre_Inspection_IC | Nombre d'avis d'inspection conditionnel | Hybride | 100 |
| Nombre_Inspection_IP | Nombre d'avis d'inspection préventive | Hybride | 100 |
| Ratio_1 | Ratio du nombre d'ordres correctif par rapport au nombre d'avis correctif | Hybride | 100 |
| Ratio_2 | Ratio du nombre d'ordres conditionnels par rapport au nombre d'avis systématique et conditionnel | Hybride | 100 |
| Retard_inspection_moy_mensuel | Retard d'inspection moyen (en mois) | Hybride | 100 |
| Fréquence_inspection | Fréquence d'inspection | Hybride | 100 |
| Coefficient_realisation | Rapport entre le nombre d'inspections et l'âge de l'équipement | Hybride | 100 |

L'un des objectifs de ce projet est de déterminer l'impact des politiques de maintenances sur le nombre de défaillances des équipements. Pour ce faire, des variables hybrides ont été développées afin de représenter ces processus de maintenance. La Figure 6.1 illustre la variable Ratio_Ordre_Avis. L'idée derrière la création de cette variable est de voir si les inspections effectuées permettent réellement de détecter des problèmes et si cela influence la fiabilité. Par exemple, un équipement qui a plusieurs ordres d'inspection et peu d'avis aura un ratio élevé. Si cette variable a une influence sur le taux de défaillance, et qu'un équipement a un ratio élevé, cela indique que les méthodes de détections sont peut-être inadéquates. À l'inverse, un ratio bas peut indiquer que la périodicité des inspections peuvent être à revoir.

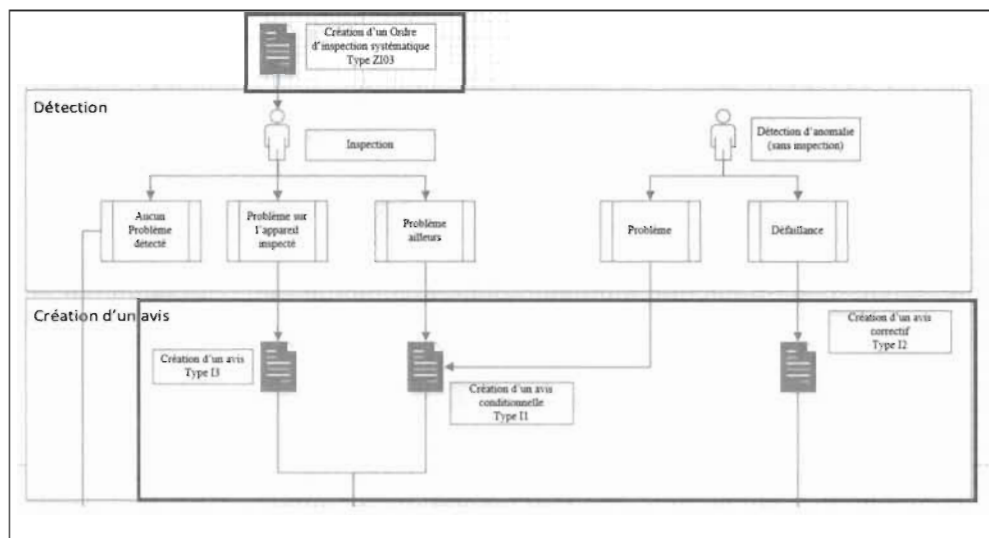


Figure 6.1: Illustration de la variable *Ratio_Ordre_Avis*

De la même façon, les variables *Ratio_1* et *Ratio_2* illustrent l'efficacité du processus d'inspection et d'ordonnancement.

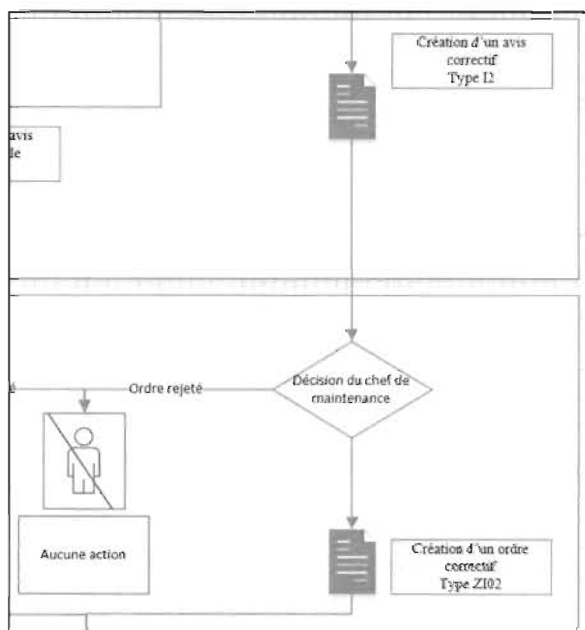


Figure 6.2: Illustration de la variable *Ratio_1*

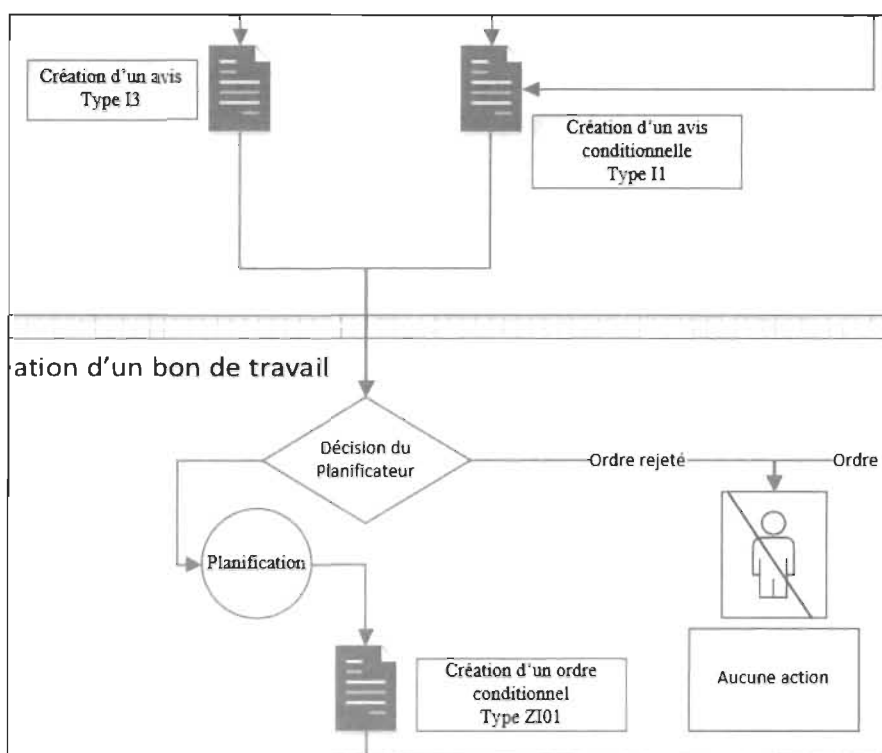


Figure 6.3: Illustration de la variable Ratio_2

6.2 Analyse Préliminaire

Dans l'intention d'obtenir un modèle satisfaisant, toutes les variables seront injectées dans le modèle, afin d'obtenir une analyse préliminaire. Cette première analyse servira à analyser l'impact de certaines variables, tandis que d'autres variables pourront être modifiées ou retirées. Les variables hybrides, par exemple, sont des constructions faites à partir des variables d'avis et d'ordre. Inclure ces deux types de variables n'ajoute pas de valeur au modèle final, puisqu'elles sont dépendantes l'une de l'autre. Le Tableau 6.7 présente les résultats obtenus pour le modèle préliminaire. Les variables sont présentées avec la valeur exponentielle de son coefficient et sa « *p-value* » ($\Pr(>Z)$). La « *p-value* » permet d'évaluer si la variable a un effet significatif sur le taux de défaillance. Dans ce cas-ci, l'effet de la variable est significatif lorsque la valeur est inférieure à un seuil de 5%.

Tableau 6.7: Sortie du modèle préliminaire PWP-TT

| # | Variable anonymisée | Exp(coeff) | Pr(>z) |
|----|----------------------------|------------|--------|
| 1 | Nombre_Ordre_Conditionnels | 1.0173 | 0.1480 |
| 2 | Nombre_Ordre_Inspection | 0.6821 | 0.0519 |
| 3 | Nombre_Avis_Conditionnels | 1.0371 | 0.0188 |
| 4 | Nombre_Inspeç_IC | 1.7388 | 0.0102 |
| 5 | Nombre_Inspeç_IP | 1.6538 | 0.0163 |
| 6 | Retard_inspect_moy_mensu | 1.0085 | 0.3227 |
| 7 | Frequence_inspection | 0.9787 | 0.1349 |
| 8 | Coefficien_realisation | 0.5452 | 0.2238 |
| 9 | AnneeConstruction | 0.9999 | 0.7316 |
| 10 | CPC | 0.6236 | 0.0005 |
| 11 | Latitude | 1.0275 | 0.0956 |
| 12 | Longitude | 1.0180 | 0.0952 |
| 13 | Cote 1 | 1.0059 | 0.3024 |
| 14 | Cote 2 | 1.6576 | 0.0000 |
| 15 | Cote 3 | 0.9387 | 0.3335 |
| 16 | Cote 4 | 0.7281 | 0.0327 |
| 17 | Cote 5 | 1.0221 | 0.7938 |
| 18 | Cote 6 | 1.1923 | 0.2578 |
| 19 | Cote 7 | 0.9138 | 0.3144 |
| 20 | Cote 8 | 1.8009 | 0.0000 |
| 21 | Classe d'actif 1 | NA | NA |
| 22 | Classe d'actif 2 | NA | NA |
| 23 | Tension 1 | 2.8267 | 0.0452 |
| 24 | Tension 2 | 0.2363 | 0.0022 |
| 25 | Tension 3 | 3.0184 | 0.0070 |
| 26 | Tension 4 | 0.9918 | 0.9839 |
| 27 | Tension 5 | 1.2834 | 0.4729 |
| 28 | Tension 6 | 1.0397 | 0.8654 |
| 29 | Tension 7 | 1.2263 | 0.2391 |
| 30 | Fabricant 1 | 0.7313 | 0.1225 |
| 31 | Fabricant 2 | 1.1451 | 0.7811 |
| 32 | Fabricant 1 (nom 2) | 1.0110 | 0.9272 |
| 33 | Fabricant 1 (nom 3) | 0.5509 | 0.2977 |
| 34 | Fabricant 3 | 0.8585 | 0.8534 |

Tableau 6.8: Sortie du modèle préliminaire PWP-TT (suite)

| # | Variable anonymisée | Exp(coeff) | Pr(>z) |
|----|---------------------|------------|--------|
| 35 | Fabricant 4 | 0.0014 | 0.0000 |
| 36 | Fabricant 5 | 0.5319 | 0.4177 |
| 37 | Fabricant 6 | 0.6276 | 0.1146 |
| 38 | Fabricant 7 | 0.0002 | 0.0000 |
| 39 | Fabricant 7 (nom 2) | 1.2785 | 0.1489 |
| 40 | Fabricant 8 | 0.0002 | 0.0000 |
| 41 | Fabricant 8 (nom 2) | 0.8989 | 0.4729 |
| 42 | Fabricant 9 | 0.7259 | 0.5620 |
| 43 | Fabricant 10 | 1.0898 | 0.8487 |
| 44 | Fabricant 11 | 1.7494 | 0.2783 |
| 45 | Fabricant 12 | 1.2268 | 0.5236 |
| 46 | Fabricant 13 | 0.4117 | 0.3059 |
| 47 | Fabricant 14 | 0.8113 | 0.4873 |
| 48 | Fabricant 15 | 0.9465 | 0.8057 |
| 49 | Fabricant 16 | NA | NA |
| 50 | Monophasé | 0.6951 | 0.0631 |
| 51 | Triphasé | NA | NA |

Pour commencer, la sortie du modèle permet de faire plusieurs constatations :

1. Les variables hybrides n'ont pas d'impact, comparativement aux variables d'ordre et d'avis (à un seuil de signification de 5%)
2. Le modèle n'a pas réussi à calculer de coefficient pour la variable classe d'actif et sur le nombre de phases.
3. Plusieurs fabricants similaires sont entrés sous des noms différents
 - a. De ce fait, certaines des variables tendent à l'infini

La Figure 6.4 illustre les courbes d'intensité en fonction du temps pour chaque strate. Chaque strate correspond au nombre d'événements cumulés, tel que présenté au CHAPITRE 2. Cela permet de mettre en évidence un nombre important de courbes (16), ce qui est anormal compte tenu de la fiabilité élevée des transformateurs de puissance. En effet, la présence de 16 courbes signifierait que certains transformateurs ont subi jusqu'à 16 défaillances pour la période observée.

Fonction de risque cumulative par nombre d'événement

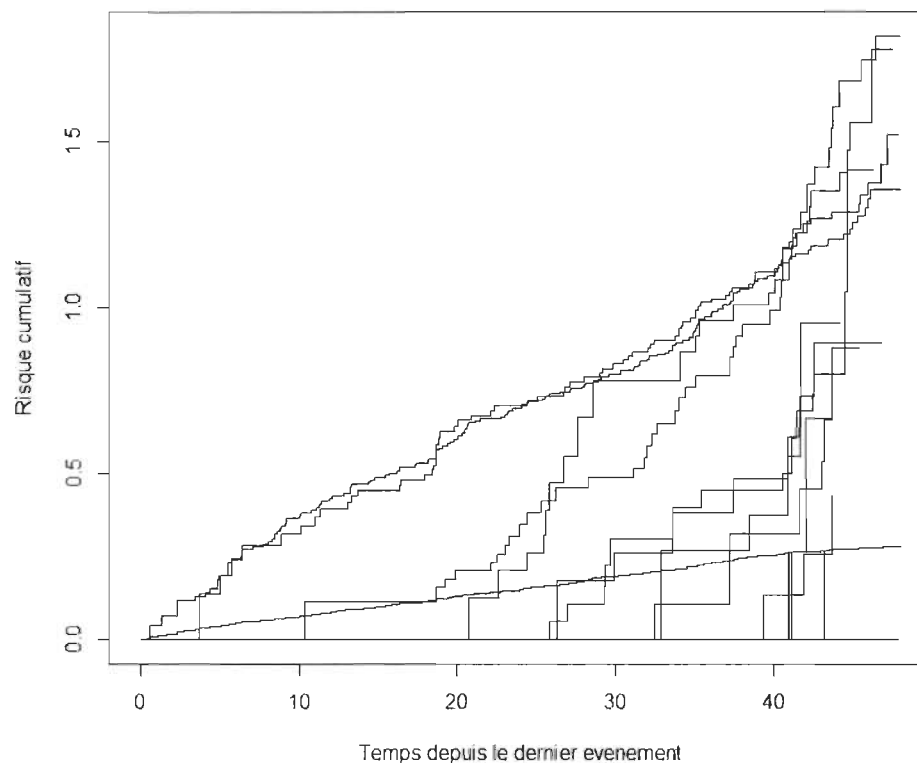


Figure 6.4: Graphique de la fonction de risque cumulative

Une analyse plus approfondie a permis de déterminer les causes de ces anomalies, notamment en retraçant certains champs texte. Comme démontré par le Tableau 6.9, plusieurs des événements qui ont été enregistrés comme des ordres correctifs ne correspondent pas à une défaillance. Comme défini, une défaillance cause l'arrêt des fonctions du système. Plusieurs de ces champs ont une référence « à vérifier », ce qui ne correspond vraisemblablement pas à une perte de fonction. De plus, les champs du Correctif 14 sont intéressants. Le champ 1 désigne que l'action effectuée est l'achat d'un composant de rechange supplémentaire, tandis que le champ 2 fait état d'un bris majeur à corriger. L'interprétation de ces deux champs est donc contradictoire. Bien que ces données sont peu nombreuses, elles peuvent avoir un impact important sur la modélisation.

Tableau 6.9: Extraction de champs textes pour les Ordres de défaillances anormales

| Type ordre | Champs texte 1 | Champs texte 2 |
|--------------|---|--|
| Correctif 1 | Changer câble relais de gas | Transfo de puissance a verifier |
| Correctif 2 | radiateur t16 bouche chauffe | TRANSPU/Sys. refroidissement a corriger |
| Correctif 3 | Changer tuyau et dessicateur | Transfo de puissance a verifier |
| Correctif 4 | traverse T7 brisee | TransPU-Transfo avec CPC a verifier |
| Correctif 5 | T26 Incendie Panne condensateur | TRANSPU/Probl. non identifie, a corriger |
| Correctif 6 | t9 ventilateur defectueux deuxieme | TRANSPU/Sys. refroidissement a corriger |
| Correctif 7 | 84078635 T13 silicagel a changer | TRANSPU/Dessicateur a corriger |
| Correctif 8 | Verification relais de gas | Transfo de puissance a verifier |
| Correctif 9 | inspection tete de câble T13 | Transfo de puissance a verifier |
| Correctif 10 | tete de câble T6 | Transfo de puissance a verifier |
| Correctif 11 | panne syst. inc. t7 | TRANSPU/Malt CC a corriger |
| Correctif 12 | T2: Alarme cond. anorm. refroidissement | Transfo de puissance a verifier |
| Correctif 13 | T2: Alarme de refroidissement anormal | TRANSPU/Probl. non identifie, a corriger |
| Correctif 14 | Achat relais de gas pour futur | TRANSPU/Bris majeure a corriger |

6.3 Résultats

Préparation des données

Les résultats préliminaires permettent de faire une première itération pour déceler les problèmes relatifs aux données et aux variables qui n'auraient pas été détectées avant. Ainsi, quelques règles peuvent être ajoutées pour produire un modèle plus robuste et plus représentatif de la réalité. Les modifications suivantes ont été appliquées au jeu de données :

1. Retrait des variables hybrides
2. Retrait de la variable classe d'actif
3. Retrait des variables Nombre_Inspec_IC et Nombre_Inspec_IP
 - a. Redondance avec Nombre_Ordre_Inspection
4. Uniformisation des noms de fabricant
5. Limite de 7 défaillances par équipements pour la période étudiée

Plutôt que d'exclure les équipements qui semblaient problématiques, il a été décidé de limiter le nombre de défaillances à 7. De cette façon, les valeurs aberrantes sont exclues, sans toutefois éliminer certaines valeurs des variables.

Modèle Andersen-Gill

Le modèle Anderson-Gill a été appliqué sur le nouveau jeu de données des défaillances des transformateurs de puissance de HQT. Le Tableau 6.10 présente les résultats obtenus par l'application du modèle AG.

Tableau 6.10: Résultats de l'application du modèle Andersen-Gill

| Variable | exp(Coefficient) | Pr(>z) |
|----------------------------|------------------|--------|
| Nombre Ordre Conditionnels | 1.06 | 0.000 |
| Nombre Ordre Inspection | 1.09 | 0.041 |
| Nombre Avis Conditionnels | 1.04 | 0.090 |
| CPC Bin | 0.53 | 0.000 |
| Cote A | 1.87 | 0.000 |
| Cote B | 0.76 | 0.101 |
| Cote C | 2.32 | 0.000 |
| Tension 1 | 2.14 | 0.200 |
| Tension 2 | 0.11 | 0.000 |
| Tension 3 | 4.31 | 0.002 |
| Tension 4 | 1.40 | 0.479 |
| Tension 5 | 1.56 | 0.263 |
| Tension 6 | 1.30 | 0.309 |
| Tension 7 | 1.49 | 0.079 |
| Fabricant 1 | 1.24 | 0.752 |
| Fabricant 2 | 0.61 | 0.513 |
| Fabricant 3 | 0.87 | 0.886 |
| Fabricant 4 | 0.43 | 0.028 |
| Fabricant 5 | 1.49 | 0.048 |
| Fabricant 6 | 1.06 | 0.753 |
| Fabricant 7 | 0.61 | 0.330 |
| Fabricant 8 | 1.25 | 0.719 |
| Fabricant 9 | 3.73 | 0.044 |
| Fabricant 10 | 1.22 | 0.590 |
| Fabricant 11 | 0.26 | 0.128 |
| Fabricant 12 | 0.64 | 0.173 |
| Fabricant 13 | 1.15 | 0.616 |
| Fabricant 14 | 1.22 | 0.204 |

Tableau 6.11: Résumé de l'impact des variables significatives pour la période terminant en 2018

| Variable | exp(Coefficient) | Pr(>z) | Analyse |
|----------------------------|------------------|--------|---|
| Nombre_Ordre Conditionnels | 1.06 | 0.000 | Plus il y a d'ordre de maintenance préventive sur un appareil, plus le taux de défaillance est élevé |
| Nombre_Ordre Inspection | 1.09 | 0.041 | Plus il y a d'ordre d'inspection sur un appareil, plus le taux de défaillance est élevé |
| CPC_Bin | 0.53 | 0.000 | Le taux de défaillance est plus élevé pour les appareils munis d'un CPC |
| Cote A | 1.87 | 0.000 | Les cotes A et C ont un impact sur le taux de défaillance: plus la cote est élevée, plus le taux de défaillance est élevé |
| Cote C | 2.32 | 0.000 | |
| Tension 2 | 0.11 | 0.000 | Les Transformateurs ayant une Tension 2 ont un plus bas taux de défaillance |
| Tension 3 | 4.31 | 0.002 | Les Transformateurs ayant une Tension 3 ont un plus haut taux de défaillance |
| Fabricant 4 | 0.43 | 0.028 | Les équipements du Fabricant 4 sont plus fiables que ceux du fabricant 0 |
| Fabricant 5 | 1.49 | 0.048 | Les équipements du Fabricant 5 sont moins fiables que ceux du fabricant 0 |
| Fabricant 9 | 3.73 | 0.044 | Les équipements du Fabricant 9 sont moins fiables que ceux du fabricant 0 |

Le

Tableau 6.11 permet de résumer l'impact de chaque variable sur la fonction de base du taux de défaillance. La colonne $exp(Coefficient)$ correspond au facteur exponentiel de β , qui a un effet multiplicatif sur le taux de défaillance $\lambda_i(t)$. La colonne $P_{r>z}$ permet de donner la signification statistique pour chaque variable β . Lorsque la valeur est inférieure à 5%, le coefficient de la variable est statistiquement différent de 0.

Tableau 6.12: Résumé des coefficients et intervalle de confiance pour le modèle AG

| factor. id | Variable | Intervalle de confiance | | | Pr(>z) |
|---------------|---------------------------------|-------------------------|------------------|--------------------|--------|
| | | Borne Inférieur | exp(Coefficient) | Borne Supérieur | |
| 1 | Nombre_Ordre_Conditionnels_ZI01 | 1.04 | 1.06 | 1.09 | 0.000 |
| 2 | Nombre_Ordre_Inspection_ZI03 | 1.03 | 1.09 | 1.16 | 0.041 |
| 4 | CPC_Bin | 0.43 | 0.53 | 0.64 | 0.000 |
| 5 | Cote A | 1.55 | 1.87 | 2.27 | 0.000 |
| 7 | Cote C | 1.97 | 2.32 | 2.73 | 0.000 |
| 9 | Tension 2 | 0.04 | 0.11 | 0.33 | 0.000 |
| 10 | Tension 3 | 1.64 | 4.31 | 11.35 | 0.002 |
| 18 | Fabricant 4 | 0.25 | 0.43 | 0.74 | 0.028 |
| 19 | Fabricant 5 | 1.14 | 1.49 | 1.96 | 0.048 |
| 23 | Fabricant 9 | 1.62 | 3.73 | 8.54 | 0.044 |

Le Tableau 6.12 présente les coefficients pour chacune des variables, ainsi que l'intervalle de confiance à un niveau de 95%. Bien que ces variables soient significatives (« *p-value* »), l'intervalle de confiance est assez grand, en particulier pour les variables catégorielles. Pour les variables continues, l'IC semble assez près du coefficient, ce qui suggère que le modèle performe mieux dans l'estimation de ce type de paramètres.

Les variables concernant les nombres d'ordres soulèvent plusieurs questionnements intéressants. Effectivement, il semble que plus il y ait de maintenance préventive et d'inspection, plus le nombre de défaillances observé est élevé. Normalement, un équipement entretenu plus fréquemment devrait être plus fiable. Le nombre de défaillances que subit un transformateur chaque année est très bas, et sa durée de vie s'étend sur des décennies. Qui plus est, l'effet de la maintenance n'est pas direct sur le taux de défaillance, mais plutôt diffère au fil du temps, en retardant l'usure globale du système. En conséquence, il est difficile de croire que les politiques de quatre années permettent de conclure à une augmentation ou même à une diminution du taux de

défaillance. La plupart des équipements observés sont en phase de vieillissement, ce qui fait en sorte qu'ils ont nécessairement plus de défaillance. De plus, les équipements vieillissants sont ceux qui nécessitent le plus d'activités de maintenance. Ainsi, plus la fréquence des défaillances augmente, plus la périodicité des entretiens augmente. Il en a de même pour la fréquence d'inspection, qui est plus élevée pour les équipements en phase de vieillissement.

Aussi, on remarque que les cotes d'impact ont un fort effet sur le taux de défaillance. Le jeu de donnée utilisé ne représente que la dernière cote attribuée, plutôt que l'évolution de ces cotes au fil du temps. Cela laisse donc croire que l'effet sur le taux de défaillance est fausement induit. Une évolution des cotes au fil du temps serait donc nécessaire pour observer si celles-ci influencent véritablement la fiabilité des équipements. Le seul constat possible actuellement, c'est que les transformateurs ayant une cote élevée en 2018 ont un taux de défaillance plus élevé dans la période d'étude. C'est un résultat qui est attendu; plus un équipement est vieillissant, plus ces cotes d'états devraient être élevées avec le temps. Les résultats du modèle démontrent que les transformateurs munis d'un CPC ont un taux de défaillance plus élevé que ceux qui n'en possèdent pas. Finalement, la fiabilité varie significativement pour certains niveaux de tension ainsi que certains fabricants. Pour interpréter les variables catégoriques, comme la tension et le fabricant par exemple, le modèle de Cox produit une estimation des paramètres, pour chaque niveau d'une variable. L'estimation se fait en comparaison du premier niveau de la variable. Ainsi, ce modèle démontre que les équipements du fabricant 5 sont moins fiables que ceux du fabricant 0 par exemple.

La Figure 6.5 résume les tests statistiques effectués pour valider le modèle. Il s'agit de tests permettant de vérifier la signification statistique globale du modèle. Chacun de ceux-ci évalue l'hypothèse nulle que toutes les covariables β sont égales à 0. Chacun des « *p-value* » est inférieur à 5%; l'hypothèse nulle est rejetée et permet de conclure que le modèle est significatif.

```

Concordance= 0.724 (se = 0.013 )
Likelihood ratio test= 534.5 on 30 df, p=<2e-16
Wald test              = 310.1 on 30 df, p=<2e-16
Score (logrank) test = 605.2 on 30 df, p=<2e-16, Robust = 110 p=5e-11

(Note: the likelihood ratio and score tests assume independence of
observations within a cluster, the wald and robust score tests do not).

```

Figure 6.5: Test statistique du modèle AG

Ensuite, la Figure 6.6 présente la courbe du risque cumulé en fonction du temps. Lorsque le taux de défaillance des équipements est constant, la fonction de risque cumulé est représentée graphiquement par une droite. Dans ce cas-ci, la courbe augmente légèrement au fil du temps, ce qui démontre que le risque de défaillance augmente, plus le temps à la dernière défaillance est grand. De plus, cela vient appuyer que le taux de défaillance ne soit pas constant, comme démontré par le modèle NHPP. La figure illustre aussi l'intervalle de confiance lié à l'estimation de la fonction de risque.

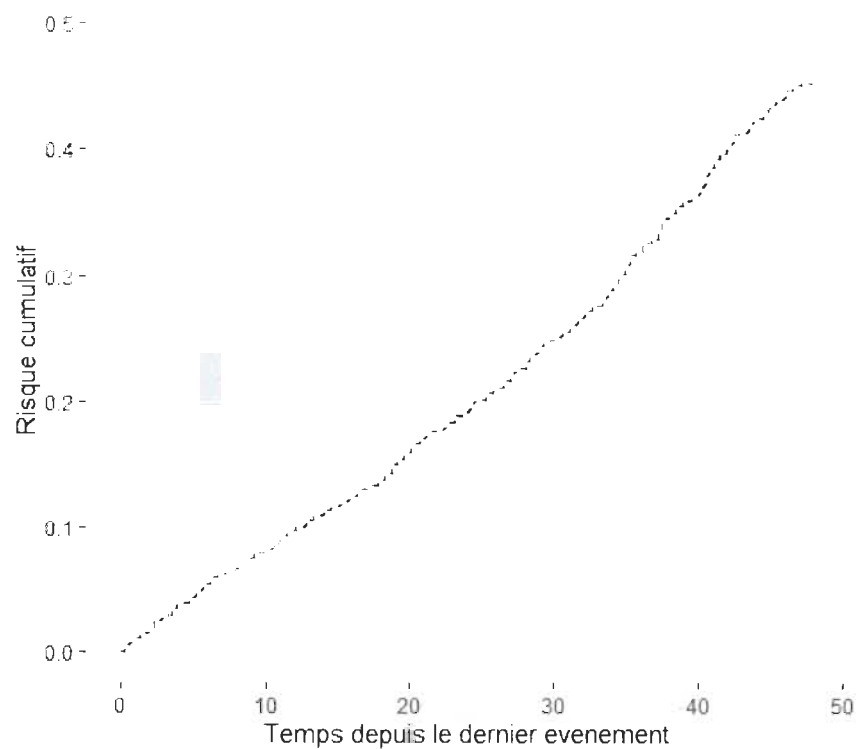


Figure 6.6: Graphique du risque cumulé pour le modèle AG

Modèle Prentice-Williams-Peterson

De la même façon que le modèle AG, le modèle PWP a été appliqué sur le nouveau jeu de données. Le Tableau 6.13 présente les résultats obtenus et le Tableau 6.14 présente une comparaison des résultats des deux modèles :

Tableau 6.13: Résumé de l'impact des variables significatives pour le modèle PWP

| Variable ID | Variable | exp(Coefficient) | Pr(>z) |
|--------------|------------------------------|------------------|--------|
| β_2 | Nombre_Ordre_Inspection_ZI03 | 1.0922 | 0.0020 |
| β_3 | Nombre_Avis_Conditionnels_I1 | 1.0418 | 0.0020 |
| β_4 | CPC_Bin | 0.6539 | 0.0001 |
| β_5 | Cote A | 1.6573 | 0.0000 |
| β_7 | Cote C | 1.8181 | 0.0000 |
| β_9 | Tension 2 | 0.1610 | 0.0001 |
| β_{10} | Tension 3 | 3.9280 | 0.0005 |
| β_{14} | Tension 7 | 1.3662 | 0.0454 |
| β_{19} | Fabricant 5 | 1.3749 | 0.0223 |
| β_{23} | Fabricant 9 | 2.3609 | 0.0482 |

Tableau 6.14: Comparaison entre les deux modèles

| Variable ID | Variable | AG | | PWP | |
|--------------|---------------------------|------------------|--------|------------------|--------|
| | | exp(Coefficient) | Pr(>z) | exp(Coefficient) | Pr(>z) |
| β_1 | Nombre Ordre Conditionnel | 1.06 | 0.000 | 1.0176 | 0.1209 |
| β_2 | Nombre Ordre Inspection | 1.09 | 0.041 | 1.0922 | 0.0020 |
| β_3 | Nombre Avis Conditionnels | 1.04 | 0.090 | 1.0418 | 0.0020 |
| β_4 | CPC_Bin | 0.53 | 0.000 | 0.6539 | 0.0001 |
| β_5 | Cote A | 1.87 | 0.000 | 1.6573 | 0.0000 |
| β_7 | Cote C | 2.32 | 0.000 | 1.8181 | 0.0000 |
| β_9 | Tension 2 | 0.11 | 0.000 | 0.1610 | 0.0001 |
| β_{10} | Tension 3 | 4.31 | 0.002 | 3.9280 | 0.0005 |
| β_{14} | Tension 7 | 1.49 | 0.079 | 1.3662 | 0.0454 |
| β_{18} | Fabricant 4 | 0.43 | 0.028 | 0.5678 | 0.0572 |
| β_{19} | Fabricant 5 | 1.49 | 0.048 | 1.3749 | 0.0223 |
| β_{23} | Fabricant 9 | 3.73 | 0.044 | 2.3609 | 0.0482 |

Le modèle PWP permet d'obtenir des résultats très près de ceux du modèle PWP. Effectivement, mis à part pour les variables du nombre d'avis et d'ordre conditionnel et pour un fabricant particulier, les mêmes covariables sont significatives à un seuil de 5%. En addition, les « *p-value* » des variables β_3 et β_{14} du modèle AG et β_{18} de PWP sont respectivement de 0.09, 0.079 et 0.0572, ce qui est près du seuil de signification. En outre, les coefficients de chacun des modèles sont assez similaires.

Tout comme le modèle AG, chacun des « *p-value* » est inférieur à 5% ce qui permet de conclure que le modèle est significatif, tel que démontré par la Figure 6.7.

| | | | | | |
|--|--|--|--|--|--|
| Concordance= 0.673 (se = 0.013) | | | | | |
| Likelihood ratio test= 224.3 on 30 df, p=<2e-16 | | | | | |
| wald test = 274.6 on 30 df, p=<2e-16 | | | | | |
| score (logrank) test = 230.6 on 30 df, p=<2e-16, Robust = 151.5 p=<2e-16 | | | | | |
| (Note: the likelihood ratio and score tests assume independence of observations within a cluster, the wald and robust score tests do not). | | | | | |

Figure 6.7: Test statistique du modèle PWP

Tableau 6.15: Résumé des coefficient et intervalle de confiance pour le modèle PWP

| factor. id | Variable | Intervalle de confiance | | | Pr(>z) |
|---------------|------------------------------|-------------------------|------------------|--------------------|--------|
| | | Borne Inférieur | exp(Coefficient) | Borne Supérieur | |
| β_2 | Nombre_Ordre_Inspection_ZI03 | 1.0286 | 1.0922 | 1.1598 | 0.0020 |
| β_3 | Nombre_Avis_Conditionnels_I1 | 1.0100 | 1.0418 | 1.0746 | 0.0020 |
| β_4 | CPC_Bin | 0.5346 | 0.6539 | 0.7997 | 0.0001 |
| β_5 | Cote A | 1.3675 | 1.6573 | 2.0084 | 0.0000 |
| β_7 | Cote C | 1.5388 | 1.8181 | 2.1483 | 0.0000 |
| β_9 | Tension 2 | 0.0534 | 0.1610 | 0.4848 | 0.0001 |
| β_{10} | Tension 3 | 1.4895 | 3.9280 | 10.3588 | 0.0005 |
| β_{14} | Tension 7 | 0.9919 | 1.3662 | 1.8819 | 0.0454 |
| β_{19} | Fabricant 5 | 1.0409 | 1.3749 | 1.8159 | 0.0223 |
| β_{23} | Fabricant 9 | 1.0440 | 2.3609 | 5.3391 | 0.0482 |

Le Tableau 6.15 présente les coefficients pour chacune des variables, ainsi que l'intervalle de confiance à un niveau de 95%. Tout comme pour le modèle AG, même lorsque la « *p-*

« *value* » démontre que l'effet de la variable est significatif, l'intervalle de confiance est assez grand pour les variables catégorielles. Par contre, l'estimation semble meilleure qu'avec le modèle AG, puisque cet écart semble moins important. Pour ce qui est des variables continues, l'IC semble aussi être assez près de du coefficient.

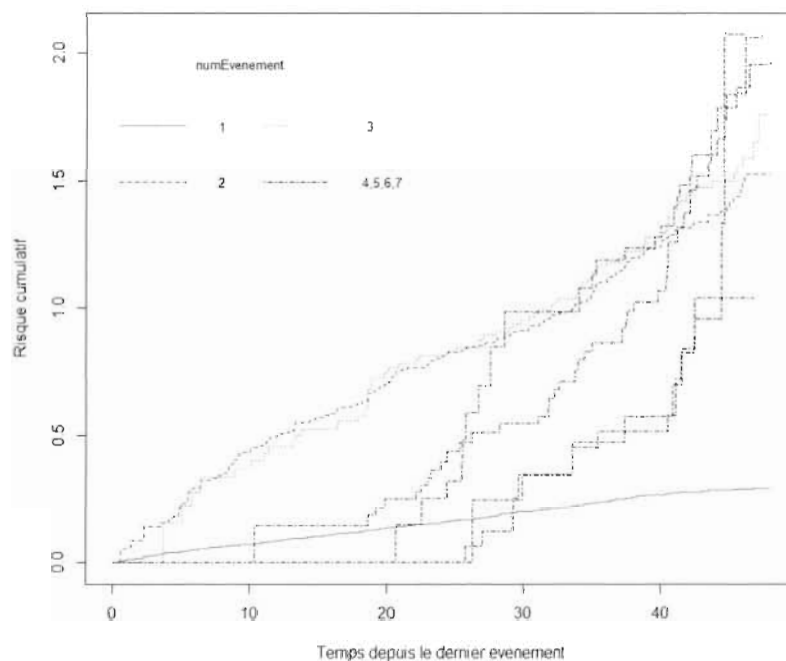


Figure 6.8: Graphique de la fonction de risque cumulative pour le modèle PWP

Le graphique du risque cumulatif pour le modèle PWP illustre les courbes de risque pour chaque strate. Les courbes montrent entre autres que le temps avant le premier événement (en rouge) est généralement plus long que les événements subséquents.

En outre, les résultats démontrent que les modèles de survie sont appropriés pour déterminer des covariables affectant la fiabilité des transformateurs. Les caractéristiques d'équipements permettent notamment d'ajouter de l'information importante qui permettra de prendre des décisions adaptées. De plus, le coefficient des covariables peut être directement appliqué à la fonction d'intensité calculée par le modèle NHPP (voir Tableau 2.3). De cette façon, le taux de défaillance prend en compte l'âge des équipements et l'ensemble des covariables significatives. Le Tableau 6.16 montre les paramètres d'une telle fonction d'intensité pour le modèle AG :

Tableau 6.16: Paramètre de la fonction d'intensité

| Paramètre | Variable | Valeur |
|--------------|---------------------------------|-------------------------|
| $\lambda(t)$ | Fonction d'intensité | $1.338*0.054*t^{0.339}$ |
| β_1 | Nombre_Ordre_Conditionnels_ZI01 | 1.06 |
| β_2 | Nombre_Ordre_Inspection_ZI03 | 1.09 |
| β_4 | CPC_Bin | 0.53 |
| β_5 | Cote A | 1.87 |
| β_7 | Cote C | 2.32 |
| β_9 | Tension 2 | 0.11 |
| β_{10} | Tension 3 | 4.31 |
| β_{18} | Fabricant 4 | 0.43 |
| β_{19} | Fabricant 5 | 1.49 |
| β_{23} | Fabricant 9 | 3.73 |

CONCLUSION

Ce projet de recherche avait pour objectif d'appliquer de nouvelles méthodes statistiques, afin d'améliorer la compréhension des facteurs qui influencent la fiabilité des transformateurs de puissance. L'étude de cas fourni par Hydro-Québec a permis de modéliser la fiabilité des équipements de transport d'électricité à partir de données réelles, comportant des biais et des incertitudes.

La première étape du projet a été d'analyser le jeu de données, dans l'intention d'effectuer un nettoyage de ces dernières. Ensuite, des modèles spécifiques aux systèmes réparables ont été appliqués. Le modèle NHPP a permis de déterminer un taux de défaillance, avec l'utilisation de la méthode de la « *Mean Cumulative Function* ». Les modèles Andersen-Gill et Prentice-Williams-Peterson ont été testés, afin de déterminer les cofacteurs qui affectent le taux de défaillance.

Le nettoyage des données a permis de diminuer les incertitudes relatives à leur qualité. Le modèle NHPP a, quant à lui, permis d'obtenir rapidement les équations de fiabilité prenant en compte les spécificités des systèmes réparables. De plus la méthode de la MCF traite les censures, inhérente au jeu de données d'HQ. L'application de ces méthodes permettra d'avoir une estimation plus précise de la fonction d'intensité, et ainsi améliorer la gestion des actifs. De plus, l'application de modèle de survie a donné de l'information supplémentaire; les modèles ont démontré que certaines caractéristiques d'équipement avaient un effet significatif sur le taux de défaillance. Ces résultats démontrent l'applicabilité de ces méthodes et permettront de mettre en place des politiques de GDA plus ciblé.

RECOMMANDATIONS

À partir des modèles, il est possible de formuler plusieurs recommandations, que ce soit par rapport à la qualité des données, aux variables, aux résultats ou à l'applicabilité même des modèles.

Qualité des données

Comme le démontre l'analyse préliminaire, la qualité des données est intrinsèque à la qualité du modèle. La duplication du nom de fabricant, par exemple, mène à des erreurs d'évaluation des coefficients, ce qui engendre une fausse interprétation. Afin de faciliter l'utilisation de ces modèles par l'équipe PGA, il serait important d'appliquer les recommandations suivantes :

- Appliquer les règles et les filtres en amont du traitement des données
 - Les données doivent être bien identifiées selon le type d'ordre et d'avis
 - Les caractéristiques d'équipements doivent être uniformisées

Il est important d'effectuer ces actions en amont; il faut que tous les analystes aient accès aux bonnes données (et aux mêmes données), pour s'assurer que les résultats sont valides et reproductibles. Ensuite, afin d'éliminer les problèmes à la source, il importe d'adopter des mesures préventives en augmentant la formation et la sensibilisation des intervenants qui crée ces données. Effectivement, comme le montre le Tableau 6.9, la codification des types d'ordres ne semble pas tout à fait maîtrisée.

Variables et modèles

En ce qui concerne les variables, il serait intéressant de recueillir les données sur les cotes de santé et sur leur évolution au fil du temps. Cela permettrait d'identifier une possible adéquation entre l'évolution de ces indicateurs et la fiabilité des équipements. Qui plus est, cette étude ne considère que les données de 2014 à 2018. Il serait donc avantageux de récupérer les données des systèmes antérieurs, dans le but d'obtenir un historique plus complet des politiques de maintenance. Les données actuelles permettent donc difficilement de déterminer l'effet de la maintenance sur la fiabilité, puisque l'impact

d'une politique de maintenance a un effet différé sur la fiabilité des équipements. Pour ce qui est du modèle NHPP, la période d'étude permet tout de même d'arriver à de bons résultats et que le modèle réussit à bien gérer les censures. Les modèles de survie offrent de bons résultats pour identifier les variables relatives aux caractéristiques d'équipement. Par contre, les variables qui portent sur l'effet des politiques de maintenance ainsi que les variables d'état ne permettent pas d'obtenir des résultats cohérents. Effectivement, le changement d'une politique de maintenance a un effet différé sur la fiabilité des équipements. Une période aussi courte (relativement à la vie de l'équipement), ne permet pas aux modèles de déceler ces changements. Pour que ces modèles perçoivent ces changements de façon cohérente, il serait important d'étendre la période d'étude, afin d'en tirer les bonnes conclusions.

Prochaines étapes

Tel que présenté à la Figure 3.5, l'application de ces méthodes statistiques nécessite plusieurs itérations, afin d'obtenir de bons résultats à long terme. Ce projet a présenté les premières itérations et mis en évidence les prochaines étapes à mettre en œuvre. Dans un premier temps, l'organisation devrait mettre les efforts pour améliorer la qualité des données. L'étude réalisée par Berti-Équille (2018) offre une méthodologie complète, qui gagnerait à être appliquée. La qualité doit être améliorée et uniformisée en amont du processus de traitement. De plus, il est indispensable que les données soient bien étiquetées; il faut bien distinguer chaque activité de maintenance, afin d'inférer correctement l'effet qu'elles ont sur la fiabilité. Qui plus est, il faut récupérer un maximum d'information en retraçant les données d'ancien système, pour avoir l'historique le plus complet possible. Enfin, il faut considérer les données comme un actif de haute importance. Après tout, la qualité des données est garante d'une prise de décision éclairée. Une fois que ces activités seront en place, l'application des modèles apportera encore plus d'information pour éclairer la prise de décision.

LISTE DE RÉFÉRENCES

- Amadi-Echendu, J. E., Brown, K., Willett, R., & Mathew, J. (2011). *Definitions, Concepts and Scope of Engineering Asset Management*. London: Springer-Verlag London.
- Amadi-Echendu, J. E., & Mafutsana, J. M. (2016). A bibliographic review of trends in design and management of electrical power transmission transformers. *Proceedings of the International Conference on Industrial Engineering & Operations Management*, 2010-2018.
- Amorim, L. D., & Cai, J. (2014). Modelling recurrent events: a tutorial for analysis in epidemiology. *International Journal of Epidemiology*, 44(1), 324-333. <http://dx.doi.org/10.1093/ije/dyu222>
- Berti-Équille, L. (2018). Qualité des données. *Techniques de l'ingénieur Bases de données, base documentaire : TIB309DUO*(ref. article : h3700).
- Carneiro, J. C. (2013, 13-17 May 2013). *Substation power transformer risk management: Predictive methodology based on reliability centered maintenance data*. Communication présentée 4th International Conference on Power Engineering, Energy and Electrical Drives.
- Davis, J. (2007). What is asset management and where do you start? *Journal: American Water Works Association*, 99(10), 26-34.
- Gámiz, M. L., Kulasekera, K. B., Limnios, N., & Lindqvist, B. H. (2011). *Applied Nonparametric Statistics in Reliability*. London: Springer London, London.
- Gorjian, N., Ma, L., Mittinty, M., Yarlagadda, P., & Sun, Y. (2010). A review on reliability models with covariates *Engineering Asset Lifecycle Management* (pp. 385-397): Springer.
- Ho, M., Hodkiewicz, M. R., Pun, C. F., Petchey, J., & Li, Z. (2015). *Asset Data Quality—A Case Study on Mobile Mining Assets*. Communication présentée, Cham.
- Hydro-Québec. (2019a). Appareils de mesure : pour la qualité et la sécurité. *Comprendre l'électricité*. Page consultée à Appareils de mesure : pour la qualité et la sécurité

- Hydro-Québec. (2019b). En route vers la transformation. *Comprendre l'électricité*. Page consultée à http://www.hydroquebec.com/comprendre/transport/construction-poste/pop_up_poste3.html#puissance
- Hydro-Québec. (2019c). Transport de l'électricité: Le transport de l'électricité au Québec. *Comprendre l'électricité*. Page consultée à <http://www.hydroquebec.com/comprendre/transport/grandes-distances.html>
- Hydro-Québec. (2019d). Transport de l'électricité:Fonctionnement d'un poste. *Comprendre l'électricité*. Page consultée à <http://www.hydroquebec.com/comprendre/transport/construction-poste/>
- IAM. (2015). Asset Management - an anatomy. *Institute of Asset Management, V3*.
- International Organisation for Standardization. (2015). ISO 55000, 55001 and 55002 Asset Management Standards. BSI Standards.
- internationale, C. é. (2002). Transformateurs de mesure *Partie 1: Transformateurs de courant* (Vol. 2.1.2). Genève,Suisse: CEI.
- internationale, C. é. (2003). Transformateurs de mesure *Partie 2: Transformateurs inductifs de tension* (Vol. 2.1.1). Genève,Suisse: CEI.
- internationale, C. é. (2011). Transformateurs de puissance *Partie 1 : Généralités* (Vol. 3.1.1). Genève,Suisse: CEI.
- King, J. P., & Jewett, W. S. (2010). *Robustness development and reliability growth: Value adding strategies for new products and processes*: Pearson Education.
- Lazzaroni, M., Cristaldi, L., Peretto, L., Rinaldi, P., & Catelani, M. (2011). *Reliability Engineering: Basic Concepts and Applications in ICT*. Berlin, Heidelberg: Berlin, Heidelberg: Springer Berlin Heidelberg.
- Leemis, L. M. (2009). *Reliability: Probabilistic Models and Statistical Methods* (2 éd.): Lawrence M. Leemis.

- Levitt, J. (2011). 11.1 What is Predictive Maintenance? *Complete Guide to Preventive and Predictive Maintenance (2nd Edition)*: Industrial Press.
- Lewis, E. E. (1996). *Introduction to reliability engineering* (2nd ed. éd.). New York, N.Y.: J. Wiley and Sons.
- Murphy, G. D. (2009). Improving the quality of manually acquired data: Applying the theory of planned behaviour to data quality. *Reliability Engineering & System Safety*, 94(12), 1881-1886.
<http://dx.doi.org/https://doi.org/10.1016/j.res.2009.05.008>
- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, 47(1), 90-100.
- Nelson, W. B. (2003). *Recurrent events data analysis for product repairs, disease recurrences, and other applications* (Vol. 10): SIAM.
- O'Connor, P. D. T., & Kleyner, A. (2011). *Practical Reliability Engineering*. Chichester, UK: Chichester, UK: John Wiley & Sons, Ltd.
- Percy, D. F., & Alkali, B. M. (2006). Generalized proportional intensities models for repairable systems. *IMA Journal of Management Mathematics*, 17(2), 171-185.
<http://dx.doi.org/10.1093/imaman/dpi034>
- Prasad, K. H., Faruque, T. A., Joshi, S., Chaturvedi, S., Subramaniam, L. V., & Mohania, M. (2011, 29 March-2 April 2011). *Data Cleansing Techniques for Large Enterprise Datasets*. Communication présentée 2011 Annual SRII Global Conference.
- Rausand, M., & Høyland, A. (2003). *System reliability theory: models, statistical methods, and applications* (Vol. 396): John Wiley & Sons.
- Reliability, O. o. E. D. a. E. (2012). Large Power Transformers And The U.S. Electric Grid. Page consultée à https://www.energy.gov/sites/prod/files/Large%20Power%20Transformer%20Study%20-%20June%202012_0.pdf

- Rexhepi, V. (2017). An Analysis of Power Transformer Outages and Reliability Monitoring. *Energy Procedia*, 141, 418-422. <http://dx.doi.org/https://doi.org/10.1016/j.egypro.2017.11.053>
- Shwu-Tzy, J., Landers, T. L., & Rhoads, T. R. (2006). Assessment of repairable-system reliability using proportional intensity models: a review. *IEEE Transactions on Reliability*, 55(2), 328-336. <http://dx.doi.org/10.1109/TR.2006.874938>
- Singh, A. K., Sinha, A., & Joshi, D. (2016). Asset Management Strategy to Improve Substation Performance. *Journal of Pure & Applied Science & Technology*, 6(2), 7-13.
- Technology, N. I. o. S. A. (2003). *NIST/SEMATECH e-Handbook of Statistical Methods Engineering statistics handbook*. National Institute of Standards and Technology. Repéré à <http://www.itl.nist.gov/div898/handbook/index.htm>
- Therneau, T., & Grambsch, P. (2000). *Modeling Survival Data: Extending The Cox Model* (Vol. 48). New York: Springer.
- Velasquez-Contreras, J. L., Sanz-Bobi, M. A., & Galceran Arellano, S. (2011). General asset management model in the context of an electric utility: Application to power transformers. *Electric Power Systems Research*, 81(11), 2015-2037. <http://dx.doi.org/https://doi.org/10.1016/j.epsr.2011.06.007>
- Wildi, T., & Sybille, G. (2005). *Électrotechnique* (4e éd. éd.). Sainte-Foy: Presses de l'Université Laval.