



UTILISATION DE L'INTELLIGENCE ARTIFICIELLE POUR LE REGROUPEMENT DES SEGMENTS DE CHAUSSEE DONT LE COMPORTEMENT EST SIMILAIRE

Projet R783.1

Rapport final

Chaussées en enrobé ayant subies une intervention palliative
et chaussées mixtes et en béton de ciment

Chargée de projet

Josianne Ouellette, Responsable du secteur statistique et
traitement des données, Direction du Laboratoire des chaussées,
Ministère des Transports, de la Mobilité durable et de
l'Électrification des transports Tél. : 418 644-0890, poste 24051
Courriel : josianne.ouellette@transports.gouv.qc.ca

Chercheurs

Mohamed-Salah Ouali et Soumaya Yacout, Professeurs titulaires,
Département de mathématiques et de génie industriel,
École Polytechnique de Montréal
Tél. : 514 340-4711, postes 5929 et 4517
Courriels : msouali@polymtl.ca, soumaya.yacout@polymtl.ca

AVERTISSEMENT

La présente étude a été réalisée à la demande du ministère des Transports, de la Mobilité durable et de l'Électrification des transports du gouvernement du Québec et a été financée par la Direction de l'environnement et de la recherche.

Les opinions exprimées dans le présent rapport n'engagent que la responsabilité de leurs auteurs et ne reflètent pas nécessairement les positions du ministère des Transports, de la Mobilité durable et de l'Électrification des transports du gouvernement du Québec.

Ce rapport est **CONFIDENTIEL**.

Collaborateurs

Luc Adjengué, Professeur agrégé, Département de mathématiques et de génie industriel, École Polytechnique de Montréal.

Ahmed Elsheikh, Étudiant au doctorat, École Polytechnique de Montréal.

Karim Yacout, Travailleur autonome, École Polytechnique de Montréal.

Mohamed Reda Tsouli, Travailleur autonome, École Polytechnique de Montréal.

Ahmed Ragab, Stagiaire post-doctoral, École Polytechnique de Montréal.

Daniel Buades-Marcos, Étudiant en échange, École Polytechnique de Montréal.

SOMMAIRE

Le projet s'intéresse à la problématique de regroupements de segments de chaussée dont le comportement est similaire à l'aide de l'intelligence artificielle. Il vise, spécifiquement, la recherche des profils qui décrivent le mieux les segments de chaussée ayant un comportement semblable et l'appréciation de ce comportement au cours du temps. Chaque segment de chaussée est décrit par : un identifiant unique, plusieurs indices de performance (IP), un ensemble de facteurs caractéristiques et son âge au moment de l'observation.

Plusieurs modèles sont développés pour établir les regroupements de segments. Ils exploitent plusieurs mesures de similarité partant de la recherche des facteurs qui discriminent une famille de segments d'une autre famille à un âge donné aux calculs de distances multidimensionnelles permettant de regrouper des segments ayant un comportement semblable selon la séquence d'observations, d'âges ou sous-séquence d'âges. Ces modèles sont utilisés pour toutes les catégories de chaussée à l'étude.

Les profils obtenus des familles de segment par IP et par âge sont bien identifiés et les plages de variation des facteurs caractéristiques sont bien définies. Ces profils sont décrits par une conjonction de facteurs pour discriminer un regroupement de segments d'un autre, à chaque âge. Les regroupements de segments ayant un comportement similaire à travers les âges sont établis à l'aide de plusieurs modèles.

Les résultats obtenus par le regroupement dynamique exploitant les pentes des indices de performance et la sous-séquence d'âges sont les plus probants. Ces résultats démontrent la validité conceptuelle des modèles proposés.

Par ailleurs, trois autres problèmes sont étudiés et solutionnés. Il s'agit de la recherche des profils de regroupements à travers les âges. Les règles de classification utilisent en majorité un seul facteur discriminant. Ce qui montre que les familles sont bien distinctes. Les modèles de prédiction de l'âge de revêtement obtenus permettent de classer les segments dont les âges sont inconnus avec un degré d'appartenance assez élevé comparativement à d'autres modèles communément utilisés dans le domaine. Enfin, l'identification des segments présentant des anomalies particulières montre que toutes les catégories de chaussée sont touchées à différents niveaux par au moins un des 8 types d'anomalies retracés. L'IRI est le plus touché par des faibles valeurs IP à jeune âge suivi de IP-Orniérage et indice de performance stable au cours du temps. Les pourcentages moyens des autres types d'anomalies par catégorie de chaussée restent en moyenne très faibles.

Au terme de ce projet, cinq recommandations sont énoncées afin de permettre la validation opérationnelle des solutions proposées : installer un mécanisme pour s'assurer davantage de l'intégrité des données ; trouver un mécanisme automatique de calibration de l'âge zéro et de la valeur IP à 100 à chaque réhabilitation ; arrimer les données de maintenance et les valeurs des IPs afin de mieux expliquer la hausse ou la stabilité des valeurs IPs d'un âge à l'autre ; analyser chaque facteur et ne retenir que celui qui caractérise le phénomène physique représenté la valeur IP ; et enfin, considérer, lorsque cela est possible, des facteurs qui expriment des écarts entre les valeurs de conception et celles mesurées en exploitation.

Comme perspective, il est important de calibrer les modèles. Cela permet de profiter pleinement des résultats du projet actuel. La réalisation d'un outil intégré d'aide à la décision constitue le meilleur cadre de développement futur de ce projet.

TABLE DES MATIÈRES

1.	INTRODUCTION.....	1
1.1	Objectifs de l'étude.....	1
1.2	Résumé des mandats.....	1
1.3	Problématique de l'étude.....	2
1.4	Solutions proposées.....	3
1.5	Résumé des rapports intermédiaires.....	6
1.6	Organisation du rapport final.....	7
2.	SYNTHÈSE DES RÉSULTATS OBTENUS.....	8
2.1	Traitement des données.....	8
2.2	Profils de familles de segments par âge.....	8
2.3	Comportements de familles de segments à travers les âges.....	10
2.4	Profils des regroupements à travers les âges.....	12
2.5	Prédiction de l'âge de revêtement.....	13
2.6	Identification des anomalies.....	14
3.	CONCLUSIONS ET RECOMMANDATIONS.....	16
3.1	Principales conclusions.....	16
3.2	Principales recommandations.....	18
3.3	Perspectives de développement.....	19
4.	SYNTHÈSE DES CONCLUSIONS ET RECOMMANDATIONS (750 MOTS) .	20

GLOSSAIRE

MTMDET : Ministère des Transports, de la Mobilité durable et de l'Électrification des transports

GCH : Système ministériel de gestion des chaussées

IP : Indice de performance de la chaussée

: Numéro

Nb : Nombre

1. INTRODUCTION

1.1 Objectifs de l'étude

Le système de gestion des chaussées (GCH) du Ministère des Transports, de la Mobilité Durable et de l'Électrification des Transports (MTMDET) utilise des modèles empiriques de comportement, regroupés en familles de chaussées, pour simuler la dégradation du réseau routier. Ces modèles servent à établir le bilan annuel de l'état du réseau et à estimer les budgets d'intervention selon différents scénarios d'investissements. Certaines de ces familles de chaussées ne sont pas assez bien définies par leurs profils (classes fonctionnelles, DJMA, type de surface, type de réhabilitation, etc.) pour discriminer adéquatement le comportement des chaussées.

Le MTMDET souhaite approfondir l'étude des familles de segments afin de réduire la variabilité des indices de performance (IP) dans une même famille en fonction de l'âge et du profil de la chaussée. Ainsi, le projet s'intéresse à la problématique de regroupements de segments de chaussée dont le comportement est similaire à l'aide de l'intelligence artificielle. Il vise, spécifiquement, la recherche des profils qui décrivent le mieux les segments de chaussée ayant un comportement semblable et l'appréciation de ce comportement au cours du temps. Chaque segment de chaussée est décrit par : un identifiant unique, plusieurs indices de performance, un ensemble de facteurs caractéristiques et son âge au moment de l'observation.

Plus spécifiquement, l'étude poursuit deux principaux objectifs :

- Établir les profils qui caractérisent les familles de segments ayant un comportement semblable à un âge donné. Ce profil est défini par les facteurs caractéristiques des segments se trouvant au même âge et appartenant à une classe d'IP donnée ;
- Appréhender le comportement des familles de segments à travers les âges. Il s'agit particulièrement d'effectuer des regroupements de segments en tenant compte de la tendance des valeurs IP de chaque segment à travers les âges.

1.2 Résumé des mandats

Deux principaux mandats sont formulés. Le premier mandat concerne l'élaboration d'une méthodologie de classification, basée sur l'analyse logique des données (LAD), permettant d'identifier, à un âge donné, les facteurs caractéristiques qui discriminent une famille de segments d'une classe IP par rapport aux segments des autres classes. Le deuxième mandat concerne le développement d'une approche permettant de regrouper de manière dynamique les segments de chaussée ayant un comportement similaire selon la tendance des valeurs IP à travers les âges.

Le premier mandat est demeuré le même tout au long du projet. Alors que le second mandat s'est ajusté au fur et à mesure que le projet avance afin d'aboutir à des résultats concordants avec les attentes de l'équipe du projet.

1.3 Problématique de l'étude

L'étude porte sur 16 702 segments de chaussée répartis en huit catégories de chaussée en enrobé C1 – neuve (17%), C2 – ayant subi un décohesionnement (6%), C3 – retraitée en place (1%), C4 – resurfacée (64%), C5 - recyclée/renforcée (5%), C6 - ayant subi une intervention palliative (3%), C7 – en béton de ciment (1%) et C8 – mixte (3%). Les huit catégories sont regroupées en quatre groupes : G1 (C1, C2 et C3), G2 (C4 et C5), G3 (C6) et G4 (C7 et C8).

Chaque segment de chaussée est décrit par 14 IPs calculés à partir des mesures de profilage de la chaussée et 39 facteurs caractéristiques relevant de la conception de la chaussée, de son exploitation et de sa maintenance. Les observations couvrent les années de 2000 à 2015. Le nombre d'enregistrements par segment varie entre 1 et 187 avec un mode de 8.

La valeur numérique de chaque IP est mesurée sur une échelle décroissante de 100 à 0. L'échelle est subdivisée en 2 à 5 classes, la classe 0 correspond au meilleur état de la chaussée. L'étendue de chaque classe dépend de la « Classe fonctionnelle » de la chaussée. Pour les besoins de l'étude, 6 IPs sont retenus après analyse de tous les IPs disponibles : IP1(IRI), IP2(Ornéage), IP3(Fissuration globale), IP5(Delta-IRI), IP7.1(Fissuration fatigue) et IP9(Fissuration globale).

Pour établir les profils de familles de segments de chaussée, une procédure de découpage et d'extraction des segments de chaussée, catégorie par catégorie, IP par IP, âge par âge est appliquée. Ce découpage est consigné dans des fichiers séparés de simulation. Pour l'ensemble des 8 catégories de chaussée, 1824 fichiers de simulation sont extraits. Ce nombre est assez grand vu l'étendu des âges de segments allant de 0 à plus de 50 ans.

Pour l'ensemble des fichiers extraits, des procédures de manipulation des données et traitement des données manquantes sont appliquées. Ces procédures permettent de supprimer les facteurs redondants ou non pertinents, fusionner ou manipuler certains facteurs et traiter les données manquantes. Ainsi, les fichiers obtenus contiennent les facteurs pertinents à l'étude sans aucune valeur manquante. Les facteurs retenus sont sensiblement les mêmes pour chaque catégorie de chaussée. La liste compte 25 ou 26 facteurs répartis entre facteurs de construction (17 ou 18), de maintenance (4) et d'exploitation (4).

Par ailleurs, une seconde extraction des segments de chaussée, catégorie par catégorie, IP par PI, tous les âges est faite pour appréhender le comportement des familles de segments à travers les âges. Un total de 37 fichiers sont extraits pour accomplir ce second objectif. Pour les catégories C7 et C8, seulement les fichiers pour IP1 et IP2 sont conservés. Les fichiers des IPs restants présentent plusieurs valeurs manquantes.

Ainsi, pour chaque catégorie de segments, deux principaux problèmes sont considérés dans l'étude :

- **Problème 1 - Regroupement de segments selon l'âge.** Le premier consiste à exploiter les profils des segments ayant le même âge pour différencier ceux qui appartiennent à une même classe d'IP par rapport aux autres segments se trouvant dans une classe IP différente. Ce problème est considéré à chaque

âge de segments. L'hétérogénéité des profils des segments ou le pouvoir discriminant de chaque facteur caractéristique d'un profil est un critère très important dans le regroupement des segments à un âge donné.

- **Problème 2 – Regroupement de segments à travers les âges.** Le second concerne le regroupement de segments ayant un comportement similaire à travers les âges en utilisant les valeurs IP des segments. Les valeurs IPs enregistrées au cours du temps peuvent donner plusieurs tendances à la baisse, représentant plusieurs allures de dégradation de la chaussée, ou à la hausse, représentant les effets des interventions de maintenance. L'arrimage des comportements de segments sur une longue période de temps avec des enregistrements partiels des valeurs IPs constitue le principal défi de ce problème.

Au cours de l'avancement du projet, trois autres problèmes spécifiques sont également considérés dans l'étude :

- **Problème 3 - Profils des regroupements à travers les âges.** Ce problème considère les regroupements de segments ayant un comportement similaire selon les valeurs IPs à travers les âges. Il cherche à identifier les facteurs caractéristiques qui expliquent ou différencient un regroupement de segments d'un autre. La résolution de ce problème est tributaire du nombre de regroupements trouvés pour chaque catégorie de segments.
- **Problème 4 - Prédiction de l'âge de revêtement.** Ce problème concerne la prédiction de l'âge de revêtement d'un segment en fonction des différents IPs et des facteurs caractéristiques de chaque segment. Comme l'âge de revêtement est une variable indispensable pour effectuer les regroupements et que certains segments peuvent avoir des âges entachés d'erreurs ou non actualisés, il est important d'offrir une alternative pour corriger automatiquement ces âges. Pour une catégorie de segments, il s'agit de trouver les variables qui discriminent une classe d'âges par rapport à une autre en utilisant les IPs et les facteurs explicatifs.
- **Problème 5 - Identification des anomalies.** Ce problème concerne l'identification des segments présentant des anomalies particulières suite à l'étude des données brutes ou des résultats de regroupements. Huit principaux types d'anomalies sont répertoriées. Ils concernent les segments ayant : une faible valeur IP à jeune âge; une forte valeur IP à vieil âge; une hausse d'IP sans maintenance ou changement de catégorie; une tendance positive d'IP au cours du temps, une baisse intempestive d'IP au cours du temps; un IP stable au cours du temps; différentes valeurs IPs à un même âge; et enfin les segments solitaires.

1.4 Solutions proposées

Pour établir les regroupements de segments âge par âge (problème 1) et à travers les âges (problème 2), sept solutions sont développées et testées. Les solutions 1 et 2 adressent le problème 1. Les cinq autres solutions concernent le problème 2. Toutes les solutions sont décrites ci-après ainsi que leurs mises en application.

Solution 1 - Utilisation de l'analyse logique des données. Cette solution est basée sur les concepts de Logical Analysis of Data (LAD). Elle met en œuvre un processus d'optimisation combinatoire basé sur la logique booléenne qui permet d'extraire des connaissances à partir des observations sous forme de règles d'association « pattern ». Le logiciel *cbmLAD* est utilisé pour déterminer les facteurs qui distinguent une famille de segments des autres à un âge donné, et ainsi effectuer les regroupements pour chaque catégorie de chaussée. Cette solution exploite la classe IP et les facteurs caractéristiques retenus, selon l'âge des segments;

Solution 2 - Utilisation de la classification ascendante hiérarchique. Cette solution se base sur le calcul de similarité par la distance euclidienne qui sépare les familles de segments. Le critère d'association de deux familles est donné par une méthode d'agrégation qui minimise la variance entre les deux familles, donc ayant le minimum de distance entre elles. Elle exploite uniquement les facteurs caractéristiques des segments, âge par âge, sans égard à la valeur IP ou à sa classe. Un algorithme de classification itératif est développé et implémenté à l'aide du langage R.

Solution 3 - Regroupement selon les valeurs IPs et la séquence d'observations. Cette solution applique le principe de regroupement dynamique de segments selon la séquence d'observations des valeurs IP au cours du temps. Un modèle de programmation dynamique, appelée Dynamic Time Warping (DTW) est programmé afin de regrouper les segments ayant un comportement similaire de leurs valeurs IPs d'une observation à l'autre. Plus précisément, cette solution exploite les valeurs numériques de l'IP et la séquence de leurs observations à travers les âges des segments, sans égard à l'âge de chaque segment, ni à ses facteurs caractéristiques. À remarquer qu'ici, la chronologie temporelle est donnée par la séquence des observations.

Solution 4 - Regroupement selon les pentes IPs et la séquence d'âges. Cette solution est une première extension de la solution précédente (solution 3). Elle permet de regrouper les segments ayant des valeurs de pentes très voisines les unes des autres selon la séquence d'âges. Dans cette solution, la chronologie temporelle est donnée par la séquence d'âges et non pas la séquence d'observations. Ce qui la rend plus sensible aux écarts d'âges entre les segments. De plus, les valeurs des pentes sont calculées segment par segment en faisant le rapport entre la différence des valeurs IPs et la différence des âges associée à ces mêmes valeurs IPs. Il s'agit d'utiliser la fonction dérivée de l'IP par rapport à l'âge. Ensuite, les regroupements sont établis de façon à minimiser la distance entre les valeurs des pentes au cours du temps.

Solution 5 - Regroupement selon les écarts IPs et la séquence d'âges. C'est une deuxième extension de la solution 3. Elle utilise uniquement l'écart des valeurs IPs d'un segment d'un âge à un autre afin d'effectuer les regroupements de segments présentant des écarts similaires selon la séquence d'âges. Cette solution partage tout d'abord le même principe que la solution 4 par rapport à la séquence d'âges. Toutefois, les regroupements sont faits de façon à minimiser la distance des segments ayant des écarts similaires au cours du temps.

Solution 6 - Regroupement selon les pentes IPs, les valeurs IPs à l'origine et la séquence d'âges. Cette solution peaufine le regroupement des segments ayant une tendance de la dérivée d'IP au cours du temps (solution 4) afin de considérer les valeurs IPs à l'origine. Cette amélioration permet de comparer les segments ayant des durées de vie comparables. Des segments peuvent présenter une même tendance de dégradation au cours du temps sans qu'ils aient des durées de vie comparables.

Solution 7 - Regroupement selon les pentes IPs la sous-séquence d'âges. Cette solution considère des sous-séquences d'âges pour regrouper les segments ayant des valeurs de pentes d'IP très voisines les unes des autres. Plus, spécifiquement, elle considère des portions de pentes qui peuvent être évaluées pour un seul ou plusieurs segments à des moments différents. Ainsi, un segment est regroupé avec son similaire s'ils possèdent des pentes similaires même s'ils ne partagent pas la même séquence d'âges. Cette solution permet de réduire le nombre de regroupements de segments. Ce qui facilite leur exploitabilité par le GCH du MTMDET.

Les solutions 3 à 7, proposées pour la résolution du problème de regroupement de segments ayant un comportement similaire au cours du temps (problème 2), sont programmées à l'aide du langage Python.

Les résultats de chacune des solutions précédentes sont consignés dans des fichiers « Excel » listant les regroupements de segments, famille par famille. Ces listes sont facilement exploitables. Elles utilisent l'identifiant unique ID de chaque segment afin de pouvoir retracer son appartenance âge par âge ou selon les âges. De plus, des fichiers « Texte » sont édités, présentant l'ensemble des profils discriminant chaque famille d'une autre selon la solution 1. Des représentations sous forme de Dendrogramme sont obtenus par la solution 2. Des fichiers « Excel » et « Image » sont produits, donnant les listes des familles de segments et montrant la représentation graphique des comportements des segments de chaque famille selon les solutions 3 à 7.

En ce qui concerne la recherche des facteurs qui expliquent les regroupements établis selon les valeurs IPs au cours du temps (problème 3) et la prédiction de l'âge de revêtement (problème 4), le logiciel *cbmLAD* (solution 1) est utilisé moyennant une préparation spécifique des données pour chaque problème.

Pour le problème 3 - *Profils des regroupements à travers les âges*, la solution 1 est appliquée en prenant les familles comme classes et les facteurs comme moyens d'explication des différences entre les classes de familles. Vu le nombre important de familles établies pour chaque groupe, une comparaison par paire de familles est appliquée.

Pour le problème 4 - *prédiction de l'âge de revêtement du segment*, des classes d'âges de revêtement sont identifiées. Ces classes couvrent la totalité des âges de segments dans toutes les catégories. Les profils caractéristiques de chaque classe d'âges représentent des modèles de prédiction de la classe d'âges de revêtement. Ces profils sont décrits par un certain nombre de facteurs d'exploitation et de maintenance importants y compris les classes des IPs retenus.

La résolution du problème 5 a nécessité l'usage de techniques de filtrage et le développement d'algorithmes de recherche basés sur les critères caractérisant chaque type d'anomalie. Les résultats sont fournis sous format « Excel ».

1.5 Résumé des rapports intermédiaires

Au cours de la réalisation du projet, trois rapports intermédiaires sont produits. Le rapport intermédiaire #1 est consacré à l'étude des segments de chaussée en enrobé et celles ayant subies une reconstruction (Groupe 1 - C1, C2 et C3). Il dresse les éléments de la problématique et propose une méthodologie de résolution pour les deux principaux problèmes : âge par âge afin d'établir les profils de segments similaires (problème 1), et à travers les âges afin d'effectuer les regroupements sur la base du comportement des IP de segments similaires.

Tout d'abord, le rapport décrit en détail les procédures de traitement préliminaire de données brutes et d'extraction des fichiers de simulation. Il développe les trois premières solutions 1, 2 et 3, basées respectivement, sur les techniques d'analyse logique des données, de classification ascendante hiérarchique et de regroupement dynamique selon la séquence d'observations, et les applique aux catégories C1, C2 et C3. Les résultats obtenus ont permis de valider la méthodologie proposée pour la résolution du problème principal 1. En effet, il est possible d'expliquer les regroupements de segments à l'aide des facteurs caractéristiques. Les règles d'association générées par le logiciel *cbmLAD* sont interprétables, les facteurs sont identifiés avec leurs plages de variation. Les résultats de la technique de regroupement dynamique (DTW) du comportement des segments selon la séquence d'âges sont prometteurs. Il est recommandé de reconduire l'usage de ces deux techniques pour l'étude des catégories C4 à C5. Toutefois, la classification ascendante hiérarchique a été abandonnée pour le reste des catégories de C4 à C8.

Le rapport intermédiaire #2 est consacré à l'étude des catégories de chaussée en enrobé réhabilitée et ayant subi une intervention palliative (Groupe 2 - C4 et C5). Il maintient la technique d'analyse logique des données (LAD); développe des modèles de prédiction de l'âge de revêtement; améliore la technique de regroupement dynamique (DTW) pour la rendre sensible à la séquence d'âges des IPs; et enfin retrace les segments présentant des anomalies particulières suite à l'étude des données brutes et à l'application des solutions proposées.

Les résultats obtenus pour le groupe 2 donnent des profils détaillés au même titre que pour le groupe 1. Ils montrent une prédominance des mêmes facteurs caractéristiques identifiées dans les familles de segments du groupe 1 (longueur du segment, surface totale revêtue, %Camion et DJMA) selon l'IRI – tous les âges. Le modèle de prédiction de l'âge de revêtement pour la catégorie C5 permet de classer les segments dont l'âge est inconnu avec un degré d'appartenance 74%. Les regroupements de segments selon la séquence d'âges comparativement à l'usage de la séquence d'observations sont plus homogènes. L'identification des segments présentant des anomalies particulières montre que toutes les catégories du groupe 1 et groupe 2 sont touchés par deux types d'anomalies : faibles valeurs des indices de performance à jeune âge et stabilité de ces derniers au cours du temps.

Le rapport intermédiaire #3 a reconduit la même technique d'analyse logique des données (LAD) pour la recherche des profils caractérisant les familles de

segments âge par âge pour les chaussées en enrobé ayant subies une intervention palliative et chaussées mixtes et en béton de ciment (Groupes 3-C6 et 4 - C7 et C8). La technique de regroupement dynamique (DTW) pour le comportement similaire à travers les âges a été améliorée afin de prendre en compte la valeur d'IP à l'origine. Cette solution améliorée a été appliquée de nouveau à toutes les catégories de segments C1 à C5. Également, le rapport présente les résultats d'une comparaison par paire des familles trouvées pour C6 à C8. Les modèles de prédiction de l'âge de revêtement ainsi que la liste des segments présentant des anomalies particulières ont été établis pour toutes les catégories non étudiées dans les rapports intermédiaires #1 et #2.

Les résultats obtenus pour les groupes 3 et 4 donnent des profils détaillés, selon la technique d'analyse logique des données (LAD), au même titre que pour les groupes 1 et 2. Ils montrent une prédominance des mêmes facteurs caractéristiques identifiées dans les familles de segments des groupes 1 et 2 (*longueur du segment, surface totale revêtue, %Camion et DJMA*) selon l'IRI – tous les âges. Les regroupements de segments selon la séquence d'âges et prenant en compte la valeur d'IP à l'origine sont de meilleure qualité comparativement à ceux obtenus sans tenir compte des valeurs à l'origine, et ce pour toutes les catégories C1 à C8. Les modèles de prédiction de l'âge de revêtement pour les catégories C1 à C5 et C6 à C8 permettent de classer les segments dont l'âge est inconnu avec un degré d'appartenance variant de 60% à 79%. L'identification des segments présentant des anomalies particulières pour C6 à C8 montre des pourcentages plus faibles que ceux obtenus pour les catégories des groupes 1 et 2, excepté le pourcentage de l'anomalie – Faible IP à jeune âge qui dépasse 84%. Le reste des pourcentages se situent dans la moyenne par rapport aux catégories C1 à C5.

1.6 Organisation du rapport final

La section 2 présente une synthèse des résultats obtenus pour toutes les catégories de chaussée. Cette synthèse est organisée en fonction des 5 problèmes décrits précédemment : 1) profils de familles de segments par âge, 2) comportement de familles de segments à travers les âges, 3) profils des regroupements à travers les âges, 4) prédiction de l'âge de revêtement de la chaussée et enfin 5) l'identification des anomalies.

La section 3 présente les principales conclusions et recommandations du projet. Les avantages et limites des solutions proposées pour résoudre les différents problèmes à l'étude sont discutés. Les leviers d'amélioration des données et des résultats obtenus sont également présentés. Les principales recommandations se rapportant aux données, solutions proposées et perspectives de développement sont présentées et commentées.

La section 4 présente une synthèse de conclusions et recommandations du projet.

2. SYNTHÈSE DES RÉSULTATS OBTENUS

2.1 Traitement des données

Le traitement des données brutes a permis d'extraire les fichiers de simulation pour les deux principaux problèmes de l'étude. Pour le problème 1, une procédure en 5 étapes est développée : codification des facteurs lexicaux, suppression des facteurs redondants ou non pertinents, fusion ou manipulation de certains facteurs, traitement des données manquantes et enfin l'extraction des fichiers de simulation, catégorie par catégorie, IP par IP et âge par âge.

Ainsi, les fichiers de simulation Catégorie-IP-âge des groupes 1 (C1, C2, C3), 3 (C6) et 4 (C7, C8) sont extraits selon la procédure précédente. Seulement pour le groupe 2 (C4, C5), une décomposition supplémentaire fichiers Catégorie-IP-âge par rapport à RTSQ a été appliquée afin d'améliorer la qualité des regroupements. La procédure est expliquée en détail dans le Rapport-intermédiaire#1.

À remarquer que le choix pris pour traiter les données manquantes consiste à supprimer toutes les observations dont l'IP est non connu. Ensuite, selon le pourcentage des données manquantes, facteur par facteur, une décision est prise pour supprimer le facteur ou re-compléter ses valeurs manquantes à l'aide de la méthode des voisins similaires. À ce titre, le facteur %*Camion* a été retenu dans l'étude alors que son pourcentage de données manquantes dépasse le seuil de 30% pour certaines catégories.

Par ailleurs, l'IP8.1 (Fissuration gel) n'a pas été retenu pour toutes les catégories. Cet IP comporte deux classes : 0 ($IP \geq 40$) et 1 ($IP < 40$). Le nombre de segments dans la classe 1 est très faible pour C1 (0,3 %) et C2 (0,1%), et nul pour C3 (0%). Ainsi, il n'est pas possible d'effectuer l'apprentissage.

Pour le problème 2 - *Regroupement de segments à travers les âges*, l'IP8.1 n'est pas retenu pour toutes les catégories, non plus IP5 pour C6. Pour les catégories C7 et C8, seulement, les IP1 (IRI) et IP2 (Orniérage) sont retenus. Les IPs non retenus présentent plusieurs valeurs manquantes de sorte qu'il n'est pas possible de reproduire l'évolution chronologie des valeurs IPs consignées. Ces valeurs sont, soient très éloignées les unes des autres, ou tout simplement, des valeurs isolées.

2.2 Profils de familles de segments par âge

La recherche de profils de familles de segments par âge relève du problème 1. L'utilisation de l'analyse logique des données LAD (solution 1) et la classification ascendante hiérarchique (solution 2) appréhendent de deux façons ce dernier. Comme la technique LAD est appliquée, catégorie par catégorie, IP par IP et âge par âge, les profils des familles obtenus pour chaque classe d'IP sont très détaillés du fait qu'ils identifient explicitement les facteurs caractéristiques ainsi que leurs plages de variation. Ces profils sont interprétables, car ils sont construits à l'aide des facteurs caractéristiques retenus. Chaque profil permet de regrouper les segments tout en connaissant les facteurs discriminants et leurs plages de variation. Ainsi, il est tout le temps possible de vérifier, en revenant au fichier source, que les facteurs caractéristiques des segments regroupés dans une famille donnée satisfassent le

profil déjà établi.

Par ailleurs, la connaissance de ces profils permet d'établir une prédiction de son appartenance à une classe d'IP à un âge donné sans que son IP ne soit mesuré, juste en se basant sur le recoupement de ses facteurs caractéristiques avec les profils déjà établis à cet âge en particulier.

En ce qui concerne l'analyse des profils établis sur la base des 25 ou 26 facteurs caractéristiques retenus, plusieurs algorithmes sont programmés afin d'extraire le plus d'informations possibles sur le pouvoir discriminant de chacun d'eux. La fréquence calculée sur la base du nombre de fois où le facteur, en question, participe à la construction d'un profil, est utilisée comme indicateur de son pouvoir discriminant entre les profils à un âge donné.

Il est important de rappeler que chaque profil doit être considéré dans sa totalité, c'est un tout. Les facteurs constituant un profil ne sont pas dissociables, ni séparément interprétables. L'usage de la fréquence des facteurs doit être vu comme un moyen pour rendre compte de leur pouvoir discriminant. Les facteurs qui sont moins fréquents ont aussi un pouvoir discriminant. Ils interviennent plus spécifiquement lorsque les premiers facteurs discriminants n'arrivent plus à départager les profils. Il est important de faire attention aux profils qui utilisent ces facteurs non fréquents à un âge donné pour appréhender des problèmes spécifiques en lien avec la construction, l'exploitation ou la maintenance de la chaussée.

Pour les catégories C1, C2 et C3, la revue de tous les profils, catégorie par catégorie, IP par IP et âge par âge montrent clairement que la *longueur* du segment, l'*épaisseur de revêtement*, l'*épaisseur totale*, le *%Camion* et le *DJMA* sortent le plus souvent dans les profils de familles à jeune et vieux âges. Ces facteurs gardent leur pouvoir discriminant dans les profils au cours du temps. Les facteurs tels que *Type de maintenance* et *Dernier code travaux* sont moins fréquents dans les profils étudiés. Ils ont un pouvoir discriminant plus spécifique.

Pour les catégories C4 et C5, une décomposition supplémentaire des fichiers de simulation est faite par rapport RSTQ0 et RSTQ1. Cette décomposition est proposée pour augmenter la qualité des profils trouvés. L'analyse des profils montrent clairement que la *longueur* du segment, la *surface totale revêtue*, le *%Camion* et le *DJMA* sont le plus souvent utilisés par les profils pour discriminer une famille de segments d'une autre à un âge donné. Les profils de familles à jeune et vieil âges sont les plus concernés par ces facteurs. Ces derniers gardent, aussi, leur pouvoir discriminant dans les profils au cours du temps. De même que pour les catégories précédentes, le *Type de maintenance* et *Dernier code travaux* sont les moins utilisés pour discriminer les familles à un âge donné. Cela pourrait relater le fait que très peu de segments de ces catégories utilisent des méthodes de maintenance différentes.

Pareillement aux résultats des groupes 1 et 2, les profils des familles obtenus pour chaque classe IP pour les groupes 3 et 4 sont très détaillés. Chaque profil permet de regrouper les segments tout en connaissant les facteurs discriminants et leurs plages de variation. À noter, que les facteurs constituant un profil ne sont pas dissociables, non plus séparément interprétables. L'usage de la fréquence des facteurs doit être vu comme un moyen pour rendre compte de leur pouvoir discriminant. Les facteurs qui sont moins fréquents ont aussi un pouvoir discriminant plus spécifique lorsque les premiers facteurs discriminants n'arrivent plus à départager les profils.

À travers l'ensemble des profils obtenus, la longueur représente un des facteurs fréquemment utilisés pour discriminer les segments de même âge, mais appartenant à des classes différentes d'IP. Rappelons tout d'abord que la longueur n'a jamais été utilisée toute seule pour discriminer une famille d'une autre. La longueur fait partie toujours d'une conjonction de facteurs. La fréquence d'utilisation de la longueur par les règles de regroupement peut caractériser son importance dans les règles mais jamais sa dominance. Ce raisonnement est aussi valable pour tous les facteurs ayant une fréquence d'utilisation élevée dans les règles de regroupement de segments par âge.

La solution 2 (classification ascendante hiérarchique) est utilisée uniquement pour les catégories C1, C2 et C3. Elle regroupe les segments, se trouvant à un âge donné, en minimisant la distance entre les facteurs caractéristiques sans égard à la valeur ou la classe d'IP. Comme chaque segment possède son propre profil au départ, il y en a autant de familles que de profils. Sur la base du calcul de similarité entre les profils, de nouveaux profils sont déduits formant ainsi de nouvelles familles de segments.

L'application de cette solution nécessite le choix au préalable du nombre souhaité de familles. Pour se faire, ce nombre est fixé à 5 afin de relater la répartition des valeurs IPs en 5 classes différentes. L'objectif était d'examiner s'il était possible de détecter une certaine dynamique dans les regroupements proposés de façon à pouvoir identifier le passage d'un segment de chaussée ou un groupe de segments d'une famille à une autre au cours du temps.

À partir des résultats préliminaires obtenus par cette solution pour C1, C2 et C3, il paraît clairement qu'un nombre très important de segments se trouvent dans une ou deux familles avec une grande variabilité de ce dernier d'un âge à un autre. À noter que plus le nombre de segments par famille est faible, plus les segments qui les constituent sont isolés. C'est-à-dire, ces segments sont trop distants des autres segments. Il faut aussi souligner qu'il ne sera pas possible d'expliquer les résultats de regroupements de manière explicite puisque le nombre de segments par famille dépend du choix du nombre de familles a priori.

Par ailleurs, cette solution ne permet pas a posteriori d'identifier les facteurs qui influencent le calcul des distances entre deux profils. Pour y arriver, il sera tout à fait possible d'utiliser les matrices intermédiaires de calculs de distances.

Comme il n'était pas possible d'identifier le passage d'un segment de chaussée d'une famille à l'autre au cours du temps, cette solution n'a pas été reconduite pour les catégories de chaussée des groupes 2, 3 et 4. Toutefois, les tests effectués sur les âges 0 à 36, pour C1, C2 et C3, nous permettent de conclure que les facteurs caractéristiques sont très semblables de sorte que les distances qui séparent les profils de segments ne sont pas suffisamment discriminatoires entre les familles.

2.3 Comportements de familles de segments à travers les âges

Le comportement de familles de segments à travers les âges est décrit par le problème 2. Il est appréhendé à l'aide des cinq solutions 3 à 7. Tout d'abord, la solution 3 (*Regroupement selon les valeurs IPs et la séquence d'observations*) recherche les familles de segments sur la base du comportement, en termes de tendance, des valeurs d'IP selon leurs séquences d'observations. Les segments se trouvant dans

une même famille ont tous vu un comportement similaire de leurs valeurs IP (augmentation, diminution ou stabilité) d'une observation à l'autre ou d'une sous séquence à l'autre, prises consécutivement au cours du temps.

Le nombre de familles pour chaque IP est choisi de manière à maximiser la similarité des comportements des valeurs d'IP mesurées selon la même séquence d'observations. Pour les catégories C1, C2 et C3, les résultats obtenus montrent une grande étendue du nombre de segments par famille, avec une médiane de quelques segments. Les représentations graphiques montrent clairement les allures comportementales des segments d'une même famille. Ces résultats sont encourageant. Toutefois, certains segments d'une même famille peuvent avoir un comportement non régulier durant une séquence d'observations. Ainsi, certaines figures montrent des segments qui divergent à la fin d'une séquence des autres segments de la même famille, alors qu'ils avaient le même comportement ou presque à partir du début.

Pour pallier à cette situation, deux extensions de la solution 3 sont proposées. La solution 4 regroupe les segments ayant des valeurs de pentes très voisines les unes des autres selon la séquence d'âges. Alors que la solution 5 utilise uniquement l'écart des valeurs IPs d'un segment d'un âge à un autre afin d'effectuer les regroupements de segments présentant des écarts similaires selon la séquence d'âges. Ces deux solutions sont alors appliquées sur les catégories C4 et C5.

Pour C4 et C5, les regroupements de segments selon la séquence d'âges comparativement à l'usage de la séquence d'observations sont plus homogènes. Il y en a moins de divergences à la fin des figures. Les segments ayant une séquence d'âges plus courte que d'autres ne figureront qu'avec leurs similaires une fois ils atteignent leurs âges limites. Ainsi, l'usage des tendances et des écarts des valeurs IP au lieu des valeurs IP donnent de meilleurs résultats. D'ailleurs, les regroupements obtenus par ces deux variantes sont sensiblement les mêmes.

Toutefois, comme les solutions 4 et 5 ne considèrent pas les valeurs à l'origine des IPs, alors certains regroupements comportent des segments qui possèdent la même pente ou écart d'IP que ses semblables mais ont une durée de vie plus courte du fait que leurs états initiaux étaient plus dégradés que leurs similaires. Pour remédier à cet inconvénient, la solution 4 a été améliorée pour inclure les valeurs à l'origine dans les regroupements. Les résultats de cette nouvelle solution (solution 6) sont de très bonnes qualités.

Les résultats obtenus pour C4 montrent clairement que l'apport de la valeur à l'origine a une importance capitale lors de la recherche de segments ayant une tendance similaire. Pour atteindre une meilleure qualité des regroupements, trois tests ont été effectués sur les segments de C4 avec une variation des valeurs IPs à l'origine de 1, 2.5 et 5 sur 100. Principalement, un faible écart à l'origine diminue la dispersion des segments dans une même famille et donne des regroupements plus homogènes. Toutefois, il augmente significativement le nombre de familles. Pour C4 seulement, le nombre de familles varie entre 4000 et 800 lorsque l'écart IP à l'origine passe de 1 sur 100 à un écart indéfini.

Après analyse de la qualité des regroupements de C4 à l'aide des indices de silhouette et de dispersion selon Davies-Bouldin modifié, un écart IP à l'origine de +/-

2.5 sur 100 donne le meilleur compromis en termes de qualité des regroupements et de nombre de familles. Ce choix a été appliqué à toutes les catégories de segments. Ainsi, le travail est repris pour les catégories C1 à C3 et complété pour C5 à C8.

L'analyse des résultats obtenus pour toutes les catégories montre clairement que les regroupements contiennent moins de segments ayant une divergence prononcée à la fin des figures tel que remarqué dans les résultats de la solution 4 (*Regroupement selon les pentes IPs la sous-séquence d'âges*). Les segments ayant une séquence d'âges plus courte partagent les mêmes tendances que leurs similaires sur tous les âges communs. Ainsi, il est possible de prédire leurs comportements sur les âges subséquents tel que vécus par leurs similaires, si les conditions d'exploitation demeurent les mêmes. De plus, le nombre moyen de familles par rapport au nombre initial de segments dans chaque catégorie varie entre 3.6 et 4.3 familles pour 100 segments.

La solution 7 (*Regroupement selon les pentes IPs la sous-séquence d'âges*) est venue réduire le nombre de familles en regroupant les segments ayant des pentes similaires par sous-séquences d'âges. Les résultats de cette solution ont été obtenus suite à la présentation du rapport intermédiaire #3. Pour chaque catégorie, le dossier « C#-Pentes-Sous-séquences-âges » contient les fichiers « image » et « Excel » formatés de la même manière que pour les autres solutions. Tous les dossiers sont déposés avec le présent rapport.

L'analyse qualitative des fichiers « image » montre de meilleures regroupements. Le nombre de familles est réduit de 70 à 73%. Ainsi, la catégorie C4 qui contient presque 80000 segments initialement, la solution 7 (*Regroupement selon les pentes IPs la sous-séquence d'âges*) permet de les regrouper en 721 familles comparativement à la solution 6 (*Regroupement selon les pentes IPs, les valeurs IPs à l'origine et la séquence d'âges*) qui offre initialement 3067 familles. Cette réduction notable est remarquée pour toutes les catégories. Ainsi, le nombre moyen de familles par rapport au nombre initial de segments dans chaque catégorie n'est que de 0.9 à 1.4 familles pour 100 segments. Ce sont des résultats très probants.

2.4 Profils des regroupements à travers les âges

Dans un premier test, il était question de déterminer les familles de segments de la catégorie C3, IP1 seulement, à l'aide de la technique de regroupement dynamique (DTW) de segments selon les valeurs de IPs et la séquence d'observations, pour ensuite appliquer la technique d'analyse logique des données (LAD) en remplaçant la classe de l'IP par le numéro de la famille. Ainsi, pour chacune des familles, il était possible de déterminer les facteurs caractéristiques qui la distinguent d'une famille.

Ce test montre clairement que les facteurs caractéristiques dont le pouvoir discriminant est le plus élevé sont presque les mêmes que ceux trouvés à l'aide de la technique d'analyse logique des données pour C3, IP1, tous les âges. Il s'agit particulièrement de : *Longueur* du segment, *Superf-totale-revêtue*, *Épaisseur-totale* et *%Camion*. Pour se faire, les 30 familles de C3 ont été utilisées comme classes dans le logiciel *cbmLAD* et l'ensemble des autres facteurs caractéristiques excluant la classe IP comme variables explicatives.

Avec le développement de la solution 6 qui permet de regrouper dynamiquement, selon la séquence d'âges, les segments ayant des valeurs de pentes d'IPs similaires incluant une variation de 2.5 sur 100 des valeurs IPs à l'origine, un second test est conduit avec les résultats des catégories C6 à C8, IP par IP applicable. Vu le nombre de familles trouvées, l'usage de *cbmLAD* en multi-classes n'est pas pratique. Une comparaison par paire de familles est proposée pour établir les profils qui distinguent une famille d'une autre, deux par deux. Ainsi, une procédure permettant de construire les fichiers de comparaison est programmée. Chaque fichier contient deux libellés, un pour chaque famille, suivi de tous les facteurs retenus selon la catégorie de chaussée. Seulement pour la catégorie C6, plus de 52000 fichiers sont ainsi comparés.

Pour les catégories C6, C7 et C8, les facteurs les plus utilisés par les règles de classification sont relatifs aux caractéristiques de conception de la chaussée telles que *Superf_totale_revêtue*, *Superficie_acc_droit_revêtue* et *Largeur_voies*. Le *%Camion* et *Épaisseur_total* ont été moins utilisés que les facteurs relatifs à la conception. Les attributs *Type_of_Maintenance* et *RSTQ* gardent les mêmes importances dans tous les fichiers de comparaison. À remarquer que les règles de classification utilisent en majorité un seul facteur discriminant. Ce qui montre que les familles sont bien distinctes. Il s'agit d'excellents résultats.

Toutefois, vu le nombre phénoménal de fichiers de comparaison, la décision a été de ne pas étendre cette solution aux autres catégories. Bien que la recherche des facteurs discriminants est presque instantanée, la manipulation de nombreux fichiers reste non commode. Mais, ce test a bien démontré la faisabilité d'une telle étude.

2.5 Prédiction de l'âge de revêtement

Le développement de modèles permettant de prédire l'âge du revêtement d'un segment vise la correction de certains âges de revêtement de segments lorsque ces derniers sont entachés d'erreurs. Le modèle repose essentiellement sur la sélection des attributs IPs et certains facteurs de conception, d'exploitation et de maintenance ayant un pouvoir discriminant élevé pour l'apprentissage; et le regroupement des âges de revêtement choisis.

Une procédure de sélection est appliquée afin de retenir les attributs les plus importants, ceux qui permettent une meilleure discrimination entre les classes d'âges. La technique de forêt d'arbres décisionnels « Random Forest » est utilisée. Suite à plusieurs essais, les attributs suivants sont retenus : IP1, IP2, IP3, IP5, IP7.1, IP9, Épaisseur totale, Épaisseur revêtement, DJMA et *%Camion*. Pour trouver le meilleur regroupement des âges en classes, plusieurs essais ont été réalisés en prenant à chaque fois un intervalle d'âges différent. En définitive, 5 intervalles (classes) comprenant chacun 6 âges consécutifs pour toutes les catégories, à l'exception de la 5^e classe d'âges qui regroupe tous les âges supérieurs à 24.

L'application de *cbmLAD* permet de générer les règles d'association de segments ayant des âges rapprochés par rapport à d'autres en fonction des profils définis par les variables explicatives utilisées. Ainsi, connaissant ces règles d'association qui discriminent une classe d'âges par rapport aux autres, il est facile d'associer un segment à une classe d'âges dès que son profil respecte la règle d'association prédéterminée dans la phase d'apprentissage.

La majorité des tests d'apprentissage ont été réalisés avec 75% des segments choisis aléatoirement selon la technique k-folder. Le reste des segments, soit 25%, est utilisé pour la validation des règles. La qualité des modèles de prédiction est mesurée par le degré d'appartenance qui caractérise le pourcentage moyen des bonnes classifications de segments à une classe d'âge donné. Plus ce pourcentage est grand mieux sera la qualité de prédiction des règles. Les degrés d'appartenance obtenus pour toutes les catégories sont au-dessus de 60%. Certaines catégories atteignent 78%. Par ailleurs, les modèles de prédiction utilisant la classe d'IP sont plus performants que ceux utilisant la valeur d'IP. Cela est tout à fait compréhensible.

En ce qui concerne les attributs utilisés par les modèles de prédiction, il paraît clairement que tous les attributs sont utilisés de manière presque uniforme et dispersée dans le cas d'usage de la valeur IP. Cependant, les attributs caractérisant l'usage et la maintenance prennent le dessus par rapport à l'importance des attributs mesurant la performance en termes de classe d'IP. Les règles de prédiction utilisent souvent les attributs d'exploitation pour différencier les âges. Cela nous semble tout à fait plausible.

Par ailleurs, un autre test a été effectué pour s'assurer de la qualité des modèles de prédiction de l'âge de revêtement. Ce test compare les résultats d'estimation de l'âge de revêtement en utilisant tous les IPs, à l'aide de plusieurs techniques de classification et de régression communément utilisées en intelligence artificielle (Support Vector Regression (SVR), Support Vector Machine (SVM), Tree Regression, Ensemble Tree Regression, Gaussian Process Regression (GPRm), K-Nearest Neighbor (KNN) classification) et celle de cbmLAD. L'analyse des résultats montre que cbmLAD donne la meilleure performance, suivi de SVR.

2.6 Identification des anomalies

Les anomalies au niveau des ~~caractérisent des situations anormales dans les~~ données ou les résultats obtenus sont détectées à partir de l'étude des données brutes ou déduites des résultats obtenus par les techniques de modélisation utilisées. Les données manquantes ne sont pas considérées comme des anomalies. Huit types d'anomalies sont identifiées : (1) Faible valeur IP à jeune âge, (2) Forte valeur IP à vieil âge, (3) Hausse d'IP sans maintenance ou changement de catégorie, (4) Tendance positive d'IP au cours du temps, (5) Baisse intempestive d'IP au cours du temps, (6) IP stable au cours du temps, (7) Segments solitaires, et (8) Différents IPs à une même séquence âge.

Les seuils utilisés pour classer IP comme faible (anomalie 1) ou fort (anomalie 2) sont extraits du document « *Doc 1-9 Seuils de defcience utilises.pdf* » fourni par le MTMDDET. L'anomalie 3 vérifie si une augmentation de IP est observée entre deux âges consécutifs sans aucune intervention de maintenance ou changement de catégorie. Quant à l'anomalie 4, une augmentation de IP est observée sur plusieurs âges consécutifs, sans raison. L'anomalie 5 retrace les segments qui enregistrent une baisse de plus de 40 % de leurs IPs entre deux âges consécutifs. L'anomalie 6, IP stable au cours du temps, est détectée lorsque la différence en valeur absolue d'un IP et IP moyen pour ce segment est inférieur ou égale à 4 pour 100.

Toutes les anomalies ont été identifiées IP par IP, catégorie par catégorie. Le nombre de segments retracés pour chaque type d'anomalie ainsi que le pourcentage

d'anomalie par catégorie de segments par rapport au nombre total de segments de la même catégorie sont calculés. La catégorie C6 présente le pourcentage d'anomalie le plus élevé de toutes les catégories, en particulier à cause de l'anomalie 1- Faible valeur IP à jeune âge qui dépasse 84%. Alors que les pourcentages d'anomalie 1 de C7 et C8 ne dépassent pas 14%. Pour cette même anomalie, les pourcentages varient de 10 à 21% pour les catégories C1 à C5. Le reste des pourcentages d'anomalie pour C6 à C8 sont assez similaires ne dépassant pas 1 à 3 % et dans certains moins élevés que ceux des catégories C1 à C5 de 2 à 5 %.

L'analyse des anomalies par IP, montre que IP1 est le plus touché par une faible valeur IP à jeune âge suivi de IP3. Plusieurs segments sont touchés par une forte valeur IP3 à vieil âge. L'anomalie 6 – IP stable au cours du temps présente des pourcentages de plus de 10% pour les catégories C1, C3 et C5, alors que ce pourcentage est en deçà de 3.5% pour les autres catégories. En résumé, les anomalies 1 et 6 sont les plus fréquentes suivies de 2, 5 et 3, et enfin par 4, 7 et 8. Ce classement apparait plus pour les catégories C1 à C5 alors que pour les catégories C6 à C8, les anomalies les plus fréquentes sont 1 et 2.

3. CONCLUSIONS ET RECOMMANDATIONS

3.1 Principales conclusions

Au terme de ce projet, plusieurs modèles basés sur des techniques d'intelligence artificielle ont été développés pour établir les regroupements de segments ayant : (1) un profil similaire à chaque âge et (2) un comportement similaire à travers les âges. Les modèles proposés exploitent plusieurs mesures de similarité partant de la recherche des facteurs qui discriminent une famille de segments d'une autre famille à un âge donné aux calculs de distances multidimensionnelles permettant de regrouper des segments ayant un comportement semblable selon la séquence d'observations, d'âges ou sous-séquence d'âges. Ces modèles ont été appliqués sur toutes les catégories de chaussée à l'étude.

Bien évidemment, les résultats obtenus pour chaque catégorie de chaussée dépendent particulièrement de la qualité des données utilisées par chaque modèle. Les procédures de traitement préliminaire des données ont permis la sélection des facteurs les plus renseignés dans la base des données, à compléter les valeurs manquantes et à extraire les fichiers de simulation. Toutefois, voulant améliorer la représentativité des modèles de classification, tant pour rechercher les profils similaires que pour la prédiction de l'âge de revêtement, plusieurs techniques d'analyse des données et de sélection de facteurs ayant un pouvoir discriminant ont été appliquées. Certains facteurs catégoriels caractérisant les profils de segments ne sont pas facilement séparables. Le nombre et leurs niveaux d'agrégation peuvent être à l'origine de cette nuisance dans les données.

D'ailleurs, bien que les profils obtenus pour la caractérisation des familles de segment par IP et par âge sont bien identifiés et les plages de variation des facteurs caractéristiques bien définies, leur nombre reste relativement élevé. Cela découle principalement de l'obligation des modèles utilisés de couvrir tous les segments par au moins une règle de regroupement. Le nombre de règles peut facilement diminuer lorsqu'un choix plus approprié des facteurs caractéristiques serait fait. Cependant, les résultats obtenus montrent clairement la contribution de certains facteurs de construction, d'exploitation et de maintenance dans la définition des profils des familles âge par âge. Ce sont des résultats importants compte tenu de la diversité et la multitude de facteurs retenus dans l'étude.

De plus, l'occurrence de ces facteurs dans les règles de regroupement démontrent leur pouvoir discriminant entre les familles de segments. Les facteurs les moins fréquents ne sont pas pour autant moins importants dans les règles. Au contraire, ils ont un pouvoir discriminant plus spécifique. Ils interviennent lorsque les facteurs les plus fréquents n'ont plus un pouvoir discriminant. Ainsi, tous les facteurs contribuent à la définition des profils des familles. Aucun facteur retenu n'était inutile.

Les solutions proposées pour le regroupement de segments ayant un comportement similaire à travers les âges ont permis d'appréhender ce problème principal de plusieurs manières. Tout d'abord, l'usage des valeurs IP et la séquence d'observations pour définir les comportements de segments similaires donnaient des résultats encourageants pour les premières catégories. De ce fait,

plusieurs extensions du modèle initial ont été apportées pour améliorer la qualité des regroupements en ce qui a trait à l'usage de la séquence d'âges, au lieu de la séquence d'observations, des pentes et écarts d'IP au lieu de la valeur initiale IP et la prise en compte des valeurs IP à l'origine. Les résultats de ces améliorations sont probants en particulier celle qui exploite les pentes à travers la séquence d'âges et les valeurs à l'origine avec un écart de 2.5 sur 100.

Toutefois, le nombre de segments par famille reste relativement élevé pour certaines catégories, en particulier pour C4. Bien qu'en général, le nombre moyen de familles par rapport au nombre initial de segments dans chaque catégorie varie entre 3.6 et 4.3 familles pour 100 segments. Pour faciliter l'usage de ces regroupements dans le système GCH, une nouvelle solution est proposée pour augmenter le nombre de segments par famille, et en conséquence réduire le nombre de famille(s). Cette solution combine les avantages des solutions précédentes et celle de l'usage des sous-séquences d'âges et non pas la séquence d'âges. Les résultats obtenus réduisent de plus de 70% le nombre de familles. Ainsi, le nombre moyen de familles par rapport au nombre initial de segments dans chaque catégorie n'est que de 0.9 à 1.4 familles pour 100 segments.

Les résultats de la comparaison par paire de familles montrent une dominance des facteurs relatifs à la conception de la chaussée. Les facteurs relatifs à l'exploitation et à la maintenance ont un pouvoir discriminant plus spécifiques lorsque les facteurs de conception ne peuvent plus discriminer les segments de familles différentes. Ce qui nous semble important dans ces résultats de comparaison par paire de familles, le fait qu'il existe des facteurs qui expliquent les particularités d'une famille par rapport à une autre. L'inconvénient reste celui du nombre important de fichiers de comparaison, il se compte en dizaine de milliers, ce qui prend un temps énorme pour compléter toutes les combinaisons.

Les modèles de prédiction de l'âge de revêtement obtenus pour toutes les catégories de segments permettent de classer les segments dont les âges sont inconnus avec un degré d'appartenance allant de 60% à 79%. Ces modèles sont très importants pour corriger les âges de certains segments, comme par exemples, ceux ayant une faible valeur IP à jeune âge ou une forte valeur IP à vieil âge. La qualité de prédiction est meilleure lorsque la classe d'IP est utilisée comparativement à la valeur d'IP. Les attributs caractérisant l'exploitation (usage) et la maintenance prennent le dessus par rapport à l'importance des attributs mesurant la performance de la chaussée en termes de classe d'IP. Toutefois, le nombre de modèles de prédiction reste assez élevé pour un usage facile dans le système GCH.

Enfin, l'identification des segments présentant des anomalies particulières montre que toutes les catégories sont touchées à différents niveaux par les faibles valeurs des indices de performance à jeune âge. La catégorie C6 est particulièrement concernée par cette anomalie par rapport aux autres. L'IP1(IRI) est le plus touché par des faibles valeurs IP à jeune âge suivi de IP3(Orniérage). Plusieurs segments des catégories C1, C3 et C5 ont des IPs stables au cours du temps, les segments des autres catégories sont concernés aussi, mais à moindre ampleur.

En résumé, les faibles valeurs IPs à jeune âge et valeurs IPs stables au cours du temps sont les plus fréquentes suivies de fortes valeurs IPs à vieil âge, baisse intempestive d'IP au cours du temps et hausse d'IP sans maintenance ou changement de catégorie, et enfin par tendance positive d'IP au cours du temps, segments

solitaires et différents IPs à une même séquence âge. Ce classement apparaît plus pour les catégories C1 à C5 alors que pour les catégories C6 à C8, les anomalies les plus fréquentes sont à faible valeur IP à jeune âge et forte valeur IP à vieil âge.

La correction de ces anomalies permet d'améliorer la qualité des données. Cela augmentera sans doute la représentativité des résultats des modèles proposés tout au long de ce projet.

3.2 Principales recommandations

L'appréhension des deux principaux problèmes de regroupements de segments par âge et à travers les âges, des problèmes de détermination des profils de familles, de prédiction de l'âge de revêtement et d'identification des anomalies ainsi que l'analyse des résultats des solutions proposées pour résoudre ces problèmes permettent de formuler les cinq recommandations suivantes.

Recommandation 1. Installer un mécanisme pour s'assurer davantage de l'intégrité des données. Cette intégrité concerne principalement deux aspects : la complétude des données manquantes et la correction des données erronées ou entachées d'erreur. Les listes de segments comportant des anomalies peuvent servir comme point de départ pour s'assurer de cette intégrité.

Recommandation 2. Trouver un mécanisme automatique de calibration de l'âge zéro et de la valeur IP à 100 à chaque réhabilitation. Cet âge est assez important dans tous les problèmes étudiés. Cette calibration doit se faire tout d'abord, par rapport aux critères qui définissent la réhabilitation de la chaussée. Ensuite, par rapport aux seuils de chacun des critères et enfin par rapport à la valeur IP à 100. Ce mécanisme devrait également tenir compte du délai entre les interventions de maintenance et la mise à jour des valeurs IPs.

Recommandation 3. Arrimer les données de maintenance et les valeurs des IPs afin de mieux expliquer la hausse ou la stabilité des valeurs IPs d'un âge à l'autre. Bien que le niveau de réhabilitation de la chaussée est censé affecter d'une manière ou d'une autre l'âge de revêtement, le fait de considérer seulement deux types de maintenance, un qui remet l'âge à zéro, aussi bon que neuf (as good as new) et un autre qui n'a aucun effet, aussi mauvais qu'avant (as bad as before) ne permettent pas d'expliquer la hausse ou la stabilité des valeurs IPs d'un âge à l'autre. Les facteurs caractéristiques de maintenance ne sont pas souvent utilisés pour établir les regroupements. Une analyse détaillée des résultats obtenus pourrait aider à développer un modèle plus précis pour chaque niveau de réhabilitation.

Recommandation 4. Analyser chaque facteur et ne retenir que celui qui caractérise le phénomène physique représenté ici par la valeur IP. Par exemple, lorsque la longueur est déjà utilisée pour normaliser les valeurs IPs des segments, elle n'est pas pertinente pour expliquer pourquoi 2 segments de même âge se trouvent dans 2 classes différentes d'IP. La longueur est démunie de son pouvoir discriminant puisqu'elle devient une variable endogène et non exogène au phénomène physique.

Recommandation 5. Considérer, lorsque cela est possible, des facteurs qui expriment des écarts entre les valeurs de conception et celles mesurées en exploitation. Ces facteurs seront plus porteurs d'informations interprétables par

rapport au phénomène physique modélisé par les valeurs ou classes d'IPs. Par exemple, les écarts du DJMA ou de tous autres facteurs pourraient avoir un pouvoir discriminant plus performant lors de l'extraction des règles de regroupements de segments. Toutefois, il est recommandé de ne pas agréger les facteurs de manière à perdre l'information primaire contenue dans chaque facteur, moins d'interventions humaines mieux le phénomène physique à l'étude serait interprétable.

3.3 Perspectives de développement

L'ensemble des solutions techniques proposées ont permis d'appréhender plusieurs facettes des différents problèmes à l'étude et ainsi démontrer que l'intelligence artificielle donne des résultats très probants. À travers les tests des différents modèles, une validation conceptuelle est d'ores et déjà acquise pour l'ensemble des catégories de chaussée. L'évaluation des performances des regroupements de segments par âge ou à travers les âges moyennant la mesure du degré d'appartenance a permis de vérifier l'adéquation entre les résultats de simulation et ceux prévus par les modèles.

D'un point de vue méthodologique, l'ensemble des recommandations décrites précédemment visent principalement la validation opérationnelle des résultats obtenus. Ainsi, une phase de calibration des modèles est nécessaire pour profiter pleinement des résultats des simulations des regroupements de segments. Il s'agit d'une phase qui conjugue l'expertise opérationnelle et l'expertise en modélisation.

À cet égard, la réalisation d'un outil intégré d'aide à la décision permettant de mettre, automatiquement, à jour les regroupements de segments par âge et selon les âges constitue le meilleur cadre des développements futurs de ce projet.

4. SYNTHÈSE DES CONCLUSIONS ET RECOMMANDATIONS (750 MOTS)

Le projet s'intéresse à la problématique de regroupements de segments de chaussée dont le comportement est similaire à l'aide de l'intelligence artificielle. Il vise, spécifiquement, la recherche des profils qui décrivent le mieux les segments de chaussée ayant un comportement semblable et l'appréciation de ce comportement au cours du temps. L'étude porte sur 16 702 segments de chaussée répartis en huit catégories. Après traitement préliminaire des données brutes, chaque segment de chaussée est décrit par : un identifiant unique, six indices de performance (IP-IRI, IP-Ornéage, IP-Fissuration globale, IP-Delta-IRI, IP-Fissuration fatigue et IP-Fissuration globale), un ensemble de 26 facteurs caractéristiques et son âge au moment de l'observation.

Plusieurs modèles sont développés pour établir les regroupements de segments ayant : un profil similaire à chaque âge et un comportement similaire à travers les âges. Les modèles exploitent plusieurs mesures de similarité partant de la recherche des facteurs qui discriminent une famille de segments d'une autre famille à un âge donné aux calculs de distances multidimensionnelles permettant de regrouper des segments ayant un comportement semblable selon la séquence d'observations, d'âges ou sous-séquence d'âges. Ces modèles sont utilisés pour toutes les catégories de chaussée à l'étude.

Les profils obtenus pour la caractérisation des familles de segment par IP et par âge sont bien identifiés et les plages de variation des facteurs caractéristiques bien définies. Ces profils sont décrits par une conjonction de facteurs pour discriminer un regroupement de segments d'un autre, à chaque âge. Les regroupements de segments ayant un comportement similaire à travers les âges sont établis à l'aide de plusieurs modèles. La variabilité des résultats montre la complexité d'établir des regroupements à partir de mesures de dégradation observées par morceaux distants au cours du temps. Toutefois, les résultats du dernier modèle, celui qui exploite les sous-séquences d'âges et les pentes IPs sont les plus probants. Le nombre de regroupements de segments passe de 4 à 1.2 familles en moyenne pour 100 segments.

Les modèles développés démontrent que l'intelligence artificielle donne des résultats très probants pour établir les regroupements selon l'âge et à travers les âges, et ce, pour toutes les catégories de segments. Par ailleurs, trois autres problèmes sont étudiés et solutionnés. Il s'agit de la recherche des profils de regroupements à travers les âges. Les règles de classification utilisent en majorité un seul facteur discriminant. Ce qui montre que les familles sont bien distinctes, ce sont d'excellents résultats. Les modèles de prédiction de l'âge de revêtement obtenus pour toutes les catégories de segments permettent de classer les segments dont les âges sont inconnus avec un degré d'appartenance allant de 60% à 79%. La qualité de prédiction dépasse celle des autres techniques communément utilisées dans le domaine d'intelligence artificielle. Enfin, l'identification des segments présentant des anomalies particulières montre que toutes les catégories sont concernées par au moins un des 8 types d'anomalies retracés. L'IRI est le plus touché par des faibles valeurs IP à jeune âge suivi de IP-Ornéage et indice de performance stable au cours du temps. Les pourcentages moyens des autres types d'anomalies par catégorie de chaussée restent en moyenne en deçà de 5%.

Toutefois, cinq recommandations sont énoncées afin de permettre la validation opérationnelle des solutions proposées : installer un mécanisme pour s'assurer davantage de l'intégrité des données, spécifiquement, la complétude des données manquantes et la correction des données erronées ou entachées d'erreurs; trouver un mécanisme automatique de calibration de l'âge zéro et de la valeur IP à 100 à chaque réhabilitation, ce mécanisme doit tenir compte du délai entre les interventions de maintenance et la mise à jour des valeurs IPs; arrimer les données de maintenance et les valeurs des IPs afin de mieux expliquer la hausse ou la stabilité des valeurs IPs d'un âge à l'autre; analyser chaque facteur et ne retenir que celui qui caractérise le phénomène physique représenté la valeur IP; et enfin, considérer, lorsque cela est possible, des facteurs qui expriment des écarts entre les valeurs de conception et celles mesurées en exploitation.

Comme perspective de développement, il est important de poursuivre le projet avec une phase dédiée explicitement pour la calibration des modèles. Elle permet de profiter pleinement des résultats des simulations des regroupements de segments. Cette phase requière la conjugaison des avis des experts de la chaussée et de modélisation. Pour ce faire, la réalisation d'un outil intégré d'aide à la décision permettant de mettre, automatiquement, à jour les regroupements de segments par âge et selon les âges constitue le meilleur cadre des développements futurs de ce projet.