

Titre: Déterminants de l'état des infrastructures routières : une analyse
Title: économétrique du réseau québécois

Auteur: Bruno Marmoux
Author:

Date: 2016

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Marmoux, B. (2016). Déterminants de l'état des infrastructures routières : une analyse économétrique du réseau québécois [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/2238/>
Citation:

Document en libre accès dans PolyPublie Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/2238/>
PolyPublie URL:

Directeurs de recherche: Marcelin Joanis
Advisors:

Programme: Maîtrise recherche en génie industriel
Program:

UNIVERSITÉ DE MONTRÉAL

DÉTERMINANTS DE L'ÉTAT DES INFRASTRUCTURES ROUTIÈRES :
UNE ANALYSE ÉCONOMÉTRIQUE DU RÉSEAU QUÉBÉCOIS

BRUNO MARMOUX

DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INDUSTRIEL)

AOÛT 2016

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

DÉTERMINANTS DE L'ÉTAT DES INFRASTRUCTURES ROUTIÈRES :
UNE ANALYSE ÉCONOMÉTRIQUE DU RÉSEAU QUÉBÉCOIS

présenté par : MARMOUX Bruno

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

Mme BEAUDRY Catherine, Ph. D., présidente

M. JOANIS Marcellin, Ph. D., membre et directeur de recherche

Mme DE MARCELLIS-WARIN Nathalie, Doctorat, membre

REMERCIEMENTS

Je tiens à remercier sincèrement mon directeur de recherche, Monsieur Marcelin Joanis, Ph. D., pour son appui continu, sa disponibilité et sa réactivité. Ses remarques, suggestions et critiques m'ont permis d'avancer rapidement sur ce mémoire tout en améliorant continuellement son contenu.

Je remercie Madame Catherine Baudry, Ph. D., pour l'honneur qu'elle me fait en présidant le jury, de même que Madame Nathalie de Marcellis-Warin, Doctorat, d'avoir accepté d'en faire partie.

Enfin, je souhaite dire merci au centre de recherche sur les infrastructures en béton (CRIB) et au centre de recherche interuniversitaire de recherche en analyse des organisations (CIRANO) pour les ressources, logistiques et financières, qu'ils ont mis à ma disposition et qui ont, non seulement, facilité mon travail mais aussi rendu mon séjour à Montréal plus agréable.

RÉSUMÉ

NOMBREUSES SONT LES ÉTUDES QUI FONT LE LIEN ENTRE INVESTISSEMENTS PUBLICS EN FAVEUR DES INFRASTRUCTURES ET CROISSANCE ÉCONOMIQUE. PARMI LES PROJETS BÉNÉFIQUES, FIGURENT NOTAMMENT LES OUVRAGES ROUTIERS. SI L'ACCENT EST SOUVENT MIS SUR LA CONSTRUCTION, LES POLITIQUES DE GESTION ET D'ENTRETIEN CONSTITUENT ÉGALEMENT UN SUJET DE PRÉOCCUPATION MAJEUR. TROP INEFFICIENTES DANS LE CAS DU QUÉBEC, ELLES SONT, PAR EXEMPLE, À L'ORIGINE D'UN DÉFICIT D'ENTRETIEN DES INFRASTRUCTURES DE L'ORDRE DE 15,1 MILLIARDS DE DOLLARS, D'APRÈS LES CHIFFRES DE LA COMMISSION ROBILLARD. LEUR ÉTAT POSE AUJOURD'HUI PROBLÈME ET LA SITUATION DEVRAIT EMPIRER DANS LES PROCHAINES ANNÉES SI AUCUN EFFORT SIGNIFICATIF N'EST FAIT POUR REDRESSER LA BARRE.

Ce projet se positionne donc dans un contexte délicat et tente de fournir aux décisionnaires les outils nécessaires pour anticiper les dégradations et planifier plus efficacement les travaux de réhabilitation. C'est tout du moins l'objectif d'un projet plus global dans lequel s'inscrit ce mémoire.

La littérature existante fournit de nombreux exemples dans lesquels les auteurs modélisent, de différentes façons, le processus de dégradation des infrastructures. Néanmoins, ces études restent au stade théorique car elles se heurtent toutes au même problème : le manque de cas concrets permettant d'éprouver les résultats obtenus. L'approche de ce projet se veut diamétralement opposée dans la mesure où les modèles sont élaborés uniquement à partir de données réelles relatives aux infrastructures routières du réseau québécois.

Ainsi, la création d'une base de données exploitable constitue la première grande étape du projet. Un code informatique est développé et utilisé pour obtenir, directement depuis Internet, une première salve d'éléments intéressants. S'ensuit alors un important travail de recherche et de collecte d'informations connexes dans le but d'enrichir cette base de données et la rendre plus complète. Les méthodes statistiques prennent alors le relai pour décrire les données à disposition, déterminer les facteurs dégradants, quantifier leurs effets ou encore prédire l'état d'une structure.

Au final, l'objectif de ce mémoire n'est pas, compte-tenu de l'ampleur de la tâche, de fournir un outil d'aide à la décision fonctionnel et directement utilisable mais de contribuer à son élaboration en dégageant des tendances, des éléments de réponse ou des pistes de réflexion qu'il s'agirait d'étudier plus en détail.

Beaucoup de travail reste donc à accomplir avant la mise au point d'un outil prédictif complet, performant et pouvant véritablement rendre service dans les processus de prise de décisions liées à l'entretien des actifs routiers. Toutefois, ce mémoire peut légitimement être considéré comme une première pierre portée à l'édifice.

ABSTRACT

The link between economic growth and public investments for infrastructures, including road works, is made in many studies. Not only construction issues must be a subject of concern but also maintenance and rehabilitation ones. In Quebec, ineffective policies lead to a \$ 15,1 billion infrastructure rehabilitation gap, according to the Robillard commission report. Their condition is now something we should worry about and it's going to get worse in the next few years if nothing is done to straighten things up.

This project is developed taking into account this difficult context and its aim is to provide decision makers with tools which could be helpful to anticipate deterioration and to plan effectively rehabilitation works. It's at least the goal of a more important project, a part of which is this master's thesis.

Existing literature shows many cases in which authors try to model, in different ways, infrastructure deterioration process. However, these studies remain at the stage of theories because they all face to the same problem: the lack of concrete examples to validate them. The approach of this project is diametrically opposed to these cases because of the choice to build models directly based on real data from the Quebec road infrastructures.

Therefore, creating an exhaustive and workable database can be seen as an important first step. Once the task is completed, statistical methods are used to describe data, identify factors for degradation, quantify their effects or even predict future states of structures using their values.

Concretely, given the size of the task, this master's thesis won't result in the elaboration of a fully functioning decision-making tool but it will contribute to its development, thanks to some found elements of a response or some highlighted tracks of reflexion which deserve to be examined in detail.

Thus, there is much work to be done before having a complete, powerful and useful predictive model but this master's thesis could nevertheless be considered as the foundation stone of the building.

TABLE DES MATIÈRES

REMERCIEMENTS	III
RÉSUMÉ	IV
ABSTRACT.....	VI
TABLE DES MATIÈRES	VII
LISTE DES TABLEAUX	XI
LISTE DES FIGURES	XIV
LISTE DES SIGLES ET ABRÉVIATIONS	XVII
LISTE DES ANNEXES	XIX
CHAPITRE 1 INTRODUCTION.....	1
1.1 Contexte du sujet	1
1.1.1 Importance des infrastructures publiques	1
1.1.2 Des chiffres inquiétants	3
1.1.3 Problématiques soulevées	7
1.2 Littérature existante	8
1.2.1 Modèles de prédiction des dégradations.....	8
1.2.2 L'existence de l'option privée pour les travaux d'entretien	16
1.3 Vers une nouvelle approche statistique	17
1.3.1 Contexte de création de la base de données.....	17
1.3.2 Un indicateur d'état pour les structures routières	18
1.3.3 Recherche des variables d'entrée	19
1.4 Objectifs du projet de recherche	22
CHAPITRE 2 CRÉATION DE LA BASE DE DONNÉES	23
2.1 Détails sur le site du MTQ.....	23

2.2 Extraction des données	24
2.3 Élargissement et réorganisation de la base de données	26
2.3.1 Tri des données	26
2.3.2 Les régions administratives	27
2.3.3 Les zones climatiques.....	28
2.3.4 L'affiliation politique des gouvernements.....	35
2.4 Base de données finale	36
2.4.1 Variables exploitables	36
2.4.2 Observations exclues	37
CHAPITRE 3 MODÉLISATIONS STATISTIQUES.....	39
3.1 Généralités	39
3.1.1 Les variables	39
3.1.2 Modèles statistiques	39
3.2 Traitement des variables quantitatives – Modèles de régression.....	40
3.2.1 Modèles linéaires de régression	40
3.2.2 En cas de multicolinéarité.....	49
3.2.3 Résultats.....	53
3.2.4 Critique	55
3.2.5 Modèles logistiques de régression	56
3.3 Traitement combiné des entrées quantitatives et qualitatives	63
CHAPITRE 4 RÉSULTATS	65
4.1 Préliminaires	65
4.1.1 Variables explicatives.....	65
4.1.2 Démarche	66

4.2 Statistiques descriptives.....	67
4.2.1 Traitement des variables catégoriques.....	67
4.2.2 Traitement des variables quantitatives	82
4.3 Méthodes de régression	86
4.3.1 ICG considéré comme continu.....	86
4.3.2 ICG considéré comme discret.....	96
4.4 Modèle général de régression	107
4.4.1 Résultats obtenus.....	107
4.4.2 Étude des variables significatives	109
4.5 Interactions entre l'âge et les facteurs catégoriques significatifs.....	110
4.5.1 Transformation de l'âge en facteur catégorique	111
4.5.2 Étude des interactions.....	112
4.6 Surévaluation des prédictions	114
4.6.1 Rectification du problème.....	115
4.6.2 Validation	116
4.6.3 Application des modèles prédictifs à de nouvelles données.....	117
CHAPITRE 5 CONCLUSIONS ET RECOMMANDATIONS.....	119
5.1 Réponses aux objectifs	119
5.1.1 Pertinence de la nouvelle approche	119
5.1.2 Identification des facteurs influents	119
5.1.3 État des lieux.....	120
5.2 Extensions et utilités possibles.....	121
5.3 Limites de l'étude.....	124
5.3.1 Critiques.....	124

5.3.2 Axes d'amélioration	124
BIBLIOGRAPHIE ET LISTE DE RÉFÉRENCES	126
ANNEXES	132

LISTE DES TABLEAUX

Tableau 1.1: Codage numérique de l'ICG.....	18
Tableau 1.2: Variables d'entrée des différents modèles de la littérature.....	21
Tableau 2.1: Répartition selon le type de structures	27
Tableau 2.2: Correspondance entre municipalités « hors MRC » et régions administratives	28
Tableau 2.3: Répartition des structures par zone climatique (Köppen-Geiger).....	30
Tableau 2.4: Répartition des structures par zone climatique (Litynski)	34
Tableau 2.5: Liste des variables exploitables.....	36
Tableau 3.1: Tableau ANOVA traduisant la qualité d'une régression	43
Tableau 3.2: Exemple type d'une matrice de corrélation retournée par le logiciel	48
Tableau 3.3: Observations exprimées dans l'espace des composantes principales	51
Tableau 3.4: Vecteurs propres.....	51
Tableau 3.5: Synthèse de la régression.....	53
Tableau 3.6: Synthèse de la régression.....	59
Tableau 3.7: Matrice de contingence.....	61
Tableau 3.8: Estimation des paramètres de la régression logistique	63
Tableau 4.1: Test comparant les valeurs moyennes d'ICG selon le type de ponts.....	68
Tableau 4.2: Test comparant les valeurs moyennes d'ICG selon le matériau (Ponts à poutres) ...	71
Tableau 4.3: Test comparant les valeurs moyennes d'ICG selon le matériau (Ponts en arc).....	72
Tableau 4.4: Test comparant les valeurs moyennes d'ICG selon l'obstacle franchi	78
Tableau 4.5: Classement par ICG moyen des différents types de climat	80
Tableau 4.6: Test comparant les valeurs moyennes d'ICG selon le parti	81
Tableau 4.7: Année de construction moyenne et ICG moyen selon le parti	81
Tableau 4.8: Description des données	82

Tableau 4.9: Matrice de corrélation pour les ponts	83
Tableau 4.10: Signes des coefficients de corrélation	85
Tableau 4.11: Paramètres estimés	87
Tableau 4.12: Analyse des facteurs d'inflation de la variance	88
Tableau 4.13: Analyse des indices de conditionnement.....	89
Tableau 4.14: Sélection pas-à-pas avant.....	90
Tableau 4.15: Sélection pas-à-pas arrière	90
Tableau 4.16: Sélection de variables – Paramètres estimés.....	91
Tableau 4.17: Valeurs propres de la matrice de corrélation	91
Tableau 4.18: ACP – Paramètres estimés	92
Tableau 4.19: Ridge – Paramètres estimés	93
Tableau 4.20: Synthèse et comparatif des résultats pour un ICG continu	95
Tableau 4.21: Défaut – Critères de qualité	96
Tableau 4.22: Ratios des cotes	97
Tableau 4.23: Significativité des variables explicatives.....	98
Tableau 4.24: Niveau 4 de l'ICG – Critères de qualité.....	99
Tableau 4.25: Niveau 4 – Significativité des variables explicatives	100
Tableau 4.26: Niveau 3 de l'ICG – Critères de qualité	101
Tableau 4.27: Niveau 3 – Significativité des variables explicatives	102
Tableau 4.28: Niveau 2 de l'ICG – Critères de qualité	102
Tableau 4.29: Niveau 2 – Significativité des variables explicatives	103
Tableau 4.30: Référence égale à 5 – Critères de qualité.....	104
Tableau 4.31: Multinom. – Significativité des variables explicatives.....	104
Tableau 4.32: Référence égale à 2 – Critères de qualité.....	105

Tableau 4.33: Synthèse et comparatif des régressions logistiques	106
Tableau 4.34: Améliorations apportées par le modèle général	108
Tableau 4.35: Test comparant les valeurs moyennes d'ICG selon l'âge	112
Tableau 4.36: Valeurs moyennes d'âge selon le type de ponts	113
Tableau 4.37: Répartition des structures selon les valeurs de l'ICG	115
Tableau 4.38: Intervalles de confiance pour le modèle général de régression	115
Tableau 4.39: Intervalles de confiance pour le modèle logistique binomial	116
Tableau 4.40: Comparaison entre le modèle général "normal" et l'équivalent corrigé	117
Tableau A.1: Avantages et inconvénients des différents types de ponts	138
Tableau D.1: Classification primaire de Köppen-Geiger	146
Tableau D.2: Classification secondaire de Köppen-Geiger	147
Tableau D.3: Climats résultants de la classification secondaire	147
Tableau D.4: Classification tertiaire de Köppen-Geiger	148
Tableau E.1: Classes climatologiques du Québec (<i>D'après Gerardin et McKenney, 2001</i>)	153
Tableau G.1: Application du modèle général	159
Tableau G.2: Application du modèle logistique de détection de défauts	160

LISTE DES FIGURES

Figure 1.1: Effet de l'investissement public sur le PIB (<i>Boccanfuso, Joanis, Richard et Savard, 2014</i>)	1
Figure 1.2: Évolution du stock de capital public en pourcentage du PIB (<i>FMI, 2014</i>).....	2
Figure 1.3: Dégradation des infrastructures publiques (<i>FMI, 2014</i>).....	2
Figure 1.4: Infrastructures routières – Répartition des dépenses (<i>Conseil du trésor, 2016</i>)	3
Figure 1.5: Âge moyen des ponts et viaducs ramené à la durée de vie utile (2007) (<i>D'après Gagnon, Gaudreault et Overton, 2008</i>).....	4
Figure 1.6: Coût annuel des interventions sur les infrastructures routières (<i>D'après la Direction des transports de Montréal, 2010</i>)	5
Figure 1.7: Évolution de la proportion d'éléments de voirie en bon état (<i>D'après la Direction des transports de Montréal, 2010</i>).....	6
Figure 1.8: Évolution de la proportion de ponts et tunnels en bon état (<i>D'après la Direction des transports de Montréal, 2010</i>).....	6
Figure 1.9: Fonction de survie (<i>D'après Herz, 1996</i>).....	8
Figure 1.10: Taux de défaillance (<i>D'après Herz, 1996</i>).....	9
Figure 1.11: Évolution de l'état du stock de structures (<i>Kleiner, 2001</i>)	12
Figure 1.12: Évolution des coûts estimés (<i>Kleiner, 2001</i>).....	14
Figure 2.1: Exemple de fiche individuelle (<i>Site du MTQ</i>).....	23
Figure 2.2: Récupération des données d'une fiche individuelle	25
Figure 2.3: Mots-clés et décalages associés.....	25
Figure 2.4: Logique de tri des données.....	26
Figure 2.5: Découpage climatologique du Québec (<i>Site du MDDELCC</i>).....	29
Figure 2.6: Zones climatiques au Québec (<i>Gerardin et McKenney, 2001</i>)	31
Figure 2.7: Carte des régions administratives du Québec (<i>Site du MDDELCC</i>).....	32

Figure 2.8: Superposition des cartes des zones climatiques et des régions administratives.....	32
Figure 2.9: Repérage des municipalités du Nord-du-Québec	33
Figure 2.10: Frise chronologique des différents partis au pouvoir au Québec	35
Figure 3.1: Processus de mise en place d'un modèle.....	40
Figure 3.2: Variance relativement constante	44
Figure 3.3: Variance non constante.....	45
Figure 3.4: Tracé des résidus sur l'échelle gaussienne	45
Figure 3.5: Diagramme de Pareto	54
Figure 3.6: Prédictions de l'ICG en fonction des valeurs observées.....	55
Figure 3.7: Adéquation des données aux lois normale et logistique – Comparatif.....	56
Figure 3.8: Courbe ROC (<i>Revue médicale suisse</i>)	62
Figure 4.1: Récapitulatif des variables utilisées.....	65
Figure 4.2: Répartition des ponts selon leur type.....	67
Figure 4.3: ICG moyen en fonction du type de ponts	68
Figure 4.4: Répartition des ponts selon le matériau utilisé.....	69
Figure 4.5: ICG moyen en fonction du matériau (Ponts à poutres)	70
Figure 4.6: ICG moyen en fonction du matériau (Ponts en arc)	72
Figure 4.7: Répartition des ponts selon la classe de routes	73
Figure 4.8: ICG moyen en fonction du type de routes	74
Figure 4.9: Répartition des ponts selon la région administrative.....	75
Figure 4.10: ICG moyen en fonction de la région.....	76
Figure 4.11: Répartition des ponts selon le type d'obstacles	77
Figure 4.12: ICG moyen en fonction de l'obstacle.....	78
Figure 4.13: ICG moyen en fonction du climat	79

Figure 4.14: ICG moyen en fonction du parti.....	81
Figure 4.15: Sélection exhaustive – Critère du R ²	89
Figure 4.16: Ridge – Choix du paramètre biaisant.....	93
Figure 4.17: PLS – Choix du nombre de composantes	94
Figure 4.18: Tracés des prédictions en fonction des valeurs observées	99
Figure 4.19: Distribution de l'âge des ponts	111
Figure 4.20: Répartition des ponts selon leur âge	112
Figure 4.21: ICG moyen selon l'âge et le type de ponts.....	112
Figure 4.22: ICG moyen en fonction de l'âge et du parti	113
Figure 4.23: Tracé des prédictions en fonction des observations pour les nouveaux ponts	117
Figure 5.1: Probabilité d'apparition d'un défaut selon l'âge	122
Figure 5.2: Probabilité d'apparition d'un défaut significatif	123
Figure 5.3: Probabilité d'apparition d'un défaut majeur	123
Figure A.1: Terminologie des ponts (<i>Wikipédia</i>).....	132
Figure A.2: Les différents ponts (<i>Wikipédia</i>)	133
Figure A.3: Différenciation selon la position du tablier	134
Figure A.4: Pont à bêquilles.....	134
Figure A.5: Pont à poutres bow-string.....	135
Figure A.6: Pont à poutres-caissons	136
Figure A.7: Portique double (<i>Wikipédia</i>).....	136
Figure A.8: Pont-dalle (<i>Wikipédia</i>)	137
Figure A.9: Nappe en éventail	138
Figure A.10: Nappe en harpe	138
Figure D.1: Planisphère des différentes zones climatiques (<i>Université de Melbourne, AUS</i>)	149

LISTE DES SIGLES ET ABRÉVIATIONS

Cette liste présente, dans l'ordre alphabétique, les sigles et abréviations utilisés dans le mémoire ainsi que leur signification.

ACP	Analyse en composantes principales
AIC	Akaike information criterion
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
BHP	Béton haute performance
BUP	Béton ultra haute performance
CEP	Circonscription électorale provinciale
CEVC	Commission d'enquête sur le viaduc de la Concorde
CFST	Concrete filled steel tube
CIRANO	Centre interuniversitaire de recherche en analyse des organisations
CRIB	Centre de recherche sur les infrastructures en béton
DJMA	Débit journalier moyen annuel
EA	Évaporation annuelle
FMI	Fonds monétaire international
IA	Indice d'accessibilité
IC	Indice de conditionnement
ICG	Indice de condition générale
ISO	International Organization for Standardization
MDDELCC	Ministère du Développement durable, de l'Environnement et de la Lutte contre les changements climatiques
MERN	Ministère de l'Énergie et des Ressources naturelles

MPC	M ois le(s) p lus c haud(s)		
MPF	M ois le(s) p lus f roid(s)		
MPH	M ois le(s) p lus h umide(s)		
MPS	M ois le(s) p lus s ec(s)		
MRC	M unicipalité r égionale de c omté		
MTQ	M inistère des T ransports du Q uébec		
PA	P récipitations a nnuelles		
PIB	P roduit i ntérieur b rut		
PLQ	P arti L ibéral du Q uébec		
PLS	P artial l east s quares		
PQ	P arti Q uébécois		
PQI	P lan q uébécois des i<td>ROC</td><td>Receiver operating characteristic</td>	ROC	R eceiver o perating c haracteristic
UN	U nion N ationale		
URL	U niform r esource l ocator		
VIF	V ariance i nflation f actor		
UN	U nion N ationale		
URL	U niform r esource l ocator		
VIF	V ariance i nflation f actor		

LISTE DES ANNEXES

ANNEXE A – GÉNÉRALITÉS SUR LES PONTS	132
ANNEXE B – DÉTAIL DES PROCÉDURES D'EXTRACTION	139
ANNEXE C – DÉTAIL DES PROCÉDURES DE TRI	142
ANNEXE D – CLASSIFICATION DE KÖPPEN-GEIGER.....	146
ANNEXE E – CLASSIFICATION DE LITYNKSI.....	150
ANNEXE F – ÉQUATIONS PRÉdictives DES RÉgressions	152
ANNEXE G – APPLICATION AUX NOUVELLES OBSERVATIONS	159

CHAPITRE 1 INTRODUCTION

1.1 Contexte du sujet

L'effet positif des investissements publics sur l'économie d'une région, d'un état ou d'un pays a déjà été démontré et souligné dans plusieurs études. En 2014, un rapport du Fonds monétaire international (FMI) met ainsi en avant une corrélation positive entre l'investissement public et la productivité à court terme, grâce à un effet sur la demande, mais aussi à long terme, grâce à un effet sur l'offre. Pour le FMI, l'investissement public est donc un catalyseur pour l'économie.

Cette conclusion est également partagée par Boccanfuso *et al.* (2014) dans une étude similaire appliquée au Québec. On peut notamment y retrouver l'évolution du produit intérieur brut (PIB) dans le temps par rapport à un scénario de référence basé sur les estimations des économistes et pour lequel la croissance annuelle québécoise serait de 2% (Figure 1.1). Si les différentes courbes représentées caractérisent chacune un mode de financement différent de l'investissement, la tendance qui se dégage est commune à toutes : le PIB augmente jusqu'à atteindre un niveau maximal quatre ans après l'investissement initial (réalisé durant l'année t).

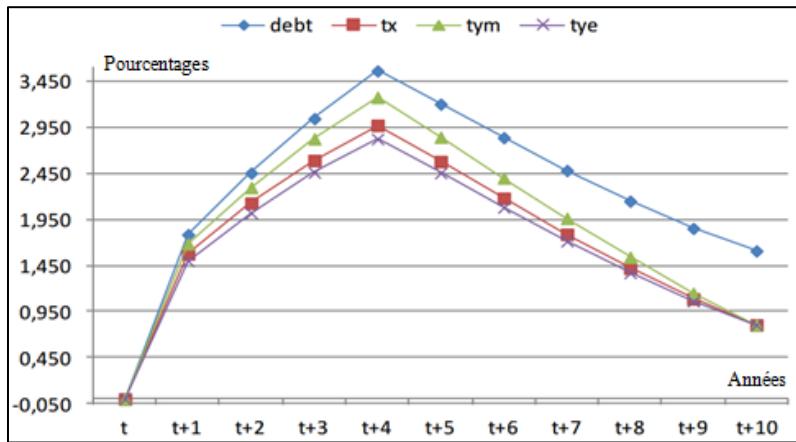


Figure 1.1: Effet de l'investissement public sur le PIB

(Boccanfuso, Joanis, Richard et Savard, 2014)

1.1.1 Importance des infrastructures publiques

Concrètement, l'investissement public se manifeste principalement par des projets de construction, d'entretien, d'amélioration ou de destruction d'infrastructures publiques. Si de tels projets peuvent jouer en faveur de la croissance, leur trop faible nombre peut avoir l'effet inverse.

C'est du moins ce que laisse penser le rapport du FMI puisque les insuffisances en termes d'investissements pour les infrastructures sont identifiées comme une des causes de la récession mondiale de 2008. Ces insuffisances se retrouvent à deux niveaux : une diminution du niveau d'infrastructures publiques par habitant, dans les pays avancés notamment (Figure 1.2), et une dégradation générale de leur état, dans les pays du G7 (Figure 1.3). Autrement dit, des lacunes en termes de construction et d'entretien sont à déplorer.

Le rapport du FMI ne considère pas directement le niveau d'infrastructures mais plutôt le stock de capital public en pourcentage du PIB. Il a néanmoins été démontré que ce dernier était un bon estimateur du nombre d'infrastructures par habitant.

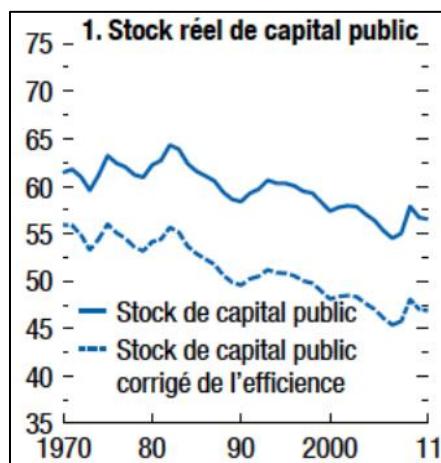


Figure 1.2: Évolution du stock de capital public en pourcentage du PIB (FMI, 2014)

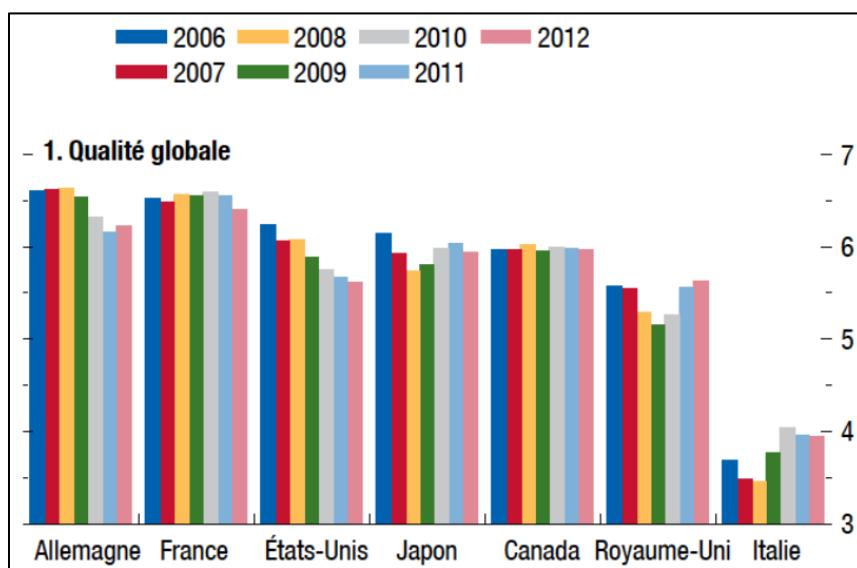


Figure 1.3: Dégradation des infrastructures publiques (FMI, 2014)

Cette dernière figure donne l'évolution de l'état de l'infrastructure dans les pays du G7, grâce à la mise en place d'un système de notes allant de 1 à 7 et augmentant avec la qualité. Elle permet de voir que la situation au Canada n'est pas si mauvaise car l'état des infrastructures reste relativement constant. Néanmoins, le problème précédemment évoqué est bien réel. Dans le cas du Québec, le rapport de la commission Robillard (2015) fait état d'un déficit d'entretien des infrastructures provinciales de l'ordre de 15,1 milliards de dollars. Plusieurs études ont tenté de chiffrer, de façon plus précise, l'ampleur du problème.

1.1.2 Des chiffres inquiétants

1.1.2.1 Importance relative des infrastructures routières

Parmi les nombreux domaines auxquels peuvent se rapporter les infrastructures, Gagné et Haarman (2011) insistent, chiffres à l'appui, sur l'importance de celles qui sont liées au transport. En effet, au Québec, en 2009, la part des investissements publics annuels totaux consacrée aux infrastructures de transport était de 81%, tandis que celles-ci représentaient environ 73% du stock total d'infrastructures publiques. Les auteurs précisent également que les infrastructures routières (réseau routier, ponts et structures) en constituent le poste de dépenses principal.

Pour aller plus loin dans l'analyse, le Conseil du trésor donne, dans son dernier rapport annuel sur le Plan québécois des infrastructures (PQI), la répartition prévue et détaillée des dépenses publiques pour l'année 2016-2017 en ce qui concerne les infrastructures routières (Figure 1.4).

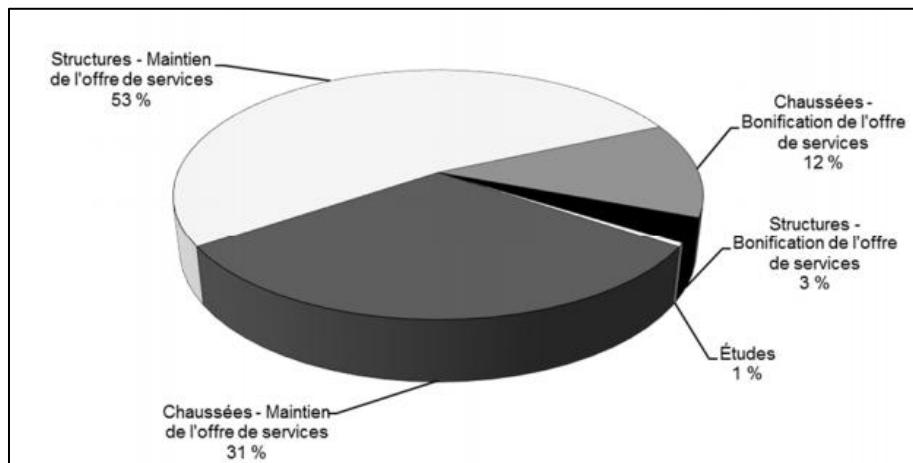


Figure 1.4: Infrastructures routières – Répartition des dépenses
(Conseil du trésor, 2016)

On peut notamment y voir que 53% des dépenses prévues sont consacrées exclusivement au maintien des structures. De manière plus générale, l'entretien et le maintien du réseau routier actuel, structures et chaussées comprises, semble être une préoccupation majeure avec 84% de l'ensemble du budget alloué. Cette inquiétude quant à l'état des infrastructures routières québécoises est donc d'actualité et constitue le sujet principal de ce mémoire. Deux raisons viennent principalement la justifier : la vétusté des structures provinciales et des investissements historiquement trop faibles pour changer la donne.

1.1.2.2 Des infrastructures routières vieillissantes

Dans une étude cherchant à comparer, entre les provinces, l'âge moyen du stock d'infrastructures publiques, Gagnon, Gaudreault et Overton (2008) trouvent des résultats préoccupants pour le Québec. Certes, la tendance générale est au rajeunissement entre 2001 et 2007, avec un âge moyen passant de 18,6 ans à 17,2 ans, mais elle n'est pas valable dans tous les domaines. Ainsi, dans le cas des infrastructures routières, les chiffres sont éloquents : l'âge moyen est en constante hausse depuis 1969 (de 18,1 ans à 31,0 ans en 2007).

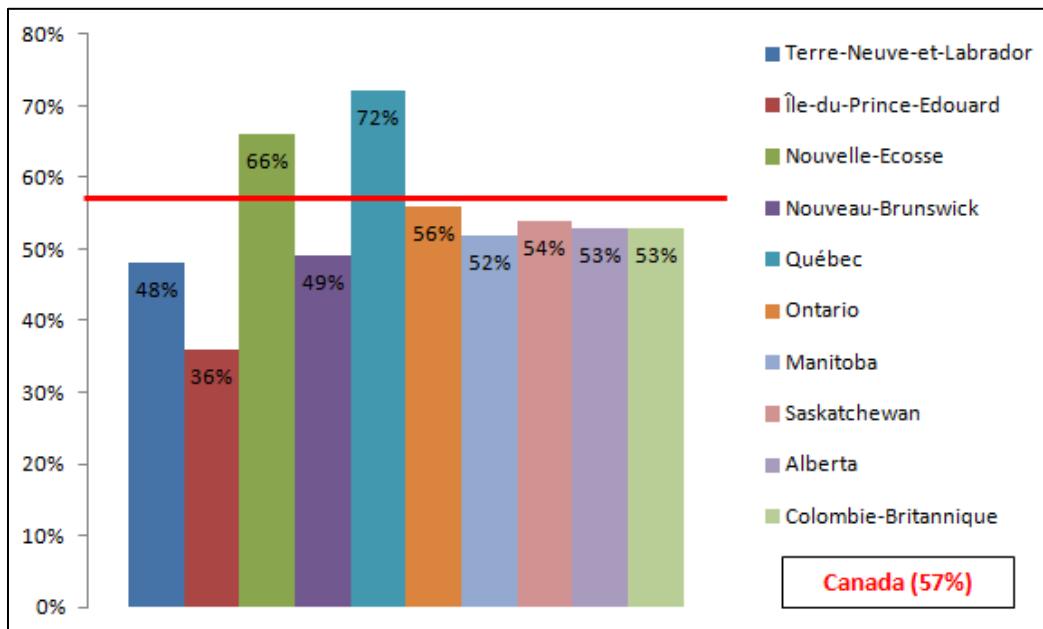


Figure 1.5: Âge moyen des ponts et viaducs ramené à la durée de vie utile (2007)
(D'après Gagnon, Gaudreault et Overton, 2008)

La Figure 1.5, inspirée de l'étude en question, montre que les infrastructures routières québécoises ont même atteint un âge moyen correspondant à 72% de leur durée de vie utile moyenne, ce qui les propulsait, en 2007, à la première place du classement interprovincial des structures les plus anciennes, assez nettement au-dessus des autres et de la moyenne nationale (17 points de plus).

1.1.2.3 Un sous-investissement problématique

Par ailleurs, une étude de la direction des transports de Montréal (2010) fournit des résultats particulièrement intéressants, notamment du point de vue financier, car elle cherchait à convaincre les commissions conjointes de la nécessité d'investir à long terme dans les infrastructures routières de la ville. Les commissions conjointes ont pour rôle de donner leur avis et faire des recommandations quant à l'aménagement et l'urbanisme du territoire visé. Cette étude traduit ainsi une prise de conscience par rapport à la situation actuelle car elle dénonce très clairement « une dégradation des infrastructures résultant du sous-investissement dans la protection (maintien et réhabilitation) des actifs routiers ».

Précisément, l'objectif de cette étude est de démontrer que le budget que la ville propose de débloquer dans cette optique est nettement insuffisant au regard des besoins réels. Celui-ci s'élevait, en effet, à 200 millions de dollars par an sur la période 2010 - 2020 mais ne suffisait même pas à financer les interventions minimales requises pour réhabiliter les structures dont l'état est critique (Figure 1.6). Il est également possible de constater qu'il faudrait rajouter près de 210 millions de dollars par an, soit une hausse de budget de 105%, pour maintenir l'état des infrastructures à leur niveau actuel.

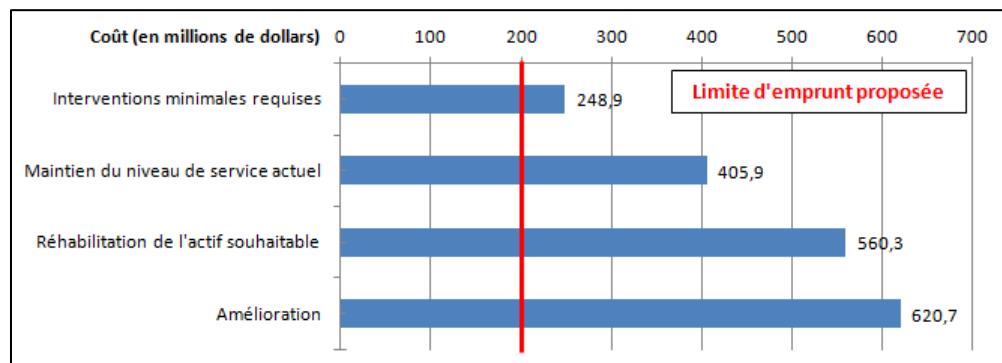


Figure 1.6: Coût annuel des interventions sur les infrastructures routières
(*D'après la Direction des transports de Montréal, 2010*)

La Figure 1.7 et la Figure 1.8 sont issues de ce même rapport et donnent ainsi la proportion de structures en bon état en fonction du temps pour, respectivement, les éléments de voirie et les ponts et tunnels.

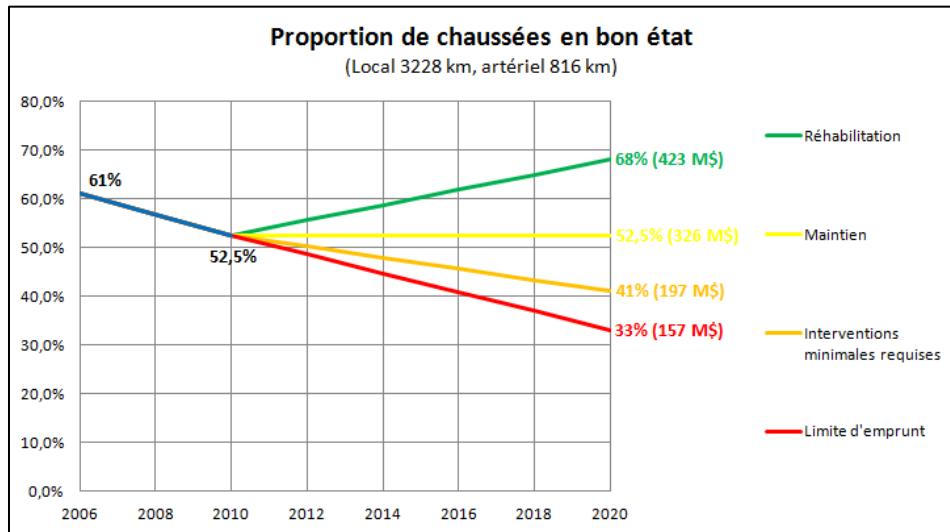


Figure 1.7: Évolution de la proportion d'éléments de voirie en bon état
(D'après la Direction des transports de Montréal, 2010)

Il est alors possible de constater qu'avec les investissements prévus initialement, seulement 33% des chaussées de Montréal seront encore en bon état d'ici 2020. Cela correspondrait à une perte de 19,5 points en dix ans et placerait le niveau 8 points en-dessous du niveau minimal requis.

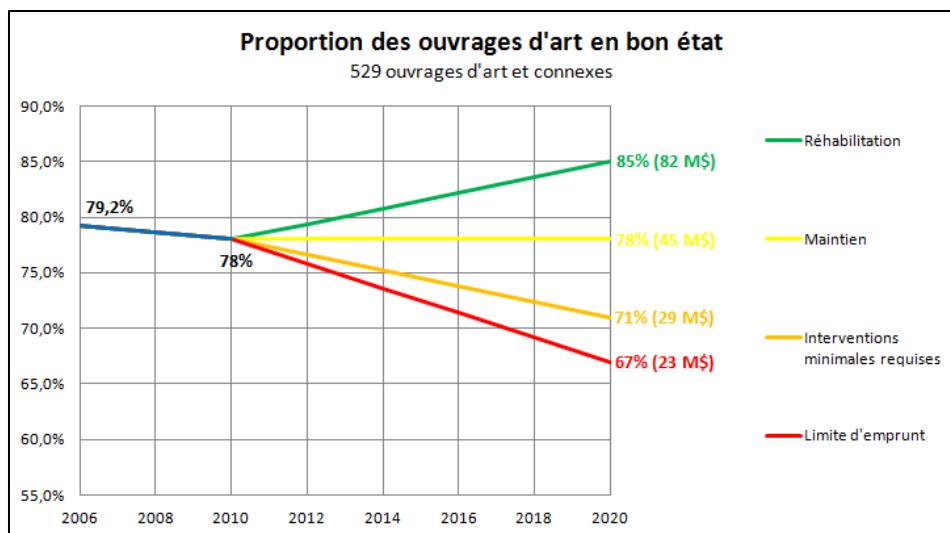


Figure 1.8: Évolution de la proportion de ponts et tunnels en bon état
(D'après la Direction des transports de Montréal, 2010)

Pour ce qui est des ponts et tunnels, le constat global est le même. Ainsi, plus que 67% d'entre eux seront en bon état en 2020, contre 78% aujourd'hui. L'écart avec le niveau minimal requis est bien moindre que dans le cas précédent mais il reste néanmoins négatif (- 4 points).

La conclusion de cette étude est donc sans appel : les efforts financiers consentis par la ville de Montréal pour ses infrastructures sont très inférieurs à ce qu'ils devraient être pour assurer une qualité ne serait-ce que convenable du réseau routier. Cette conclusion peut, en outre, se généraliser à l'ensemble du Québec et rend compte d'une certaine inefficacité du secteur public pour entretenir de façon satisfaisante ses structures. C'est en tout cas l'avis que partage Boyer (2012) sur la question.

1.1.3 Problématiques soulevées

Le Québec accuse ainsi un réel retard sur les autres provinces canadiennes. Cette analyse est confortée par celle de Boyer qui précise que ce retard se retrouve aussi par rapport à d'autres partenaires économiques tels que les États-Unis, par exemple. De plus, la vétusté des infrastructures québécoises, qui tendent vers un âge correspondant à leur durée de vie utile moyenne, laisse penser que le problème pourrait prendre encore plus d'ampleur dans les années à venir.

Pour le résoudre, il précise que de grands changements doivent être opérés et il va même jusqu'à comparer la tâche à l'un des douze travaux d'Hercule. Les investissements prévus ne suffisant pas à renverser la vapeur et améliorer la situation, se pose alors la question de l'efficience de l'entretien des infrastructures routières : est-il possible de mieux utiliser les ressources afin de satisfaire davantage le besoin en réhabilitation ?

Dans cette optique, plusieurs chercheurs ont tenté de mettre en place des modèles mathématiques d'aide à la décision à destination des organismes publics. L'utilité de telles études est de proposer des méthodes qui peuvent être suivies pour que l'entretien des infrastructures se fasse de façon optimale et pas de manière non planifiée, en fonction des urgences.

1.2 Littérature existante

1.2.1 Modèles de prédition des dégradations

Historiquement, plusieurs approches ont été tentées et chacune d'elles repose sur une discréétisation de l'état, c'est-à-dire la définition d'un nombre fini et restreint de modalités d'état que peuvent prendre des infrastructures qui se dégradent (bon, moyen ou mauvais, par exemple). Des modèles mathématiques sont alors élaborés pour prédire le passage d'une modalité à l'autre et caractériser l'évolution de la structure dans chacun des états.

1.2.1.1 Modèle de survie par cohorte (Herz, 1996)

1.2.1.1.1 Principe

En 1996, Herz s'inspire du modèle de survie par cohorte dont l'objet est d'étudier les probabilités de survie d'individus au sein d'un groupe. Il fait ainsi le parallèle entre une cohorte et le stock d'infrastructures. Herz part du principe qu'une réparation, une réhabilitation ou un remplacement de la structure peuvent être perçus comme la création d'une nouvelle structure (naissance d'un individu). Celle-ci va se dégrader dans le temps (vieillissement de l'individu) jusqu'à ce qu'il y ait défaillance (mort de l'individu) et qu'elle doive être remplacée à son tour. Dès lors, Herz met en place une fonction de survie qui correspond à l'évolution de la probabilité que la structure étudiée soit en bon état en fonction de son âge (Figure 1.9).

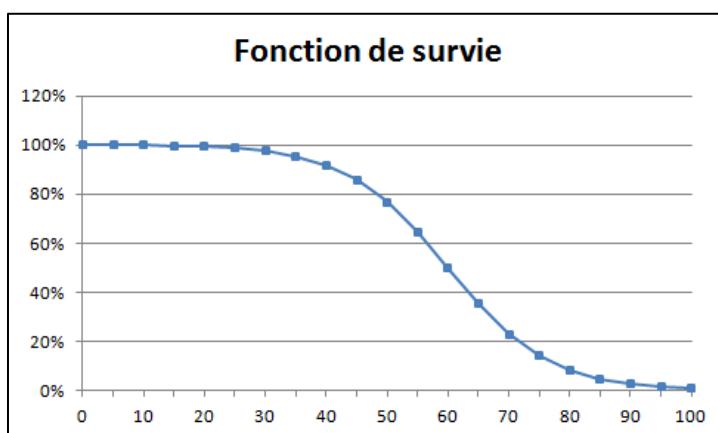


Figure 1.9: Fonction de survie (D'après Herz, 1996)

Il est à noter que deux types de défaillances sont distingués : les défaillances initiales qui sont dues à des défauts de construction ou de conception et les défaillances naturelles dues à l'œuvre du temps sur la structure. Herz définit ainsi une période de résistance en-dessous de laquelle d'éventuelles défaillances ne doivent pas être prises en compte car elles s'expliquent par des défauts initiaux et la correction de ces derniers relève du service de garantie.

Pour un stock d'infrastructures aux caractéristiques similaires, cette courbe peut également être interprétée comme la proportion de celles qui sont en bon état.

À partir de cette fonction de survie, il évalue la durée de vie restante d'une structure en faisant le rapport entre la probabilité de survie à la date de construction et celle à la date d'aujourd'hui. Plusieurs autres fonctions intéressantes peuvent également être définies. En dérivant la fonction de survie, il détermine la densité de probabilités de durée de vie des infrastructures. En lui appliquant la formule de l'espérance mathématique pour les variables continues à densité, la valeur moyenne de durée de vie est trouvée. Enfin, pour une année $t + 1$, il est possible d'évaluer, par itération, le stock d'infrastructures en bon état à partir du stock et du taux de défaillance (Figure 1.10) de l'année t .

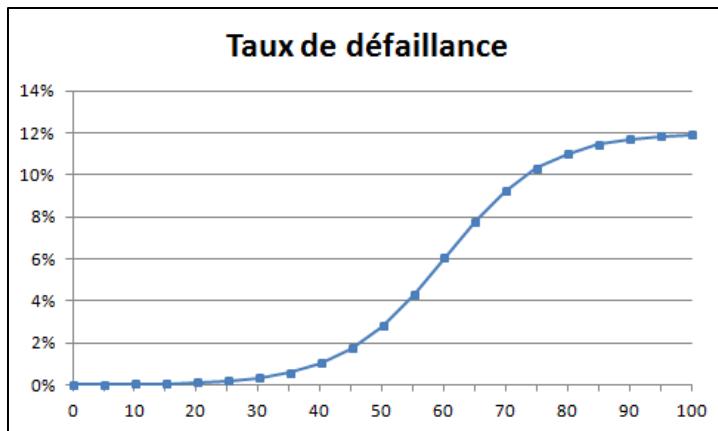


Figure 1.10: Taux de défaillance (*D'après Herz, 1996*)

Le besoin de renouvellement pour l'année $t + 1$ est alors défini comme la remise à neuf de l'ensemble des structures s'étant dégradées au cours de l'année t .

Herz semble donc définir, dans son étude, un cadre théorique assez complet pour anticiper et quantifier les dégradations du stock d'infrastructures. Le problème principal est désormais de voir comment les outils qu'il propose peuvent être utilisés en pratique.

1.2.1.1.2 Estimation des paramètres

L'application pratique passe en premier lieu par la détermination des différents paramètres intervenant dans les fonctions définies. Ils dépendent du type de structures considéré, du matériau utilisé ainsi que du type de chargement appliqué à la structure. Leur valeur peut être trouvée à partir d'observations sur le terrain.

Cependant, du fait du manque de données relatives à l'état et à l'entretien d'infrastructures publiques, Herz précise qu'il est, dans la plupart des cas, impossible d'estimer les coefficients par quelque méthode statistique que ce soit. Il propose donc de s'en remettre principalement à l'avis d'experts sur le sujet et d'ajuster ensuite les paramètres estimés en fonction des futures données collectées.

1.2.1.1.3 Conclusion et critiques

Cette étude marque un premier pas important dans le processus d'aide à la gestion des infrastructures mais elle présente une limite majeure : si elle permet de quantifier les opérations d'entretien nécessaires et leurs conséquences, elle ne dit rien sur la faisabilité et la priorisation des projets de maintenance.

D'autres éléments de cette étude peuvent également être remis en question. On pense notamment au fait que l'auteur considère qu'une réhabilitation est équivalente à un remplacement de la structure défaillante puisqu'il s'agit dans les deux cas du point de départ d'un nouveau cycle de vie. La différence entre les deux situations est en fait prise en compte dans les valeurs attribuées aux paramètres du modèle mais l'interprétation est bien plus délicate que si la distinction avait été faite d'entrée.

Ces paramètres posent également problème dans la compréhension des facteurs jouant sur la vitesse de dégradation des structures. En effet, si l'âge est reconnu comme étant le principal facteur dégradant, le type de structures, le matériau ainsi que le chargement peuvent également jouer un rôle. Le problème est que l'effet de chacun de ces facteurs est capté par la valeur des paramètres et il est alors difficile de quantifier leur impact réel sur la dégradation de l'infrastructure.

Enfin, le dernier défaut que l'on peut trouver à cette étude, et non des moindres, est le fait qu'elle ne fasse qu'évoquer la possibilité de relier la méthode proposée à d'éventuels coûts. La non prise

en compte de l'aspect financier dans la planification des opérations est donc à déplorer. C'est d'autant plus dommageable que l'on sait que la contrainte budgétaire est le principal frein aux projets publics d'entretien des infrastructures. Une méthode alternative, proposée par Kleiner (2001) et basée sur la chaîne de Markov, a notamment le mérite de pallier ce manque et de prendre en compte trois types de coûts différents : les coûts dus à la défaillance, les coûts d'inspection et les coûts d'intervention.

1.2.1.2 Modèle de la chaîne de Markov

1.2.1.2.1 Les bases de la théorie (Abraham et Wirahadikusumah, 1999)

Tout comme la méthode précédente, le procédé de Markov se base sur des valeurs discrètes d'état. L'idée est à nouveau de décrire le processus de dégradation des infrastructures mais, ce coup-ci, l'approche est résolument probabiliste. La chaîne de Markov est définie par la probabilité de transition entre un état $x = i$ et un état $x = j$. Cette probabilité est notée $p_{i,j}^{t,t+1}$, où i et j sont les variables d'espace d'états et t et $t + 1$ sont les variables de temps associées au passage de i à j .

Le principe essentiel de la méthode de Markov est le fait que le futur ne dépende que du présent et pas du passé. Autrement dit, si i est l'état actuel (temps t) et j l'état futur (temps $t + 1$) :

$$p_{i,j}^{t,t+1} = p(x_{t+1} = j | x_0 = i_0, x_1 = i_1, \dots, x_t = i) = p(x_{t+1} = j | x_t = i)$$

Il est alors possible de mettre en place une matrice de transition P telle que :

$$P^{t,t+1} = \begin{pmatrix} p_{1,1}^{t,t+1} & \dots & p_{1,n}^{t,t+1} \\ \vdots & \ddots & \vdots \\ p_{n,1}^{t,t+1} & \dots & p_{n,n}^{t,t+1} \end{pmatrix} \text{ avec } \sum_{j=1}^n p_{i,j}^{t,t+1} = 1, \forall i \in [1, n]$$

La lecture de P se fait ligne par ligne. En prenant l'exemple de la première ligne, l'état initial de la structure est au niveau 1. Il y a une probabilité de $p_{1,1}$ qu'il y reste, une probabilité $p_{1,2}$ qu'il passe au niveau 2, une probabilité $p_{1,3}$ qu'il passe au niveau 3, etc.

Il est également possible de définir un vecteur de masse de probabilité $A(t)$ pour une structure. Ses composantes a_i^t sont les probabilités que ladite structure soit dans l'état i à l'instant t . Pour un stock d'infrastructures, a_i^t représente donc la part de structures dont l'état est au niveau i à

l'instant t . Concrètement, si t est l'instant présent, il suffit d'observer l'état du stock aujourd'hui pour obtenir $A(t)$.

Par itération et grâce aux matrices de transition, il est alors possible de prévoir l'état probable d'une structure ou d'un stock dans plusieurs années. Graphiquement, cela se représente par le tracé de l'évolution des probabilités associées à chacun des niveaux de dégradation en fonction du temps (Figure 1.11).

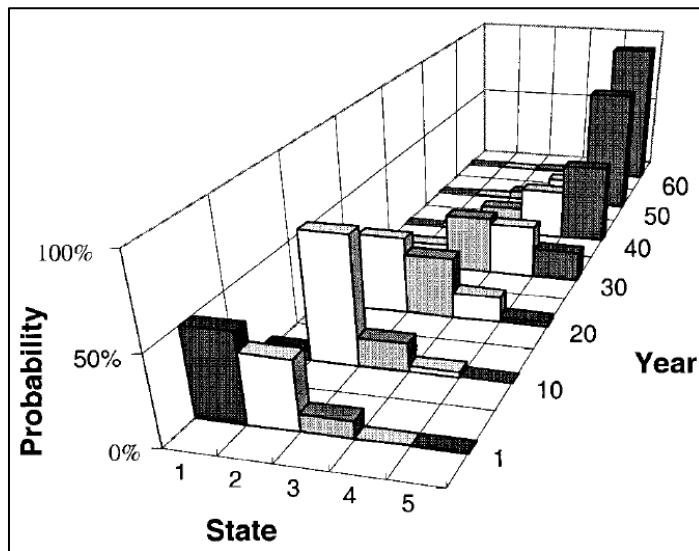


Figure 1.11: Évolution de l'état du stock de structures
(Kleiner, 2001)

1.2.1.2.2 Développement du modèle par Kleiner (2001)

Tout d'abord, deux hypothèses sont faites par Kleiner pour simplifier l'approche initiale selon la méthode de Markov. La première est de considérer que la dégradation n'a lieu que par étapes successives : il est impossible que l'état passe directement d'un niveau i à $i + 2$, par exemple. La seconde consiste à dire que l'évolution de l'état des structures ne peut se faire que dans un sens, celui de la dégradation. Ainsi, si le niveau 1 représente le meilleur état, l'évolution se fera nécessairement de 1 vers 2, puis de 2 vers 3, etc. La matrice de transition s'en trouve alors grandement modifiée puisque seule la diagonale de l'identité et la diagonale immédiatement supérieure ne sont pas remplies de zéros.

S'il est difficile de remettre en cause le bien fondé de la seconde hypothèse, la première est plus contestable dans la mesure où, en réalité, rien ne permet d'affirmer avec certitude que la

dégradation est forcément progressive (un état à la fois). Cette remarque s'inscrit en fait dans le cadre d'une critique plus générale, qui peut être émise envers un postulat implicite et commun aux deux méthodes étudiées. Celles-ci ne tiennent compte que des dégradations naturelles des structures, principalement dues au temps. Il s'agit d'une simplification importante de la réalité car les structures ne sont pas à l'abri d'un évènement causant directement leur défaillance (bris pour des canalisations souterraines, section d'un câble de télécommunication, etc.). Globalement, on peut donc reprocher aux études de Herz (1996) et Kleiner (2001) de ne pas avoir pris en compte le risque d'incidents à travers un coefficient d'ajustement ou la définition d'un intervalle de confiance pour les valeurs trouvées, par exemple.

Le vecteur de masse de probabilité $A(t)$ se déduisant directement de l'observation, la "seule" difficulté de la méthode est de mettre en place les probabilités de transition. Pour résoudre ce problème, Kleiner introduit la notion de « temps d'attente ». Il est possible de le voir comme la durée que va mettre une structure, dans un état donné, à se dégrader au point de passer à l'état suivant. Ce « temps d'attente », assimilable à un temps de survie donc, est noté T_{ij} ou encore T_i sachant que j est forcément égal à $i + 1$ compte-tenu des hypothèses simplificatrices vues précédemment.

Kleiner définit ensuite la loi que la variable aléatoire T_i suit. Il s'agit d'une loi de distribution de Weibull à deux paramètres, λ_i et β_i . Le choix semble particulièrement pertinent dans le sens où elle permet de tester indirectement un grand nombre de lois en jouant sur les valeurs des paramètres. En découlent alors une densité de probabilité, une fonction de survie ainsi qu'une estimation de la durée de vie de la structure. Cependant, cette démarche ne fait que déplacer le problème : au lieu d'estimer les probabilités de transition, il faut désormais estimer les paramètres λ_i et β_i du modèle de Weibull. La difficulté à laquelle est confronté Kleiner est alors le même que pour la méthode de survie par cohorte : le manque de bases de données relatives à l'état d'un stock d'infrastructures empêche une estimation statistique. L'avis des experts sur le sujet est donc à nouveau nécessaire à la détermination des valeurs de paramètres à retenir.

Si elle s'arrêtait là, cette étude n'apporterait rien de plus que la précédente car le type de résultats obtenus et la limite majeure sont les mêmes. Néanmoins, avec cinq ans de recul sur les travaux de Herz, Kleiner prend en compte la critique principale émise à l'encontre des travaux du premier

cité et consacre une partie entière de son étude à l'aspect financier du problème de gestion des infrastructures.

Il définit ainsi trois coûts différents dont il faut tenir compte, les coûts de défaillance C_F , les coûts d'inspection C_I et les coûts d'entretien C_R , dans les processus de prise de décision liés à la maintenance. Kleiner cherche alors à estimer le coût total qui en résulterait dans le futur en tenant compte du taux d'actualisation.

La mission du décisionnaire est de trouver le moment idéal pour réaliser les travaux d'entretien, de réhabilitation ou de renouvellement. Pour cela, il peut se baser, en théorie, sur la courbe représentative du coût total estimé, qui est la somme des différents coûts mentionnées (Figure 1.12). La date t^* , correspondant à la minimisation du coût total, est ainsi la date idéale. Elle est suffisamment lointaine, temporellement parlant, pour profiter de l'effet réducteur de l'actualisation, tout en limitant les risques de défaillance encourus, captés par C_F .

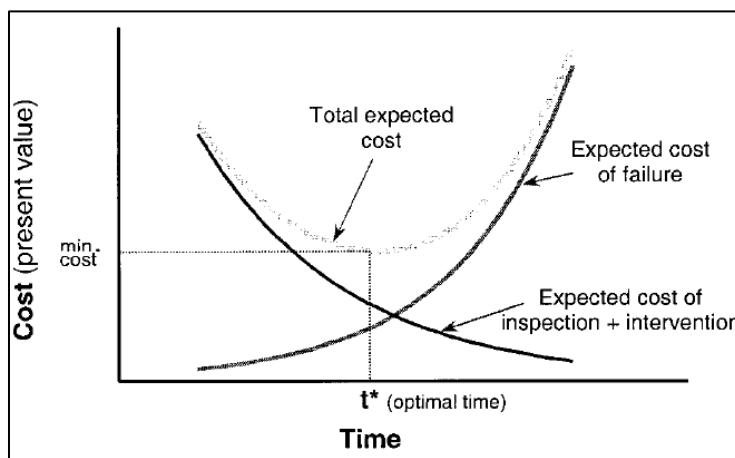


Figure 1.12: Évolution des coûts estimés (Kleiner, 2001)

Si cette étude corrige la principale limite de celle de Herz grâce à la prise en compte des coûts associés aux opérations d'entretien, elle ne va néanmoins pas aussi loin que l'on aurait pu espérer dans la mesure où elle ne fournit aucune indication sur la façon de les calculer. Elle ne fait, effectivement, que définir la stratégie de planification qui doit être adoptée à partir des coûts et de leur évolution dans le temps.

C'est dans cette optique que peut être utilisée l'étude de De Marcellis-Warin et Peignier (2015) sur l'évaluation des coûts socio-économiques dans le cas de bris d'infrastructures souterraines. Bien que celle-ci se limite aux canalisations, conduits et câbles enfouis (réseaux d'eau, de

télécommunication, d'électricité et de gaz), la logique mise en avant et les méthodes de détermination de coûts peuvent être adaptées à d'autres types d'infrastructures. L'intérêt n'est pas, ici, de lister tous les coûts recensés par De Marcellis-Warin et Peignier mais juste de préciser que des travaux de ce genre existent et sont disponibles. Il est également intéressant de noter que l'étude en question se distingue des autres car elle met particulièrement l'accent sur les coûts indirects, c'est-à-dire les coûts liés aux conséquences des défaillances.

Le côté très mathématique de la méthode mise en place par Kleiner est également critiquable, non pas pour sa justesse, mais pour son "non-rattachement" à une réalité plus tangible. Ainsi, si le processus de dégradation des structures est modélisé et peut être quantifié, rien n'est dit sur les facteurs physiques qui induisent ladite dégradation.

1.2.1.3 Conclusion

Les méthodes évoquées sont orientées vers des applications aux réseaux de canalisations. L'approche qu'elles proposent est néanmoins intéressante car elle répond à un réel besoin de méthodes quantitatives pour la planification de l'entretien des structures.

En effet, la seule étude chiffrée menée jusqu'alors sur le problème était celle de l'International Organization for Standardization (1997). La méthode dite « du facteur » qui y est décrite consiste simplement à appliquer un coefficient multiplicateur, k , à la durée de vie de référence d'un élément de structure pour obtenir une estimation de sa durée de vie probable en réalité. Si la structure fait l'objet d'une maintenance poussée, k sera supérieur à 1 tandis que pour un environnement très agressif, il sera inférieur à 1. Hormis sa simplicité, qui constitue évidemment un inconvénient majeur, cette méthode nécessite une documentation très complète (pour connaître les durées de vie de référence) et est bien trop lourde à mettre en œuvre pour les ouvrages les plus complexes (une durée de vie estimée pour chaque élément de l'ouvrage).

Les méthodes de Herz et Kleiner sont donc prometteuses mais le manque de bases de données relatives à l'état des infrastructures publiques ne permet pas réellement de les éprouver. Si des efforts importants sont nécessaires quant à la politique de maintenance des infrastructures, un énorme travail d'inventaire doit donc aussi être réalisé par les organismes publics. De Marcellis-Warin est bien consciente de cela et propose ainsi, en 2014, sur la demande de l'Union des municipalités du Québec, un guide pour les aider à répertorier de manière simple et efficace les structures dont elles ont la charge.

Toutefois, le ministère des Transports du Québec (MTQ) avait, contraint par les circonstances il est vrai (cf. partie 1.3.1 – Contexte de création de la base de données), pris les devants en se lançant dans le recensement de ses propres structures routières. Les résultats de cet inventaire sont aujourd’hui disponibles sur Internet et on peut y voir une excellente opportunité de développer un modèle statistique basé sur des observations concrètes. Cela permettrait ainsi de s’affranchir du problème lié au manque de données mis en avant dans les études de Herz et Kleiner.

1.2.2 L’existence de l’option privée pour les travaux d’entretien

Une solution alternative pour que l’entretien des infrastructures publiques soit mieux réalisé est de laisser les partenaires privés s’occuper de cette tâche. C’est l’option qui a notamment été privilégiée pour la construction du nouveau pont Champlain à Montréal. En effet, au cours d’une conférence ayant eu lieu le 23 février 2016 dans les locaux de l’École Polytechnique de Montréal, G. Mailhot, ingénieur en chef du projet, révélait que l’entretien du nouveau pont incomberait, dans un premier temps, au partenaire privé. Cette méthode est intéressante dans le sens où l’expertise du secteur privé est généralement plus élevée. De plus, le partenaire privé aura un budget de maintenance alloué spécifiquement à ce projet tandis que le gouvernement aurait, lui, dû faire face à l’habituel problème d’allocation de ses ressources : quels choix faire en termes d’entretien compte-tenu de sa contrainte budgétaire ? Il faut ajouter à cela le fait que le pont fera l’objet d’une inspection complète d’ici trente-cinq ans, avant d’être rétrocédé au gouvernement.

En outre, l’introduction du privé dans le domaine de la maintenance d’infrastructures est présentée par Boyer (2012) comme un enjeu crucial pour le développement économique du Québec. Ce changement important est bien entendu contesté car il modifie la façon de faire classique. Il est néanmoins nécessaire de prendre conscience des nombreux avantages, autres que ceux cités précédemment, qu’il apporterait :

- ✓ Une hausse de l’efficacité des méthodes d’entretien grâce à la concurrence entre les entreprises ;
- ✓ L’apprentissage du public au contact du privé, c’est-à-dire des compétences et des connaissances accrues pour le secteur public ;

- ✓ Une comptabilité gouvernementale facilitée car les services de l'entrepreneur sont tarifés mais le gouvernement n'a plus à craindre et à assumer, en théorie, dépendamment du type de contrat passé, un éventuel dépassement de coûts ;
- ✓ La diminution de l'influence politique inhérente à la prise de décision publique.

Cependant, cette solution présente un inconvénient majeur qui est la dépendance du secteur public vis-à-vis du privé. Il serait alors nécessaire pour le gouvernement de gérer efficacement ce type de partenariats pour éviter la capture du public par le privé.

1.3 Vers une nouvelle approche statistique

1.3.1 Contexte de création de la base de données

Suite à l'effondrement du viaduc de la Concorde le 30 septembre 2006 à Laval, le gouvernement du Québec a mis en place un plan d'inspection complète de l'ensemble des infrastructures routières de son réseau. Une telle inspection avait plusieurs finalités. Tout d'abord, elle a permis d'effectuer un bilan exhaustif de l'état des structures et d'effectuer ensuite les travaux nécessaires pour éviter une nouvelle catastrophe comme celle de Laval. En effet, plusieurs raisons ont été évoquées pour justifier l'effondrement du pont, dont des défaillances au niveau de la qualité et du contrôle de son entretien mises en lumière par le rapport de la Commission d'enquête sur le viaduc de la Concorde (CEVC). Une vaste base de données listant les structures et certaines de leurs caractéristiques a également pu être créée grâce à cette enquête. Enfin, le dernier objectif poursuivi par ce plan d'inspection était de regagner la confiance de la population en rendant consultable, sur le site du MTQ, cette base de données. Ainsi, la transparence des informations permet au public de se tenir informé de l'état des structures routières québécoises et des travaux prévus pour les améliorer.

Avec 5359 structures répertoriées au 15 juin 2015, l'idée du projet est d'utiliser cette large base de données, disponible en libre accès, comme support pour la mise en place de modèles statistiques de prédiction basés uniquement sur les observations. On s'affranchit ainsi du principal problème des modèles de la littérature, à savoir une quantité de données insuffisante pour faire le lien entre la théorie et la réalité.

1.3.2 Un indicateur d'état pour les structures routières

Cette idée est d'autant plus pertinente que l'on peut trouver, dans cette base de données, un indicateur de l'état des structures, l'indice de condition générale (ICG). Il est établi en fonction de deux facteurs : la dégradation du matériau et le fait que la structure puisse encore remplir le rôle pour lequel elle a été conçue. Quatre modalités pour l'ICG sont alors définies et retrouvent les différents états dans lesquels peut être l'ouvrage :

- ✓ Un remplacement de la structure est nécessaire car son état est préoccupant ;
- ✓ L'état de structure requiert des travaux majeurs de réhabilitation ;
- ✓ Des réparations sont à effectuer ;
- ✓ La structure est en bon état : aucune intervention n'est nécessaire.

Il peut alors être intéressant de mettre en place une échelle numérique pour traduire ces différentes modalités. La solution retenue est d'attribuer aux structures une note allant de 2, pour celles dont l'état est le pire, à 5 (Tableau 1.1).

Tableau 1.1: Codage numérique de l'ICG

Modalité	Codage numérique
Structure ne nécessitant aucune intervention	5
Structure nécessitant des réparations	4
Structure nécessitant des travaux majeurs	3
Structure nécessitant un remplacement	2

Cette solution présente, outre le fait de rendre l'ICG directement exploitable comme variable de sortie dans les modèles statistiques, l'avantage de décrire de façon logique l'évolution de l'état des structures. En effet, un ouvrage en bon état se verra attribuer une note maximale de 5, note qui diminuera ensuite en fonction des dégradations subies. En modélisant l'état comme une variable numérique discrète ordonnée, on va ainsi dans le sens de ce qui a été fait dans la littérature existante.

Remarque :

Les notes 0 et 1 ont été attribuées, respectivement, et à titre indicatif, aux structures dont l'inspection n'a pas encore eu lieu et à celles pour lesquelles elle est en cours. Elles n'ont donc pas été reportées dans le Tableau 1.1 car elles ne suivent pas l'évolution logique dont il est question précédemment. En effet, cette notation est arbitraire et remplacer 0 et 1 par 6 et 7, par exemple, est tout aussi valable.

1.3.3 Recherche des variables d'entrée

Dans un premier temps, il est donc nécessaire de récupérer les données issues du site du MTQ. Il faut ensuite les compléter à l'aide d'informations pertinentes que l'on peut rechercher parmi un large éventail de domaines tels que l'ingénierie, la géographie, la climatologie ou encore la politique, par exemple. La finalité de ce travail préliminaire est d'obtenir une base de données la plus complète possible, à partir de laquelle vont être développés les modèles.

Outre les facteurs intrinsèquement liés aux structures tels que les dimensions ou les matériaux de construction utilisés, il est possible de s'inspirer des différentes études menées sur la dégradation des infrastructures pour établir une première liste de variables explicatives habituellement prises en compte.

Tout d'abord, Madanat et Wan Hashim (1995) se basent sur quelques ponts de l'Indiana, États-Unis, pour essayer d'évaluer les probabilités de transition issues du modèle de la chaîne de Markov. Il est intéressant de noter que, dans la modélisation qui en résulte, la détérioration est exprimée comme une fonction de l'âge de la structure et des sollicitations. Il en est de même dans l'étude de DeStefano et Gravas qui, en 1998, appliquent ce raisonnement à certains ponts de l'état de New-York, États-Unis. Le nombre de facteurs pris en compte reste donc très faible. D'autres types d'infrastructures, à savoir les canalisations et réseaux d'égout, conduisent, en revanche, à des études dans lesquelles plus de variables interviennent dans le processus de dégradation. On peut alors tenter de faire le parallèle entre ces variables et d'éventuels équivalents pour les ponts et autres structures routières.

En plus de l'âge, Abraham et Wirahadikusumah (1999) prennent ainsi en compte les agressions du milieu sur l'infrastructure. Dans le cas des canalisations, cela renvoie à la nature du sol mais il est possible d'étendre cette réflexion, pour les infrastructures routières, à l'ensemble des

conditions extérieures pouvant intervenir. On pourrait notamment penser aux conditions climatiques et météorologiques telles que les températures, les précipitations, le vent, l'humidité, etc. Une première variable climatologique s'imposerait donc. Cela est d'ailleurs confirmé par une étude de Gagné et Haarman (2011) dans laquelle les auteurs définissent le climat comme l'un des enjeux déterminants et spécifiques au Québec quant à l'entretien des infrastructures routières.

Par ailleurs, la prise en considération du milieu extérieur peut également se traduire, pour les ponts et les ponceaux, au niveau de l'obstacle franchi. En effet, il est légitime de penser qu'un fleuve n'aura pas le même impact sur l'état de la structure qu'une route de campagne, par exemple.

D'autres facteurs retiennent également l'attention d'Abraham et Wirahadikusumah et peuvent trouver écho dans les informations fournies par la base de données du MTQ. Le régime hydraulique, caractérisé, notamment, par le débit du fluide, peut logiquement être assimilé au débit journalier moyen annuel (DJMA), celui-ci étant une estimation du débit total de circulation sur une journée moyenne de l'année. Par ailleurs, l'effet de la taille des déchets sur la canalisation peut être comparé à l'impact de la taille des différents véhicules sur la structure. L'information sur le pourcentage de camions dans le DJMA présente alors un intérêt certain.

Toujours en ce qui concerne l'état des infrastructures souterraines, diverses études mettent en avant d'autres facteurs potentiellement influents. La localisation joue un rôle d'après McDonald et Zhao (2001) tandis que d'autres, à l'image de Rostrum *et al.* (1999), insistent plutôt sur l'importance de l'historique de dégradation et d'entretien.

Plus récemment, plusieurs travaux permettent de recentrer le sujet sur les infrastructures routières. On peut notamment citer les rapports de Gagné et Haarman, dont il était déjà question pour la variable climatique, ou ceux de l'Association mondiale de la route (2014). Plutôt que de chercher les facteurs ayant un effet direct sur la structure, ces deux études ont en commun de présenter des paramètres qui impactent les opérations de maintenance et donc, de manière indirecte, l'état de ladite structure. Gagné et Haarman insistent ainsi sur l'importance de l'aspect politique et sur celle du contexte d'investissements. Pour illustrer leurs propos, ils prennent l'exemple de l'essor du secteur de la construction dans les années 60 à 80 au Québec¹, période

¹ Sous l'impulsion notamment de l'Exposition universelle de 1967 et des Jeux olympiques de 1976, à Montréal.

propice aux investissements en faveur des infrastructures routières, que ce soit construction ou entretien. Ils l'opposent alors aux périodes de ralentissement économique liées aux récessions de 1981 à 1983 et de 1990 à 1994, durant lesquelles la maintenance des actifs routiers aurait été "négligée". Quant au rapport de l'Association mondiale de la route, l'accent est mis sur des facteurs davantage en rapport avec la structure elle-même tels que le type de réseau routier.

Parmi toutes les variables possiblement explicatives fournies par la littérature, certaines peuvent être directement obtenues depuis le site du MTQ. D'autres, en revanche, sont à ajouter et nécessitent un travail supplémentaire de recherche et d'obtention de données. Le Tableau 1.2 récapitule ce qu'il en est pour chacune d'elles.

Tableau 1.2: Variables d'entrée des différents modèles de la littérature

Variables possiblement explicatives	Disponibles sur le site du MTQ ?
Dimensions	Oui, directement
Matériaux de construction	Oui mais indirectement
Âge	Oui mais indirectement
Type de sollicitations	Oui mais indirectement
Climat	Non
Obstacle franchi	Oui mais indirectement
Débit de véhicules	Oui, directement (DJMA)
Pourcentage de camions	Oui, directement
Localisation	Oui, directement (latitude et longitude)
Historique d'entretien	Non
Aspect politique	Non
Contexte d'investissements	Non
Type de réseau routier	Oui, directement (classe de route)

1.4 Objectifs du projet de recherche

Enrichir, dans la mesure du possible, la base de données du MTQ grâce aux différents éléments dont les chercheurs se servent dans la littérature constitue ainsi le premier objectif de ce projet de recherche. De plus amples détails sur la manière de procéder sont fournis par la suite (partie 2.3 – Élargissement et réorganisation de la base de données).

Par ailleurs, ce projet s'inscrit dans la lignée des différentes études de la littérature dans la mesure où l'un des objectifs finaux est d'anticiper la dégradation des structures, grâce à l'ICG, et donc offrir aux organismes publics chargés de l'entretien un nouvel outil d'aide à la planification. Néanmoins, il s'en démarque totalement par la démarche suivie. En effet, plutôt que de penser des modèles mathématiques théoriques qu'il faut ensuite valider, on s'appuie directement sur la réalité pour élaborer des modèles statistiques. Le projet peut alors être décomposé en deux parties aussi importantes l'une que l'autre.

L'un des intérêts majeurs des méthodes statistiques réside dans la possibilité d'analyser les données à disposition. À partir de la base de données créée, il est alors possible de dresser un état des lieux du réseau québécois d'infrastructures routières. La mise en lumière de certaines tendances intéressantes quant à la situation actuelle peut ainsi être visée dans le cadre de ce projet.

Vérifier que la nouvelle approche proposée est viable constitue également un objectif crucial. Autrement dit, les méthodes statistiques permettent-elles de prédire la dégradation des structures de manière satisfaisante ? Pour répondre à cette question par l'affirmative, les modèles développés devront fournir des résultats satisfaisants et être robustes.

Enfin, la volonté affichée de déterminer les variables jouant un rôle véritable dans le processus de détérioration est également une finalité propre à ce projet. Cela pourrait servir de point de départ à d'autres projets visant à étudier l'impact de l'un de ces facteurs de manière plus spécifique.

CHAPITRE 2 CRÉATION DE LA BASE DE DONNÉES

Pour la mise en place des modèles statistiques, on s'appuie, comme mentionné dans le Chapitre 1, sur les données du MTQ relatives aux infrastructures routières québécoises, l'objectif étant de les utiliser directement ou les relier à d'autres variables intéressantes. Il faut donc, dans un premier temps, créer un code permettant de récupérer et organiser les informations du site du ministère. Celles-ci sont, en effet, publiques mais pas forcément toujours exploitables en l'état.

2.1 Détails sur le site du MTQ

Avant de rentrer dans les détails de la méthodologie de codage, il est intéressant de savoir à quoi correspondent les données répertoriées sur le site du ministère et de comprendre comment elles y sont organisées. C'est d'ailleurs cette organisation particulière des données qui impose certaines méthodes de codage plutôt que d'autres.

Les caractéristiques de chacune des structures recensées sont présentées sous forme de fiches individuelles (Figure 2.1).

Identification du pont	
Identification	00006
Numéro du pont	-
Nom	-
Type de structure	Pont à poutres en béton armé
Route	
Nom	Route 109
Classe route	Nationale
Municipalité	
Municipalité	Amos
MRC	Abitibi
CEP	Abitibi-Ouest
Obstacle	
Nom	Rivière Chevalier
Localisation	
Latitude	48,537387
Longitude	-78,131118
Carte	
Dimension	
Longueur totale	18,3 m
Longueur tablier	12,2 m
Largeur hors tout	10,8 m
Largeur carrossable	9,9 m
Superficie tablier	131,0 m ²
Année	
Construction	1959



Indicateurs de gestion	
Indicateur	
Indice de condition générale	Structure ne nécessitant aucune intervention
Indice d'accèsibilité	Vert

Fréquence des inspections	
Inspections générales	
Dernière inspection générale	Juillet 2014
Prochaine inspection générale	2017

Rapport d'inspection	
La structure a fait l'objet d'au moins un rapport d'inspection	

Caractéristiques de la route	
Circulation	
DJMA	2 800 % camion 10%
Chaussée	
Nombre de voies	2

Figure 2.1: Exemple de fiche individuelle (*Site du MTQ*)

Leur format est toujours le même tandis que le fond varie légèrement d'un type de structure à l'autre. En effet, les données relatives à l'identification et aux dimensions de la structure, notamment, ne sont pas les mêmes selon le type de structure considéré (pont, ponceau, mur ou tunnel). Pour un pont, sont présentées les longueurs totale et de tablier, les largeurs hors tout et carrossable ainsi que la superficie du tablier. Pour ce qui est des ponceaux, les données enregistrées sont l'ouverture totale, la longueur, la superficie totale et l'épaisseur du remblai. Enfin, la longueur totale, la hauteur moyenne et la superficie sont les dimensions caractéristiques des murs, tandis que les longueur et largeur totales, la largeur carrossable et la superficie totale sont celles des tunnels.

Remarque :

L'Annexe A donne plus de détails quant à la terminologie employée pour les structures et leurs dimensions.

2.2 Extraction des données

Il faut maintenant extraire les données du site pour les importer sur Excel afin de faciliter leur traitement. Un code réalisé grâce au module Visual Basic du logiciel est donc mis au point. Les détails de son contenu sont fournis en Annexe B et on choisit de ne présenter, dans cette partie, que les principes importants de la méthode retenue.

Une technique de repérage par mots-clés est utilisée. Les fiches individuelles des structures sont parcourues une à une, à partir de leur adresse URL. Pour chacun des liens, les données web sont extraites sous le même format que celui présenté sur la Figure 2.2. Le code va alors détecter certains termes spécifiques et retourner les valeurs contenues dans les cellules adjacentes. À titre d'exemple, si l'information intéressante est le type de structure, on voit qu'elle est contenue dans une cellule décalée d'une colonne vers la droite par rapport au mot-clé « Type de structure ».

USAGERS			
PARTENAIRES PRIVÉS			
PARTENAIRES PUBLICS			
MINISTÈRE			
Accueil > Infrastructures de transport > Structures > Fiche de la structure			
Note : L'indice de condition générale (ICG) indiqué dans la présente fiche ne sera modifié (ou mis à jour) que lorsque les travaux seront effectués.			
Identification du pont			
Identification			
Numéro du pont	2		
Nom	-		
Type de structure	Pont acier-bois		
	Pont acier-bois		
Route			
Nom	Route 395		
Classe route	Collectrice		
Municipalité			
Municipalité	Amos		
MRC	Abitibi		
CEP	Abitibi-Ouest		
Obstacle			
Nom	Ruisseau Alexandre		
Localisation			
Latitude	48,646627	Longitude	-77,996859
Carte	Localiser la structure		
Dimension			
Longueur totale	20,2 m		
Longueur tablier	19,8 m		
Largeur hors tout	7,7 m		
Largeur carrossable	7,3 m		

Figure 2.2: Récupération des données d'une fiche individuelle

La Figure 2.3 met en avant les différents mots-clés² et décalages³ voulus. Ainsi, on retrouve bien le décalage défini pour le type de structure⁴.

A	B	C	D	E	F	G
1						
2						
3	Décalage					
4	Mots Clés	0-1	2-1	0-1	1-1	0-1
5	Adresse	Numéro du pont	Identification de structure	Route	Classe route	
6						
7						
8						
9						
10						
11						
12						
13						
14						

Figure 2.3: Mots-clés et décalages associés

² cf. encadré inférieur sur la Figure 2.3.

³ cf. encadré supérieur sur la Figure 2.3.

⁴ « 0-1 » signifie 0 ligne en-dessous et 1 colonne vers la droite.

2.3 Élargissement et réorganisation de la base de données

2.3.1 Tri des données

Les données du site sont désormais extraites sur un fichier Excel et prêtes à être utilisées. Une première manipulation intéressante à réaliser est alors de scinder ce fichier en quatre autres plus petits, pour chacun des quatre types de structures répertoriées. Pour ce faire, il est possible de s'appuyer sur les informations d'identification relevées. En effet, si la procédure d'extraction de données trouve, pour une structure particulière, une référence sous la forme d'un numéro de pont, alors ladite structure est forcément un pont. En revanche, si la mention « non trouvé » est retournée, elle peut appartenir à n'importe quel groupe à l'exception de celui des ponts. Il est ainsi possible de trier toutes les données grâce à un code en trois étapes dont la logique est explicitée sur la Figure 2.4 et le détail fourni en Annexe C.

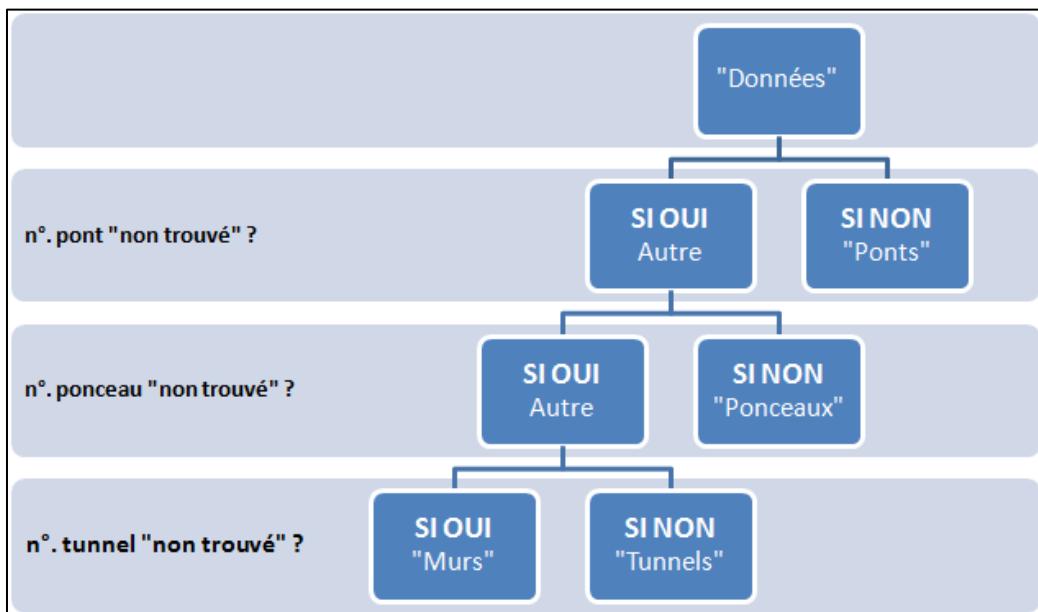


Figure 2.4: Logique de tri des données

Après le tri, il s'avère que les 5359 structures recensées par le MTQ sont réparties de la manière indiquée par le Tableau 2.1.

Tableau 2.1: Répartition selon le type de structures

Type de structures	Nombre de structures
Ponts	3991
Ponceaux	802
Murs	556
Tunnels	10
TOTAL	5359

Cependant, pour que cette base de données soit réellement exploitable, quelques modifications supplémentaires sont nécessaires pour rendre les variables existantes plus parlantes et/ou créer de nouvelles variables pertinentes.

2.3.2 Les régions administratives

Parmi les données extraites du site du ministère, figurent les municipalités régionales de comté (MRC). Elles sont au nombre de quatre-vingt-neuf dans la province du Québec et il est facile de les relier aux régions administratives. Cette opération présente trois intérêts principaux. Tout d'abord, elle permet de réduire le nombre de modalités envisagées pour une variable d'entrée : au lieu de quatre-vingt-dix modalités possibles pour les MRC (une pour chaque MRC et une pour les structures considérées comme « hors MRC »), on passe à dix-sept pour les régions administratives.

De plus, cette transformation est intéressante car elle permet de travailler sur l'ensemble des structures, ce qui n'aurait pas été le cas en ne gardant que la variable MRC. En effet, au Québec, quatre-vingt-onze municipalités sont classées « hors MRC », c'est-à-dire qu'elles n'appartiennent à aucune MRC. Cela concerne 1043 ponts sur 3991 recensés (26,1%), 138 ponceaux sur 802 (17,2%), 312 murs sur 556 (56,1%) et 10 tunnels sur 10 (100%), soit une grande partie des structures listées. En revanche, ces municipalités appartiennent bien à des régions administratives (Tableau 2.2).

Tableau 2.2: Correspondance entre municipalités « hors MRC » et régions administratives

Région administrative	Municipalités « hors MRC »
Saguenay-Lac-Saint-Jean	Saguenay
Capitale-Nationale	L’Ancienne-Lorette, Québec, Saint-Augustin-des-Desmaures
Mauricie	La Tuque, Lac-Edouard, Shawinigan, Trois-Rivières
Estrie	Sherbrooke
Montréal	Baie d’Urfé, Beaconsfield, Dorval, Kirkland, Mont-Royal, Montréal, Montréal-Est, Montréal-Ouest, Pointe-Claire, Sainte-Anne-de-Bellevue, Senneville, Westmount
Outaouais	Gatineau
Abitibi-Témiscamingue	Rouyn-Noranda
Nord-du-Québec	Chibougamau, Baie-James, Lebel-sur-Quévillon, Matagami
Gaspésie-Îles-de-la-Madeleine	Grosse-Île, Les-Îles-de-la-Madeleine
Chaudière-Appalaches	Lévis
Laval	Laval
Laurentides	Mirabel
Montérégie	Boucherville, Brossard, Longueuil, Saint-Bruno-de-Montarville, Saint-Lambert

Enfin, le dernier avantage de cette opération réside dans le fait qu'il est plus simple de trouver des données (démographique, économique, etc.) relatives aux régions administratives qu'aux MRC.

2.3.3 Les zones climatiques

Une donnée, semble-t-il, importante quant à l'état des infrastructures routières est le climat. Il paraît, en effet, évident que selon la zone climatologique les sollicitations auxquelles elles devront faire face ne seront pas les mêmes. Leur tenue dans le temps pourra alors en être affectée. Ainsi, profiter du fait d'avoir la localisation (longitude et latitude) des structures pour les positionner dans les différentes zones climatiques du Québec peut apporter un vrai plus à l'étude.

2.3.3.1 Classification climatologique de Köppen-Geiger

Plusieurs modèles de classification climatologique existent mais le plus reconnu est celui de Köppen-Geiger. Cette classification propose un découpage du monde en cinq grandes zones de climat, elles-mêmes subdivisées en plusieurs sous-zones. Plus de détails sont fournis en Annexe D. Pour le Québec, le ministère du Développement durable, de l'Environnement et de la Lutte contre les changements climatiques (MDDELCC) propose un découpage en trois zones climatiques dont les limites peuvent être approximées par les degrés de latitude 51N et 58N (Figure 2.5).

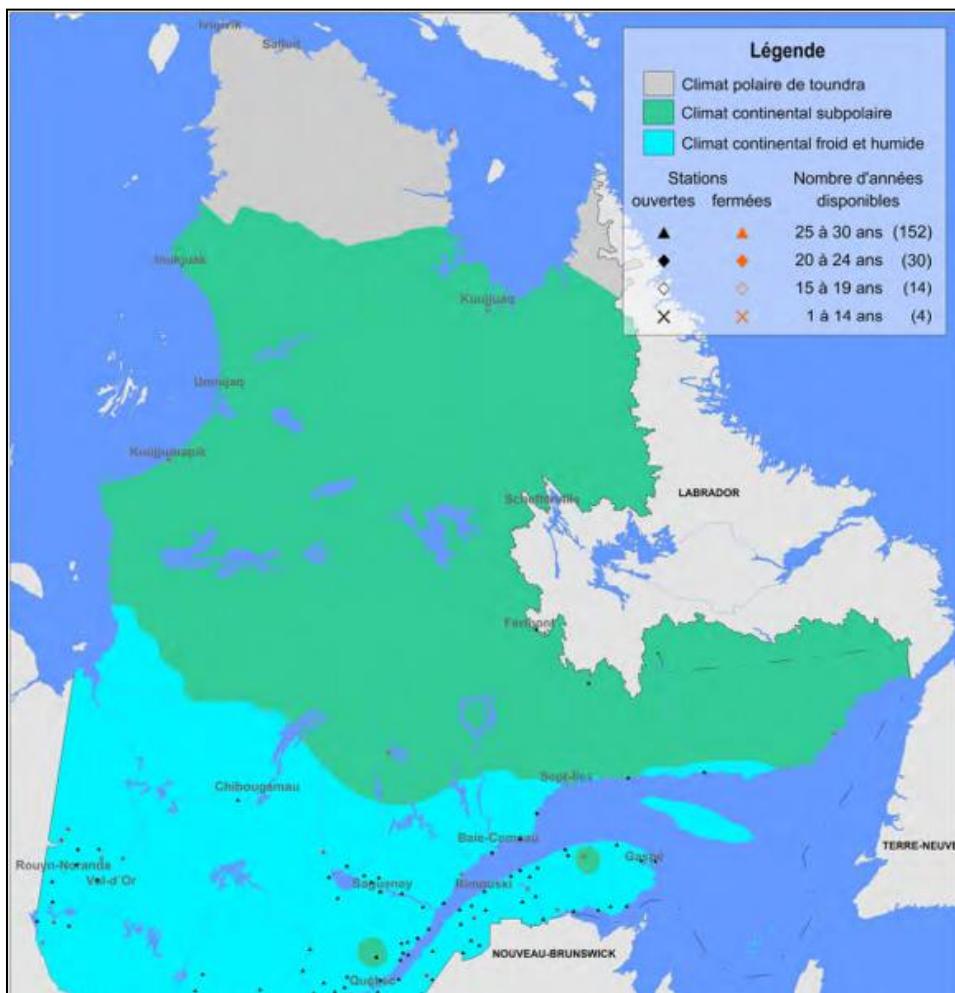


Figure 2.5: Découpage climatologique du Québec (*Site du MDDELCC*)

Le sud du Québec est la partie se situant sous le 51^{ème} degré nord de latitude. Le climat y est qualifié de "continental froid et humide", ce qui correspond à la dénomination « Dfb » du modèle de classification de Köppen-Geiger.

Entre les 51^{ème} et 58^{ème} degrés nord de latitude, le climat est celui correspondant à une végétation du type taïga, c'est-à-dire un climat "continental subpolaire" (dénomination « Dfc »).

Si les deux zones précédentes sont parfois regroupées, le nord du Québec possède, en revanche, un climat bien particulier : le climat "polaire de la toundra". Il couvre la province à partir du 58^{ème} degré nord de latitude et correspond à l'appellation « ET » du modèle de Köppen-Geiger.

En adoptant le découpage détaillé ci-dessus, les 5359 structures répertoriées par le MTQ sont très inégalement réparties entre les différentes zones climatologiques, comme on peut le voir dans le Tableau 2.3.

Tableau 2.3: Répartition des structures par zone climatique (Köppen-Geiger)

Zone	Ponts	Ponceaux	Murs	Tunnels	TOTAL	
Dfb	3956	766	556	10	5288	98,7%
Dfc	34	36	0	0	70	1,3%
ET	1	0	0	0	1	0%
TOTAL	3991	802	556	10	5359	-

La variable mise en place n'est donc pas aussi pertinente qu'on l'aurait souhaité dans le sens où il sera difficile de dégager une tendance pour les 1,3% des structures de la zone « Dfc » et impossible de le faire pour la seule structure appartenant à la zone « ET ».

De plus, la classification de Köppen-Geiger coupe le Québec en trois grandes zones climatiques mais ne prend pas en compte les particularités locales de chaque zone. À titre d'exemples, dans la zone continentale froide et humide, les Laurentides et les Appalaches relèvent plutôt du climat subpolaire, tandis que dans l'est du Québec, le Saint-Laurent rend le climat un peu plus "maritime".

On va donc se tourner vers un autre type de classification climatique, celle de Litynski, qui donne un découpage beaucoup plus détaillé et précis de la province.

2.3.3.2 Classification climatologique de Litynski

Cette nouvelle classification s'appuie sur une étude publiée par la direction du patrimoine écologique et du développement durable, sous la tutelle du ministère de l'environnement, visant à distinguer les différentes zones climatiques du Québec. Dans celle-ci, Gerardin et McKenney (2001) se basent sur le modèle de Litynski et répertorient douze climats différents sur le territoire provincial. Le principe de classification ainsi que son application au Québec sont détaillés en Annexe E. Dans cette partie, on ne s'intéresse, en effet, qu'à l'utilisation des résultats de l'étude pour la détermination d'une nouvelle variable climatique.

S'il apparaît évident que les douze classes de climat définies vont conduire à douze modalités possibles pour cette nouvelle variable, il est en revanche plus compliqué et fastidieux d'attribuer une modalité à chacune des structures. Ce travail est réalisé en trois étapes consistant à superposer, à trois échelles différentes, des cartes du Québec et la carte des zones climatiques définies par Gerardin et McKenney (Figure 2.6).

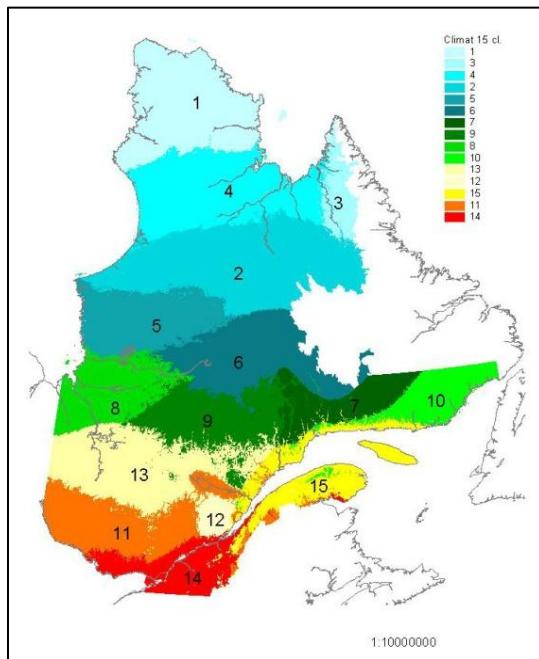


Figure 2.6: Zones climatiques au Québec (*Gerardin et McKenney, 2001*)

La première superposition est celle de la carte des régions administratives (Figure 2.7) avec la carte des zones climatiques. L'intérêt est de pouvoir attribuer rapidement à toutes les structures d'une même région une zone climatique dans le cas où celle-ci engloberait ladite région dans son ensemble.

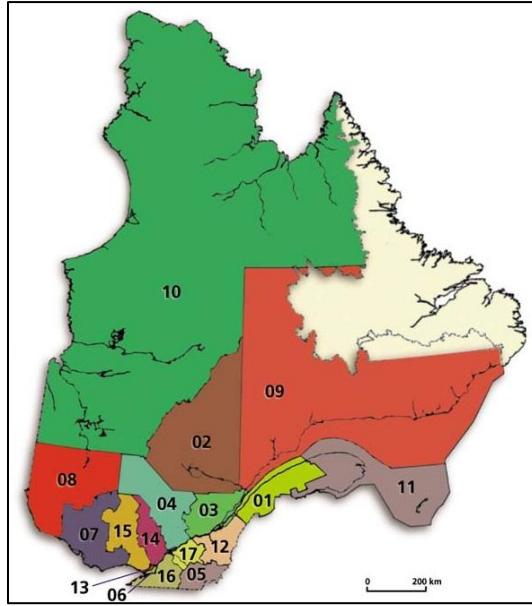


Figure 2.7: Carte des régions administratives du Québec (*Site du MDDELCC*)

L'opération de superposition des deux cartes est réalisée par transparence (Figure 2.8).

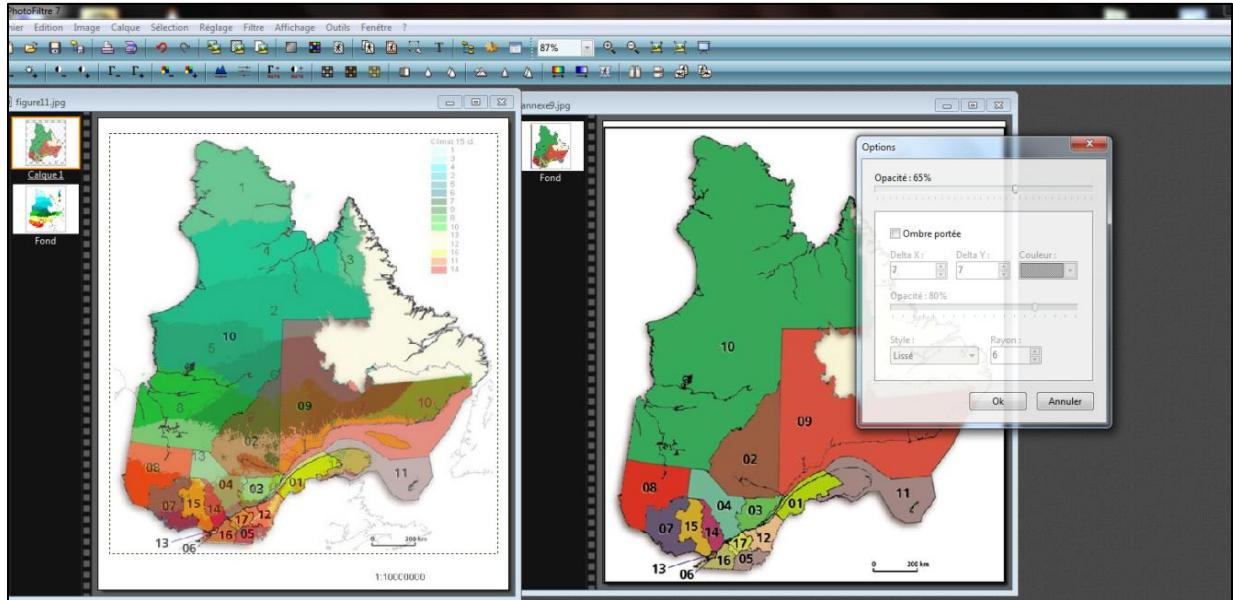


Figure 2.8: Superposition des cartes des zones climatiques et des régions administratives

Ainsi, il est possible d'attribuer directement une zone climatique à l'ensemble des structures des régions du Bas Saint-Laurent et de la Gaspésie-Îles-de-la-Madeleine (zone 12), de la Capitale-Nationale (zone 10), du Centre-du-Québec, de l'Estrie, de la Montérégie, de Montréal et de Laval (zone 11).

Pour les autres régions administratives, le même principe est appliqué mais à l'échelle des MRC. Pour les MRC dans lesquelles plusieurs zones climatiques sont encore représentées, les municipalités hébergeant des structures sont repérées sur une carte (Figure 2.9: exemple pour les municipalités du Nord-du-Québec) et leur appartenance à telle ou telle zone climatique est déterminée au cas par cas, par superposition.



Figure 2.9: Repérage des municipalités du Nord-du-Québec

Ainsi, toutes les structures se voient attribuer une zone climatique. Il faut cependant noter que si douze classes de climat sont représentées sur le territoire québécois, seules huit d'entre elles⁵ hébergent une des structures répertoriées par le MTQ (Tableau 2.4).

⁵ Les zones 4, 6, 7, 8, 9, 10, 11, 12

Tableau 2.4: Répartition des structures par zone climatique (Litynski)

Zone	Ponts	Ponceaux	Murs	Tunnels	TOTAL	
1	0	0	0	0	0	0,0%
2	0	0	0	0	0	0,0%
3	0	0	0	0	0	0,0%
4	2	0	0	0	2	0,0%
5	0	0	0	0	0	0,0%
6	30	16	0	0	46	0,9%
7	74	57	2	0	134	2,5%
8	31	9	0	0	40	0,7%
9	374	81	25	0	480	9,0%
10	536	81	85	0	702	13,1%
11	2322	380	328	10	3040	56,7%
12	622	178	116	0	915	17,1%
TOTAL	3991	802	556	10	5359	-

La majorité des structures se situent dans la zone climatique 11. Il s'agit en fait d'une zone dont le climat est modéré en termes de températures, qualifié de subhumide au regard des précipitations annuelles et dont la saison de croissance est longue.

Trois autres zones abritent un nombre relativement conséquent de structures. Les zones 10 et 12 correspondent à des climats subpolaires à période de croissance moyenne dont la différence est faite au niveau des précipitations : 10 est une zone humide tandis que 12 est plutôt qualifiée de subhumide. Dans la zone 9, le climat est subpolaire doux et subhumide avec une longue période de croissance.

Enfin, le reste des structures est réparti entre quatre autres zones climatologiques :

- ✓ La zone 4 qui correspond à un climat polaire avec des précipitations modérées et une très courte saison de croissance ;

- ✓ La zone 6 qui correspond à un climat subpolaire froid et subhumide avec une courte saison de croissance ;
- ✓ La zone 7 qui correspond à un climat subpolaire froid et subhumide avec une saison de croissance moyenne ;
- ✓ La zone 8 qui correspond à un climat subpolaire et humide avec une courte saison de croissance.

2.3.4 L'affiliation politique des gouvernements

Parmi les variables extraites du site du ministère, l'année de construction constitue une information cruciale. En la retranchant à la date d'inspection, il est en effet possible de déterminer l'âge des structures. Elle est également utile pour tenir compte de la dimension politique, aspect considéré, dans l'introduction, comme l'un de ceux faisant l'originalité du projet. En effet, grâce aux dates de mandat des différents premiers ministres, il est possible de déterminer le parti politique québécois au pouvoir au moment de la construction.

Sur la période couverte par la base de données (1900 – 2014), vingt-quatre premiers ministres se sont succédé. La moitié d'entre eux étaient affiliés au Parti Libéral Québécois (PLQ). Il s'agit du plus vieux parti politique encore existant aujourd'hui au Québec car il a été créé dès 1867. Il s'agit d'un parti traditionnellement libéral et progressiste. Deux autres partis ont été au pouvoir au Québec. Tout d'abord, l'Union Nationale (UN) est un parti créé en 1935 de la fusion du Parti Conservateur du Québec et de l'Action Libérale Nationale. Il s'agit d'un parti dont les idéologies sont le nationalisme, comme son nom l'indique, ou encore l'autonomisme. Il cesse cependant d'exister en 1989. Enfin, le dernier parti représenté est le Parti Québécois (PQ). Il s'agit du parti dont la fondation est la plus récente (1968). Il est de la mouvance nationaliste et prône l'indépendance du Québec. Chacun de ces deux derniers partis politiques a vu six de ses membres occuper la fonction de premier ministre.

Il est alors possible de relier les structures au parti au pouvoir (Figure 2.10).

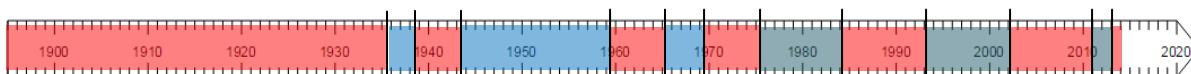


Figure 2.10: Frise chronologique des différents partis au pouvoir au Québec

2.4 Base de données finale

2.4.1 Variables exploitables

En cumulant les variables nouvellement obtenues et celles récupérables plus ou moins directement sur le site du MTQ, on couvre désormais un éventail de domaines suffisamment vaste (Tableau 2.5) pour pouvoir obtenir de premiers résultats intéressants.

Tableau 2.5: Liste des variables exploitables

Catégorie	Variable	Type		Pont	Ponceau	Mur	Tunnel
Ouvrage	Type ⁶	Entrée	CAT	X	X	-	-
	Matériau(x) ⁶	Entrée	CAT	X	X	X	X
	Âge ⁷	Entrée	CONT	X	X	X	X
Obstacle	Obstacle franchi ⁶	Entrée	CAT	X	X	-	X
Localisation	Latitude ⁸	Entrée	CONT	X	X	X	X
	Longitude ⁸	Entrée	CONT	X	X	X	X
	Région admin. ⁷	Entrée	CAT	X	X	X	X
Dimensions	Longueur totale ⁸	Entrée	CONT	X	-	X	X
	Longueur tablier ⁸	Entrée	CONT	X	-	-	-
	Largeur hors tout ⁸	Entrée	CONT	X	-	-	-
	Largeur carrossable ⁸	Entrée	CONT	X	-	-	X
	Superficie tablier ⁸	Entrée	CONT	X	-	-	-
	Ouverture totale ⁸	Entrée	CONT	-	X	-	-
	Longueur ⁸	Entrée	CONT	-	X	-	-

⁶ Les informations sont indirectement disponibles sur le site du MTQ ; il a notamment fallu travailler sur la définition des différentes modalités que peuvent prendre les variables concernées.

⁷ L'obtention de ces variables a été traitée dans la partie 2.3 – Élargissement et réorganisation de la base de données.

⁸ Les fiches individuelles des structures donnent directement ces informations.

Tableau 2.5: Liste des variables exploitables (suite et fin)

Dimensions	Superficie totale ⁸	Entrée	CONT	-	X	-	X
	Épaisseur remblai ⁸	Entrée	CONT	-	X	-	-
	Hauteur moyenne ⁸	Entrée	CONT	-	-	X	-
	Superficie ⁸	Entrée	CONT	-	-	X	-
	Longueur totale ⁸	Entrée	CONT	-	-	-	X
Chaussée	Classe de route ⁸	Entrée	CAT	X	-	X	X
	DJMA ⁸	Entrée	CONT	X	X	-	X
	% camion ⁸	Entrée	CONT	X	X	-	X
	Nombre de voies ⁸	Entrée	DISC	X	X	-	X
Climat	Type de climat ⁷	Entrée	CAT	X	X	X	X
Politique	Parti au pouvoir ⁷	Entrée	CAT	X	X	X	X
État	ICG ⁷	Sortie	DISC	X	X	X	-
Nombre de variables				19	17	12	16

Remarque :

« CAT » signifie que la variable associée est de type catégorique. « CONT » et « DISC » servent à qualifier des variables numériques respectivement continues et discrètes (cf. Chapitre 3 – Modélisations statistiques, pour connaître la signification de ces termes).

2.4.2 Observations exclues

Cependant, on peut constater, dans le Tableau 2.5, qu'aucune information sur l'état des tunnels ne figure dans les données du MTQ. En outre, leur faible nombre (seulement dix tunnels répertoriés, d'après le Tableau 2.1) ne permet pas d'obtenir des résultats significatifs, statistiquement parlant. Ces structures ne présentent donc pas d'intérêt dans le cadre du projet et sont ainsi laissées de côté pour la suite.

Par ailleurs, rien ne peut être tiré des structures dont les inspections n'ont pas encore été faites ou sont en cours (ICG prenant les valeurs 0 et 1). Cela concerne 400 ponts sur 3991, 36 ponceaux sur 802 et 5 murs sur 556.

Au final, ce sont donc 451 observations qui sont exclues pour arriver à un total de 4908 structures (3591 ponts, 766 ponceaux et 551 murs) dont les données sont parfaitement exploitables.

CHAPITRE 3 MODÉLISATIONS STATISTIQUES

3.1 Généralités

3.1.1 Les variables

Il existe, entre les variables, plusieurs différences intrinsèques qui ont un impact sur la manière dont elles peuvent être traitées.

En premier lieu, on parle de variables catégoriques lorsque celles-ci prennent des "valeurs" qualitatives appelées modalités. À contrario, les variables quantitatives prennent des valeurs chiffrées. En outre, au sein de cette dernière famille de variables, il est possible de distinguer deux sous-catégories : les variables continues, qui peuvent prendre n'importe quelle valeur numérique, et les variables discrètes, pour lesquelles le nombre de possibilités est limité.

Pour illustrer ce point, on peut penser à l'âge des structures. Cette variable est à première vue continue. En effet, l'âge est exprimé par un chiffre "quelconque" traduisant l'ancienneté de la structure, en années par exemple (17, 45, 96, etc.). Cependant, cette variable peut également être catégorique si on décide de répartir les âges en trois modalités : « récent » pour moins de 30 ans, « moyen » pour des valeurs comprises entre 31 ans et 60 ans et « ancien » pour plus de 61 ans.

On différencie également les variables selon la manière dont elles sont obtenues : passivement pour des valeurs observationnelles ou activement pour des valeurs expérimentales.

Enfin, on distingue les variables de réponse, également appelées sorties ou variables dépendantes, Y_j , des variables d'entrée explicatives notées X_i .

3.1.2 Modèles statistiques

Tout modèle correspond alors à la mise en place d'une fonction de transfert φ telle que $Y_j = \varphi(X_i) + \varepsilon$. Cette logique est reprise sur la Figure 3.1 pour un modèle ne possédant qu'une sortie, ce qui est le cas du projet avec l'ICG comme seule réponse.

ε est l'erreur expérimentale, et donc une source de variabilité, qui peut être due à des données erronées (mensonges lors d'un sondage de collecte par exemple) ou à une perte d'informations (lors d'un regroupement de données par exemple). Elle suit une loi de distribution normale $N(0, \sigma^2)$.

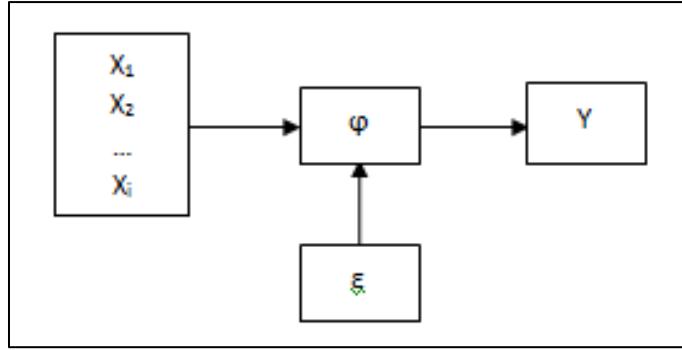


Figure 3.1: Processus de mise en place d'un modèle

L'élaboration de modèles peut satisfaire deux objectifs distincts. Pour les méthodes de régression, il s'agit de la détermination exacte de la fonction φ , utilisée ensuite à des fins prédictives. En effet, tout l'intérêt de ce genre de méthodes réside dans la possibilité de prévoir les valeurs de sortie par application de φ aux valeurs d'entrée. En revanche, l'objectif des méthodes d'analyse de la variance (ANOVA) n'est pas d'obtenir une équation prédictive mais d'évaluer l'influence des variables explicatives sur la réponse.

3.2 Traitement des variables quantitatives – Modèles de régression

3.2.1 Modèles linéaires de régression

3.2.1.1 Généralités

En pratique, la première étape est de choisir un modèle adapté aux données. Les modèles linéaires sont les plus employés et plusieurs méthodes existent pour rendre linéaires ceux qui, initialement, ne le sont pas. À titre d'exemple, l'équation qui suit est celle d'un modèle linéaire à deux variables explicatives sans effet d'interaction.

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2$$

Il est possible de la généraliser en tenant compte des effets d'interaction (avec l'ajout d'un terme $\beta_{12} \cdot X_1 \cdot X_2$ dans ce cas-ci) ou en considérant un nombre plus élevé de variables explicatives. Il existe néanmoins, en fonction de la situation et/ou des besoins, de nombreux autres modèles non linéaires.

Il est à noter que le principe sur lequel se basent les régressions linéaires est celui des moindres carrés. Concrètement, dans le cas du modèle ci-dessus, par exemple, cela revient à minimiser la somme des $Y - \beta_0 - \beta_1 \cdot X_1 - \beta_2 \cdot X_2$ au carré.

Les termes β_i sont des paramètres statistiques inconnus. Leur estimation, grâce au logiciel, est importante car elle permet de trouver l'équation prédictive. Cependant, pour pouvoir se fier à celle-ci afin d'estimer les valeurs de la sortie, un certain nombre de critères sont à valider ; le modèle doit être adapté et de qualité. Avant de lister ces critères, un petit point sur les notations et certaines grandeurs statistiques "clés" semble toutefois s'imposer.

3.2.1.2 Définitions et notations

On note :

- ✓ x_k la valeur que prend la variable X pour la k -ième observation ;
- ✓ $\bar{x} = \frac{\sum x_k}{n}$ la valeur moyenne empirique (estimée à partir de l'échantillon) de X ;
- ✓ $SS_X = \sum (x_k - \bar{x})^2$ la somme des carrés de X ;
- ✓ $MS_X = \frac{SS_X}{ddl}$ le carré moyen de X où ddl est en fait le nombre de degrés de liberté ;
- ✓ \hat{A} la valeur estimée de la grandeur A .

En restant sur l'exemple précédent d'un modèle linéaire à deux variables explicatives, l'équation de prédiction obtenue est donc de la forme :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_1 + \hat{\beta}_2 \cdot X_2$$

À partir de là, plusieurs grandeurs peuvent être définies :

- ✓ $e_k = \hat{y}_k - y_k$: le résidu, c'est-à-dire l'écart entre la réponse prédictive et la valeur réelle ;
- ✓ $SS_{\text{resid}} = \sum e_k^2 = \sum (\hat{y}_k - y_k)^2$: la somme des carrés résiduelle ;
- ✓ $MS_{\text{resid}} = MSE = \frac{SS_{\text{resid}}}{ddl}$: le carré résiduel moyen ou erreur quadratique moyenne ;
- ✓ $SS_{\text{reg}} = \sum (\hat{y}_k - \bar{y})^2$: la somme des carrés de régression ;
- ✓ $MS_{\text{reg}} = \frac{SS_{\text{reg}}}{ddl}$: le carré de régression moyen ;
- ✓ $SS_{\text{tot}} = SS_{\text{reg}} + SS_{\text{resid}} = \sum (y_k - \bar{y})^2$: la somme des carrés totale ;

Le nombre de degrés de liberté de SS_{reg} est égal au nombre de variables explicatives prises en compte dans le modèle, p par exemple. Celui de SS_{tot} est égal au nombre d'observations (effectif total) n . Ainsi, le nombre de degrés de liberté de SS_{resid} est égal à la différence $n - p$.

En outre, il faut savoir que toute valeur estimée \hat{y}_k peut être décomposée en deux parties. Une première partie de la valeur est expliquée par le modèle de régression. L'autre partie est quant à elle inexpliquée : c'est la part résiduelle, c'est-à-dire la variance des résidus.

Cela se traduit également sur la variance, grandeur caractérisant la dispersion des valeurs autour de leur moyenne. La variance est égale à l'espérance (c'est-à-dire la moyenne) des carrés des écarts à la moyenne (c'est-à-dire le carré moyen). Ainsi, une part de la variance est expliquée par la régression et peut être caractérisée par le SS_{reg} et l'autre part correspond à la variance résiduelle inexpliquée traduite par le SS_{resid} .

3.2.1.3 Critères de qualité du modèle

3.2.1.3.1 Coefficient de détermination

Le coefficient de détermination, noté R^2 , est donné par la formule :

$$R^2 = \frac{SS_{\text{reg}}}{SS_{\text{tot}}}$$

Il correspond donc à la part de variance totale expliquée par la régression et traduit la qualité de la régression. En effet, le but étant d'avoir le moins de variance inexpliquée possible, plus le R^2 est proche de 1 et plus la régression donne des résultats satisfaisants. Il présente néanmoins le problème d'évoluer avec le nombre de variables explicatives. Ainsi, si ce dernier augmente, le R^2 augmentera, même si un trop grand nombre de variables explicatives réduit la robustesse du modèle, c'est-à-dire la capacité à généraliser les conclusions dans le temps ou à une population. Pour tenir compte de ce problème, on privilégie le $R^2_{\text{ajusté}}$, qui est défini de la façon suivante :

$$R^2_{\text{ajusté}} = R^2 - \frac{p \cdot (1 - R^2)}{n - p - 1}$$

Bien que le but soit d'avoir une valeur la plus élevée possible, aucun seuil n'est véritablement fixé. C'est à l'utilisateur de juger la valeur obtenue selon le type ou le domaine d'application ou encore l'objectif poursuivi.

3.2.1.3.2 Tableau d'analyse de la variance ANOVA

La qualité d'une régression s'estime donc en comparant deux variances (régressive et résiduelle). Peut intervenir alors le test de Fisher-Snedecor ; à l'instar d'un test de Student pour la comparaison de deux moyennes, celui de Fisher a pour finalité, pour un certain niveau de confiance, de tester l'égalité de deux variances. Si, classiquement, il est utilisé pour comparer les variances de deux échantillons différents, il est tout à fait possible de l'appliquer pour deux composantes d'une même variance.

L'hypothèse H_0 testée est celle de variances égales entre elles mais l'objectif, dans le cas d'une régression, est de rejeter cette hypothèse. En effet, ce qui caractérise une bonne régression est le fait que la variance expliquée soit supérieure à la variance inexpliquée. Aussi, H_0 est rejetée si la valeur F_0 ($= \text{MS}_{\text{reg}} / \text{MS}_{\text{resid}}$) est supérieure à une valeur critique F obtenue grâce à la table de Fisher.

L'utilisation de la p-value est également très répandue et découle directement des tests statistiques. Il s'agit en fait de la probabilité que F soit supérieur à F_0 , autrement dit la probabilité que H_0 soit vraie.

Dans le cadre de la régression, on cherche donc à avoir une valeur de F_0 supérieure à F et une p-value la plus faible possible (seuil classiquement fixé à 0,05) pour s'en assurer. Ces informations sont généralement fournies par les logiciels de statistiques sous la forme d'un tableau appelé tableau ANOVA (Tableau 3.1).

Tableau 3.1: Tableau ANOVA traduisant la qualité d'une régression

SOURCE	ddl	SS	MS	F_0	p-value
Régression	a	α	α / a	$(b \cdot \alpha) / (a \cdot \beta)$	p-value
Résiduelle	b	β	β / b	-	-
TOTALE	$a + b$	$\alpha + \beta$	-	-	-

3.2.1.4 Critères de vérification du modèle

Toujours avant de pouvoir utiliser l'équation du modèle, il faut également s'assurer de la justesse de la régression et de son adéquation aux données.

3.2.1.4.1 Indépendance des observations

Les observations doivent être indépendantes entre elles. Concrètement, une valeur x_k ne doit pas être influencée par les valeurs x_{k-j} , $j \geq 1$, qui la précèdent. On va supposer que c'est bien le cas dans cette étude : les valeurs prises par une structure n'influencent pas, à priori, les valeurs prises par les autres.

Remarque :

Cette hypothèse est discutable dans le sens où le budget d'entretien et de maintenance est limité. Ainsi, le choix d'entretenir un pont x (effet sur ICG_x) peut aussi être vu comme le choix de ne pas entretenir un pont y (effet, ou plutôt non effet, sur ICG_y).

3.2.1.4.2 Analyse de résidus

Une partie importante du "model-checking" est l'analyse des résidus. Dans un premier temps, il faut s'assurer de la relative constance de la variance. Concrètement, on cherche surtout à vérifier qu'aucune logique ne peut-être constatée dans l'évolution de la variance. Cela s'observe sur la courbe des résidus en fonction des valeurs prédictes de la sortie. L'idéal serait que tous les résidus soient situés dans une bande horizontale "resserrée" autour de 0 (Figure 3.2 et Figure 3.3).

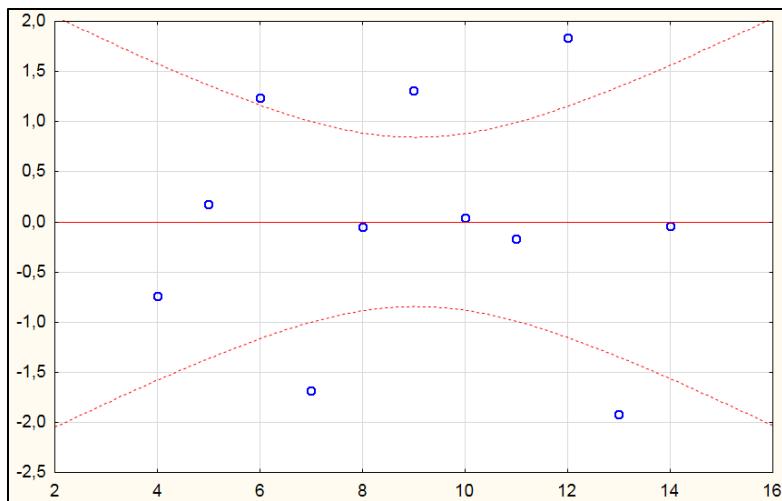


Figure 3.2: Variancerelativement constante

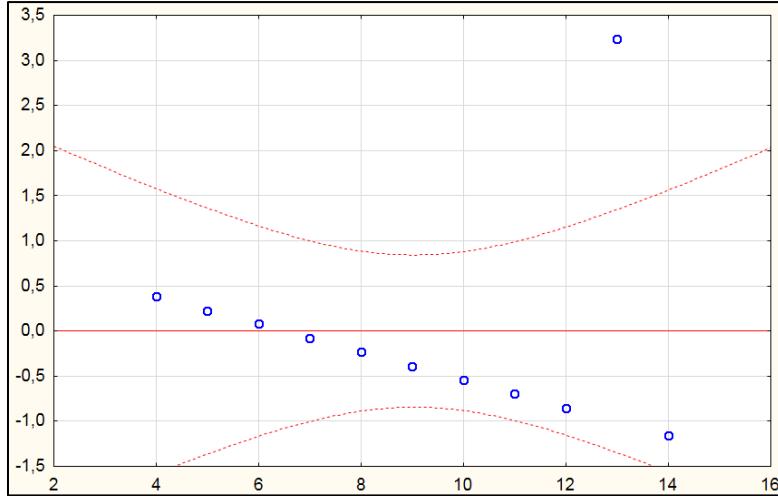


Figure 3.3: Variance non constante

Le deuxième critère relatif à l'analyse des résidus est la normalité. En effet, les résidus doivent suivre une loi de distribution normale. Le tracé sur l'échelle gaussienne (Figure 3.4) est un outil graphique efficace pour s'en assurer. Il permet de comparer les résidus aux valeurs qu'ils auraient eues s'ils suivaient parfaitement la loi normale. En outre, ce genre de graphiques permet également de mettre en évidence les valeurs aberrantes de la réponse Y : un résidu e_k très différent des autres peut, par exemple, conduire à se demander si la valeur observée y_k n'est pas le résultat d'une erreur de mesure, de report, etc. Ces valeurs aberrantes ou atypiques sont appelées "outliers".

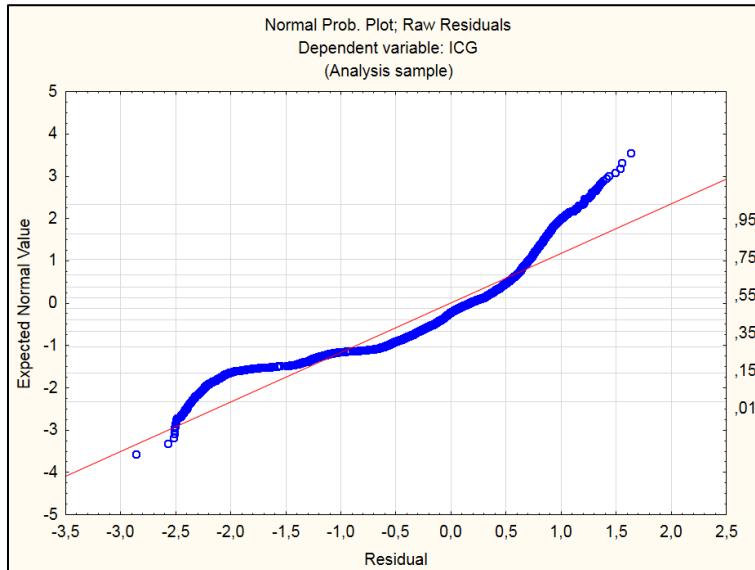


Figure 3.4: Tracé des résidus sur l'échelle gaussienne

3.2.1.4.3 Analyse de sensibilité

L'analyse de sensibilité tend justement à mettre en avant les "outliers" qui ont trop d'influence sur la régression. Il est important de procéder à cette analyse car ceux-ci peuvent fausser les résultats. En effet, la méthode des moindres carrés est très sensible aux points influents. Trois caractéristiques permettent alors de détecter une observation trop influente.

La première est la grandeur des résidus. Ainsi, si la valeur d'un résidu e_k est trop élevée par rapport aux autres, il est possible que la k-ième observation, trop atypique et ayant une influence sur la régression supérieure à celles de valeurs plus "normales", fausse la régression. En valeur absolue, on considère qu'un résidu est problématique lorsqu'il dépasse, une fois studentisé, la valeur seuil de 2,5.

La seconde caractéristique est la distance de Cook D_k , dont la définition mathématique est la suivante :

$$D_k = \frac{\sum_l (\hat{y}_l - \hat{y}_{l(k)})^2}{\mu \cdot MSE}$$

Où D_k est la distance de Cook pour la k-ième observation ;

\hat{y}_l est la valeur prédictive de la sortie pour la l-ième observation ;

$\hat{y}_{l(k)}$ est la valeur prédictive de la sortie pour la l-ième observation si la k-ième observation est exclue de la régression ;

μ est le nombre de paramètres du modèle.

Concrètement, si D_k est trop grand, cela veut dire que la k-ième observation est trop influente et apporte trop de changements à la régression. Elle est considérée comme telle si sa valeur est supérieure à 1.

Enfin, la dernière caractéristique est la distance de Mahalanobis. Elle correspond à l'écart, dans l'espace des variables explicatives (une variable = une dimension), entre une observation x_k et la valeur moyenne des observations \bar{x} . Elle met ainsi en évidence ce qu'on appelle l'effet de levier, c'est-à-dire l'importance de l'influence d'une observation x_k par rapport aux influences des autres. Concrètement, pour une observation, à partir du moment où la valeur de distance est plus de dix fois supérieure à la valeur moyenne des distances, on considère ladite observation comme trop influente.

Remarque :

Tous les seuils critiques définis sont fixés arbitrairement. Ainsi, si une valeur est très proche du seuil sans le dépasser, la considérer quand même comme problématique peut être pertinent.

Une procédure prévue pour l'occasion a été mise en place sur Excel. Elle permet, en fonction de seuils préalablement définis, de retourner directement les observations problématiques.

Au final, si un "outlier" est détecté, l'exclure peut conduire à un accroissement de la qualité de la régression. Cependant, il peut être intéressant de comprendre pourquoi cette valeur est différente. Il est également à noter que l'analyse de sensibilité peut être effectuée avant la mise en place du modèle afin d'exclure d'entrée les observations qui peuvent altérer le résultat.

3.2.1.4.4 Multicolinéarité

On parle de multicolinéarité quand les variables explicatives sont corrélées entre elles. Un tel cas de figure est problématique car cela veut dire que l'équation de régression linéaire n'a pas de solution. Il est plus simple d'appréhender ce problème en raisonnant sur les formes matricielles.

Soit X la matrice des variables explicatives et Y la matrice des variables de réponse. On considère un problème à n observations, p entrées et une seule sortie. X est donc une matrice à n lignes et $p + 1$ colonnes (la première colonne renvoie à la constante β_0) et Y est une matrice à n lignes et une seule colonne.

$$X = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & x_{1,3} & \dots & x_{1,p} \\ 1 & x_{2,1} & x_{2,2} & x_{2,3} & \dots & x_{2,p} \\ 1 & x_{3,1} & x_{3,2} & x_{3,3} & \dots & x_{3,p} \\ 1 & x_{4,1} & x_{4,2} & x_{4,3} & \dots & x_{4,p} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n,1} & x_{n,2} & x_{n,3} & \dots & x_{n,p} \end{pmatrix} \text{ et } Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \dots \\ y_n \end{pmatrix}$$

La matrice des coefficients estimés est obtenue par la formule :

$$\hat{\beta} = (X^t \cdot X)^{-1} \cdot X^t \cdot Y = C \cdot Y$$

Celle des sorties estimées est alors :

$$\hat{Y} = X \cdot \hat{\beta} = X \cdot (X^t \cdot X)^{-1} \cdot X^t \cdot Y = H \cdot Y$$

Il s'agit en fait d'une manière alternative d'écrire l'équation de régression.

En cas de multicolinéarité, situation dans laquelle deux ou plusieurs variables explicatives linéairement liées entre elles, le déterminant de la matrice $X^t \cdot X$ est donc nul. Celle-ci n'est donc pas inversible et il est alors impossible de déterminer les matrices $\hat{\beta}$ et \hat{Y} .

De telles corrélations conduisent également à une répétition de l'information, contraire au critère de concision d'un modèle robuste.

Comme pour l'analyse de sensibilité, trois critères permettent de détecter un éventuel problème de multicolinéarité. Le premier d'entre eux consiste à vérifier qu'il n'y ait pas de coefficient de corrélation supérieur à 0,95. Ces coefficients se retrouvent dans la matrice de corrélation R (Tableau 3.2) et traduisent les relations linéaires qui peuvent exister entre les variables explicatives deux à deux. Ils sont notés r_{ij} et plus ils sont élevés, plus les deux variables concernées, X_i et X_j dans ce cas, sont liées entre elles.

Tableau 3.2: Exemple type d'une matrice de corrélation retournée par le logiciel

Variable	Means	Std.Dev.	Longueur totale	Longueur tablier	Largeur hors tout	Largeur carrossable	Superficie tablier
Longueur totale	69,8065	96,374	1,000000	0,991677	0,003423	0,110493	0,904298
Longueur tablier	59,0334	94,661	0,991677	1,000000	-0,000947	0,094123	0,908357
Largeur hors tout	14,8209	12,655	0,003423	-0,000947	1,000000	0,533945	0,145007
Largeur carrossable	11,7999	4,779	0,110493	0,094123	0,533945	1,000000	0,290479
Superficie tablier	872,9229	1760,294	0,904298	0,908357	0,145007	0,290479	1,000000

Dans l'exemple du Tableau 3.2, il semble exister de fortes corrélations entre plusieurs variables explicatives considérées dans le projet : longueur totale et la longueur de tablier, longueur de tablier et superficie de tablier, entre autres.

Si cette première condition est nécessaire, elle n'est cependant pas suffisante pour conclure. Ainsi, il faut également vérifier les valeurs prises par les facteurs d'inflation de la variance (VIF), qui sont définis comme les inverses des tolérances :

$$VIF_i = \frac{1}{1 - R_i^2}$$

R_i est le coefficient de corrélation multiple entre X_i et toutes les autres variables $X_{j,j \neq i}$. Si R_i^2 est trop élevé, on considère que la i -ème variable est trop fortement corrélée aux autres. On va alors s'assurer qu'aucun VIF_i ne soit supérieur à 2 ou 3.

Enfin, la dernière vérification s'effectue sur les indices de conditionnement (IC). Pour les obtenir, il faut rechercher les valeurs propres de la matrice R et les ranger dans l'ordre décroissant :

$$\lambda_1 = \lambda_{\max} \geq \lambda_2 \geq \lambda_3 \geq \text{etc}$$

On définit alors :

$$IC_i = \frac{\lambda_{\max}}{\lambda_i}$$

Le seuil critique est cette fois-ci fixé à 100 ; un IC supérieur traduirait alors un problème de multicolinéarité.

Si tel est le cas, deux options sont envisageables. La première consiste à enlever les variables explicatives redondantes grâce à des processus de sélection des variables. La deuxième possibilité est de conserver toutes les variables mais d'effectuer une régression "spéciale". Les analyses sur composantes principales (ACP) et les régressions ridge et "partial least squares" (PLS) en sont des exemples.

Remarque :

Il est très important de distinguer corrélation de causalité : ce n'est pas parce qu'une relation de corrélation existe entre le type et l'ICG qu'on peut en déduire que le type a un effet sur l'ICG. Cette corrélation peut, en effet, être due à un autre facteur sous-jacent au type de ponts. Une des méthodes de construction peut être plus ancienne que l'autre, donc les ponts correspondants globalement plus anciens, donc en plus mauvais état. Dans ce cas de figure, la causalité mise en avant ne serait pas type/ICG mais bien âge/ICG.

3.2.2 En cas de multicolinéarité

3.2.2.1 Méthodes de sélection des variables

Trois méthodes sont testées dans ce projet, l'objectif étant, à chaque fois, d'obtenir un modèle dans lequel seules les variables influentes interviennent.

3.2.2.1.1 Sélection exhaustive

Elle s'appuie essentiellement sur deux critères pour déterminer le nombre de variables utiles.

Le premier est le coefficient de détermination. On étudie ainsi l'évolution marginale du R² : de combien augmente-t-il lorsque le nombre de variables augmente d'une unité ? On choisit alors un nombre le plus faible possible mais suffisant pour que le R² soit relativement proche de sa valeur maximum, obtenue quand toutes les variables sont prises en compte.

Le second critère porte sur le C_p de Mallows. Il s'agit d'un coefficient comparant la qualité du modèle complet à celle d'un modèle basé sur un sous-ensemble de variables explicatives. Choisir un C_p faible revient alors à ne choisir que les entrées réellement nécessaires au modèle, c'est-à-dire un sous-ensemble avec lequel la précision de prédiction est proche de celle du modèle complet.

3.2.2.1.2 Sélections pas-à-pas

Des tests de significativité sont réalisés sur chaque variable explicative à tour de rôle. Il existe alors deux manières de procéder.

On parle de sélection pas-à-pas avant lorsqu'une nouvelle variable significative est rajoutée, à chaque étape, au modèle. Concrètement, on part d'un modèle trivial, uniquement constitué de la constante. Chacun des modèles à une variable (X₁, X₂ et X₃, par exemple) est alors testé et celui pour lequel la variable ajoutée a la plus faible p-value, la plus significative (X₁ par exemple), est retenu. Puis, on teste les modèles bivariés constitués de cette première variable et de celles qui restent (X₁ et X₂ puis X₁ et X₃ dans l'exemple). À nouveau la variable la plus significative est gardée. On procède ainsi, pas-à-pas, jusqu'à ce que la p-value correspondante devienne supérieure à un certain seuil de significativité.

Le raisonnement de la méthode de sélection pas-à-pas arrière est l'exact opposé. Le modèle de base est celui composé de l'ensemble des variables d'entrée. À chaque étape, on retire la variable dont la p-value associée est la plus élevée. Le processus est répété tant que les p-value restent au-dessus du seuil de significativité ; il prend fin quand elles deviennent inférieures à celui-ci.

3.2.2.2 Méthodes particulières de régression linéaire

Pour contourner la multicolinéarité tout en gardant l'ensemble des variables d'entrée, et donc l'intégralité de l'information disponible, plusieurs méthodes existent. Parmi celles-ci, trois, les régressions ACP, ridge et PLS, sont développées dans cette étude. À la fin, l'objectif est de disposer du plus large éventail possible de méthodes afin de pouvoir choisir la meilleure.

3.2.2.2.1 Analyse sur composantes principales

L'ACP repose sur le principe de la réduction d'échelle. Cette méthode consiste ainsi à projeter les observations de l'espace initial dans un espace dont le nombre de dimensions est moindre mais suffisant pour que la grande majorité de l'information soit conservée. Ce nouvel espace est celui des composantes principales.

Concrètement, après avoir exprimé les observations dans l'espace ainsi défini (Tableau 3.3) et trouvé les vecteurs propres, c'est-à-dire les expressions des composantes principales en fonction des variables explicatives (Tableau 3.4), il est possible d'effectuer une régression de l'ICG sur les composantes principales.

Tableau 3.3: Observations exprimées dans l'espace des composantes principales

Observation	Facteur 1	Facteur 2
1	– 1,67057	– 1,11366
2	– 1,93736	– 0,90960
3	– 1,85432	– 0,75930
Etc.	Etc.	Etc.

Tableau 3.4: Vecteurs propres

Variable	Facteur 1	Facteur 2
Longitude	0,707107	0,707107
Âge	– 0,707107	0,707107

On en déduit :

$$\begin{cases} F1 = 0,707107 \cdot \text{longitude} - 0,0707107 \cdot \text{âge} \\ F2 = 0,707107 \cdot \text{longitude} + 0,0707107 \cdot \text{âge} \end{cases} \quad \text{Équations 3.1}$$

L'équation de prédiction trouvée est, dans cet exemple, la suivante :

$$\text{ICG} = 4,474798 + 0,223862 \cdot F1 - 0,248759 \cdot F2 \quad \text{Équation 3.2}$$

Soit, en combinant les Équations 3.1 et l'Équation 3.2 :

$$\text{ICG} = 4,474798 - 0,017605 \cdot \text{longitude} - 0,334194 \cdot \text{âge}$$

Remarque :

Les variables d'entrée sont, pour la régression ACP, obligatoirement centrées-réduites.

3.2.2.2.2 Ridge

Dans la partie 3.2.1.3 – Critères de qualité du modèle, il a été vu que la qualité d'une régression pouvait être estimée par la comparaison entre la part de variance expliquée et la part résiduelle, qui avait été qualifiée d'inexpliquée. En réalité, cette dernière peut avoir deux origines différentes : elle est, en effet, la somme d'une variance "ponctuelle", propre à chaque observation, et d'une variance (ou erreur) systématique, commune à toutes les observations, appelée biais.

Le principe de la régression ridge est alors de biaiser délibérément les estimateurs, par l'intermédiaire d'un paramètre biaisant, noté k et compris entre 0 et 1. On espère ainsi faire diminuer la part ponctuelle de manière telle que l'augmentation de la part systématique serait compensée et qu'au final la variance résiduelle totale serait plus faible, et le modèle de meilleure qualité donc.

Les estimateurs biaisés sont donnés par :

$$\beta_{r,i} = k \cdot \beta_i$$

L'exemple utilisé tout au long de cette partie devient alors :

$$Y = \beta_{r,0} + \beta_{r,1} \cdot X_1 + \beta_{r,2} \cdot X_2 = k \cdot (\beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2)$$

En pratique, pour savoir quelle valeur de k choisir, il suffit de tracer l'évolution des coefficients β_r en fonction de k . La valeur retenue est celle à partir de laquelle les coefficients se stabilisent.

3.2.2.2.3 Partial Least Squares

À l'instar de la méthode ACP précédemment détaillée, les régressions PLS s'effectuent sur des composantes intermédiaires, appelées variables latentes, plutôt que sur les entrées du problème directement. Toute la difficulté de la méthode réside alors dans le choix du nombre de

composantes. Il doit être le plus faible possible tout en assurant un R^2 suffisamment important pour garantir la qualité du modèle.

3.2.3 Résultats

Quelle que soit la méthode de régression choisie parmi celles qui viennent d'être présentées, le logiciel retourne les résultats sous la même forme (Tableau 3.5 : exemple de régression linéaire de l'ICG sur les variables explicatives « âge » et « longitude »).

On y trouve notamment les valeurs des coefficients de détermination et de détermination ajusté (dans l'en-tête) mais aussi une colonne contenant les valeurs estimées des paramètres de régression, que le logiciel note b au lieu de β .

Un autre résultat intéressant que l'on trouve sur le Tableau 3.5 est la significativité, dans le modèle, de chacune des variables explicatives. Elle est déterminée en réalisant des tests de Student sur chaque paramètre β_i . L'hypothèse H_0 testée est celle d'un paramètre β_i nul. Elle est rejetée si la valeur t_0 (= estimateur du paramètre $\hat{\beta}_i$ / son écart type σ_{β_i}) est supérieure à la valeur critique t donnée par la table de Student pour un degré de liberté de $n - k - 1$. La probabilité que H_0 soit vraie, retranscrite par la p-value, fournit alors directement le résultat escompté. Le paramètre testé est pertinent si elle est inférieure à 0,05 et la variable associée est alors significative. Au contraire, si la p-value est supérieure au seuil fixé, il est possible de retirer de la régression la variable explicative correspondante sans que cela nuise à la qualité du modèle.

Tableau 3.5: Synthèse de la régression

Regression Summary for Dependent Variable: ICG (Base de Donn R= ,38000835 R ² = ,14440634 Adjusted R ² = ,14392942 F(2,3588)=302,79 p<0,0000 Std.Error of estimate: ,81119						
N=3591	b*	Std.Err. of b*	b	Std.Err. of b	t(3588)	p-value
Intercept			4,687235	0,322447	14,5365	0,000000
Age	-0,381182	0,015498	-0,015895	0,000646	-24,5958	0,000000
Longitude	-0,020080	0,015498	-0,005815	0,004488	-1,2957	0,195167

Le Tableau 3.5 indique alors que le modèle de régression est défini par un $R^2_{ajusté}$ de 0,1439 et l'équation de prédiction suivante :

$$ICG = 4,687235 - 0,015895 \cdot \text{âge} - 0,005815 \cdot \text{longitude}$$

Cependant, dans ce modèle, seul l'âge semble jouer un rôle déterminant, la p-value associée à la longitude étant supérieure à 0,05. Le diagramme de Pareto (Figure 3.5) est un outil graphique permettant de représenter de manière claire et efficace ce résultat.

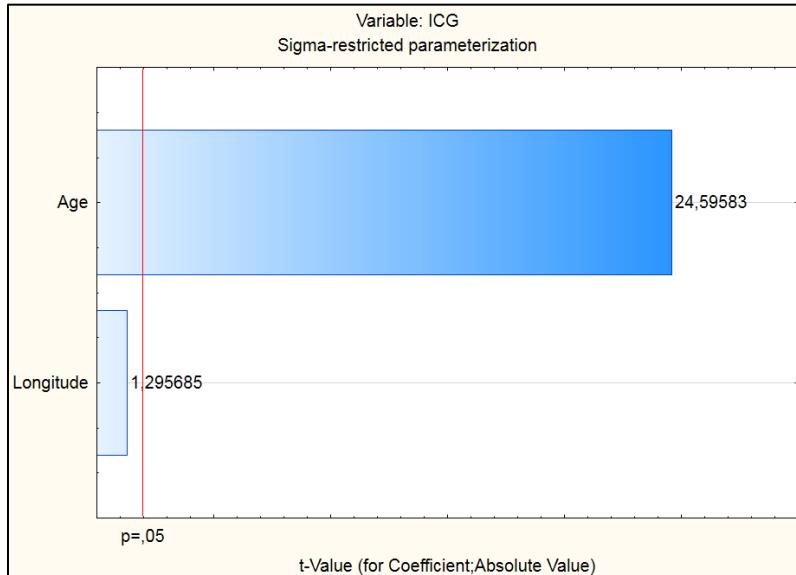


Figure 3.5: Diagramme de Pareto

On y retrouve, en abscisse, les valeurs de t et notamment la valeur critique correspondante à une p-value de 0,05. Ainsi, si la valeur t_0 , calculée et représentée pour la variable X_i , est inférieure la valeur critique (hypothèse H_0 vérifiée), β_i n'est pas significatif. En revanche, si elle est supérieure (hypothèse H_0 rejetée), β_i est significatif et la variable X_i est à prendre en compte dans le modèle.

À partir de l'équation de prédiction, il est également possible de vérifier la qualité du modèle en traçant les valeurs de l'ICG déduites de l'équation, c'est-à-dire les valeurs prédictes, en fonction des valeurs observées en réalité (Figure 3.6). L'objectif est alors d'avoir une droite de régression la plus proche possible de la première bissectrice et dont le coefficient de détermination est le plus élevé possible.

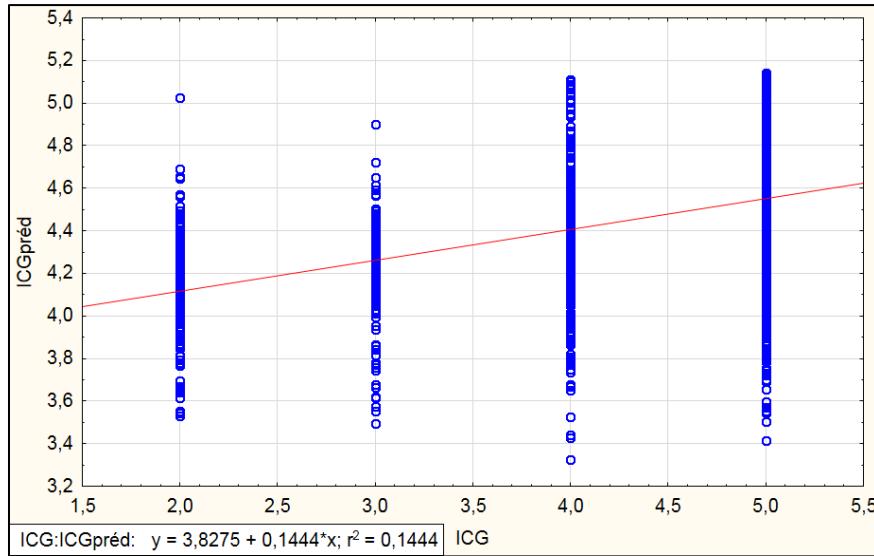


Figure 3.6: Prédictions de l'ICG en fonction des valeurs observées

Remarque :

Le module PLS du logiciel Statistica fait exception car il ne renvoie pas de p-value. Toutefois, il dispose d'un outil équivalent qui permet d'estimer les poids des différents prédicteurs suivant le nombre de composantes choisi. Autrement dit, en fixant un poids limite, il est quand même possible de savoir quelles variables d'entrée ont un effet significatif sur la réponse et lesquelles n'en ont pas.

3.2.4 Critique

On a déjà passé en revue toute une panoplie de méthodes de régression. Cependant, malgré les différences d'approche, elles avaient toutes en commun la prise en compte de l'ICG comme une variable continue suivant une loi de distribution normale. Compte-tenu de la taille de l'échantillon, 3591 observations, cette dernière hypothèse semble fondée mais la continuité de l'ICG peut, elle, être remise en question. En effet, cette variable dépendante est en réalité discrète et ne peut prendre que quatre valeurs différentes : 2, 3, 4 et 5. Le graphe de gauche sur la Figure 3.7 illustre d'ailleurs bien l'inadéquation entre la distribution réelle des valeurs de l'ICG et la loi normale. Il peut alors être intéressant de vérifier s'il n'existe pas d'approche différente, quant à la nature même de l'ICG, donnant de meilleurs résultats.

L'approche logistique semble alors particulièrement adaptée dans la mesure où elle s'applique aux variables de sortie discrètes, binomiales (deux modalités) ou multinomiales (nombre faible, mais supérieur à deux, de modalités). La partie droite de la Figure 3.7 abonde parfaitement dans ce sens.

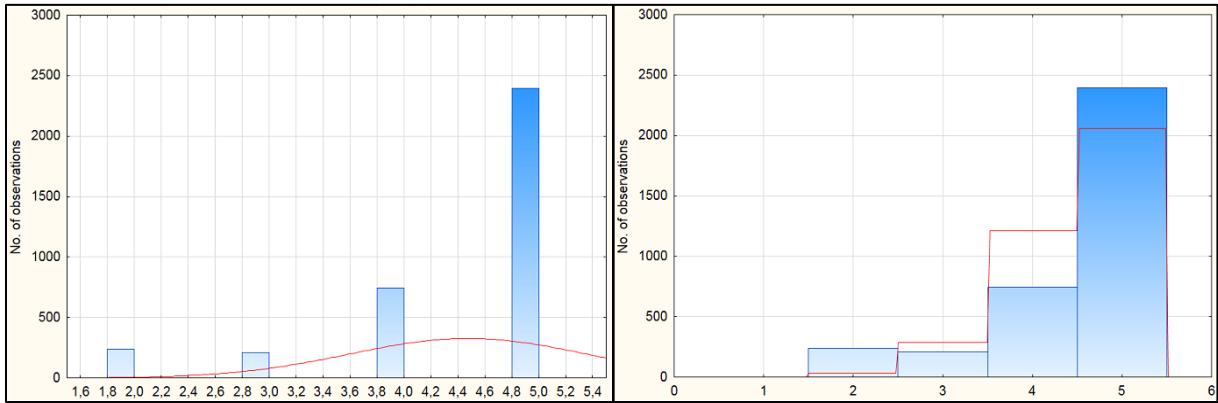


Figure 3.7: Adéquation des données aux lois normale et logistique – Comparatif

3.2.5 Modèles logistiques de régression

3.2.5.1 Généralités

Les régressions logistiques se démarquent particulièrement des autres par leur approche probabiliste du problème. Les paramètres estimés ne servent pas à prédire la valeur exacte de la sortie mais plutôt à évaluer la probabilité qu'elle a de prendre telle ou telle valeur : on parle de probabilité d'affectation. La valeur que va renvoyer le modèle en sortie est alors très facilement interprétable : l'évènement a lieu si elle est supérieure à 0,5 et il n'a pas lieu si elle est inférieure à 0,5.

Par ailleurs, une notion importante définie en régression logistique est le rapport des cotes, noté r . Il se détermine pour chaque variable explicative et correspond au ratio entre la probabilité qu'un évènement ait lieu et la probabilité qu'il n'ait pas lieu, selon les valeurs de l'entrée.

Il reste désormais à définir l'évènement recherché. Plusieurs approches sont alors envisageables selon la manière dont on veut utiliser l'ICG.

3.2.5.2 Approche binomiale

Dans ce premier cas de figure, la sortie doit être binomiale. Il faut donc transformer la variable dépendante afin de ne lui faire prendre que deux valeurs. Il est alors nécessaire d'agréger les données de façon logique. On a ainsi choisi de considérer une nouvelle sortie, Y par exemple, qui prend la valeur 1 si la structure est en bon état ($ICG = 5$) et la valeur 0 si le moindre dégât est à signaler ($ICG = 2, 3$ ou 4).

À partir de là, soit on cherche à détecter un défaut présent sur la structure étudiée ($Y = 0$), soit, au contraire, l'évènement recherché est le bon état de ladite structure ($Y = 1$). Les ratios des cotes correspondants sont respectivement :

$$r = \frac{p(Y = 0)}{p(Y = 1)} \text{ et } r = \frac{p(Y = 1)}{p(Y = 0)}$$

Dans le cas d'une sortie dichotomique, les rapports des cotes permettent de mesurer l'effet des variables d'entrée sur la probabilité d'apparition de l'évènement en question. Si le ratio est supérieur à 1 pour la variable explicative X_i , alors la probabilité augmente lorsque X_i augmente d'une unité et inversement s'il est inférieur. Un ratio égal à 1 indique que la valeur prise par la sortie est indépendante de la valeur de X_i .

En gardant l'exemple d'un modèle à deux variables explicatives X_1 et X_2 , la probabilité d'affectation est donnée par :

$$P(ICG = k) = \frac{e^{\beta_{0,k} + \beta_{1,k} \cdot X_1 + \beta_{2,k} \cdot X_2}}{1 + e^{\beta_{0,k} + \beta_{1,k} \cdot X_1 + \beta_{2,k} \cdot X_2}}$$

Remarque :

Il est possible de généraliser ce raisonnement. Concrètement, de la même façon qu'on a défini une nouvelle variable Y prenant deux modalités, 1 si l'ICG valait 5, 0 sinon, trois autres variables peuvent être créées :

$$\begin{cases} Y_4 = 1 \text{ si l'ICG vaut 4, 0 sinon} \\ Y_3 = 1 \text{ si l'ICG vaut 3, 0 sinon} \\ Y_2 = 1 \text{ si l'ICG vaut 2, 0 sinon} \end{cases}$$

Toutefois, plutôt que de créer quatre variables pour chacune des modalités de l'ICG et d'effectuer quatre régressions logistiques binomiales, le logiciel Statistica dispose d'un outil adapté aux variables discrètes à faible nombre de modalités. Il s'agit de la régression logistique multinomiale.

3.2.5.3 Approche multinomiale

Après ce qui vient d'être vu, la logique de cette dernière est simple à cerner : le modèle estime la probabilité qu'a l'ICG de prendre une valeur par rapport à une modalité K choisie comme référence. Cela est traduit par le ratio des cotes pour la k-ième modalité :

$$r_k = \frac{p(Y = k)}{p(Y = K)}$$

La question du choix de la référence et de son sens se pose alors. Deux options sont explorées dans cette étude.

La première consiste à prendre la valeur 5 de l'ICG, c'est-à-dire une structure en bon état, comme référence. Les probabilités calculées par le modèle correspondent alors aux probabilités qu'il y ait un défaut, plus ou moins graves (« 4 » correspond à un défaut mineur, « 3 » à un défaut significatif et « 2 » à un défaut majeur).

La seconde option est de choisir « 2 », soit le pire état possible, comme modalité de référence. On recherche alors les probabilités que la structure étudiée soit dans un état légèrement meilleur (ICG = 3), dans un état bien meilleur (ICG = 4) et dans un état très nettement meilleur (ICG = 5).

Cette fois-ci, la probabilité d'affectation est donnée par :

$$P(\text{ICG} = k) = \frac{e^{\beta_{0,k} + \beta_{1,k} \cdot X_1 + \beta_{2,k} \cdot X_2}}{1 + \sum_{i \neq K} e^{\beta_{0,i} + \beta_{1,i} \cdot X_1 + \beta_{2,i} \cdot X_2}}$$

Remarque :

L'approche multinomiale semble préférable dans la mesure où toute l'information disponible est conservée, ce qui n'était pas le cas pour l'approche binomiale du fait de l'agrégation des modalités de l'ICG.

3.2.5.4 Critères de qualité

Contrairement aux cas de régressions linéaires, les modèles logistiques ne sont pas estimés par la méthode des moindres carrés mais par celle du maximum de vraisemblance. Cela implique de nombreux changements quant au traitement des données et, plus particulièrement, quant à la façon d'évaluer la performance du modèle. Plusieurs nouveaux critères sont ainsi définis et fournis par le logiciel. On peut notamment les retrouver dans le tableau de synthèse de la régression (Tableau 3.6).

Tableau 3.6: Synthèse de la régression

	Df	Stat.	Stat/Df
Deviance	3579	3974,197898	1,110421
Scaled Deviance	3579	3974,197898	1,110421
Pearson Chi ²	3579	3531,477412	0,986722
Scaled P. Chi ²	3579	3531,477412	0,986722
AIC		3998,197898	
AICC		3998,285098	
BIC		4072,432130	
Cox-Snell R ²		0,152239	
Nagelkerke R ²		0,211533	
Loglikelihood		-1987,098949	

3.2.5.4.1 Logarithme de vraisemblance

Noté LL, il s'agit du logarithme népérien de la vraisemblance. L'objectif est donc, comme pour la vraisemblance, de maximiser sa valeur. En outre, il permet de déterminer un certain nombre d'autres indicateurs de qualité tels que la déviance, le critère d'information d'Akaike (AIC) ou encore le R² de McFadden.

3.2.5.4.2 Déviance

Elle peut être interprétée comme la distance du modèle à la réalité. Le but est donc de minimiser sa valeur qui est donnée par :

$$D = -2 \cdot LL$$

3.2.5.4.3 AIC

Ce critère est assez proche de la déviance dans l'idée mais il s'en différencie par la prise en compte du nombre de paramètres p du modèle. Les modèles trop complexes sont ainsi pénalisés pour tenir compte du problème de surajustement : la qualité d'un modèle augmente forcément

avec le nombre de paramètres mais cela s'accompagne aussi d'une baisse de robustesse. Un compromis est donc à trouver entre concision et qualité.

$$AIC = 2 \cdot p - 2 \cdot LL$$

Comme pour la déviance, plus la valeur de l'AIC est faible et meilleur est le modèle.

3.2.5.4.4 Pseudo R²

Il s'agit d'un indicateur mis en place pour rendre comparable entre eux les modèles. Ainsi, ce pseudo R² se veut le pendant du coefficient de détermination "classique" pour les régressions logistiques. Il existe en fait plusieurs formules de pseudo R². Le logiciel retourne, par exemple, les R² de Cox-Snell et de Nagelkerke. Mais le R² reconnu comme étant le plus adapté à un modèle logistique est celui de McFadden :

$$R^2_{\text{McFadden}} = 1 - \frac{LL}{LL_0}$$

LL₀ correspond au logarithme de vraisemblance d'un modèle trivial ne prenant en compte que la constante (sous-entendu aucune variable, à priori explicative, n'est prise en compte). Le R² de McFadden peut donc être vu comme un indicateur de la performance du modèle testé en comparaison d'un modèle de référence que l'on sait non performant.

Avec :

$$LL_0 = \begin{cases} n_t \cdot \ln(1 - p^+) + n^+ \cdot \ln\left(\frac{p^+}{1 - p^+}\right) & \text{si la sortie est binomiale} \\ \sum_k n_k \cdot \ln\left(\frac{n_k}{n}\right) & \text{si la sortie est multinomiale (k modalités)} \end{cases}$$

Où n_t : nombre total d'observations ;

n⁺ : nombre d'observations positives (observations pour lesquelles l'évènement a lieu) ;

n_k : nombre d'observations / Y = Y_k, k-ième modalité de Y ;

p⁺ : proportion d'observations positives.

Le pseudo R² ne fournit cependant qu'une vague idée de la qualité du modèle. Pour voir si celui-ci est bel et bien plus performant que le modèle trivial, le test statistique du rapport de vraisemblance est plus adapté.

3.2.5.4.5 Test du rapport de vraisemblance

On postule l'hypothèse nulle que tous les coefficients de régression sont égaux à zéro, hormis la constante. En d'autres termes, on teste l'égalité entre le modèle développé et le modèle trivial. Une p-value inférieure à 0,05 est attendue pour que l'hypothèse nulle soit rejetée et que l'on puisse conclure que le modèle développé est effectivement plus performant.

3.2.5.4.6 Matrice de contingence et taux associés

La matrice de contingence donne, pour chaque modalité observée, le nombre de fois où le modèle prédit bien et le nombre de fois où il a mal prédit. Le Tableau 3.7 est un exemple.

Tableau 3.7: Matrice de contingence

	Prédit à 0	Prédit à 1
Observé à 0	358	836
Observé à 1	259	2138

Il est alors possible de calculer le taux d'erreur et les taux de vrais positifs (probabilité qu'une observation à 1 soit bien détectée) et vrais négatifs (idem mais pour les observations à 0).

L'objectif est bien entendu de minimiser le taux d'erreur ϵ et maximiser les taux de vrais positifs et vrais négatifs, également appelés, respectivement, sensibilité et spécificité.

Dans le cas du Tableau 3.7, par exemple, on a :

$$\left\{ \begin{array}{l} \epsilon = \frac{836 + 259}{358 + 836 + 259 + 2138} \approx 0,305 \\ \text{Sensibilité} = \frac{2138}{259 + 2138} \approx 0,892 \\ \text{Spécificité} = \frac{358}{358 + 836} \approx 0,300 \end{array} \right.$$

3.2.5.4.7 Courbe "receiver operating characteristics" (ROC)

Enfin, une dernière méthode est celle de l'aire sous la courbe ROC. Il s'agit d'une courbe sur laquelle le taux de vrais positifs (la sensibilité) apparaît en ordonnée et le taux de faux positifs ($1 - \text{la spécificité}$) en abscisse (Figure 3.8).

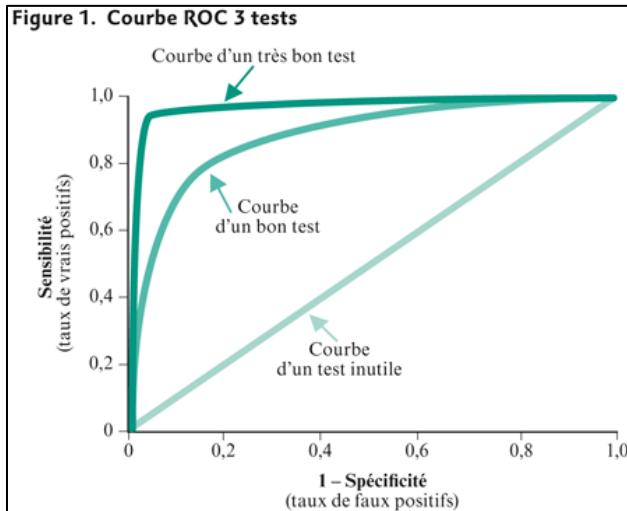


Figure 3.8: Courbe ROC (*Revue médicale suisse*)

Le modèle est intéressant quand le taux de vrais positifs augmente sans que le taux de faux positifs augmente aussi. Concrètement, cela se manifeste par une aire sous la courbe proche de 1. Au contraire, la courbe d'un modèle inintéressant est proche de la première bissectrice (aire proche de 0,5 donc) : les résultats sont dus au hasard et les taux de vrais positifs et faux positifs évoluent sensiblement dans les mêmes proportions.

3.2.5.5 Critères de vérification du modèle

L'hypothèse d'observations indépendantes les unes des autres ayant déjà été formulée, il suffit, dans ce cas, de réaliser une analyse de résidus pour vérifier l'adéquation du modèle aux données. L'idée et la manière de procéder sont donc les mêmes que dans la partie 3.2.1.4.3 – Analyse de sensibilité.

3.2.5.6 Résultats

Comme pour les modèles basés sur un ICG continu, on peut obtenir les coefficients de régression (colonne « Estimate » sur le Tableau 3.8) qui permettent d'écrire l'équation, non linéaire cette

fois, de prédiction et de tracer les probabilités prédictives en fonction de l'occurrence réelle de l'évènement étudié.

À nouveau, il est possible de déterminer les variables explicatives qui sont pertinentes dans le modèle grâce à un test de Wald visant à éprouver la significativité des coefficients estimés associés (colonne « p » sur le Tableau 3.8).

Tableau 3.8: Estimation des paramètres de la régression logistique

Effect	Level of Effect	Colu mn	Estimate	Standard Error	Wald Stat.	Lower CL 95,0%	Upper CL 95,0%	p
Intercept		1	-1.85080	0.912928	4,1100	-3,64010	-0,061491	0,042630
Longitude		2	0,01018	0,012701	0,6431	-0,01471	0,035077	0,422596
Age		3	0,04411	0,002242	387,0971	0,03971	0,048501	0,000000
Scale			1,00000	0,000000		1,00000	1,000000	

Il est également possible d'obtenir ces résultats en effectuant, sur les rapports de vraisemblance, un test du Chi-2 pour lequel on regarde à quel point le modèle développé est proche du modèle trivial si on suppose que β_i est nul. Si la différence entre les deux n'est statistiquement pas significative, on en déduit que X_i ne joue pas de rôle important quant aux valeurs prises par la sortie.

Ces différents tests sont, en quelque sorte, les équivalents du test de Student utilisé pour les méthodes de régression linéaire.

3.3 Traitement combiné des entrées quantitatives et qualitatives

L'intérêt principal du modèle général de régression présenté dans cette partie est de pouvoir prendre en compte à la fois les variables d'entrée continues et les facteurs explicatifs catégoriques. Ainsi, la nature même de cette méthode la place à mi-chemin entre les régressions et les méthodes d'analyse de la variance (ANOVA). Par conséquent, les résultats que l'on peut trouver avec l'aide du logiciel sont globalement les mêmes. L'équation prédictive, tenant compte de toutes les variables d'entrée, quelque soit leur nature, est alors particulièrement intéressante dans le cadre du projet.

Pour la mettre en place, les facteurs catégoriques doivent néanmoins être codés pour être exploitables dans la régression. Le principe retenu est alors celui du codage à effet : on crée, pour

chaque variable catégorique à k modalités, k – 1 variables. Ces variables, $C_{M,i}$, où M est le facteur catégorique codé et i la i-ème modalité de ce facteur, peuvent prendre trois valeurs possibles :

- ✓ $C_{M,i} = 1$ si $M = i$;
- ✓ $C_{M,i} = -1$ si $M = K$ (modalité de référence) ;
- ✓ $C_{M,i} = 0$ sinon.

Avec cette méthode, les critères de qualité et de vérification du modèle sont les mêmes que ceux des méthodes de régression "classiques".

Remarque :

La prise en compte des effets d'interaction est, dans le cas de l'étude, impossible. En effet, il y a trop de variables catégoriques et trop de modalités pour que chaque combinaison possible de modalités soit représentée dans les données. Cela se vérifie même pour des effets d'interaction d'ordre 2, c'est-à-dire des interactions entre seulement deux facteurs explicatifs (modalité i du premier et modalité j du second).

Compte-tenu du nombre de modalités de chaque facteur explicatif, 181 400 combinaisons possibles existent. Il est donc logique que la base de données de 3591 points ne permette pas d'avoir ne serait-ce qu'une seule équivalence pour chaque combinaison.

CHAPITRE 4 RÉSULTATS

4.1 Préliminaires

4.1.1 Variables explicatives

Un important inventaire de méthodes statistiques a donc été réalisé dans la partie précédente. Celles-ci peuvent toutes être testées sur les structures de la base de données et l'objectif est alors de déterminer laquelle est la plus adaptée et fournit les meilleurs résultats. Cependant, les ponts constituant la grande majorité des structures répertoriées (Tableau 2.1), il a été décidé de ne se focaliser, dans un premier temps, que sur ce type de structures en particulier. Ce sont ainsi dix-huit variables explicatives qui sont utilisées pour l'étude statistique (Figure 4.1) : onze quantitatives et huit qualitatives.

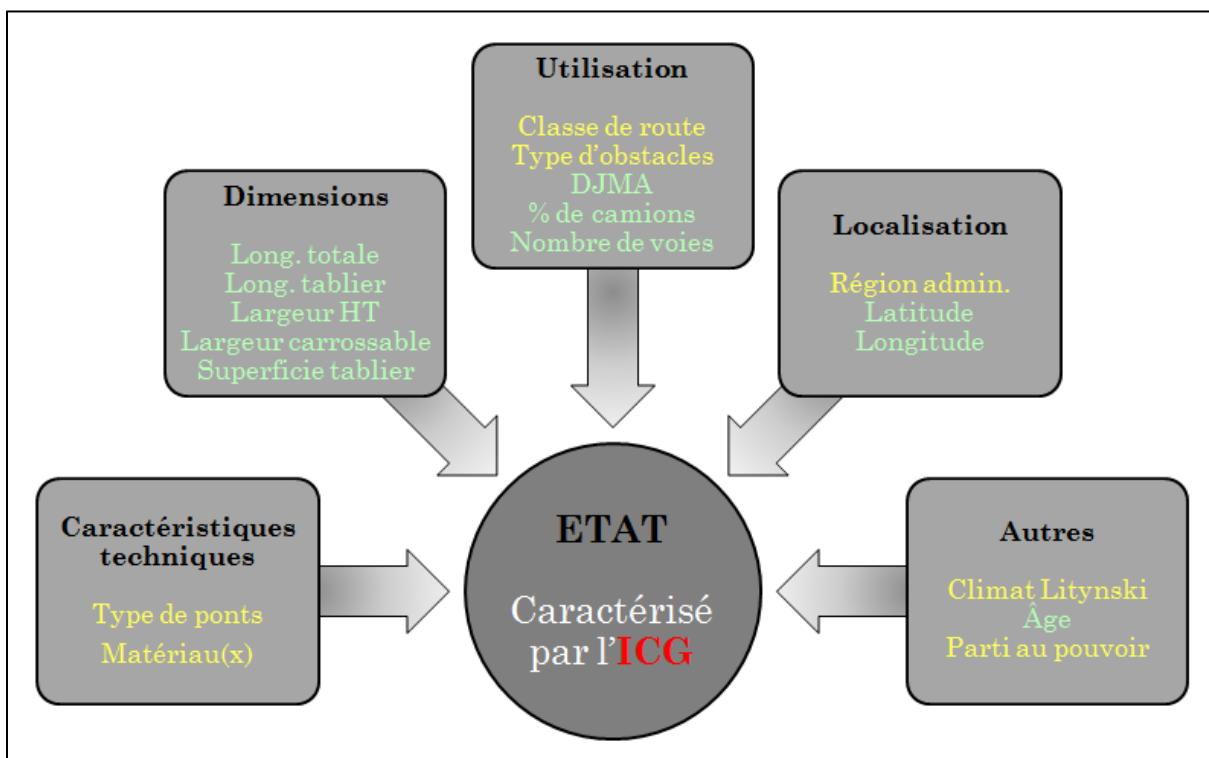


Figure 4.1: Récapitulatif des variables utilisées

Les appellations utilisées pour certaines d'entre elles peuvent néanmoins paraître "obscures". Il convient donc de préciser que :

- ✓ La longueur totale renvoie à la longueur de l'ouvrage dans son intégralité ;
- ✓ La longueur du tablier correspond à la longueur de la structure porteuse qui supporte les charges induites par le trafic ;
- ✓ La largeur hors tout désigne la distance entre les deux flancs extérieurs de l'ouvrage ;
- ✓ La largeur carrossable est égale à la largeur totale des voies de circulation ;

4.1.2 Démarche

Au niveau de la démarche suivie, on s'intéresse en premier lieu à l'analyse descriptive des données. Puis, sont mis en place plusieurs modèles de régression ne se basant que sur les variables quantitatives, le but n'étant pas tant de prévoir que d'identifier celles qui sont significatives. Enfin, les variables numériques et qualitatives sont traitées conjointement à des fins prédictives. Autrement dit, on cherche à savoir à quel point il est possible de prévoir les valeurs d'ICG à partir des données d'entrée sélectionnées pour le projet.

Cette démarche peut, bien entendu, être ensuite appliquée aux autres structures que sont les ponceaux et les murs.

Remarque :

Les différentes étapes et sous-étapes de ces analyses, réalisées grâce au logiciel Statistica, peuvent être très largement illustrées. Néanmoins, dans un souci d'allègement et de lisibilité du rapport, seuls les figures et tableaux essentiels y sont intégrés. Ainsi, les valeurs des différents critères de qualité ou d'adéquation du modèle, dont les processus d'obtention ont déjà été décrits dans le Chapitre 3, sont parfois reportées dans ce document sans que ce soit le cas des figures associées.

4.2 Statistiques descriptives

4.2.1 Traitement des variables catégoriques

4.2.1.1 Type de ponts

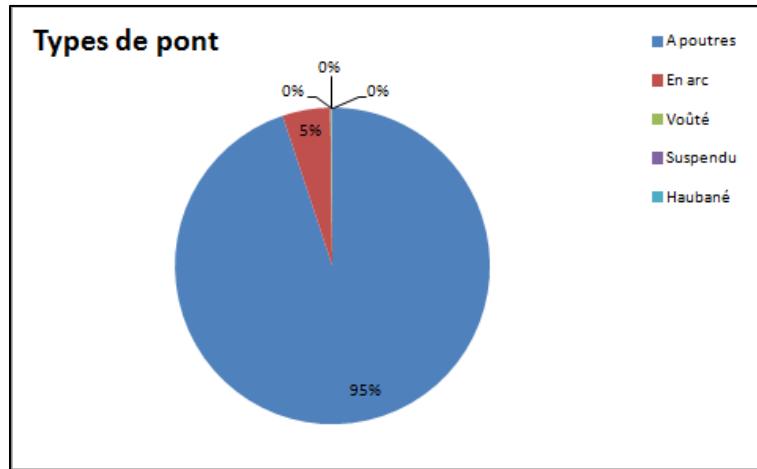


Figure 4.2: Répartition des ponts selon leur type

Comme le montre la Figure 4.2, la grande majorité des ponts possèdent en fait des éléments porteurs qui sont des poutres sollicitées en flexion. Les ponts en arc sont également bien représentés, ce qui n'est pas le cas des ponts voûtés, suspendus et à haubans. Les faibles proportions que représentent ces trois derniers types parmi les 3591 ponts étudiés vont conduire à des résultats très peu significatifs. C'est pourquoi les structures correspondantes vont être, par la suite, exclues de l'étude. Pour illustrer ce propos et mieux comprendre la raison de cette exclusion, on prend l'exemple des ponts suspendus. Seulement deux d'entre eux sont recensés sur l'ensemble des 3591 ponts. On peut alors se demander si calculer une valeur moyenne pour l'ICG sur deux éléments aurait un sens ? Statistiquement parlant, la réponse est évidemment négative.

Il est également possible de rechercher les valeurs moyennes de l'ICG pour chacune des modalités et de les représenter graphiquement. Des tendances peuvent ainsi d'ores-et-déjà se dégager, même si cela reste ensuite à vérifier. Pour ce qui est du type de ponts, l'ICG moyen est plus élevé pour les ponts à poutres que pour les ponts en arc (Figure 4.3).

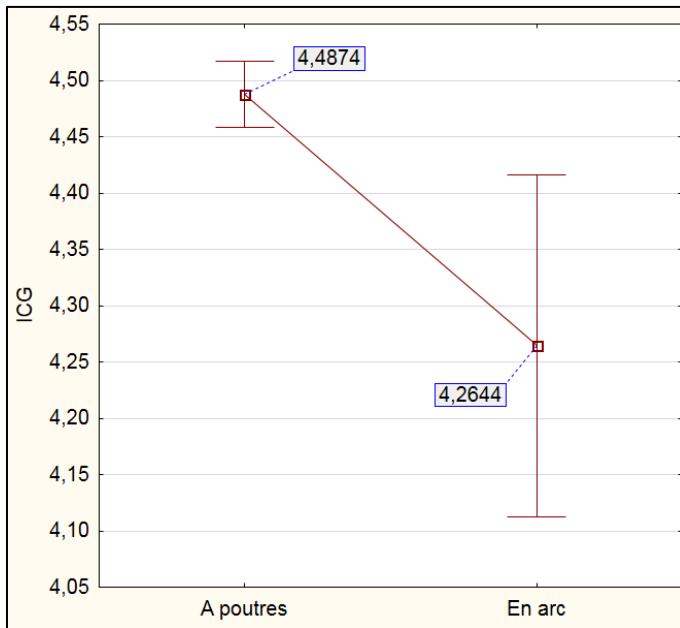


Figure 4.3: ICG moyen en fonction du type de ponts

Remarque :

Sur ce type de graphe, l'intervalle défini sur la figure et centré sur la moyenne a une longueur de quatre fois l'écart type. Statistiquement, il correspond à la plage dans laquelle 95% des valeurs se situent. Plus il est resserré autour de la moyenne et plus celle-ci est représentative. Il porte ainsi le nom d'intervalle de confiance.

Il faut cependant s'assurer que cette différence soit réellement significative. Pour cela, plusieurs tests statistiques formels permettent de comparer les moyennes entre les différentes modalités prises par le facteur explicatif. Que ce soit le test de Mann-Whitney dans ce cas de figure, où seulement deux modalités existent, ou celui de Kruskal-Wallis, pour un nombre de modalités supérieur, l'égalité entre les différents groupes constitue l'hypothèse nulle. Une différence significative entre deux catégories est donc traduite par une p-value inférieure à 0,05.

Tableau 4.1: Test comparant les valeurs moyennes d'ICG selon le type de ponts

p-value	À poutres	En arc
À poutres	–	0,005135
En arc	0,005135	–

Au regard du Tableau 4.1, une corrélation entre le type de ponts et l'ICG existerait bel et bien : les ponts dont les éléments porteurs sont sollicités en flexion (les ponts à poutres) semblent être globalement en meilleur état que ceux dont les éléments porteurs sont sollicités principalement en compression (les ponts en arc).

4.2.1.2 Matériaux utilisés

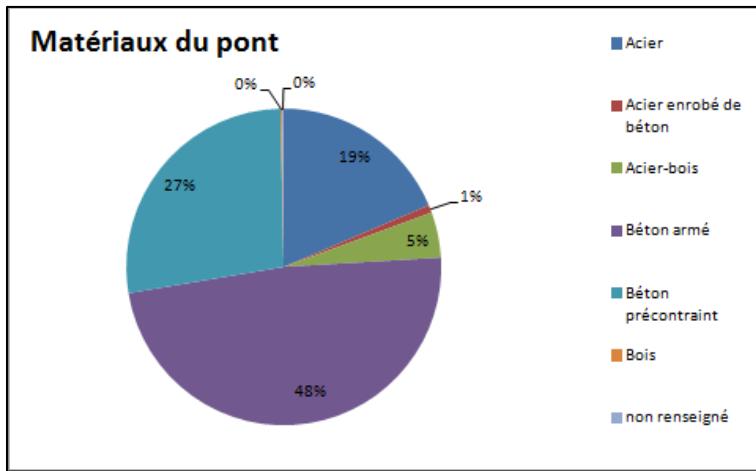


Figure 4.4: Répartition des ponts selon le matériau utilisé

La répartition des ponts par matériaux (Figure 4.4) laisse apparaître que le béton armé est le matériau le plus employé. L'acier et le béton précontraint sont aussi très utilisés tandis que l'acier enrobé de béton, le couple acier-bois et le bois seul sont des matériaux plutôt "exotiques" pour ce type de structures et ne sont que très peu utilisés. Les modalités « bois » et « non renseigné » sont d'ailleurs abandonnées car elles ne sont pas suffisamment représentées.

L'étude de la valeur moyenne de l'ICG en fonction des matériaux employés est plus intéressante si le type d'efforts auxquels sont soumises les structures est pris en compte ; le comportement et la résistance d'un même matériau ne sont pas les mêmes suivant la façon dont il est sollicité. Ainsi, deux tracés sont réalisés : un premier pour les ponts à poutres (flexion), Figure 4.5, et le second pour les ponts en arc (compression), Figure 4.6.

4.2.1.2.1 Ponts à poutres

Pour ce qui est des ponts à poutres, il peut être intéressant de commencer par faire un point sur les sollicitations en flexion et sur les propriétés mécaniques des matériaux.

Tout d'abord, il faut savoir qu'en flexion (chargement vertical descendant, par exemple) une partie de la poutre est comprimée (partie supérieure à la fibre neutre dans l'exemple) tandis que l'autre partie travaille en traction (celle qui est inférieure à la fibre neutre).

Il est également important de noter qu'un acier de construction classique (de type S235, par exemple) a une résistance en compression à peu près équivalente à sa résistance en traction (limite élastique de 235 MPa pour l'acier cité précédemment). Le béton peut, lui, avoir une résistance en compression satisfaisante, allant de 20 à 50 MPa pour les bétons classiques et jusqu'à 70 MPa pour les bétons haute performance (BHP), voire 150 MPa pour les bétons ultra hautes performances (BUP), mais a une résistance en traction beaucoup plus faible (de l'ordre de 10% de sa résistance en compression). Le principe du béton armé est alors de combiner les points forts de ces matériaux pour créer un matériau globalement plus résistant que le béton seul : on utilise la résistance à la compression du béton tandis que les armatures en acier reprennent les efforts de traction. C'est d'ailleurs pour cela que, sur les poutres sollicitées en flexion, les armatures ne se retrouvent que dans la zone étirée. Cette logique se retrouve aussi pour le béton précontraint, sauf qu'en plus de cela le béton est initialement comprimé pour améliorer sa tenue en traction. Il est donc logique de voir, sur la Figure 4.5, que les ponts à poutres en béton précontraint sont globalement en meilleur état que ceux pour lesquels elles sont en béton armé.

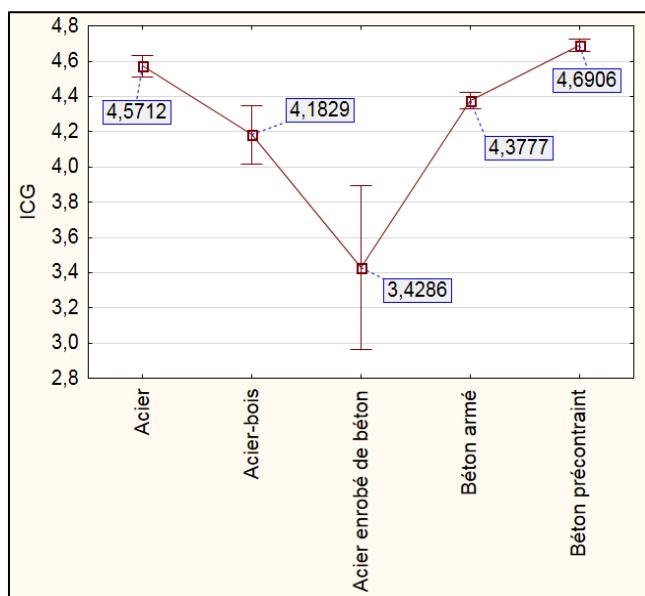


Figure 4.5: ICG moyen en fonction du matériau
(Ponts à poutres)

La Figure 4.5 indique également qu'entre ces deux types de ponts à poutres s'intercalent les ponts à poutres en acier. Il est alors possible de penser que le gain de résistance en traction engendré par l'ajout d'armatures en acier haute résistance ne suffit pas au béton armé pour atteindre une résistance en flexion équivalente à celle de l'acier mais que cela est le cas lorsque ces armatures sont précontraintes. Cela tend à être confirmé par le Tableau 4.2 : s'il n'existe pas de différence significative entre les poutres en béton précontraint et en acier, la valeur obtenue pour celles en béton armé semble, en revanche, se démarquer.

Tableau 4.2: Test comparant les valeurs moyennes d'ICG selon le matériau (Ponts à poutres)

p-value	Acier-bois	Béton armé	Acier	Béton préc.	Acier enrobé
Acier-bois	–	0,309445	0,000956	0,000009	0,046583
Béton armé	0,309445	–	0,004485	0,000000	0,000738
Acier	0,000956	0,004485	–	1,000000	0,000019
Béton préc.	0,000009	0,000000	1,000000	–	0,000002
Acier enrobé	0,046583	0,000738	0,000019	0,000002	–

Après ces trois premiers matériaux, on retrouve, par ordre d'ICG moyen décroissant, les ponts à poutres en acier-bois et en acier enrobé de béton. Il n'est pas surprenant de retrouver les poutres en acier-bois car leur usage en extérieur dans une province telle que le Québec et son rude climat constitue un facteur dégradant important.

Enfin, les ponts à poutres en acier enrobé de béton sont ceux dont la valeur moyenne est la plus faible car il s'agit d'une technologie ancienne, dépassée depuis par le béton armé.

4.2.1.2.2 Ponts en arc

En compression, contrairement à la flexion, la Figure 4.6 semble indiquer que, si les ponts en béton précontraint sont toujours ceux en meilleur état, ceux en béton armé sont globalement en meilleur état que ceux en acier. Cela pourrait s'expliquer par le fait qu'en compression les deux matériaux, le béton et l'acier, offrent le meilleur de leurs performances en termes de résistance mécanique.

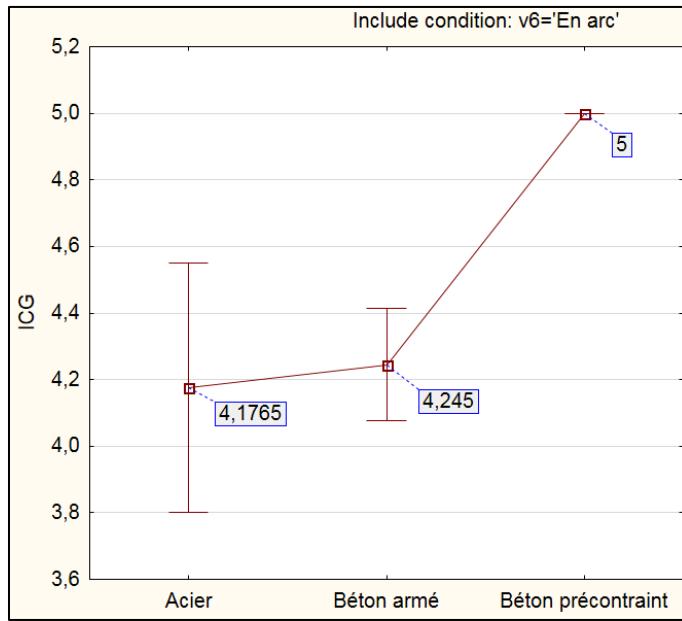


Figure 4.6: ICG moyen en fonction du matériau
(Ponts en arc)

Toutefois, aucune différence significative ne semble exister entre les différents matériaux utilisés (Tableau 4.3) ; la tendance observée ne peut donc pas être considérée comme un résultat pertinent de l'étude.

Tableau 4.3: Test comparant les valeurs moyennes d'ICG selon le matériau (Ponts en arc)

p-value	Béton armé	Acier	Béton préc.
Béton armé	–	0,926975	0,187387
Acier	0,926975	–	0,087492
Béton préc.	0,187387	0,087492	–

4.2.1.3 Classe de routes

La Figure 4.7 montre que la grande majorité des structures étudiées se trouvent sur seulement quatre types de route, qui sont :

- ✓ Les autoroutes (25,3%) : routes principales à haute vitesse et accès limité ;

- ✓ Les routes nationales (26,0%) : routes principales interrégionales liant les agglomérations urbaines importantes (généralement plus de 25 000 habitants) entre elles ;
- ✓ Les routes régionales (15,7%) : routes liant les agglomérations urbaines secondaires (entre 5000 et 25 000 habitants, globalement) entre elles ou aux agglomérations principales ;
- ✓ Les routes collectrices (21,3%) : routes liant les agglomérations rurales (moins de 5000 habitants) aux agglomérations urbaines proches.

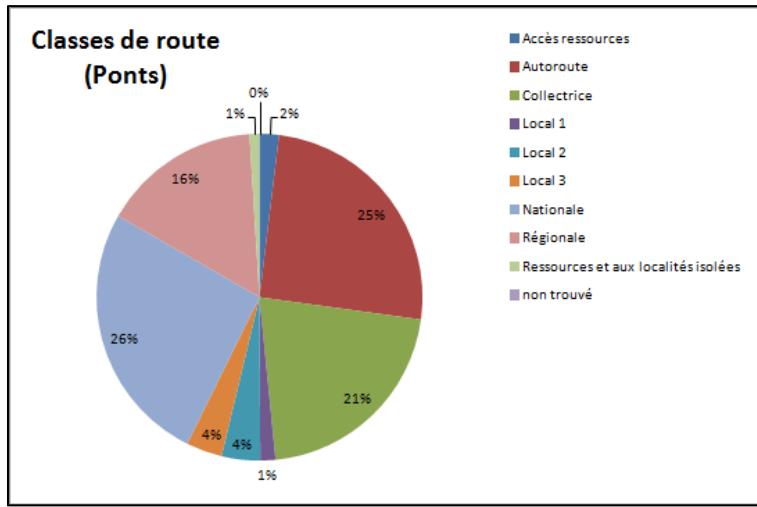


Figure 4.7: Répartition des ponts selon la classe de routes

Ces routes forment, dans l'ordre décroissant d'importance, le réseau supérieur, géré par le MTQ. À cela, il faut également ajouter les routes du réseau local gérées par les municipalités qui sont, par ordre décroissant d'importance :

- ✓ Les routes locales de type 1 : routes liant les centres ruraux entre eux ;
- ✓ Les routes locales de type 2 : routes d'accès aux propriétés rurales habitées en permanence ;
- ✓ Les routes locales de type 3 : routes d'accès aux propriétés rurales non habitées en permanence et rues urbaines résidentielles.

Enfin, il existe un dernier groupe de routes composé des routes « d'accès aux ressources » et des routes « d'accès aux ressources et aux localités isolées ». Ce sont des routes qui donnent accès à des zones d'exploitations forestière et minière, à des parcs ou encore à des réserves fauniques, par exemple. Il est difficile de les caractériser et cela se traduit aussi au niveau de leur gestion.

Certaines dépendent du MTQ certes, mais d'autres sont sous la responsabilité du ministère de l'Énergie et des Ressources naturelles (MERN) ou d'Hydro-Québec.

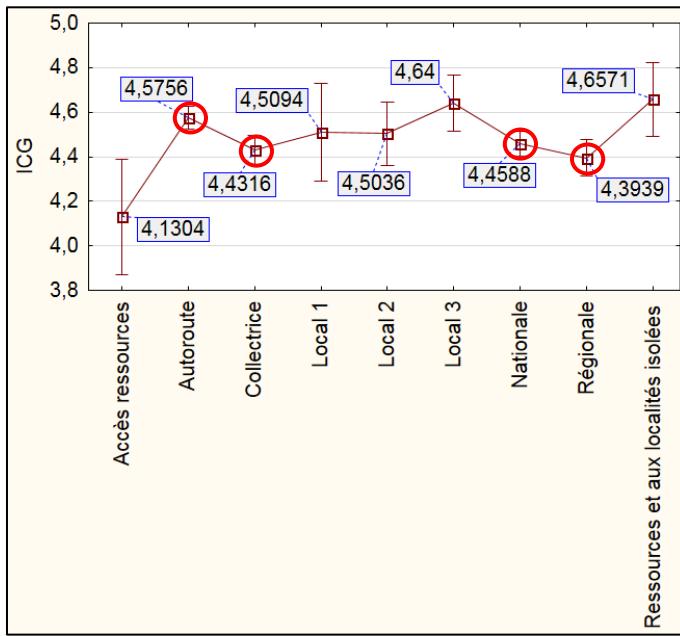


Figure 4.8: ICG moyen en fonction du type de routes

Le tracé de la valeur moyenne de l'ICG en fonction du type de routes (Figure 4.8) permet de mettre en évidence quelques tendances intéressantes. Tout d'abord, les routes gérées par les municipalités paraissent être globalement en meilleur état que les routes gérées par le ministère (valeurs entourées).

Il faut également souligner que si les moyennes sont à peu près équivalentes pour chaque classe de routes relevant des municipalités, il en est différemment pour celles du MTQ. En effet, les autoroutes semblent être en meilleur état que les routes nationales. De plus, si les routes régionales et les routes collectrices ont des valeurs très proches, elles sont en moins bon état que les routes nationales. Pour ce qui relève du ministère, il semblerait donc que plus la route est importante et plus elle est entretenue, et donc meilleur est son état.

À la lumière du test de Kruskal-Wallis, les valeurs moyennes obtenues ne sont cependant pas suffisamment différentes pour que les résultats issus de la Figure 4.8 soient significatifs.

4.2.1.4 Région administrative

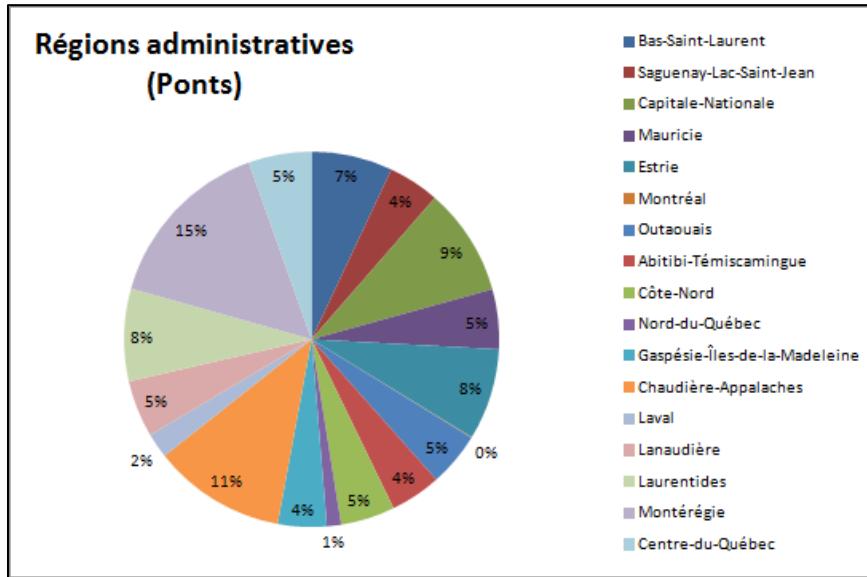


Figure 4.9: Répartition des ponts selon la région administrative

Les structures sont à peu près équitablement réparties entre les régions administratives, comme il est possible de le voir sur la Figure 4.9. Les régions de Laval et du Nord-du-Québec présentent néanmoins des densités en structures légèrement plus faibles que les autres régions. À l'inverse, la Montérégie présente une densité supérieure. Enfin, la représentation de la région de Montréal est tellement insignifiante (0,06%) que les deux structures correspondantes sont exclues dans la suite de l'étude.

D'après le graphe d'évolution de la valeur moyenne de l'ICG en fonction de la région (Figure 4.10), la grande majorité des régions (douze sur seize) ont un ICG moyen compris entre environ 4,4 et 4,7. S'il ne semble donc pas y avoir de variation significative de valeurs entre ces douze régions, quatre autres se démarquent en sortant de cet intervalle de valeurs. Les structures du Centre-du-Québec semblent ainsi être en moins bon état que les autres, derrière celles des régions d'Abitibi-Témiscamingue et de Montérégie (moyennes inférieures à 4,4). Au contraire, les structures de l'Outaouais semblent être en meilleur état que les autres avec un ICG moyen proche de 4,8.

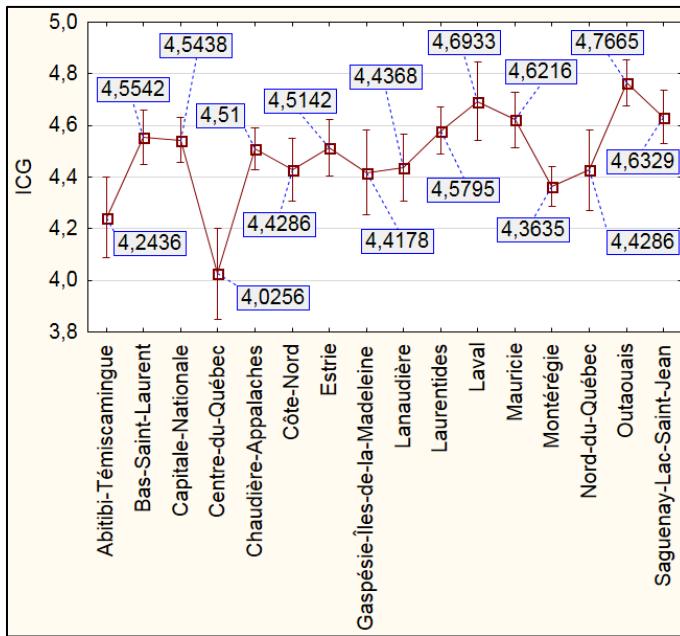


Figure 4.10: ICG moyen en fonction de la région

Cependant, le test de Kruskal-Wallis réalisé apporte une précision : seules les valeurs moyennes obtenues pour les régions de l'Outaouais et du Centre-du-Québec sont significativement différentes des autres. Les structures de l'Abitibi-Témiscamingue et de la Montérégie semblent finalement rentrer dans une plage de valeurs que l'on pourrait qualifier de "normales".

Pour mieux comprendre les raisons de ces écarts, il faudrait étudier plus en détail les caractéristiques de chacune des régions et les regrouper entre elles s'il y a des similarités.

4.2.1.5 Type d'obstacles

Comme cela est visible sur la Figure 4.11, la très grande majorité des ponts ont été construits dans le but d'enjamber soit des cours d'eau, soit des routes. Le reste des obstacles ne représente qu'une toute petite proportion des ponts. Tellement petite dans le cas des passages d'animaux, des stationnements et autres, que les observations correspondantes sont abandonnées.

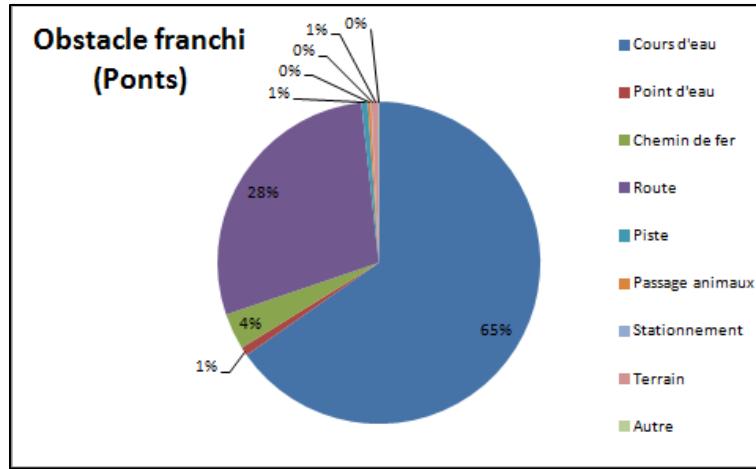


Figure 4.11: Répartition des ponts selon le type d'obstacles

En outre, il peut être intéressant de s'attarder à ces différentes modalités pour détailler ce qui se cache réellement sous de telles dénominations. La notion de cours d'eau est évidente mais il faut préciser qu'elle recoupe des cours d'eau de tout type (naturels, canaux, etc.) et de toute taille (fleuves, ruisseaux, décharges, etc.). De la même façon, les points d'eau regroupent toutes les étendues d'eau "stagnantes" (lacs, étangs, bassins, etc.) et les routes réfèrent à tous les types de chaussées (autoroutes, rues, voies pour bus, etc.). La modalité « piste » renvoie aux chemins réservés à des usagers tels que les piétons, les cycles et autres motoneiges. Enfin, le terme de terrain regroupe tous les ponts enjambant des talus, fossés ou autres variations de terrain empêchant le tracé d'une chaussée à même le sol.

Au regard de la Figure 4.12, les ponts dont l'état global est le meilleur sont ceux dont les obstacles sont des variations de terrain. Leur bon état peut s'expliquer par le fait qu'il n'y ait, à priori, pas de facteurs dégradants autres que le temps et la circulation sur le pont en jeu. Dans cette optique, il est donc logique de trouver ensuite des ponts soumis à d'autres facteurs dégradants. Ainsi, viennent, dans l'ordre, les ponts servant à franchir des routes, des pistes, des chemins de fer, des cours d'eau puis des points d'eau. Dans chacun de ces cas, un nouveau facteur dégradant est introduit par l'obstacle : circulation sous le pont, humidité/eau, par exemple.

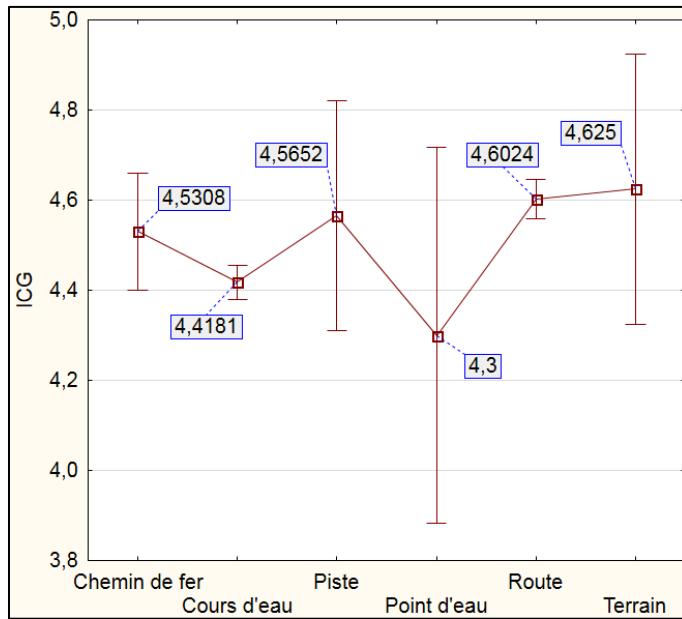


Figure 4.12: ICG moyen en fonction de l'obstacle

À la vue des résultats, l'eau semble donc être plus néfaste pour les structures que la circulation de véhicules et la pollution qui en résulte. Cela est confirmé par le test de Kruskal-Wallis (Tableau 4.4), celui-ci faisant, en effet, état d'une différence significative entre les valeurs moyennes des modalités « cours d'eau » et « route ».

Tableau 4.4: Test comparant les valeurs moyennes d'ICG selon l'obstacle franchi

p-value	Cours d'eau	Point d'eau	Route	Chemin de fer	Piste	Terrain
Cours d'eau	—	1,000000	0,010375	1,000000	1,000000	1,000000
Point d'eau	1,000000	—	1,000000	1,000000	1,000000	1,000000
Route	1,000000	1,000000	—	1,000000	1,000000	1,000000
Chemin de fer	1,000000	1,000000	1,000000	—	1,000000	1,000000
Piste	1,000000	1,000000	1,000000	1,000000	—	1,000000
Terrain	1,000000	1,000000	1,000000	1,000000	1,000000	—

4.2.1.6 Type de climat

Le nombre de structures concernées par la modalité « 4 » n'étant pas suffisant pour que les résultats soient pertinents (retour sur le Tableau 2.4), ce sont donc sept types de climat qui, au final, sont étudiés et comparés.

D'après la Figure 4.13, il est possible de classer, par ordre d'ICG moyen décroissant, les zones :

1. Zone 9 (4,554) ;
2. Zone 10 (4,488) ;
3. Zone 12 (4,484) ;
4. Zone 6 (4,467) ;
5. Zone 11 (4,461) ;
6. Zone 7 (4,406) ;
7. Zone 8 (4,294).

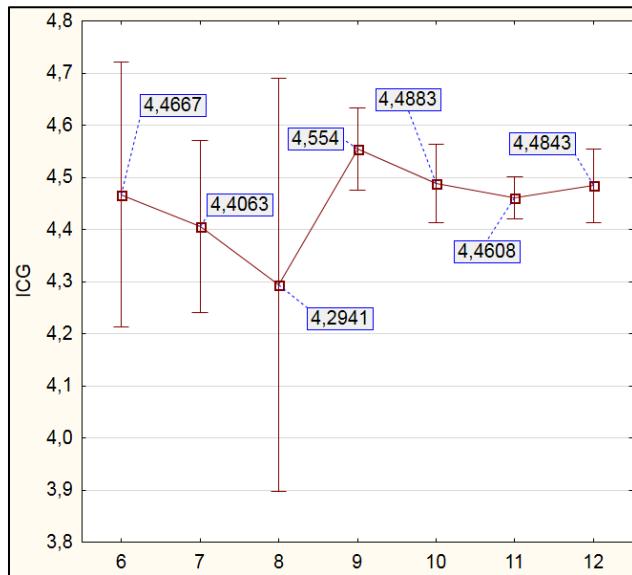


Figure 4.13: ICG moyen en fonction du climat

Remarque :

Les résultats pour les zones « 6 », « 7 » et « 8 » semblent néanmoins assez incertains.

Il est cependant très difficile de discuter de ce classement en l'état actuel. C'est pour pallier à ce problème qu'a été mis en place le Tableau 4.5. Il reprend ainsi le classement ci-dessus en précisant pour toutes les zones les trois caractéristiques fondamentales de la classification de Litynski, à savoir températures, précipitations et durée de la saison de croissance.

Tableau 4.5: Classement par ICG moyen des différents types de climat

Clt	Zone	Températures	Précipitations	Durée croissance
1	9	Subpolaires douces	Subhumide	Longue
2	10	Subpolaires	Humide	Moyenne
3	12	Subpolaires	Subhumide	Moyenne
4	6	Subpolaires froides	Subhumide	Courte
5	11	Modérées	Subhumide	Longue
6	7	Subpolaires froides	Subhumide	Moyenne
7	8	Subpolaires	Humide	Courte

Du point de vue climatique, aucune logique d'évolution de l'ICG moyen en fonction de telle ou telle caractéristique ne semble facilement identifiable. Le test de Kruskal-Wallis vient, en outre, confirmer cette conclusion dans le sens où aucune zone climatologique ne peut, au regard des seules statistiques, être distinguée des autres⁹.

4.2.1.7 Parti politique au pouvoir

La tendance qui se dégage de la Figure 4.14 est assez logique. En effet, l'UN ayant été dissoute en 1989, il est envisageable que les structures concernées soient globalement plus anciennes que celles des deux autres partis, toujours existants à ce jour. Cela expliquerait le fait que ces structures soient globalement en moins bon état, le temps étant un facteur dégradant reconnu. Ce point reste malgré tout à vérifier étant donné le fait que le PLQ ait été fondé avant l'UN. Il est en revanche beaucoup plus probable que les structures dépendantes du PQ soient plus récentes, et donc en meilleur état, que les autres.

⁹ Aucune comparaison entre les modalités ne fait état d'une différence significative.

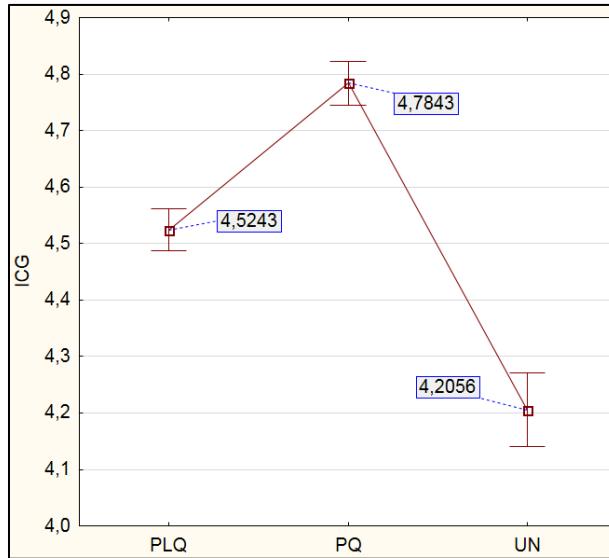


Figure 4.14: ICG moyen en fonction du parti

Tableau 4.6: Test comparant les valeurs moyennes d'ICG selon le parti

p-value	PLQ	UN	PQ
PLQ	—	0,000000	0,000004
UN	0,000000	—	0,000000
PQ	0,000004	0,000000	—

Si le Tableau 4.6 montre que les différences entre les partis sont bel et bien significatives, le Tableau 4.7, qui, pour chaque parti, donne l'année moyenne de construction des structures confirme, quant à lui, les hypothèses formulées ci-dessus. Ainsi, les structures construites sous l'affiliation du PQ seraient en meilleur état car elles sont plus récentes et celles affiliées à l'UN, de par leur ancienneté, seraient en moins bon état.

Tableau 4.7: Année de construction moyenne et ICG moyen selon le parti

Parti	Moyenne	Écart type	ICG moyen
UN	1958	8	4,2056
PLQ	1979	22	4,5243
PQ	1993	13	4,7843

4.2.2 Traitement des variables quantitatives

4.2.2.1 Descriptions statistiques des données avec les variables quantitatives

Le Tableau 4.8 présente les valeurs moyennes, minimales et maximales ainsi que l'écart-type obtenus, pour chaque variable quantitative, à partir des informations sur les structures de la base de données.

Tableau 4.8: Description des données

Variable	Min.	Moy.	Max.	Écart type	+/- 95%
Latitude	45,01	46,73	58,67	1,38	[46,68 ; 46,77]
Longitude	- 79,48	- 72,04	- 57,23	3,03	[- 72,14 ; - 71,94]
Long. totale	4,6 m	69,8 m	2 708,3 m	96,4 m	[66,7 m ; 73,0 m]
Long. tablier	4,5 m	59,0 m	2 702,4 m	94,7 m	[55,9 m ; 62,1 m]
Largeur HT	4,0 m	14,8 m	500 m	12,7 m	[14,4 m ; 15,2 m]
Largeur car.	2,0 m	11,8 m	54,4 m	4,8 m	[11,6 m ; 12,0 m]
Sup. tablier	19 m ²	872,9 m ²	48 129,0 m ²	1 760,3 m ²	[815,3 m ² ; 930,5 m ²]
Âge	0 an	39,7 ans	112 ans	21,0 ans	[39,0 ans ; 40,4 ans]
DJMA	2	9830	157 000	17 615	[9254 ; 10 406]
% camions	0,0 %	12,1 %	95,0 %	8,7 %	[11,8 % ; 12,4 %]
Nb voies	0,00	2,30	9,00	0,94	[2,27 ; 2,33]

Il permet notamment de voir que la gamme de ponts couverte par les inspections du ministère est très vaste, allant d'ouvrages modestes (longueur de 4,5 m et largeur de 2,0 m, par exemple) à d'autres beaucoup plus importants (longueur de plus de 2,7 km, par exemple). Cette diversité se répercute, en étant même amplifiée, pour la superficie de tablier et le DJMA. Les écarts-types associés sont alors trop grands en comparaison des valeurs moyennes et cela pose problème pour la mise en place des modèles de régression. C'est pourquoi, dans la suite de l'étude, ces deux grandeurs sont centrées-réduites (CR).

4.2.2.2 Corrélations entre variables quantitatives d'entrée

Concernant la présentation de la matrice de corrélation, les coefficients de corrélation supérieurs à 0,9 (qui peuvent donc poser problème pour les futures régressions mises en place) et ceux supérieurs à 0,5 (seuil arbitraire), traduisant une certaine corrélation entre les deux variables concernées, sont encadrés afin d'être facilement identifiables. On considère enfin qu'il n'y a pas de corrélation significative entre deux variables dont le coefficient est inférieur à 0,5.

Tableau 4.9: Matrice de corrélation pour les ponts

Variable	Latitude	Longitude	Longueur totale	Longueur tablier	Largeur hors tout	Largeur carrossable	Superf. tablier CR	Age	DJMA CR	% camion	Nombre de voies
Latitude	1,000000	0,472830	-0,067388	-0,059640	-0,063851	-0,185845	-0,085586	-0,047424	-0,226580	0,224709	-0,154556
Longitude	0,472830	1,000000	-0,047258	-0,046576	0,001194	-0,074037	-0,049766	-0,084690	-0,115176	-0,010806	-0,076100
Longueur totale	-0,067388	-0,047258	1,000000	0,991677	0,003423	0,110493	0,904298	-0,002312	0,230908	-0,081554	0,159375
Longueur tablier	-0,059640	-0,046576	0,991677	1,000000	-0,000947	0,094123	0,908357	0,022691	0,213215	-0,075463	0,149193
Largeur hors tout	-0,063851	0,001194	0,003423	-0,000947	1,000000	0,533945	0,145007	-0,034414	0,252326	-0,045728	0,469696
Largeur carrossable	-0,185845	-0,074037	0,110493	0,094123	0,533945	1,000000	0,290479	-0,136744	0,530108	-0,072416	0,852204
Superf. tablier CR	-0,085586	-0,049766	0,904298	0,908357	0,145007	0,290479	1,000000	-0,001688	0,320319	-0,085036	0,347935
Age	-0,047424	-0,084690	-0,002312	0,022691	-0,034414	-0,136744	-0,001688	1,000000	-0,039899	0,006705	-0,041484
DJMA CR	-0,226580	-0,115176	0,230908	0,213215	0,252326	0,530108	0,320319	-0,039899	1,000000	-0,177219	0,526183
% camion	0,224709	-0,010806	-0,081554	-0,075463	-0,045728	-0,072416	-0,085036	0,006705	-0,177219	1,000000	-0,131276
Nombre de voies	-0,154556	-0,076100	0,159375	0,149193	0,469696	0,852204	0,347935	-0,041484	0,526183	-0,131276	1,000000

1. Forte corrélation entre la longueur totale et la longueur de tablier (0,992) :

Cela paraît logique vu que la longueur totale résulte de la somme des longueurs d'accès à la structure et de la longueur de tablier (qui, en plus, représente la plus grande partie).

2. Forte corrélation entre la longueur de tablier et la superficie de tablier (0,908) :

Encore une fois, cela paraît logique, la longueur rentrant directement en compte dans le calcul de la superficie.

3. Forte corrélation entre la longueur totale et la superficie de tablier (0,904) :

Compte-tenu des deux assertions précédentes, ce résultat n'est pas une surprise.

4. Corrélation notable entre le nombre de voies et la largeur carrossable (0,852) :

Logiquement, plus le nombre de voies est élevé et plus la voie est large. Le signe, positif, de la corrélation tend à appuyer cette idée.

5. Corrélation modérée entre la largeur hors tout et la largeur carrossable (0,534) :

À l'instar du premier point, la largeur carrossable est l'un des termes de l'addition de longueurs effectuée pour obtenir la largeur hors tout de l'ouvrage. Il n'est donc pas surprenant de voir que la largeur carrossable et la largeur hors tout sont corrélées entre elles.

6. Corrélation modérée entre le DJMA et la largeur carrossable (0,530) :

Le DJMA traduit le débit de véhicules sur une chaussée. Or, lorsque la chaussée est conçue, une capacité, c'est-à-dire un débit maximal, supérieur au débit usuel estimé, est définie de manière à assurer un trafic suffisamment fluide la plupart du temps. La capacité de la chaussée est par ailleurs définie comme étant la somme des capacités de chacune des voies de circulation. Cependant, il existe des valeurs seuils que la capacité par voie ne peut pas dépasser : une voie de circulation ne peut pas supporter un débit infini de véhicules. Pour augmenter la capacité globale de la chaussée, la solution est donc d'augmenter le nombre de voies et, par conséquent, construire une chaussée plus large (cf. point 4). C'est pour cela que le DJMA est corrélé positivement, non seulement avec la largeur carrossable, mais également avec le nombre de voies.

7. Corrélation modérée entre le nombre de voies et le DJMA (0,526) :

L'explication est donnée ci-dessus.

4.2.2.3 Élimination du problème des variables trop corrélées

À la lumière de la matrice précédente (Tableau 4.9), d'importantes corrélations ont été mises à jour entre plusieurs variables. Cela concerne le triplet {longueur totale, longueur de tablier, superficie de tablier} et le doublet {DJMA, nombre de voies}. Pour que les régressions soient exactes, il faut donc ne retenir que deux de ces cinq variables. Ainsi, chaque combinaison possible est testée au cours d'une régression linéaire multiple et il apparaît que le couple « superficie de tablier » – « DJMA » fournit les meilleurs résultats. La longueur totale, la longueur de tablier ainsi que le nombre de voies sont donc mis de côté pour la suite de l'étude¹⁰.

¹⁰ La méthode d'analyse sur composantes principales (ACP) fait figure d'exception. En effet, elle permet de prendre en compte toutes les variables, corrélées entre elles ou non, sans que cela ne pose de problème.

4.2.2.4 Corrélations entre les variables d'entrées et la variable de sortie

En dehors de l'étude des valeurs des coefficients de corrélation, celle de leur signe peut également donner des résultats intéressants. En effet, elle permet de déterminer la façon dont va évoluer une variable en fonction des variations de l'autre. Il convient alors d'étudier les signes des corrélations entre les différents variables d'entrée, en tenant compte de l'exclusion précédente, et la sortie, l'ICG (Tableau 4.10).

Tableau 4.10: Signes des coefficients de corrélation

	Lat.	Long.	Larg. HT	Larg. car.	Sup. tab.	Âge	DJMA	% camion
ICG	+	+	+	+	-	-	+	+

Les cases les plus claires indiquent des corrélations pour lesquelles le coefficient est proche de 0 (inférieur à 0,05) tandis que les cases plus sombres mettent en valeur des corrélations plus marquées (coefficient supérieur à 0,05).

1. Corrélation négative entre l'ICG et l'âge de la structure :

Cela signifie que plus la structure est récente et meilleur est son état général, ce qui est logique étant donné que le temps est sans nul doute l'un des facteurs les plus dégradants pour les structures (répétitions des cycles de gel/dégel, phénomènes mécaniques de fluage et de fatigue, etc.).

2. Corrélation positive entre l'ICG et le DJMA :

Au premier abord, cette corrélation peut paraître illogique dans le sens où plus il y a de circulation sur une structure et plus celle-ci se détériore. Cependant, une route très fréquentée (DJMA élevé) est sans doute un axe routier important et fait donc l'objet d'un traitement tout particulier au niveau des contrôles et de l'entretien. C'est pourquoi cette corrélation n'est, au final, pas si improbable que cela.

3. Corrélation positive entre l'ICG et le pourcentage de camions :

Les camions roulant principalement sur les axes majeurs, l'explication du signe de cette corrélation est la même que celle du point 2 : l'état des grands axes routiers est sans doute surveillé de très près.

4. Corrélations positives entre l'ICG et les largeurs hors tout et carrossables :

La partie précédente indiquait que le DJMA, le nombre de voies, la largeur hors tout et la largeur carrossable étaient des variables corrélées entre elles. Il est donc normal, compte-tenu du point 2, que les corrélations entre l'ICG et chacune des deux largeurs soient positives.

Pour les corrélations entre l'ICG et les variables restantes (latitude, longitude et superficie de tablier), l'interprétation est plus difficile. Cela est d'ailleurs confirmé par les coefficients de corrélation, très proches de zéro. Ainsi, rien n'indique qu'en retirant ou en rajoutant quelques ponts, ces signes demeureraient inchangés.

4.3 Méthodes de régression

4.3.1 ICG considéré comme continu

4.3.1.1 Régression linéaire multiple

On obtient $R^2 = 0,154538$ et $R^2_{\text{ajusté}} = 0,152650$. Autrement dit, la part de variance de l'ICG pouvant être uniquement expliquée par la régression est un peu plus faible que 15,5%.

Par ailleurs, le Tableau 4.11 donne les valeurs estimées des coefficients de régression. Grâce à ceux-ci, l'équation de prédiction de l'ICG est facilement déduite.

Remarque :

Étant relativement lourde, elle n'est cependant pas reprise dans cette partie mais en Annexe F (Équation F.1). C'est aussi le cas de l'ensemble des équations de prédiction correspondant aux modèles mis en place par la suite.

En outre, au regard des p-value associées (Tableau 4.11), il apparaît que, sur les huit variables du modèle, seulement trois jouent un rôle significatif sur la valeur de l'ICG. Il s'agit de la largeur carrossable, de l'âge et du DJMA.

Tableau 4.11: Paramètres estimés

Variable	Param. estimé	Erreur type	p-value
Constante	4,563947	0,790593	0,000000
Latitude	- 0,002247	0,011688	0,847555
Longitude	- 0,004988	0,005116	0,329641
Largeur hors tout	0,001515	0,001263	0,230519
Largeur carrossable ***	0,019025	0,003906	0,000000
Superf. tablier CR ¹¹	- 0,022095	0,014400	0,125037
Âge ***	- 0,015369	0,000652	0,000000
DJMA CR ¹¹ ***	- 0,049160	0,016574	0,003036
% camion	0,165253	0,162162	0,308244

À partir de l'équation de prédiction, il est possible de représenter graphiquement les valeurs prédictes en fonction des valeurs observées. On a alors :

$$\text{ICG}_{\text{pré}} = 3,7833 + 0,1545 \cdot \text{ICG}_{\text{obs}} \text{ avec } R^2 = 0,1545$$

Les critères de vérification d'indépendance et liés à l'analyse de résidus sont validés. Il reste à voir ce qu'il en est pour les analyses de sensibilité et de corrélation.

4.3.1.1.1 Analyse de sensibilité

Grâce à la procédure Excel mise en place et aux différents seuils fixés, les observations problématiques sont mises en avant. Elles sont au nombre de 27 au regard du critère de Mahalanobis et 184 suivant celui des résidus studentisés. En enlevant les premières, le R² passe

¹¹ Pour rappel, les mentions « CR » renvoient aux variables qui ont été centrées-réduites car elles présentaient initialement des écarts-types trop importants.

de 0,154538 à 0,152409. En retirant les secondes, il passe de 0,154538 à 0,157332. Autrement dit, le gain de qualité du modèle est inexistant, et même négatif (perte de qualité), dans chacun des deux cas ; la base de données est donc gardée telle quelle.

4.3.1.1.2 Multicolinéarité

Se pose désormais le problème de la multicolinéarité. Pour rappel, la matrice mettait en avant trois corrélations problématiques (coefficient supérieur à 0,9) et une corrélation qui pouvait possiblement l'être aussi (coefficient proche 0,9). Ce problème a été corrigé avec le retrait de trois variables d'entrée mais deux coefficients supérieurs à 0,5 subsistent. Il peut alors être bon de vérifier que cela ne pose pas de problème grâce aux autres critères de détection.

Pour ce qui est des facteurs d'inflation de la variance (VIF), aucune valeur n'est supérieure au seuil de 2 (Tableau 4.12). Il ne semble donc pas y avoir multicolinéarité. Cette conclusion est encore renforcée avec l'étude des indices de conditionnement (IC), que l'on obtient, pour rappel, à partir des valeurs propres (Tableau 4.13).

Tableau 4.12: Analyse des facteurs d'inflation de la variance

Variable	VIF
Latitude	1,431280
Longitude	1,322107
Largeur hors tout	1,409128
Largeur carrossable	1,921457
Superf. tablier CR	1,143021
Âge	1,034499
DJMA CR	1,514096
% camions	1,097538

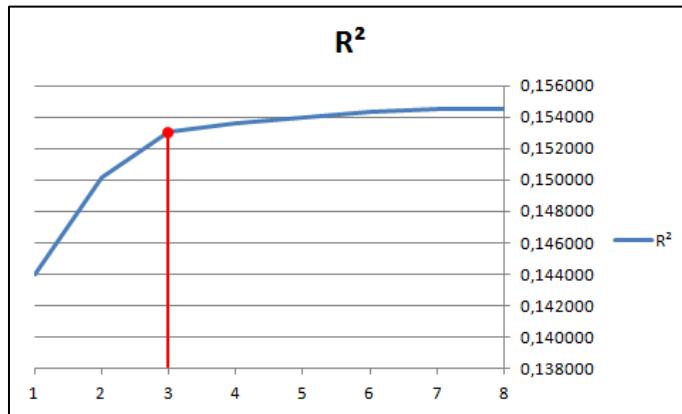
Tableau 4.13: Analyse des indices de conditionnement

Valeur propre		IC
$\lambda_1 = \lambda_{\max}$	2,253981	1,000000
λ_2	1,430874	1,575248
λ_3	1,014922	2,220841
λ_4	0,974485	2,312998
λ_5	0,874177	2,578403
λ_6	0,633366	3,558731
λ_7	0,469106	4,804841
λ_8	0,349089	6,456757

4.3.1.2 Méthodes de sélection des variables

4.3.1.2.1 Sélection exhaustive

Le tracé du R^2 en fonction du nombre de variables (Figure 4.15) permet de voir que trois variables sont réellement significatives.

Figure 4.15: Sélection exhaustive – Critère du R^2

Il s'agit de la largeur carrossable, de l'âge, du DJMA. Cela est, en outre, confirmé par le critère de minimisation du C_p de Mallows.

4.3.1.2.2 Pas-à-pas avant

La méthode de sélection pas-à-pas avant donne les mêmes résultats que ceux obtenus avec la méthode de sélection exhaustive (Tableau 4.14).

Tableau 4.14: Sélection pas-à-pas avant

Age	Step Number	4	1	557,5677	0,000000			In
Largeur carrossable			1	38,2315	0,000000			In
DJMA CR			1	12,3166	0,000454			In
Longitude			1			1,3719	0,241564	Out
Superf. tablier CR			1			2,4619	0,116724	Out
Latitude			1			0,2280	0,633055	Out
Largeur hors tout			1			1,2746	0,258982	Out
% camion			1			1,1570	0,282153	Out

4.3.1.2.3 Pas-à-pas arrière

Encore une fois, les résultats sont les mêmes (Tableau 4.15).

Tableau 4.15: Sélection pas-à-pas arrière

DJMA CR	Step Number	6	1	12,3166	0,000454			In
Largeur carrossable			1	38,2315	0,000000			In
Age			1	557,5677	0,000000			In
Superf. tablier CR			1			2,461931	0,116724	Out
Longitude			1			1,371897	0,241564	Out
Largeur hors tout			1			1,274593	0,258982	Out
% camion			1			1,157037	0,282153	Out
Latitude			1			0,227980	0,633055	Out

4.3.1.2.4 Conclusion

Les trois méthodes donnent des résultats identiques. Il est donc possible de conclure en disant que seulement trois des huit variables continues sont nécessaires à l'établissement d'un modèle prédictif de l'ICG :

- ✓ La largeur carrossable ;
- ✓ L'âge ;
- ✓ Le DJMA ;

On peut alors développer un modèle basé sur cette conclusion et ne prenant en considération que les trois variables listées ci-dessus. On trouve $R^2 = 0,153051$ et $R^2_{\text{ajusté}} = 0,152342$.

En outre, à partir des paramètres estimés (Tableau 4.16), de l'équation de prédiction (Annexe F – Équation F.2) et du tracé des valeurs prédictes en fonction des valeurs observées, il vient :

$$\text{ICG}_{\text{pré}} = 3,7899 + 0,1531 \cdot \text{ICG}_{\text{obs}} \text{ avec } R^2 = 0,1531$$

Tableau 4.16: Sélection de variables – Paramètres estimés

Variable	Param. estimé	Erreur type	p-value
Constante	4,837103	0,051847	0,000000
Largeur carrossable ***	0,020748	0,003356	0,000000
Âge ***	-0,015285	0,000647	0,000000
DJMA CR ***	-0,055801	0,015900	0,000454

4.3.1.3 Régressions particulières

4.3.1.3.1 Analyse sur composantes principales ACP

Pour sélectionner le nombre de facteurs principaux retenus pour le modèle, le critère de vérification de Kaiser est utilisé. Celui-ci précise qu'un facteur est significatif si l'inertie de la valeur propre associée, c'est-à-dire la proportion de variance expliquée, est supérieure à l'inertie moyenne.

Tableau 4.17: Valeurs propres de la matrice de corrélation

Value number	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	3,472122	31,56475	3,47212	31,5647
2	2,218200	20,16545	5,69032	51,7302
3	1,471422	13,37656	7,16174	65,1068
4	1,019708	9,27007	8,18145	74,3768
5	0,958412	8,71284	9,13986	83,0897
6	0,682912	6,20829	9,82278	89,2980
7	0,479353	4,35776	10,30213	93,6557
8	0,467368	4,24880	10,76950	97,9045
9	0,134107	1,21915	10,90360	99,1237
10	0,088770	0,80700	10,99237	99,9307
11	0,007627	0,06934	11,00000	100,0000

Dans le cas présent, l'inertie moyenne valant 9,09%, le Tableau 4.17 permet d'identifier les facteurs à conserver : F1 (31,56%), F2 (20,17%), F3 (13,38%) et F4 (9,27%). Le cinquième

facteur présente cependant une inertie relativement proche de l'inertie moyenne (8,71%). Ainsi, l'intégrer également au modèle peut être une bonne idée.

En régressant sur les composantes principales, il est alors possible d'obtenir les paramètres estimés du modèle (Tableau 4.18).

Tableau 4.18: ACP – Paramètres estimés

Facteur	Param. estimé	Erreur type	p-value
Constante	4,474748	0,013512	0,000000
F1 **	0,015916	0,007253	0,028264
F2 ***	0,057811	0,009074	0,000000
F3 ***	-0,084837	0,011141	0,000000
F4 ***	0,119073	0,013383	0,000000
F5 ***	-0,289956	0,013804	0,000000

Dès lors, il est possible d'obtenir l'équation de prédiction en fonction des variables initiales (Annexe F – Équation F.3) en suivant la démarche explicitée dans la partie 3.2.2.2.1 – Analyse sur composantes principales.

En outre, du tracé des valeurs prédites en fonction des valeurs observées, il vient :

$$\text{ICG}_{\text{prédict}} = 3,8116 + 0,1482 \cdot \text{ICG}_{\text{obs}} \text{ avec } R^2 = 0,1482$$

4.3.1.3.2 Régression ridge

La Figure 4.16 indique que le choix d'un paramètre biaisant λ égal à 0,4 est satisfaisant. En l'appliquant à la régression, il permet d'obtenir un modèle dont le R^2 vaut 0,111455 et le $R^2_{\text{ajusté}}$ 0,109470.

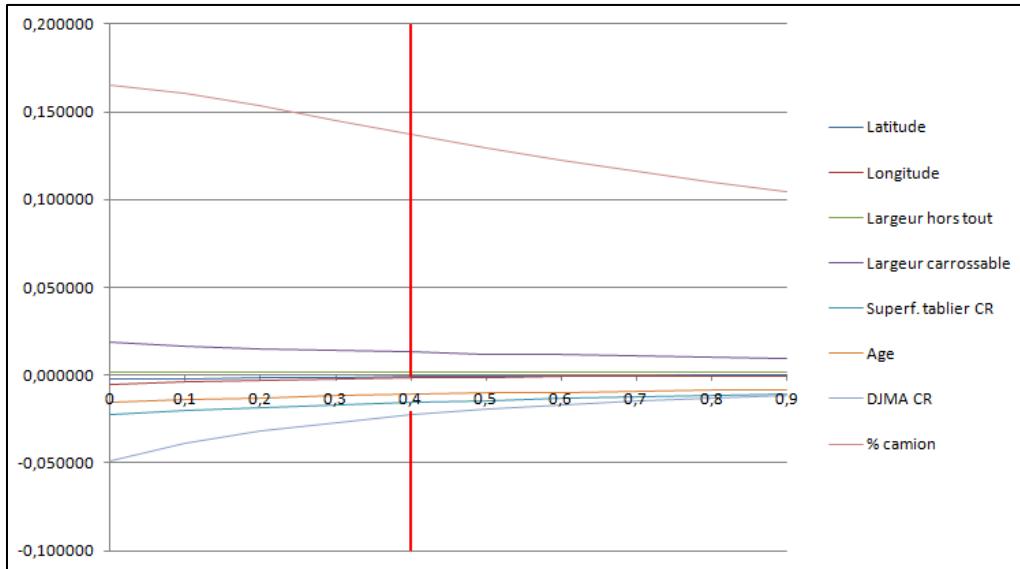


Figure 4.16: Ridge – Choix du paramètre biaisant

Par ailleurs, si l'objectif de cette méthode est de garder l'ensemble des variables, le Tableau 4.19 indique qu'uniquement deux d'entre elles ont un véritable effet sur la valeur de l'ICG en sortie : la largeur carrossable du pont et l'âge de celui-ci.

Tableau 4.19: Ridge – Paramètres estimés

Variable	Param. estimé	Erreur type	p-value
Constante	4,636132	0,600113	0,000000
Latitude	- 0,000835	0,009231	0,927962
Longitude	- 0,001674	0,004114	0,684103
Largeur hors tout *	0,001693	0,000999	0,090308
Largeur carrossable ***	0,013036	0,002852	0,000005
Superf. tablier CR	- 0,015688	0,012110	0,195230
Âge ***	- 0,011039	0,000559	0,000000
DJMA CR *	- 0,022722	0,012948	0,079361
% camions	0,137085	0,137021	0,317151

En outre, à partir de l'équation de prédiction (Annexe F – Équation F.4) et du tracé des valeurs prédictes en fonction des valeurs observées, il vient :

$$\text{ICG}_{\text{pré}} = 3,9760 + 0,1115 \cdot \text{ICG}_{\text{obs}} \text{ avec } R^2 = 0,1540$$

4.3.1.3.3 Régression PLS

D'après la Figure 4.17, choisir trois composantes semble être un bon compromis entre faible nombre de composantes et R^2 satisfaisant ($R^2 = 0,154345$). Une fois ce choix fait, il est possible de lancer la régression.

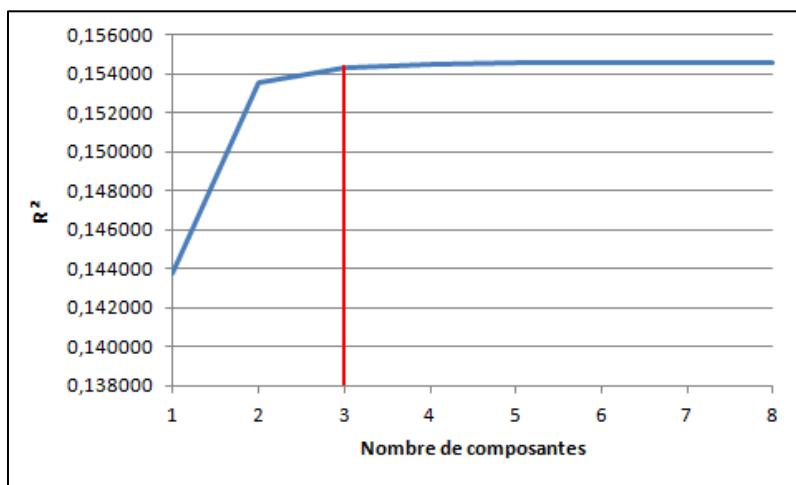


Figure 4.17: PLS – Choix du nombre de composantes

En outre, à partir de l'équation de prédiction (Annexe F – Équation F.5) et du tracé des valeurs prédictes en fonction des valeurs observées, il vient :

$$\text{ICG}_{\text{pré}} = 3,7837 + 0,1543 \cdot \text{ICG}_{\text{obs}} \text{ avec } R^2 = 0,1543$$

Remarque :

En fixant à 0,1 le poids limite de significativité, il apparaît que toutes les variables jouent un rôle significatif dans les variations d'ICG.

4.3.1.4 Synthèse

Le Tableau 4.20 reprend les résultats trouvés pour chacune des méthodes de régression mises en place. Il permet ainsi de les comparer entre elles, de choisir le meilleur modèle de prédiction ou encore de dégager les tendances qui reviennent d'une méthode à l'autre.

Tableau 4.20: Synthèse et comparatif des résultats pour un ICG continu

Méthode	R²	Nombre de variables		ICG prédis vs ICG observés		
		Initial	Signific.	Origine	Coeff. directeur	R²
Linéaire multiple	0,154538	8	3	3,7833	0,1545	0,1545
Sélection de var.	0,152342	3	3	3,7899	0,1531	0,1531
ACP	0,148206	11	11	3,8116	0,1482	0,1482
Ridge	0,111455	8	2	3,9760	0,1115	0,1540
PLS	0,154345	8	8	3,7837	0,1543	0,1543

Tout d'abord, il apparaît, selon l'ensemble des critères reportés dans le tableau, que la meilleure méthode est la régression linéaire multiple. Elle présente notamment le meilleur coefficient de détermination, représentatif de la qualité du modèle. C'est également la méthode qui présente les meilleurs résultats en ce qui concerne le tracé des valeurs prédictes d'ICG en fonction des valeurs observées. En effet, avec l'ordonnée à l'origine la plus proche de 0, le coefficient directeur le plus proche de 1 et la meilleure représentativité des données (le R² le plus élevé), la droite de régression associée est celle qui se rapproche le plus de la première bissectrice.

On peut voir cependant que les résultats fournis par les autres méthodes sont très proches. Les différentes méthodes semblent ainsi concorder et confirment alors la robustesse du modèle.

Enfin, si le nombre de variables significatives change selon le modèle développé, un "noyau" semble tout de même se dégager. Il est composé des trois variables mises en valeur par les méthodes de sélection, à savoir la largeur carrossable, l'âge et le DJMA. Seule la régression ridge fait figure d'exception en ne définissant comme significatives que deux de ces trois variables (la

largeur carrossable et l'âge). Cette exception peut être interprétée comme une précision apportée à la conclusion tirée : les variables ayant le plus d'effet sur l'ICG sont la largeur carrossable et l'âge et, à un degré moindre, le DJMA.

4.3.2 ICG considéré comme discret

4.3.2.1 Régression logistique binomiale

4.3.2.1.1 Détection d'un défaut de condition

La probabilité recherchée est, dans cette partie, celle d'apparition d'un défaut, c'est-à-dire la probabilité que la nouvelle variable binomiale créée prenne la valeur 0.

4.3.2.1.1.1 Qualité du modèle

Les valeurs obtenues pour les différents critères de qualité sont consignées dans le Tableau 4.21. Il n'est cependant pas nécessaire d'en discuter en tant que telles, l'intérêt principal étant de les comparer aux résultats obtenus avec les autres méthodes.

Tableau 4.21: Défaut – Critères de qualité

Critère	Valeur
Logarithme de vraisemblance	- 2000,45
Déviance	4000,90
AIC	4018,90
Pseudo R ²	R ² Cox-Snell
	R ² Nagelkerke
	R ² McFadden
Matrice de contingence	Taux d'erreur
	Sensibilité
	Spécificité
Aire sous la courbe ROC	0,7288

Remarque :

Pour les régressions logistiques binomiales, le nombre de variables d'entrée est limité par le nombre d'observations se rapportant à chacune des valeurs de la sortie. En général, une variable explicative est acceptée toutes les dix observations. Sachant que 2397 ponts sont en bon état et prennent donc la valeur 1 et que la sortie vaut 0 pour les 1194 ouvrages restants, il ne faudrait pas que le modèle développé contienne plus de 119 entrées, ce qui est bien le cas.

4.3.2.1.1.2 Résultats

Le Tableau 4.22 donne les valeurs des ratios de cotes pour chacune des variables d'entrée.

Tableau 4.22: Ratios des cotes

Variable	Ratio	Effet sur l'ICG ?	En accord avec corrélations ?
Latitude	1,118977	Négatif	Non
Longitude	0,991737	Positif	Oui
Largeur HT	0,991891	Positif	Oui
Largeur car.	0,963895	Positif	Oui
Sup. tablier CR	1,218801	Négatif	Oui
Âge	1,045447	Négatif	Oui
DJMA CR	1,156593	Négatif	Non
% camions	0,381776	Positif	Oui

Les effets étant désormais décrits, il peut être intéressant de regarder s'ils sont véritablement notables ou s'ils ne sont qu'anecdotiques. Dans cette optique, les résultats du test de Wald (Tableau 4.23) indiquent que six variables explicatives sur huit semblent avoir une influence sur l'ICG. Ces six variables sont la latitude, la largeur carrossable, la superficie de tablier, l'âge, le DJMA et le pourcentage de camions. Ces résultats sont, en outre, confirmés par le test du rapport de vraisemblance.

Tableau 4.23: Significativité des variables explicatives

Variable	Param. estimé	Erreur type	Test de Wald	
			Stat. Chi-2	p-value
Latitude ***	0,11241	0,03213	12,2360	0,000469
Longitude	- 0,00830	0,01403	0,3499	0,554149
Largeur hors tout	- 0,00814	0,00614	1,7614	0,184448
Largeur carrossable ***	- 0,03677	0,01377	7,1372	0,007550
Superf. tablier CR ***	0,19787	0,05127	14,9015	0,000113
Âge ***	0,04444	0,00232	366,0886	0,000000
DJMA CR ***	0,14548	0,04716	9,5180	0,002035
% camions **	- 0,96292	0,48492	3,9432	0,047061

Comme pour les méthodes de régression précédentes, une équation de prédiction, non linéaire cependant, peut être mise en place à partir des paramètres estimés du modèle (Annexe F – Équation F.6).

De celle-ci, et du tracé des valeurs prédites en fonction des valeurs observées, il vient :

$$\text{Prédiction} = 0,4278 - 0,1427 \cdot \text{Observation}, \text{avec } R^2 = 0,1391$$

4.3.2.1.2 Détection d'un bon état

Contrairement au cas précédent, on recherche ici la probabilité que la structure soit en bon état. Le problème étant simplement "retourné", les résultats obtenus sont exactement les mêmes mais les coefficients estimés sont de signe opposé. Cela a une répercussion sur l'équation de prédiction (Annexe F – Équation F.7) et sur le tracé des valeurs prédites en fonction des valeurs observées :

$$\text{Prédiction} = 0,5722 + 0,1427 \cdot \text{Observation}, \text{avec } R^2 = 0,1391$$

Cette différence est clairement visible sur la Figure 4.18. Ainsi, le tracé est, cette fois-ci, croissant (à droite) tandis qu'il était décroissant dans le cas précédent (à gauche). Néanmoins, la logique qui s'en dégage est la même puisque dire que la probabilité de prédire un défaut diminue avec les valeurs observées (tracé de gauche) équivaut à dire que la probabilité de prédire le bon état augmente avec ces dernières (tracé de droite).

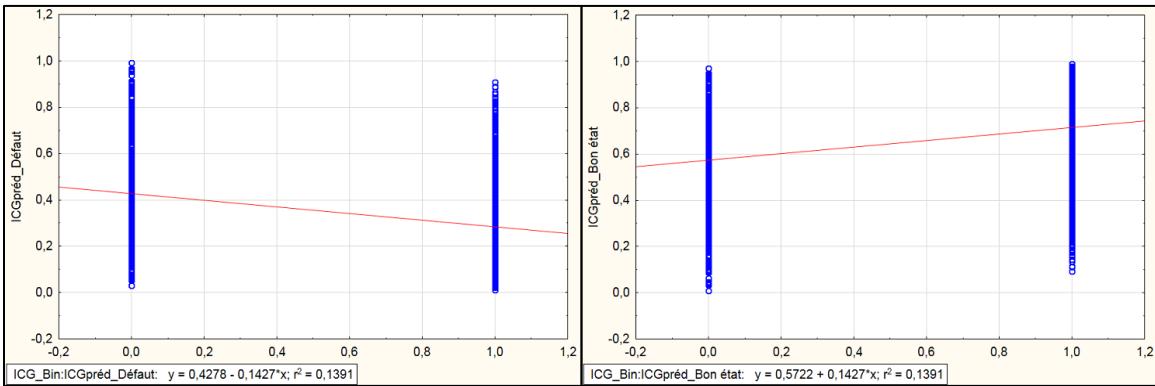


Figure 4.18: Tracés des prédictions en fonction des valeurs observées

4.3.2.2 Généralisation aux autres états

4.3.2.2.1 Détection d'un ICG égal à 4

Au niveau des valeurs observées, 742 points présentent un ICG de 4 et ont donc une valeur de 1 pour Y_4. Cette nouvelle variable dépendante prend donc la valeur 0 pour les 2849 points restants. Les "limites du 0 et du 1" sont donc respectivement de 284 et 74 variables. Une nouvelle fois, les huit variables explicatives prises en compte dans le modèle ne posent aucun problème.

Les résultats de la régression sont alors les suivants :

Tableau 4.24: Niveau 4 de l'ICG – Critères de qualité

Critère	Valeur	
Logarithme de vraisemblance	- 1749,96	
Déviance	3499,93	
AIC	3517,93	
Pseudo R ²	R ² Cox-Snell	0,043304

Tableau 4.24: Niveau 4 de l'ICG – Critères de qualité (suite et fin)

Pseudo R ²	R ² Nagelkerke	0,067767
	R ² McFadden	0,043448
Matrice de contingence	Taux d'erreur	20,94 %
	Sensibilité	0,94 %
	Spécificité	99,40 %
Aire sous la courbe ROC		0,6452

En outre, à partir de l'équation de prédiction (Annexe F – Équation F.8) et du tracé des valeurs prédites en fonction des valeurs observées, il vient :

$$\text{Prédiction} = 0,1980 + 0,0415 \cdot \text{Observation}, \text{avec } R^2 = 0,0381$$

Enfin, le test de Wald (Tableau 4.25), appuyé du test du rapport de vraisemblance, indique que cinq des huit variables continues d'entrée sont significatives : la latitude, la superficie de tablier, l'âge, le pourcentage de camions et, à un degré moindre, la longitude.

Tableau 4.25: Niveau 4 – Significativité des variables explicatives

Variable	Param. estimé	Erreur type	p-value
Latitude ***	0,2005	0,0345	0,000000
Longitude **	- 0,0323	0,0149	0,029791
Largeur hors tout	- 0,0323	0,0053	0,363507
Largeur carrossable	0,0069	0,0133	0,601142
Superf. tablier CR ***	0,1787	0,0468	0,000132
Âge ***	0,0218	0,0022	0,000000
DJMA CR	0,0571	0,0513	0,265984
% camions ***	- 1,4946	0,5491	0,006487

4.3.2.2.2 Détection d'un ICG égal à 3

Au niveau des valeurs observées, 212 ponts présentent un ICG de 3 et ont donc une valeur de 1 pour Y_3. Cette nouvelle variable dépendante prend donc la valeur 0 pour les 3379 ponts restants. Les "limites du 0 et du 1" sont donc respectivement de 337 et 21 variables. Une nouvelle fois, les huit variables explicatives prises en compte dans le modèle ne posent aucun problème.

Les résultats de la régression sont alors les suivants :

Tableau 4.26: Niveau 3 de l'ICG – Critères de qualité

Critère	Valeur
Logarithme de vraisemblance	- 734,67
Déviance	1469,35
AIC	1487,35
Pseudo R ²	R ² Cox-Snell
	R ² Nagelkerke
	R ² McFadden
Matrice de contingence	Taux d'erreur
	Sensibilité
	Spécificité
Aire sous la courbe ROC	0,7445

En outre, à partir de l'équation de prédiction (Annexe F – Équation F.9) et du tracé des valeurs prédites en fonction des valeurs observées, il vient :

$$\text{Prédiction} = 0,0566 + 0,0413 \cdot \text{Observation}, \text{avec } R^2 = 0,0342$$

Enfin, le test de Wald (Tableau 4.27), appuyé du test du rapport de vraisemblance, indique que quatre des huit variables continues d'entrée sont significatives : l'âge et, à des degrés moindres, la latitude, la largeur carrossable et le pourcentage de camions.

Tableau 4.27: Niveau 3 – Significativité des variables explicatives

Variable	Param. estimé	Erreur type	p-value
Latitude **	- 0,1352	0,0671	0,044018
Longitude	- 0,0024	0,0293	0,935526
Largeur hors tout	- 0,0083	0,0167	0,619599
Largeur carrossable **	- 0,0818	0,0328	0,012774
Superf. tablier CR	0,0781	0,0543	0,150193
Âge ***	0,0362	0,0040	0,000000
DJMA CR *	0,1703	0,0957	0,075223
% camions **	1,8312	0,8235	0,026161

4.3.2.2.3 Détection d'un ICG égal à 2

Au niveau des valeurs observées, 240 ponts présentent un ICG de 2 et ont donc une valeur de 1 pour Y_2. Cette nouvelle variable dépendante prend donc la valeur 0 pour les 3351 ponts restants. Les "limites du 0 et du 1" sont donc respectivement de 335 et 24 variables. Une nouvelle fois, les huit variables explicatives prises en compte dans le modèle ne posent aucun problème.

Les résultats de la régression sont alors les suivants :

Tableau 4.28: Niveau 2 de l'ICG – Critères de qualité

Critère	Valeur	
Logarithme de vraisemblance	- 741,34	
Déviance	1482,68	
AIC	1500,68	
Pseudo R ²	R ² Cox-Snell	0,074901
	R ² Nagelkerke	0,193130
	R ² McFadden	0,158647

Tableau 4.28: Niveau 2 de l'ICG – Critères de qualité (suite et fin)

Matrice de contingence	Taux d'erreur	6,96 %
	Sensibilité	2,08 %
	Spécificité	99,55 %
Aire sous la courbe ROC		0,8118

En outre, à partir de l'équation de prédiction (Annexe F – Équation F.10) et du tracé des valeurs prédites en fonction des valeurs observées, il vient :

$$\text{Prédiction} = 0,0607 + 0,0913 \cdot \text{Observation}, \text{avec } R^2 = 0,0776$$

Enfin, le test de Wald (Tableau 4.29), appuyé du test du rapport de vraisemblance, indique que quatre des huit variables continues d'entrée sont significatives : l'âge et, à des degrés moindres, la largeur carrossable, la superficie de tablier et le DJMA.

Tableau 4.29: Niveau 2 – Significativité des variables explicatives

Variable	Param. estimé	Erreur type	p-value
Latitude	- 0,0005	0,0650	0,993287
Longitude *	0,0479	0,0277	0,083530
Largeur hors tout	- 0,0234	0,0192	0,222326
Largeur carrossable **	- 0,0719	0,0347	0,038208
Superf. tablier CR ***	- 0,3077	0,1384	0,026156
Âge ***	0,0557	0,0044	0,000000
DJMA CR ***	0,3454	0,0868	0,000069
% camions	- 0,5283	0,9625	0,583090

4.3.2.3 Régression logistique multinomiale

4.3.2.3.1 ICG de référence pris égal à 5

Tableau 4.30: Référence égale à 5 – Critères de qualité

Critère	Valeur
Logarithme de vraisemblance	- 3018,18
Déviance	6036,36
AIC	6090,36
R ² McFadden	0,109187
Taux d'erreur	33,28 %

Le logiciel retourne une série de coefficients estimés pour chaque modalité, autre que celle de référence, prise par l'ICG. À partir de celles-ci (Annexe F – Équation F.11, Équation F.12 et Équation F.13) et des tracés des valeurs prédites en fonction des valeurs observées, il vient :

$$\text{Prédiction}_2 = 0,0607 + 0,0923 \cdot \text{Observation}, \text{ avec } R^2 = 0,0856$$

$$\text{Prédiction}_3 = 0,0565 + 0,0429 \cdot \text{Observation}, \text{ avec } R^2 = 0,0420$$

$$\text{Prédiction}_4 = 0,1967 + 0,0482 \cdot \text{Observation}, \text{ avec } R^2 = 0,0476$$

Le test de Wald (Tableau 4.31) indique que six variables d'entrée sont significatives : la latitude, la largeur carrossable, la superficie du tablier, l'âge, le DJMA et le pourcentage de camions.

Tableau 4.31: Multinom. – Significativité des variables explicatives

Variable	Wald	p-value
Latitude ***	31,1222	0,000001
Longitude	5,9762	0,112773
Largeur hors tout	3,1347	0,371323
Largeur carrossable ***	14,2939	0,002531
Superf. tablier CR ***	22,8943	0,000042

Tableau 4.31: Multinom. – Significativité des variables explicatives (suite et fin)

Âge ***	394,0040	0,000000
DJMA CR ***	23,2022	0,000037
% camions ***	12,407	0,006118

4.3.2.3.2 ICG de référence pris égal à 2

Tableau 4.32: Référence égale à 2 – Critères de qualité

Critère	Valeur
Logarithme de vraisemblance	- 3018,18
Déviance	6036,36
AIC	6090,36
R ² McFadden	0,109187
Taux d'erreur	33,28 %

Par ailleurs, à partir des équations de prédiction (Annexe F – Équation F.14, Équation F.15 et Équation F.16) et des tracés des valeurs prédites en fonction des valeurs observées, il vient :

$$\text{Prédiction}_3 = 0,0565 + 0,0429 \cdot \text{Observation}, \text{ avec } R^2 = 0,0420$$

$$\text{Prédiction}_4 = 0,1967 + 0,0482 \cdot \text{Observation}, \text{ avec } R^2 = 0,0476$$

$$\text{Prédiction}_5 = 0,5720 + 0,1430 \cdot \text{Observation}, \text{ avec } R^2 = 0,1366$$

Les résultats quant aux critères de qualité des modèles sont exactement les mêmes quelque soit l'approche multinomiale considérée. Ainsi, seul le sens que l'on souhaite donner aux résultats détermine laquelle des deux il est préférable de choisir.

4.3.2.4 Synthèse

Le Tableau 4.33 reprend les résultats pour les différentes régressions logistiques et permet ainsi de les comparer entre elles.

Tableau 4.33: Synthèse et comparatif des régressions logistiques

Critère		Binomiale		Généralisation du raisonnement binomial				Multinomiale			
		Défaut	Bon état	ICG_5	ICG_4	ICG_3	ICG_2	ICG_5	ICG_4	ICG_3	ICG_2
Log. de vrais.		- 2000,45	- 2000,45	- 2000,5	- 1749,96	- 734,67	- 741,34	- 3018,18			
Déviance		4000,90	4000,90	4000,90	3499,93	1469,35	1482,68	6036,36			
AIC		4018,90	4018,90	4018,90	3517,93	1487,35	1500,68	6090,36			
R ² McFadden		0,124007	0,124007	0,124007	0,043304	0,087922	0,158647	0,109187			
Taux d'erreur		31,05 %	31,05 %	31,05 %	20,94 %	5,96 %	6,96 %	33,28 %			
Nb var. sign.		6 / 8	6 / 8	6 / 8	5 / 8	4 / 8	4 / 8	6 / 8			
ICG _{prédict} vs ICG _{obs}	Ord.	0,4278	0,5722	0,5722	0,1980	0,0566	0,0607	0,5720	0,1967	0,0565	0,0607
	Coeff.	- 0,1427	0,1427	0,1427	0,0415	0,0413	0,0913	0,1430	0,0482	0,0429	0,0923
	R ²	0,1391	0,1391	0,1391	0,0381	0,0342	0,0776	0,1366	0,0476	0,0420	0,0856

Ce tableau permet de voir que la régression multinomiale donne globalement des résultats qui traduisent mieux la réalité des données : les ordonnées à l'origine sont un peu plus proches de 0 et les coefficients directeurs sont un peu plus proches de 1. De plus, il peut sembler plus pratique de mettre en place un modèle global plutôt que plusieurs modèles "individuels".

Cependant, les modèles binomiaux présentent l'avantage d'être moins complexes que le modèle multinomial, comme le laissent penser les valeurs de l'AIC. Les modèles binomiaux sont donc plus robustes.

En fait, il n'y a pas de conclusion générale à tirer du Tableau 4.33. Les modèles sont assez proches les uns des autres et l'utilisation de l'un plutôt que l'autre va finalement dépendre de ce que l'on cherche à faire. Si on veut mettre en évidence un défaut ou un niveau particulier d'état de la structure, le modèle binomial est bien plus adapté que le multinomial dans le sens où il a été spécialement conçu pour remplir cette tâche. Il est, par exemple, envisageable de déterminer des seuils sur les variables d'entrée au-dessus desquels ou en-dessous desquels l'ICG a des chances de se dégrader. En revanche, si on souhaite plutôt étudier l'évolution de l'état, le modèle multinomial est plus intéressant car il prend en compte les différents niveaux de l'ICG.

Par ailleurs, sur les sept régressions logistiques effectuées, l'âge du pont est la seule variable qui est toujours considérée comme significative. Plusieurs autres variables reviennent et semblent jouer des rôles plus ou moins importants selon le modèle retenu : la latitude, la largeur carrossable, la superficie de tablier et le pourcentage de camions.

4.4 Modèle général de régression

4.4.1 Résultats obtenus

Les valeurs de R^2 et $R^2_{\text{ajusté}}$, caractéristiques de la qualité du modèle mis en place, sont respectivement de 0,195319 et 0,184069.

L'équation de prédiction se déduit des coefficients estimés et ne comprend pas seulement quinze variables (huit numériques + sept catégoriques) mais quarante-neuf compte-tenu des nombreuses variables créées pour le codage des facteurs catégoriques (Annexe F – Équation F.17).

Remarque :

Pour les variables catégoriques de l'étude, les modalités de référence choisies par le logiciel pour le codage sont :

- ✓ « En arc » pour le type de ponts ;
- ✓ « Acier enrobé de béton » pour le matériau ;
- ✓ « Local 1 » pour la classe de route ;
- ✓ « Bas Saint-Laurent » pour la région administrative ;
- ✓ « Terrain » pour le type d'obstacles ;
- ✓ « 12 » pour le climat ;
- ✓ « PQ » pour le parti politique.

En outre, à partir de l'équation de prédiction et du tracé des valeurs prédictes en fonction des valeurs observées, il vient :

$$\text{ICG}_{\text{prédict}} = 3,6022 + 0,1953 \cdot \text{ICG}_{\text{obs}} \text{ avec } R^2 = 0,1953$$

Le Tableau 4.34 permet alors, en comparant ces résultats à ceux obtenus dans le cadre de la régression linéaire multiple, de quantifier l'augmentation induite par la prise en compte des facteurs explicatifs quant à la qualité de la prédiction.

Tableau 4.34: Améliorations apportées par le modèle général

Méthode	Variables considérées	R²	Nombre de variables		ICG prédicts vs ICG observés		
			Initial	Signific.	Origine	Coeff. directeur	R²
Linéaire multiple	Quantitatives	0,154538	8	3	3,7833	0,1545	0,1545
Général	Toutes	0,195319	15	9	3,6022	0,1953	0,1953

Le fait que les résultats soient meilleurs lorsque l'ensemble des variables explicatives est pris en compte dans le modèle n'est pas surprenant car la prise en compte de variables supplémentaires augmente évidemment la quantité d'information disponible. Le véritable intérêt du Tableau 4.34 est alors de montrer que, sans considérations techniques autres que le type de ponts et le matériau

utilisé, il est statistiquement possible d'expliquer près de 20% de la variance de l'ICG des ponts du réseau routier québécois.

Remarque :

Devant le nombre important de variables intervenant dans le modèle, sa robustesse peut être remise en cause. Il peut alors être intéressant de mettre en place des procédures de sélection de variables pour avoir un modèle équivalent en qualité mais probablement plus robuste.

4.4.2 Étude des variables significatives

Le traitement commun des variables quantitatives et qualitatives peut être intéressant pour confirmer ou infirmer les résultats obtenus quant à l'influence de certaines entrées sur l'ICG. Pour rappel, parmi les variables numériques, ont été principalement définis comme significatifs l'âge de la structure, la largeur carrossable et le DJMA¹². Pour ce qui est des facteurs catégoriques, les tests de comparaison des moyennes ont identifié comme statistiquement significatives les différences entre les modalités du type de ponts, du matériau utilisé, de certaines régions administratives et du parti politique.

Pour le modèle général, les p-value associées aux tests de significativité des différents coefficients de prédiction laissent penser que sur les quarante-neuf variables incluses dans le modèle, quatorze d'entre elles, seulement, ont un effet significatif sur l'ICG : la largeur carrossable ***, la superficie de tablier **, l'âge ***, le DJMA ***, les quatre variables créées pour les modalités du matériau ***, trois des variables créées pour les modalités de région administrative (« Laurentides » **, « Centre-du-Québec » *** et « Outaouais » ***) et, à des degrés moindres, la variable créée pour la modalité « Cours d'eau » du type d'obstacles **, celle créée pour la modalité « Ressources et aux localités isolées » du type de routes ** et celle créée pour la modalité « UN » du parti politique **.

¹² Les régressions binomiales semblent également faire état d'un impact plus modéré de la latitude, de la superficie de tablier et du pourcentage de camions sur l'ICG.

Autrement dit, parmi les quinze variables initialement prises en compte, six influent certainement sur les valeurs prises par la sortie :

- ✓ La largeur carrossable ;
- ✓ La superficie du tablier ;
- ✓ L'âge ;
- ✓ Le DJMA ;
- ✓ Le matériau ;
- ✓ La région administrative.

À cela, il faut rajouter trois autres variables qualitatives, pour lesquelles une des modalités semblent se démarquer :

- ✓ Le type de routes (« Accès ressources et aux localités isolées ») ;
- ✓ Le type d'obstacles (« Cours d'eau ») ;
- ✓ Le parti politique (« UN »).

Contrairement aux premières conclusions émises à partir de l'analyse des facteurs catégoriques, le type de ponts ne semble finalement pas avoir d'effet significatif sur la sortie. À un degré moindre, l'effet du parti politique, très marqué lors du test de Kruskal-Wallis effectué, ne semble finalement pas si conséquent que cela. Ces divergences sont sans doute dues au fait que ces facteurs étaient perçus comme significatifs car ils "captaient" la dimension temporelle du problème. Pour étoffer cette hypothèse, démontrer l'existence d'interaction entre ceux-ci et l'ancienneté des ponts peut être une bonne idée.

4.5 Interactions entre l'âge et les facteurs catégoriques significatifs

Il est possible de mettre en évidence des interactions entre deux facteurs explicatifs grâce au tracé des valeurs moyennes de l'ICG selon les modalités d'un premier facteur A et d'un second facteur B. Concrètement, pour chaque modalité de A, par exemple, on trace une courbe d'évolution de l'ICG moyen en fonction des modalités de B. Si ces courbes sont "parallèles", alors il n'y a pas d'interaction. À contrario, si elles se croisent, il y a interaction entre A et B et on peut possiblement parler d'un effet combiné de A et de B sur l'ICG.

Il faut donc, pour obtenir les résultats escomptés, commencer par rendre catégorique la variable explicative « âge ».

4.5.1 Transformation de l'âge en facteur catégorique

La distribution des valeurs d'âge peut être un bon indice sur la manière dont il faut procéder (Figure 4.19).

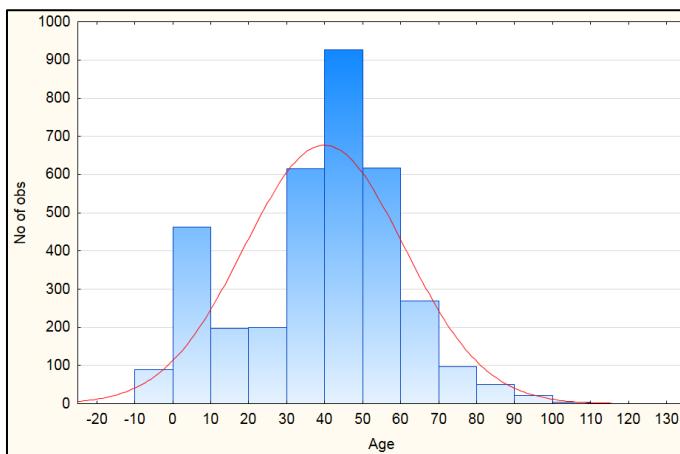


Figure 4.19: Distribution de l'âge des ponts

Trois intervalles sont assez facilement identifiables :

- ✓ Un intervalle d'âges "classiques" ou "normaux" qui contient les ponts âgés de 31 à 60 ans inclus. Il s'agit de l'intervalle central dans lequel se trouvent le plus de ponts. Il correspond à l'essor du secteur de la construction au Québec, des années 60 aux années 80 ;
- ✓ Au-delà, les ponts peuvent être qualifiés d'anciens. Leur âge est alors supérieur ou égal à 61 ans ;
- ✓ En deçà de l'intervalle central, c'est-à-dire pour des valeurs comprises entre 0 et 30, les ponts peuvent être qualifiés de récents.

Ce sont donc trois modalités d'âge qui sont mises en place : « Récent », « Normal » et « Ancien ». La Figure 4.20 permet de constater que ce découpage est adéquat dans le sens où le nombre de ponts est suffisamment élevé pour chaque modalité.

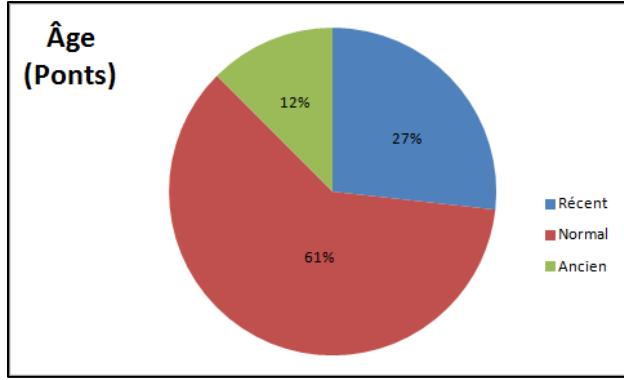


Figure 4.20: Répartition des ponts selon leur âge

Le test de Kruskal-Wallis confirme, comme prévu, que l'âge de la structure a bien un effet significatif sur son état (Tableau 4.35).

Tableau 4.35: Test comparant les valeurs moyennes d'ICG selon l'âge

p-value	Récent	Normal	Ancien
Récent	—	0,000000	0,000000
Normal	0,000000	—	0,000000
Ancien	0,000000	0,000000	—

4.5.2 Étude des interactions

4.5.2.1 Interaction entre l'âge et le type de ponts

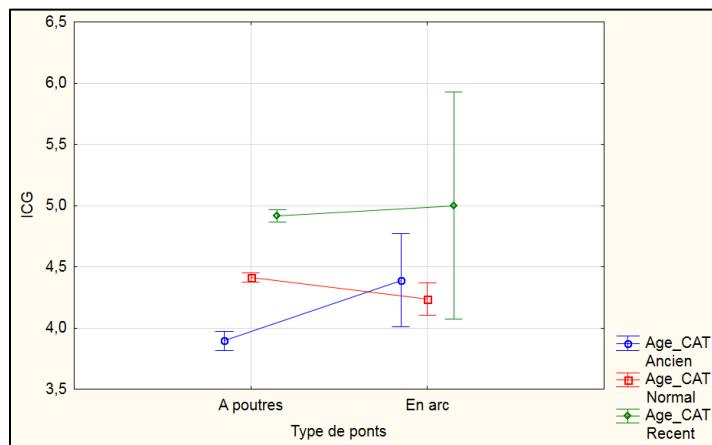


Figure 4.21: ICG moyen selon l'âge et le type de ponts

Le grand intervalle de confiance présent pour la double modalité « Récent - En arc » laisse penser qu'il n'y a que peu de ponts récents en arc. La réalisation de ponts en arc serait alors une méthode utilisée autrefois pour la construction et qui serait, aujourd'hui, tombée quelque peu en désuétude. Cela peut se vérifier sur le Tableau 4.36 qui donne la valeur moyenne de l'âge des ponts selon le type considéré.

Tableau 4.36: Valeurs moyennes d'âge selon le type de ponts

Modalité	Moyenne	Ecart Type	Intervalle 95%	
			Borne inf.	Borne sup.
En arc	48,18	0,93	46,35	50,01
À poutres	39,15	0,36	38,43	39,86

De plus, d'après la Figure 4.21, les ponts en arc anciens semblent être en meilleur état que ceux d'âge "normal". Il est donc fort probable que les plus vieux d'entre eux aient subi des opérations d'entretien visant à améliorer leur état. Si tel est le cas, alors la Figure 4.21 met en avant l'efficacité de cet entretien.

4.5.2.2 Interaction entre l'âge et le parti politique

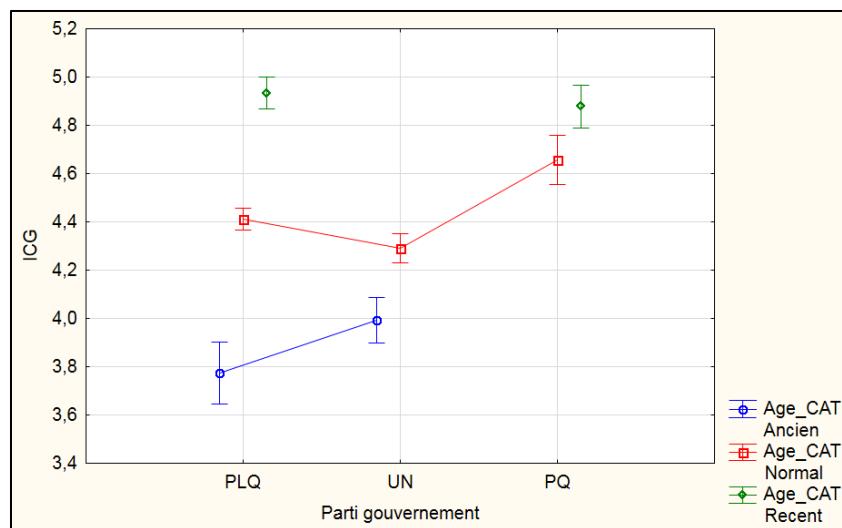


Figure 4.22: ICG moyen en fonction de l'âge et du parti

La liaison entre les variables « parti » et « âge » a déjà été établie grâce au tableau recensant les années de construction moyenne des structures selon les différents partis au pouvoir (Tableau 4.7). Si cela ne suffisait pas à s'en convaincre, la Figure 4.22 le confirme. On y voit, en effet, qu'aucun pont récent n'est recensé pour la modalité « UN », confirmant l'ancienneté globale des structures affiliées à ce parti. En ce qui concerne le PQ, les ouvrages sont, dans l'ensemble, plus récents ; aucun pont ancien n'est recensé.

4.5.2.3 Conclusion

Les différents éléments qui viennent d'être vus tendent à confirmer l'hypothèse selon laquelle le type de ponts et le parti politique, étant liés à l'âge de la structure, en captaient les effets sur l'ICG. Ces variables ne seraient donc pas significatives en tant que telles. Cependant, avant de pouvoir conclure définitivement, d'autres tests seraient nécessaires.

Il est à noter que le fait que le parti politique n'ait pas d'impact sur la tenue aux dégradations des structures corrobore les résultats de Joanis dans une étude portant sur l'influence de la dimension politique dans l'attribution des ressources financières dédiées aux infrastructures publiques (2011). Son objectif est de voir s'il existe une « prime de loyauté » pour les circonscriptions électorales provinciales (CEP) historiquement fidèles au parti politique gouvernant le Québec. S'il semble que ce soit le cas pour les travaux de construction, Joanis conclut, en revanche, que rien de semblable ne peut être mis en avant en ce qui concerne les travaux d'entretien.

4.6 Surévaluation des prédictions

Si les modèles mis en place sont robustes, compte-tenu de la répétabilité des résultats obtenus, et la part de variance expliquée par les régressions plutôt bonne, les valeurs prédites pour l'ICG ont tendance à être surévaluées. Cela est dû au fait qu'une grande partie des ponts, sur les 3591 étudiés, sont en bon état. L'ICG est effectivement égal à 5 pour près de 67% d'entre eux (Tableau 4.37) et les résultats sont donc "tirés vers le haut". C'est d'ailleurs ce qui explique les valeurs trop élevées des ordonnées à l'origine dans les tracés des prédictions en fonction des observations.

Tableau 4.37: Répartition des structures selon les valeurs de l'ICG

Valeur ICG	Ponts		Ponceaux		Murs		Total	
5	2397	66,8%	666	86,9%	500	90,7%	3563	72,6%
4	742	20,7%	44	5,7%	30	5,4%	816	16,7%
3	212	5,9%	20	2,6%	9	1,6%	241	4,9%
2	240	6,7%	36	4,7%	12	2,2%	288	5,7%
TOTAL	3591	-	766	-	551	-	4908	-

Remarque :

Ce tableau justifie également le choix fait de se concentrer exclusivement sur les ponts. En effet, pour les autres structures, que sont les ponceaux et les murs, trop de valeurs d'ICG ne sont pas suffisamment représentées pour que les résultats obtenus soient statistiquement significatifs (nombre d'observations correspondantes inférieur à trente).

4.6.1 Rectification du problème

4.6.1.1 Cas du modèle général

Pour tenir compte du problème de surévaluation mentionné, on décide de mettre en place une procédure de correction des prédictions. Elle consiste à définir les intervalles de confiance des valeurs prédites pour chacune des modalités observées de l'ICG (Tableau 4.38).

Tableau 4.38: Intervalles de confiance pour le modèle général de régression

ICG _{obs} à ...	Moyenne ICG _{prédict}	Écart-type ICG _{prédict}	Intervalle de confiance à 97,5%
5	4,584554	0,369248	[4,567585 ; 4,601522]
4	4,359593	0,300898	[4,334614 ; 4,384572]
3	4,156600	0,287426	[4,112037; 4,201162]
2	4,035605	0,298019	[3,992307 ; 4,078904]

Il s'agit d'intervalles définis de manière statistique, à partir des moyennes et des écarts-types, et dans lesquels 97,5%, dans le cas présent, des valeurs se trouvent. Plutôt que de garder les prédictions telles quelles, on va donc chercher à les situer par rapport à ces intervalles et leur attribuer les valeurs correspondantes. Si, pour une structure particulière, le modèle prédictif renvoie, par exemple, une valeur de 4,35 pour l'ICG, alors on peut, au regard du Tableau 4.38, lui donner la valeur de 4.

Si la valeur prédite se situe entre deux intervalles, on la conserve ; si elle est supérieure à 4,601522, on lui attribue logiquement la valeur 5 ; enfin, si elle est inférieure à 3,992307, c'est de façon tout aussi logique que la valeur 2 lui est conférée.

Remarque :

Le Tableau 4.38 rend le problème de surévaluation parfaitement visible. En effet, on constate, dans le pire des cas, que 97,5% des prédictions correspondant à une valeur réelle d'ICG de 2 sont comprises entre 3,992307 et 4,078904 (plus proche de 4 que de 2 donc).

4.6.1.2 Cas de la détection de défauts

Le problème se retrouve bien entendu pour les régressions logistiques. La solution retenue est la même et le Tableau 4.39 présente les intervalles de confiance définis pour le modèle de détection de défauts.

Tableau 4.39: Intervalles de confiance pour le modèle logistique binomial

ICG _{obs} à ...	Moyenne ICG _{prédi}	Écart-type ICG _{prédi}	Intervalle de confiance à 97,5%
1	0,427752	0,159128	[0,417430 ; 0,438074]
0	0,285050	0,171256	[0,277210 ; 0,292890]

4.6.2 Validation

En appliquant les mesures correctives, le tracé des valeurs prédites en fonction des valeurs observées donne un bien meilleur résultat. On se rapproche, il est vrai, assez nettement de l'équation recherchée, à savoir celle de la première bissectrice (Tableau 4.40).

Tableau 4.40: Comparaison entre le modèle général "normal" et l'équivalent corrigé

Modèle	Ordonnée à l'origine	Coefficient directeur	R ²
Général "normal"	3,6022	0,1953	0,1953
Général corrigé	2,0542	0,4779	0,1842

Si les modèles sont sensiblement améliorés par les corrections apportées, sont-ils néanmoins suffisamment précis pour prédire de manière juste ? C'est à cette question que l'on tente de répondre en les éprouvant sur les nouvelles structures recensées par le MTQ.

4.6.3 Application des modèles prédictifs à de nouvelles données

Depuis l'extraction des données du site, le 15 juin 2015, de nouvelles structures ont été répertoriées par le ministère. Au total, ce sont donc vingt-neuf nouveaux ponts dont l'état, par l'intermédiaire de l'ICG, est connu. Il est alors intéressant de voir si ces états peuvent être retrouvés grâce aux modèles prédictifs et aux données concernant les différentes variables d'entrée. Le Tableau G.1 et le Tableau G.2 en Annexe G donnent les résultats obtenus lors de l'application de la démarche et tendent à prouver la justesse des équations de prédiction développées. En effet, en tenant compte de la correction par intervalles de confiance, ce sont 81,0% et 88,9% des valeurs qui sont correctement prédites, respectivement par le modèle général et le modèle logistique binomial.

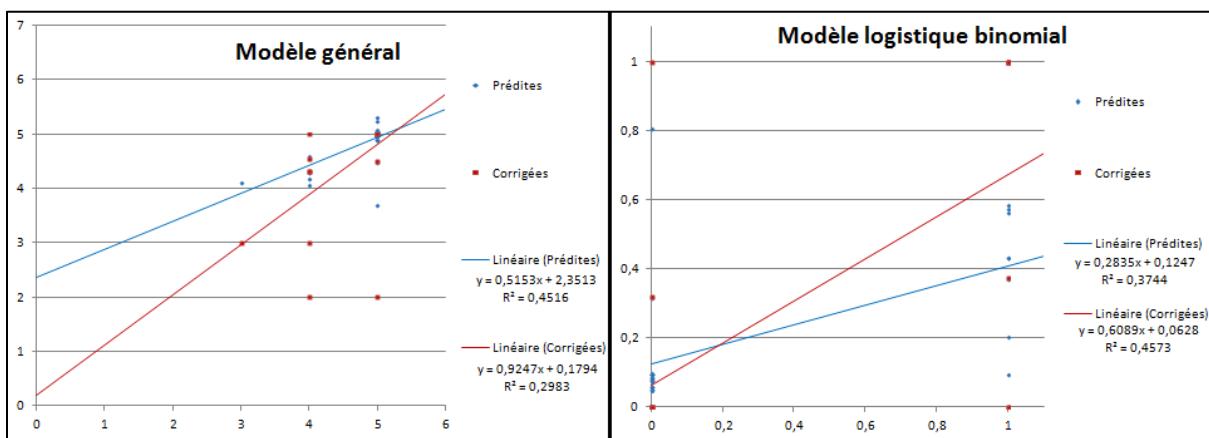


Figure 4.23: Tracé des prédictions en fonction des observations pour les nouveaux ponts

Les tracés correspondants des valeurs prédictes en fonction des valeurs observées (Figure 4.23) soulignent, une fois de plus, l'apport de qualité engendré par la méthode des intervalles de confiance.

CHAPITRE 5 CONCLUSIONS ET RECOMMANDATIONS

5.1 Réponses aux objectifs

5.1.1 Pertinence de la nouvelle approche

L'objectif principal de ce projet étant d'ouvrir la voie à une nouvelle approche du problème de planification des travaux d'entretien, on peut dire qu'il a été, de ce point de vue, atteint. En effet, la mise en place d'une base de données conséquente et exploitable a permis de développer de nouveaux modèles basés uniquement sur les observations, ce qui n'avait pas encore été fait.

Les résultats obtenus sont assez satisfaisants dans la mesure où, sans éléments véritablement techniques, on arrive à expliquer environ 20% de la variance de l'ICG. Ils sont également très similaires d'une méthode à l'autre. Cela tend à souligner la robustesse des modèles développés. Enfin, bien qu'il soit encore possible de les optimiser ou de les améliorer, on a vu, sur les ponts récemment ajoutés, que le pouvoir prédictif des modèles mis en place était déjà très intéressant. Le bilan global est donc plus qu'encourageant.

En outre, certains résultats trouvés dans l'étude sont particulièrement intéressants. On peut notamment penser aux variables identifiées comme jouant un rôle dans le processus de dégradation des structures, mais aussi à certaines tendances mises en avant lors de l'analyse descriptive. De ce point de vue aussi, le projet remplit son objectif car il débouche sur certaines pistes de réflexion qui pourraient être explorées plus en détail.

5.1.2 Identification des facteurs influents

Les méthodes statistiques ont permis de déterminer les variables, dans la base de données, qui ont un effet sur les valeurs prises par l'ICG. Si de premiers résultats ont été pressentis grâce à l'analyse descriptive des données, le modèle général a permis de les affiner en excluant certains facteurs qui ne faisaient finalement que capter l'effet d'autres variables significatives. On pense notamment au parti politique ou au type de ponts dont les effets, avérés dans un premier temps, pouvaient en fait être dus à la prise en compte, de manière implicite, de l'aspect temporel.

Ces conclusions justifient, en outre, la démarche suivie, des analyses spécifiques vers une plus générale, car ces effets "cachés" auraient été plus difficilement détectables si le raisonnement avait été mené dans l'autre sens (du tout vers le détail).

Au final, ce sont donc six éléments qui s'avèrent être véritablement significatifs :

- ✓ L'âge de la structure ;
- ✓ La largeur carrossable ;
- ✓ La superficie du tablier ;
- ✓ Le débit journalier moyen annuel ;
- ✓ La région administrative dans laquelle se trouve le pont ;
- ✓ Le matériau des éléments porteurs.

5.1.3 État des lieux

En dehors de l'identification des facteurs influents, les analyses des valeurs moyennes de l'ICG en fonction des différentes modalités (cf. partie 4.2.1 – Traitement des variables catégoriques), permettent de tirer quelques conclusions intéressantes.

Un aspect pouvant notamment être perçu comme une invitation à une réflexion plus poussée concerne les différences entre les régions administratives. Les structures de l'Outaouais semblent effectivement être en meilleur état que les autres, alors que celles du Centre-du-Québec paraissent plus dégradées. Sachant que la dimension temporelle ne semble pas intervenir, chercher à expliquer ces écarts peut être intéressant. Cela passe sans doute par l'étude de données démographiques propres aux régions et l'établissement de groupes de régions dont les caractéristiques sont assez proches.

Par ailleurs, si on pouvait penser, au regard de la littérature, que le climat jouerait un rôle important quant au processus de détérioration, les analyses statistiques s'accordent sur le fait que ce n'est pas le cas. Ce facteur semble donc être correctement pris en compte dans les choix faits lors des phases de conception préliminaire. Le fait que les organismes chargés de l'entretien soient conscients des problèmes potentiels liés au climat et adaptent leur manière de procéder pour y faire face est également une explication plausible. Dans les deux cas, la compétence des décisionnaires est mise en valeur.

5.2 Extensions et utilités possibles

Comme mentionné à plusieurs reprises, ce mémoire constitue l'élément fondateur d'un projet plus vaste et plus complet. Il convient alors de préciser les travaux complémentaires qui pourraient possiblement suivre.

Tout d'abord, au niveau des tendances mises en avant dans la partie 5.1.3 – État des lieux, il peut être particulièrement intéressant de creuser la piste selon laquelle l'état des structures diffère suivant les régions administratives. Pour ce faire, il faudrait étudier en détail les caractéristiques de chacune d'elles, les éventuelles différences de politiques en ce qui concerne la construction ou l'entretien des ouvrages, etc.

Cependant, la manière la plus utile d'aborder et d'approfondir le travail effectué est de le voir sous l'angle de l'aide à la décision. En effet, une première approche consisterait à considérer les modèles prédictifs comme de bons outils complémentaires, mis à la disposition des décisionnaires, pour valider ou questionner les choix faits quant à de futurs projets. L'objectif n'est pas de remplacer les méthodes calculatoires basées sur les théories des ponts et de résistance des matériaux, ni même l'expérience des acteurs du processus de prise de décisions, mais d'apporter une méthode supplémentaire pour s'assurer de la durabilité de l'ouvrage selon les choix effectués. Si l'ICG prédit est bon, alors c'est un nouveau voyant qui passe au vert, tandis que s'il est mauvais, il peut être bon de vérifier à nouveau les options retenues.

Toutefois, l'intérêt principal de ces modèles prédictifs résiderait dans la possibilité d'assister les décisionnaires, non pas dans les phases de construction, mais dans la planification des opérations de maintenance. Jusqu'alors, les travaux de réparation ne sont entrepris qu'après une inspection complète et détaillée de la structure. Ces inspections prennent du temps et sont généralement coûteuses. Les équations de prédiction peuvent alors être de bons substituts. Si, pour une structure, toutes les variables explicatives, à l'exception de son âge, sont fixées, on peut prévoir l'évolution de l'état de la structure en fonction du temps. Il est alors possible de déterminer l'âge à partir duquel la structure va se dégrader et prévoir des travaux de réhabilitation aux alentours de la date correspondante. Cela peut, par exemple, se faire grâce à la création d'un logiciel dont l'interface proposerait aux décisionnaires de rentrer les valeurs numériques et cocher certaines cases correspondant aux modalités relatives aux structures. Il retournerait ensuite, à partir de l'équation de prédiction, le tracé de l'ICG en fonction du temps. C'est d'ailleurs dans ce cadre

que le modèle logistique prend tout son sens, grâce à l'approche probabiliste introduite : plutôt que d'affirmer que telle structure sera dans tel état à tel moment, il est plus intéressant de s'intéresser aux chances qu'elle a de présenter un défaut, par exemple (Figure 5.1).

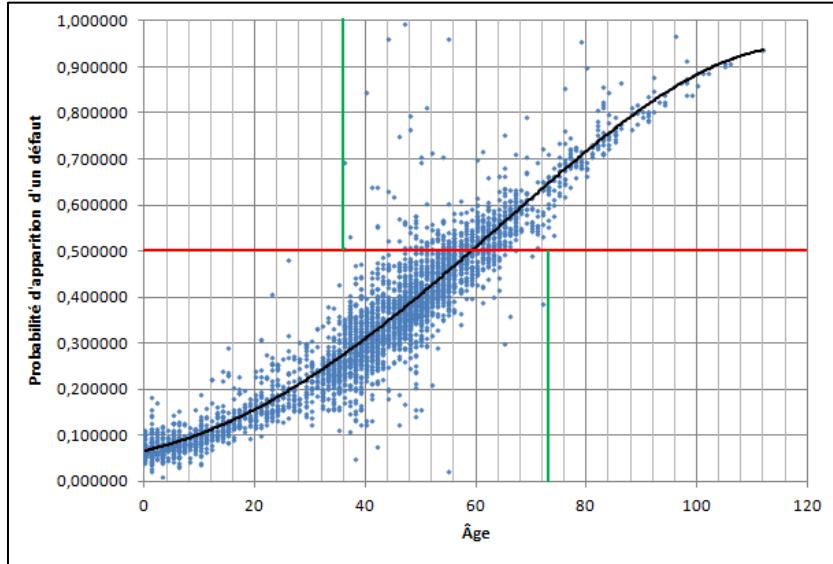


Figure 5.1: Probabilité d'apparition d'un défaut selon l'âge

Cette figure est un exemple donnant l'évolution temporelle de la probabilité d'apparition d'un défaut, pour l'ensemble des ponts étudiés (toutes modalités confondues donc). On y voit que la probabilité devient supérieure à 50% aux alentours de 60 ans. Les limites verticales indiquent, dans un cas, qu'aucune structure n'a de chance de présenter un défaut avant 36 ans et, dans l'autre, que plus aucun pont n'a encore une chance d'être en bon état au bout de 73 ans.

Il est ensuite bien entendu possible d'étendre ce raisonnement : si la Figure 5.1 donnait la probabilité d'apparition d'un défaut quelconque, on peut aussi s'intéresser à celles pour les apparitions d'un défaut significatif (Figure 5.2) ou majeur (Figure 5.3).

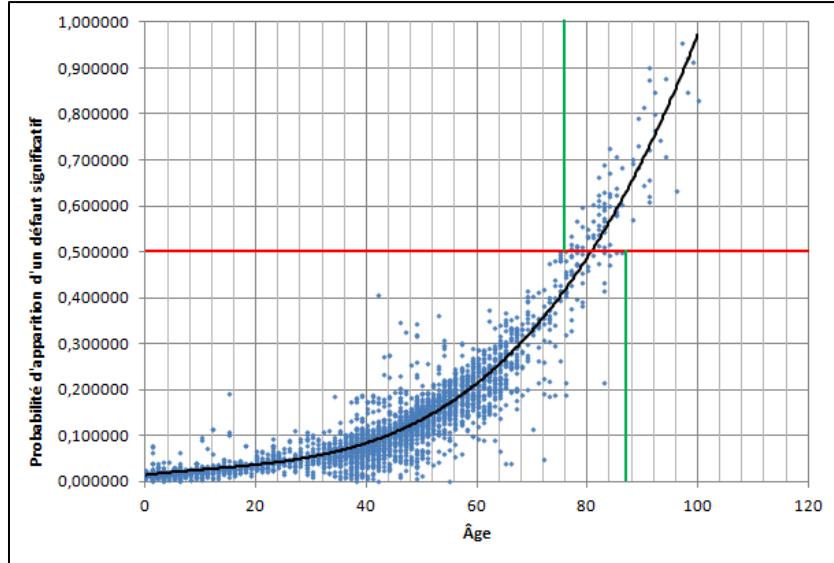


Figure 5.2: Probabilité d'apparition d'un défaut significatif

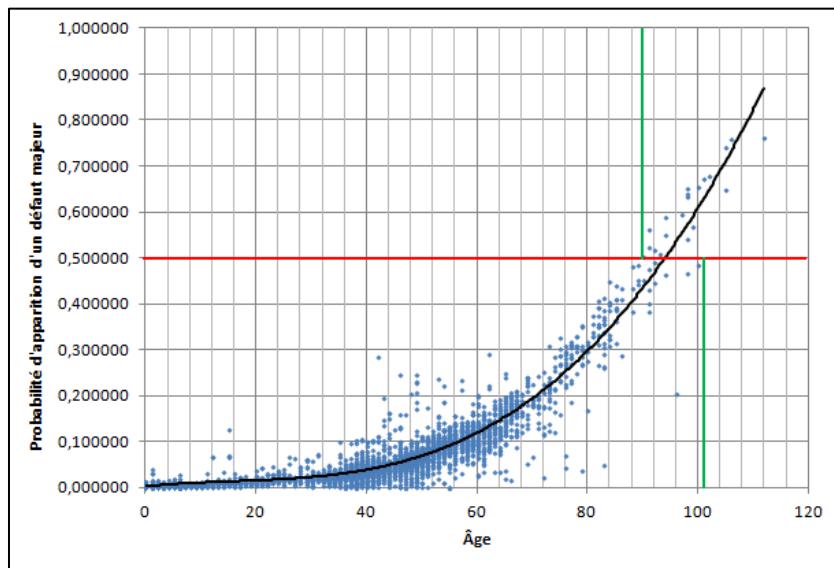


Figure 5.3: Probabilité d'apparition d'un défaut majeur

Plus le défaut recherché est grave et plus l'âge à partir duquel la probabilité d'apparition devient supérieure à 0,5 est élevé. Si pour un défaut quelconque, cet âge "critique" était de 60 ans, il est, respectivement, de 80 ans et 94 ans pour des défauts significatif et majeur. Il est rassurant de trouver des résultats conformes à la logique.

Avec de tels modèles, il est donc envisageable de prédire la dégradation des structures et de grouper, géographiquement parlant, les opérations de maintenance sur plusieurs ponts

"sensibles", au lieu d'agir au cas par cas. Cela pourrait permettre de réaliser des économies d'échelle, importantes compte-tenu des contraintes budgétaires des organismes publics.

5.3 Limites de l'étude

Ce projet présente néanmoins quelques points qu'il est possible de critiquer et qui peuvent faire l'objet d'améliorations dans le futur.

5.3.1 Critiques

Lors du traitement de variables catégoriques, il est recommandé, quel que soit la méthode statistique utilisée, d'avoir un nombre limité de modalités par facteur pour que les résultats soient les plus significatifs possibles (classiquement entre quatre et six). Cependant, dans le mémoire, des variables telles que le type de climat ou la région administrative présentent un nombre de modalités bien supérieur. Elles ont néanmoins été conservées telles quelles car la solution serait de regrouper certaines modalités entre elles mais cela nécessite une étude assez approfondie pour ne pas agréger de manière incomplète ou incorrecte.

Un autre aspect discutable de l'étude est la prise en compte de la dimension politique par l'intermédiaire des gouvernements provinciaux. Il serait, à priori, plus intéressant de se pencher sur la question à l'échelle des CEP et des députés associés. Encore une fois, cela demande une quantité de travail supplémentaire non négligeable et il était préférable de mettre ces efforts au profit d'autres aspects plus primordiaux tels que l'analyse précise et rigoureuse des données, la mise en place de premiers modèles permettant de juger la pertinence de l'étude ou la détermination des domaines d'application de celle-ci, par exemple.

5.3.2 Axes d'amélioration

Pour tenir compte du problème d'agrégation, l'analyse par grappes pourrait être utilisée. Elle permet, en effet, de regrouper les modalités selon les similarités qu'elles peuvent présenter au regard de certaines données. Pour le cas des régions administratives, on peut notamment penser aux données démographiques, économiques, etc.

Toujours en ce qui concerne la mise en place de méthodes statistiques "alternatives", il peut être intéressant d'avoir recours aux techniques de validation croisée. Elles consistent à mettre en place un modèle prédictif à partir d'un nombre limité d'observations, l'échantillon d'élaboration, et le

valider ensuite sur les autres, l'échantillon de test. On recoupe ainsi ce qui a été fait lors de la validation du modèle général sur les nouvelles structures recensées mais à une échelle plus importante. À titre d'exemple, en choisissant de manière aléatoire cinquante observations pour chaque valeur d'ICG, et en régressant uniquement sur ces deux cents cas, on a pu faire passer, dans le cas de la régression linéaire multiple, le $R^2_{\text{ajusté}}$ de 0,152650 à 0,234695. Si le résultat était ensuite bien moins convaincant sur l'échantillon test et si cela est, bien sur, plus compliqué à mettre en œuvre quand les facteurs catégoriques sont pris en compte, il peut toutefois s'agir d'un axe d'amélioration envisageable.

Une autre manière d'améliorer les modèles pourrait être de prendre en compte un plus grand nombre de variables explicatives. Ainsi, quand on regarde la liste des paramètres potentiellement influents, élaborée à partir de l'analyse des différents travaux de la littérature (Tableau 1.2), certains n'ont pas été incorporés à l'étude car obtenir les données correspondantes pour l'ensemble des 3591 ponts était trop complexe dans le cadre de ce seul mémoire de maîtrise. Toutefois, les ajouter par la suite pourrait affiner la qualité de prédiction. Dans cette optique, un de ces facteurs devrait avoir une importance majeure : l'historique d'entretien. En effet, plutôt que d'analyser l'état d'une structure à un instant précis, sans indications sur ce qui a pu se passer avant, avoir le détail des interventions de maintenance ayant eu lieu devrait permettre d'augmenter le pouvoir prédictif des modèles. De plus, il serait alors possible d'évaluer l'impact des opérations d'entretien sur l'état des structures.

Enfin, si la base de données du site du MTQ venait à s'étoffer de manière significative, il serait bien de développer des méthodes similaires pour les structures de type ponceaux ou murs, laissées de côté pour l'instant.

BIBLIOGRAPHIE ET LISTE DE RÉFÉRENCES

- Abraham D. M. & Wirahadikusumah R. (1999). Development of Prediction Models for Sewer Deterioration, *Durability of Building Materials and Components*, Vol. 8, 1257-1267.
- Akiyama M. & Frangopol D.M. (2013). Life-cycle of bridges under multiple hazards : Earthquake, tsunami, and continuous deterioration, *Safety, Reliability, Risk and Life-Cycle Performance of Structures & Infrastructures*.
- Association mondiale de la Route (2014). *Importance de l'entretien routier*.
- Aubert H. (2011). *Manuel de statistique*, Ellipses Marketing, 408 p.
- Beck J.-L. (2013). Bayesian system identification and response predictions robust to modeling uncertainty, *Safety, Reliability, Risk and Life-Cycle Performance of Structures & Infrastructures*.
- Boccanfuso D., Joannis M., Richard P. & Savard L. (2014). A comparative analysis of funding schemes for public infrastructure spending in Quebec, *Applied Economics*, 46:22, 2653-2664.
- Boyer M. (2012). Favoriser les partenariats public-privé dans le développement et le maintien des infrastructures publiques, *La performance et le développement économiques à long terme du Québec : Les douze travaux d'Hercule-Québec (mis à jour, revus et corrigés – 6^e édition)*, Chap. XII, 98-101.
- Bruxelle C. (2005). *Élaboration d'un Cadre de Système d'Aide à la Gestion et à l'Entretien (sage) des Infrastructures Civiles*, École de Technologie Supérieure – Montréal.
- Busscher T., Tillema T. & Arts J. (2015). In search of sustainable road infrastructure planning: How can we build on historical policy shifts ?, *Transport Policy*, Vol. 42, 42-51.
- Cheng T. & Pandey M. D. (2012). An accurate analysis of maintenance cost of structures experiencing stochastic degradation, *Structure and Infrastructure Engineering*, 8:4, 329-339.
- Cho T., Kim S.-S. & Kim T.-S. (2014). A quadratic hierarchical Bayesian dynamic prediction model for infrastructure maintenance, *Nonlinear Dynamics*, 76: 609-626.
- Chouquet C. (2010). *Modèles Linéaires*, Laboratoire de Statistique et Probabilités – Université Paul Sabatier – Toulouse.

Chughtai F. & Zayed T. (2008). Infrastructure Condition Prediction Models for Sustainable Sewer Pipelines, *Journal of Performance of Constructed Facilities*, September/October 2008, 333-341.

Code Civil du Québec (2016). Chapitre I-8.3 : Loi sur les infrastructures publiques, Chap. I, II, III et V.

Commission d'enquête sur le viaduc de la Concorde (2007). *Commission d'enquête sur l'effondrement d'une partie du viaduc de la Concorde*, Rapport.

Commission de révision permanente des programmes (2015). *Cap sur la performance – Sommaire*, Rapport.

Conseil du trésor du Québec (2015). Les infrastructures publiques du Québec, *Plan québécois des infrastructures 2015-2025*.

Conseil du trésor du Québec (2016). Les infrastructures publiques du Québec, *Plan québécois des infrastructures 2016-2026*, Partie II, 23-77.

DeStefano P. & Grivas D. (1998). Method for Estimating Transition Probability in Bridge Deterioration Models, *Journal of Infrastructure Systems*, 4(2), 56-62.

De Bourmont M. (2012). *La résolution d'un problème de multicolinéarité au sein des études portant sur les déterminants d'une publication volontaire d'informations : proposition d'un algorithme de décision simplifié basé sur les indicateurs de Belsley, Kuh et Welsch (1980)*.

De Marcellis-Warin N. (2014). Réaliser un inventaire et évaluer l'état des infrastructures dans votre municipalité, *Guide de gestion des risques reliés aux projets d'infrastructure municipale à destination des élus*, 1.1, 26-33.

De Marcellis-Warin N. & Peignier I. (2015). *Évaluation des coûts sociaux économiques reliés aux bris des infrastructures souterraines au Québec – Sommaire exécutif*, Rapport de projet du CIRANO No. 2015RP-14.

De Marcellis-Warin N., Peignier I., Mouchikhine V. & Mahfouf M. (2013). *Évaluation des coûts socio-économiques reliés aux bris des infrastructures souterraines au Québec*, Rapport de projet du CIRANO No. 2013RP-22.

Direction des transports de la ville de Montréal (2010). *Planification à long terme des investissements*, Présentation aux commissions conjointes.

Encyclopédie Larousse, *Climat : Les climats du monde*. [en ligne]

<http://www.larousse.fr/encyclopedie/divers/climat_les_climats_du_monde/185927>

FIB, CERIB, SNBPE, SNPB & CIMBETON (2005). Les ponts courants en béton, *Bétons et Ouvrages d'art*, Tome 1.

Fonds Monétaire International (FMI) (2014). Le moment est-il propice à une relance des infrastructures? Les effets macroéconomiques de l'investissement public, *Perspectives de l'économie mondiale: Nuages et incertitudes de l'après-crise*, Chap. 3, 79-119.

Gagné R. & Haarman A. (2011). *Les Infrastructures Publiques au Québec : Évolution des Investissements et Impact sur la Croissance de la Productivité*, Partenariat entre HEC Montréal et le Centre sur la productivité et la prospérité.

Gagnon M., Gaudreault V. & Overton D. (2008). *L'âge de l'infrastructure publique : une perspective provinciale*, Statistique Canada, No. 11-621-MIF2008067.

Galfalvy H. C. (2000). *Censored Regression Models with Applications to Infrastructure Degradations Studies*, University of Illinois – Urbana-Champaign.

Gerardin V. & McKenney D. (2001). Une classification climatique du Québec à partir de modèles de distribution spatiale de données climatiques mensuelles : vers une définition des bioclimats du Québec, *Contribution du service de la cartographie écologique*, No. 60, Direction du patrimoine écologique et du développement durable, Ministère de l'Environnement.

Guillaumot V., Durango P. L. & Madanat S. M. *Adaptative Optimization of Infrastructure Maintenance and Inspection Decisions under Performance Model Uncertainty* (Submitted for possible publication in the ASCE *Journal of Infrastructure Systems*).

Hébert B.-P. (1998). *Régression avec une variable dépendante ordinaire : comparaison de la performance de deux modèles logistiques ordinaires et du modèle linéaire classique à l'aide de données simulées*, Université Laval – Québec.

Hegazy T. & Saad D. A. (2014). A microeconomic perspective on infrastructure rehabilitation, *Construction Management and Economics*, 32:5, 433-445.

- Herz R. (1996). Dégradation et renouvellement des infrastructures [Un modèle de survie par cohorte], *Flux*, No. 23, 21-36.
- Imbert F. (2006). *Ponts à haubans – Réglages des haubans et contrôles géométriques*, Présentation du service d'Études techniques des routes et autoroutes.
- International Organization for Standardization (ISO) (1997). *Buildings: Service life planning, part 1 – General principles*, ISO/TC59/SC3.
- Joanis M. (2011). The road to power: partisan loyalty and the centralized provision of local infrastructure, *Public Choice*, 146: 117-143.
- Jones S., Hensher D. A., Rose J. & Walker R. G. (2012). Infrastructure Asset Reporting Options: A Stated Preference Experiment, *Accounting Horizons*, Vol. 26, No. 3, 465-491.
- Kleiner Y. (2001). Scheduling Inspection and Renewal of Large Infrastructure Assets, *Journal of Infrastructure Systems*, 7 (4), 136-143.
- Liao D. & Valliant R. (2012), Indices de conditionnement et décompositions des variances pour le diagnostic de la colinéarité dans l'analyse de données d'enquête au moyen de modèles linéaires, *Techniques d'enquête*, Vol. 38, 205-220.
- Madanat S., Mishalani R. & Wan Ibrahim W. H. (1995). Estimation of Infrastructure Transition Probabilities from Condition Rating Data, *Journal of Infrastructure Systems*, 1:2, 120-125.
- McDonald S. & Zhao J. (2001). Condition assessment and rehabilitation of large sewers, *Proc. Int. Conf. on Underground Infrastructure Research*, 361-369.
- Ministère des Affaires municipales et de l'Occupation du territoire du Québec (2010). *Répertoire des municipalités – Hors MRC*. [en ligne]
<<http://www.mamrot.gouv.qc.ca/repertoire-des-municipalites/fiche/mrc/000/>>
- Ministère des Affaires municipales et de l'Occupation du territoire du Québec (2010). *Répertoire des municipalités – Hors MRC (autochtones)*. [en ligne]
<<http://www.mamrot.gouv.qc.ca/repertoire-des-municipalites/fiche/mrc/001/>>
- Ministère des Transports du Québec (1995). *Manuel de conception des ponceaux*, Service de l'hydraulique.

Ministère des Transports du Québec (1993). Partage des responsabilités entre le gouvernement et les municipalités, *La voirie locale*, Guide d'information.

Ministère des Transports du Québec, Direction des structures (2007). *Manuel d'entretien des structures*.

Ministère des Transports du Québec, Direction des structures (2010). *Manuel d'inspection des structures*.

Ministère des Transports et de l'Infrastructure (2012). Entretien des immobilisations routières, *Rapport de la vérificatrice générale*, Chap. 5, 217-274.

Ministère des Transports du Québec, Projets et infrastructures – Structures et infrastructures – Infrastructures routières – Inventaire des structures. [en ligne]

< <https://www.diffusion.transports.gouv.qc.ca/ords/pes/f?p=diff:DSI> >

Ministère du Développement durable, de l'Environnement et de la Lutte contre les changements climatiques du Québec (2015). *Normales climatiques (1981 – 2010) – Climat du Québec*. [en ligne]

< <http://www.mddelcc.gouv.qc.ca/climat/normales/climat-qc.htm> >

Mirza S. (2006). Durability and sustainability of infrastructure – a state-of-the-art report, *Canadian Journal of Civil Engineering*, 33: 639-649.

Morcous G. & Akhnoukh A. (2005). Stochastic Modeling of Infrastructure Deterioration: an Application to Concrete Bridge Decks, *Joint International Conference on Computing and Decision Making in Civil and Building Engineering June 14-16, 2006 – Montréal, Canada*.

Morin A. & Findlay S. (2016). ANOVA à critères multiples, Université d'Ottawa – Bio 4518 – Biostatistiques appliquées, Diapositives de cours (mise à jour du 2016-06-06 19:19).

Rakotomalala R. (2007). *Didactiel Régression Logistique Polytomique Ordinale*, Université de Lyon 2.

Rakotomalala R. *Régression logistique polytomique*, Université de Lyon 2, Diapositives de cours. [en ligne]

< http://eric.univ-lyon2.fr/~ricco/cours/slides/regression_logistique_polytomique.pdf >

Rostrum J, Baur R., Saegrov S., Harold S. & Schilling W. (1999). Predictive service-life models for urban water infrastructure management, *Proc. 8th Int. Conf. on Urban Storm Drainage*, 594-601.

Ruwanpura J., Ariaratnam S. T. & El-assaly A. (2004). Prediction models for sewer infrastructure utilizing rule-based simulation, *Civil Engineering and Environmental Systems*, 21:3, 169-185.

Saporta G. (2011). *Probabilités, Analyse des données et Statistique*, Technip, 622 p.

Stefanini G. (2012). *Clin d'œil sur un matériel : Histoire du pont Bailey*, Musée du Génie d'Angers.

Tufféri S. (2015). *Modélisation prédictive et Apprentissage statistique avec R*, Technip, 415 p.

Yang Y. (1999). *Statistical Models for Assessing Sewer Infrastructure Inspection Requirements*, University of Alberta – Edmonton.

Yi J. (2010). Application and Comparison of Regression and Markov Chain Methods in Bridge Condition Prediction and System Benefit Optimization, *Journal of the Transportation Research Forum*, Vol. 49, No. 2, 91-110.

Younis R. & Knight M. A. (2010). A probability model for investigating the trend of structural deterioration of wastewater pipelines, *Tunnelling and Underground Space Technology*, Vol. 25, 670-680.

ANNEXE A – GÉNÉRALITÉS SUR LES PONTS

Cette annexe présente le lexique spécifique aux ponts et aux ponceaux ainsi que les principaux types définis par le MTQ et que l'on retrouve, par conséquent, dans la base de données.

Terminologie des ponts

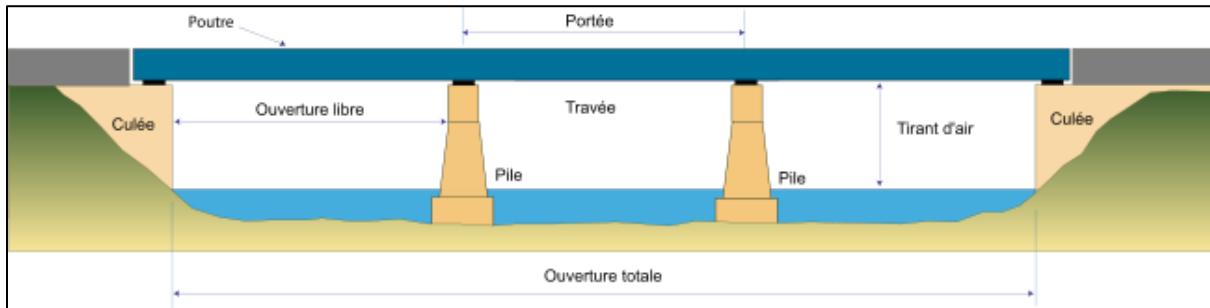


Figure A.1: Terminologie des ponts (*Wikipédia*)

Le tablier constitue l'élément "principal" du pont. Il s'agit en effet de la structure sur laquelle vont se faire les déplacements. Le tablier est supporté par des appuis qui peuvent être des piles, des culées ou des piles-culées si le tablier est discontinu.

Les piles sont des appuis intermédiaires massifs et permanents. Elles peuvent être principalement de deux types (colonne ou voile) et travaillent non seulement en compression mais aussi en flexion à cause des effets du vent. Dans tous les cas, une pile est surmontée d'un chevêtre, qui lui permet de reprendre l'effort encaissé par le tablier, et repose sur une semelle qui assure la transmission des efforts au sol et lie le pont aux fondations.

L'ouverture (libre) est l'espace situé entre deux piles ou entre une pile et une culée tandis que l'ouverture totale correspond à l'espace entre les deux piédroits des culées. L'ouverture totale est donc égale à la somme des ouvertures et des largeurs de piles.

Enfin, le tirant d'air correspond en fait à la hauteur libre sous l'ouvrage.

Plusieurs classifications sont possibles pour distinguer les ponts entre eux. Ils peuvent ainsi être différenciés par leur fonction, c'est-à-dire le type de voie portée (pont routier, pont ferroviaire, passerelle, etc.), par leurs matériaux, par leur importance, par le type de sollicitations auxquelles ils sont soumis ou encore par le type d'éléments porteurs.

Classification des ponts par éléments porteurs

Dans ce dernier cas, cinq catégories de ponts existent et sont présentées sur la Figure A.2.

Remarque :

Les flèches indiquent la reprise du chargement.

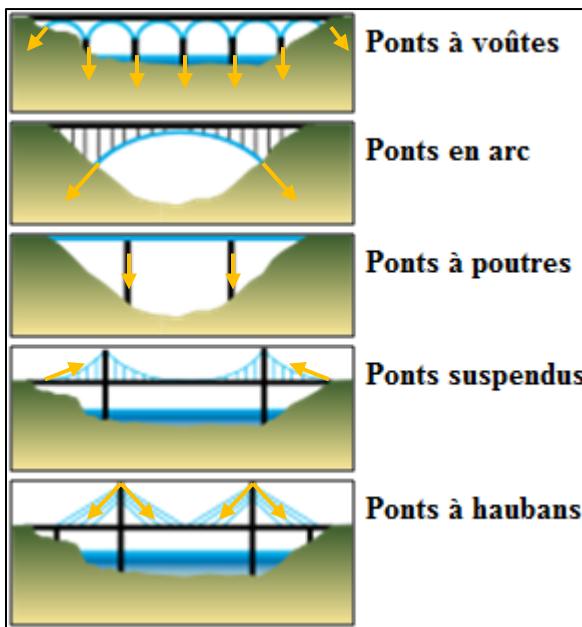


Figure A.2: Les différents ponts (*Wikipédia*)

Les ponts suspendus sont caractérisés par deux câbles principaux de retenue ancrés de part et d'autre du pont, sur les rives. Des suspentes font office de liens entre lesdits câbles et le tablier.

Dans le cas des ponts haubanés, le tablier est directement soutenu par des câbles (haubans) qui vont transmettre le chargement aux pylônes.

Ponts à voûtes

Ils sont également appelés ponts de maçonnerie et ils sont souvent considérés comme une sous-catégorie des ponts en arc plutôt que comme une catégorie à part entière. Cela est notamment dû au fait que, comme pour les ponts en arc, les éléments porteurs travaillent en compression.

Ils peuvent être subdivisés selon plusieurs critères : la forme de la voûte, le type d'appareillage de la voûte (c'est-à-dire le mode de construction) et le type d'avant-bec et d'arrière-bec (triangulaire, en amande, rectangulaire, circulaire). L'avant-bec est en fait l'élément du pont

servant à protéger la pile contre les éléments en amont qui peuvent la heurter. L'arrière-bec a le même rôle mais pour les éléments en aval qui seraient pris dans les tourbillons créés par l'écoulement d'un fluide autour de la pile, par exemple.

Ponts en arc

Un pont en arc est un pont pour lequel l'obstacle est franchi par une seule arche, sollicitée en compression. C'est d'ailleurs ce qui le différencie fondamentalement d'un pont voûté, pont pour lequel l'obstacle est franchi en plusieurs arches à l'aide de piles intermédiaires.

Plusieurs aspects permettent de catégoriser ces ponts. Tout d'abord, il est possible de les différencier entre eux par la position qu'occupe le tablier. Comme le montre la Figure A.3, celui-ci peut être en position inférieure (tablier suspendu), tenu par des suspentes entre l'arc et le tablier, intermédiaire ou supérieure (tablier porté).



Figure A.3: Différenciation selon la position du tablier

En outre, il est possible de distinguer les ponts en arc par les matériaux dont ils sont constitués. De ce point de vue, il peut être intéressant de préciser que, récemment, des ponts "concrete filled steel tube" (CFST) ont été développés. Il s'agit en fait de ponts en arc dont l'arche est réalisée en treillis de tubes métalliques remplis de béton.

Les éléments porteurs (les bêquilles) des ponts à bêquilles travaillent en compression sur le même principe que les ponts en arc. Ils sont donc considérés comme une sous-catégorie de ces derniers.



Figure A.4: Pont à bêquilles

Ponts à poutres

Les ponts à poutres sont, comme leur nom l'indique, des ponts dont les éléments porteurs sont des poutres droites sollicitées en flexion. Elles peuvent être de différentes formes et constituées de différents matériaux. Pour ce qui est de la forme des poutres, il en existe quatre grandes familles : les poutres à âmes pleines, les poutres caissons, les poutres treillis (barres métalliques appelées membrures) et les poutres bow-string.

a) Poutres bow-string

Ce dernier type de ponts à poutres est un peu particulier dans la mesure où il est caractérisé par une poutre latérale faisant office de tirant et d'un arc. Un tirant est en fait une membrure reprenant des efforts de traction et empêchant l'écartement.

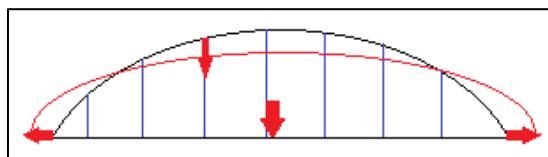


Figure A.5 : Pont à poutres bow-string

Au niveau des efforts induits par le chargement (Figure A.5), le tablier travaille en flexion. Les suspentes verticales vont donc se tendre et entraîner une "compression" de l'arc (silhouette "compressée"). Cette dernière se traduit par une traction du tirant sur lequel est ancré l'arc. L'effet de cette traction se répercute au niveau des points d'ancrage et s'oppose à celui de la flexion initiale.

b) Poutres treillis

À l'instar de l'arche des ponts en arc, le tablier peut également être en position supérieure, intermédiaire ou inférieure par rapport aux treillis et il existe aussi des ponts à poutres treillis provisoires. Il s'agit en fait de ponts de type Bailey. L'origine de ces ponts est militaire et date de la Première Guerre Mondiale, même si son utilisation s'est démocratisée durant la Seconde. Le principe était de permettre le transport et la construction simples et rapides de ponts permettant de franchir des obstacles et poursuivre une éventuelle offensive.

c) Poutres caissons

Par rapport aux poutres simples, les caissons offrent une meilleure résistance en torsion, en cisaillement ou encore en flambement. Ils présentent donc un avantage non négligeable étant donné que la flexion (type d'effort auquel sont soumises les poutres) s'accompagne souvent d'un effort tranchant de cisaillement.



Figure A.6: Pont à poutres-caissons

d) Autres types de ponts à poutres

Bien que ne possédant pas forcément de poutres, d'autres types de ponts sont régulièrement classés dans cette catégorie de ponts. Cela s'explique par le fait que leurs éléments porteurs sont également sollicités en flexion. C'est notamment le cas des portiques (Figure A.7) ou des ponts à dalle (Figure A.8).

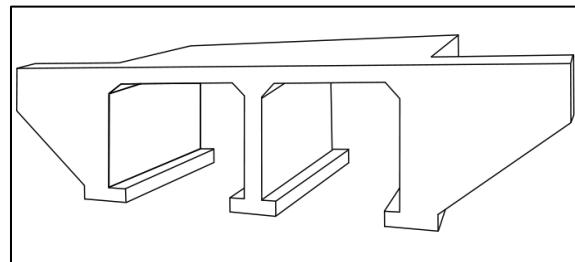


Figure A.7: Portique double (*Wikipédia*)

Les portiques sont constitués d'une dalle horizontale encastrée sur des piédroits et des piles. Il s'agit généralement de structures préfabriquées en béton armé ou en béton précontraint. Si le principe des ponts-dalles est sensiblement le même, la méthode de construction est différente car la dalle supérieure est coulée sur place et simplement posée sur les appuis.

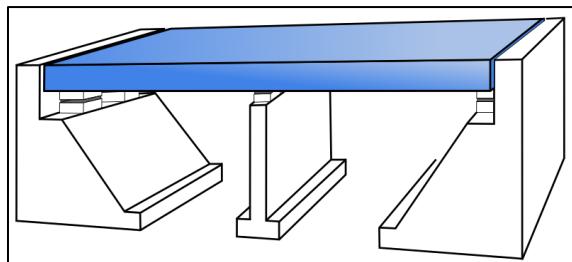


Figure A.8: Pont-dalle (*Wikipédia*)

Ponts suspendus

Les éléments porteurs des ponts suspendus travaillent en traction. En effet, la reprise des efforts dus au chargement du tablier se fait par les suspentes, puis par les câbles porteurs (les câbles paraboliques). Les efforts sont alors encaissés par les pylônes (compression) mais surtout par les câbles de retenue (traction) qui sont les câbles situés entre les ancrages et les pylônes. Ces câbles de retenue peuvent être amarrés soit dans des massifs d'ancrage, soit sur le tablier lui-même dans le cas d'ouvrages auto-ancrés.

Outre cette différenciation selon le type d'ancrages, il est possible de distinguer différents types de ponts suspendus selon le type de suspensions. Ainsi, les suspensions peuvent être des câbles métalliques, dans la plupart des cas, mais également des chaînes ou des cordages dans le cas d'ouvrages plus "rudimentaires". Il existe également des ponts caténaires dont le tablier est placé directement le long d'un ou plusieurs câbles porteurs.

Ponts haubanés

Il s'agit de la seconde grande famille de ponts dont les éléments porteurs, les haubans, sont sollicités en traction (même s'il est vrai que les pylônes travaillent, eux, en compression). L'ensemble des haubans est appelé la nappe et le hauban le plus proche de la culée est appelé hauban de retenue. La nappe peut être en éventail, en semi éventail ou en harpe.

Le terme "en éventail" renvoie à la situation dans laquelle il n'y a qu'une seule accroche pour tous les haubans, au sommet du pylône (Figure A.9). Les nappes en semi éventail disposent, au contraire, de plusieurs accroches sur une longueur limitée en tête de pylône tandis que celles en harpe ont des haubans exclusivement parallèles entre eux (Figure A.10).

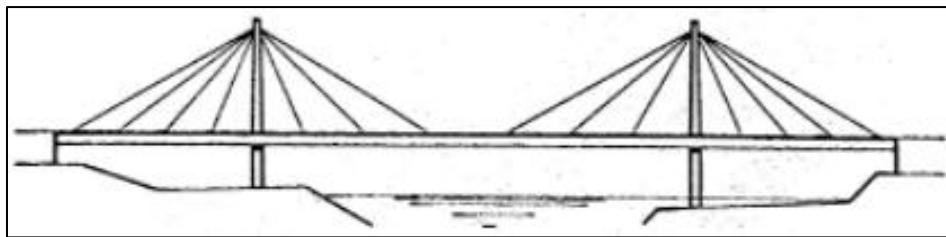


Figure A.9: Nappe en éventail

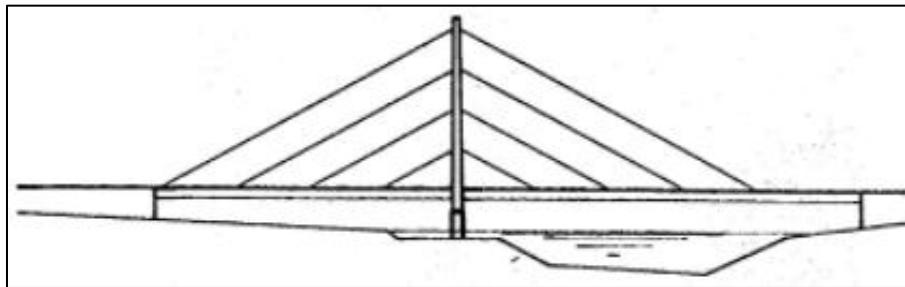


Figure A.10: Nappe en harpe

Comparatif des différents types de ponts

Tableau A.1: Avantages et inconvénients des différents types de ponts

Type de ponts	Avantages	Inconvénients
En arc	Tirant d'air important Longueur importante possible	Obligation d'avoir des appuis solides de part et d'autre
Voûté	/	Distance limitée entre les piles Coût de construction important Temps de construction long
À poutres	Solidité et légèreté Simplicité de construction Large choix de matériaux utilisables	Importantes dilatations thermiques des poutres Portée limitée et/ou obligation d'avoir régulièrement des piles
Suspendu	Les plus grandes portées	Obligation d'avoir des massifs d'ancre Entretien et remplacement des câbles long et nécessitant fermeture
Haubané	Distance très importante	Haubans fragiles et sensibles au vent et aux vibrations dues à la circulation

ANNEXE B – DÉTAIL DES PROCÉDURES D'EXTRACTION

Procédure de transformation des hyperliens

L'intérêt de cette première procédure est de convertir les hyperliens disponibles sur la page d'accueil du site du MTQ en des adresses URL exploitables pour l'extraction des données.

```
Sub TransfoLiens()
    Définition du nom de la procédure
    Dim Cell As Range
        Déclaration de la variable « Cell » (plage)
    On Error Resume Next
        Routine de gestion des erreurs : si une erreur d'exécution se produit, la procédure reprend directement là où elle s'était arrêtée
    derlig = ActiveSheet.Cells(Rows.Count, "A").End(xlUp).Row
        Attribution du nom « derlig » au numéro de la dernière ligne non-vide de la colonne A de la feuille active (« Liens »)
    For Each Cell In Range("$A$3:A" & derlig)
        Pour chacune des cellules « Cell » de la plage allant de la ligne 3 à la dernière ligne non-vide pour la colonne A
        Cell.Offset(0, 1) = Cell.Hyperlinks(1).Address
            Récupération dans la colonne immédiatement à droite des adresses URL correspondant aux hyperliens de chacune des cellules « Cell »
    Next Cell
        Passage à la cellule « Cell » suivante
    Columns(2).AutoFit
        Réajustement automatique de la largeur de la colonne B (où il y a les URL)
End Sub
Fin de la procédure
```

Procédure d'extraction des données

```
Sub ExtractionDonnées()
    Définition du nom de la procédure
    Dim i As Long, ws As Worksheet
        Déclaration des variables « i » (nombre) et « ws » (feuille de travail)
    Dim c As Range, Decalage As String
        Déclaration des variables « c » (plage) et « Decalage » (chaîne de caractères)
    Set ws = Worksheets("Données")
        Attribution de la feuille « Données » à la variable « ws »
```

```

Worksheets("Traitement").Select
Sélection de la feuille « Traitement »
Worksheets("Traitement").[$A$1].Select
Sélection de la cellule A1 de la feuille « Traitement »
derlig = ws.Cells(Rows.Count, "B").End(xlUp).Row
Attribution du nom « derlig » au numéro de la dernière ligne non-vide de la colonne B de la feuille « Données »
For i = 5 To derlig
    Itération pour « i » variant de 5 à « derlig »
    With ActiveSheet.QueryTables.Add(Connection:="URL;" & ws.Cells(i, 2).Value,
    Destination:=Range("$A$1"))
        À partir du lien inscrit dans la cellule [ligne i, colonne 2] de la feuille « Données », extraction des données web dans la cellule A1 de la feuille active (« Traitement »)
        .FieldNames = True
        .RowNumbers = False
        .FillAdjacentFormulas = False
        .PreserveFormatting = True
        .RefreshOnFileOpen = False
        .BackgroundQuery = False
        .RefreshStyle = xlInsertDeleteCells
        .SavePassword = False
        .SaveData = True
        .AdjustColumnWidth = False
        .RefreshPeriod = 0
        .WebSelectionType = xlEntirePage
        .WebFormatting = xlWebFormattingNone
        .WebPreformattedTextToColumns = True
        .WebConsecutiveDelimitersAsOne = True
        .WebSingleBlockTextImport = False
        .WebDisableDateRecognition = False
        .WebDisableRedirections = False
        .Refresh BackgroundQuery:=True
    End With
    Réglages des options de l'extraction de données Web
    With ws
        Sélection de la feuille « Données »
        For col = 3 To .Cells(4, Columns.Count).End(xlToLeft).Column

```

Pour les colonnes de la colonne 3 à la dernière colonne non vide de la ligne 4 (itération sur les colonnes prises une à une ; la colonne considérée est appelée « col »)

```
Set c = Range("A$1:I$50").Find(.Cells(4, col), LookIn:=xlValues, LookAt:=xlWhole)
```

Recherche, dans la plage allant d'A1 à I50 de la feuille « Traitement », de la cellule dans laquelle se retrouve le mot clé de la 4^{ème} ligne et de la colonne « col » de la feuille « Données ». Cette cellule est nommée « c ». La plage définie ici correspond l'emplacement occupé par chaque nouvelle fiche individuelle importée

```
If Not c Is Nothing Then
```

S'il y a une valeur dans « c »

```
Decalage = .Cells(3, col)
```

Attribution du nom « Decalage » à la cellule [ligne 3, colonne col]

```
.Cells(i, col) = c.Offset(Split-Decalage, "-") (0), Split-Decalage, "-" (1))
```

Dans la cellule [ligne i, colonne col], report de la valeur inscrite dans la cellule décalée vers le bas de x lignes et vers la droite de y colonnes par rapport à la cellule « c »

x étant le nombre indiqué à gauche du séparateur « - » dans la cellule « Decalage »

y étant le nombre indiqué à droite du séparateur « - » dans la cellule « Decalage »

```
Else
```

S'il n'y a pas de valeur dans « c »

```
.Cells(i, col) = "non trouvé"
```

Mention "non trouvé" dans la cellule [ligne I, colonne col]

```
End If
```

Fin de la condition Si

```
Next col
```

Passage à la colonne « col » suivante

```
End With
```

Fin de sélection de la feuille « Données »

```
Next i
```

Passage à la ligne « i » suivante

```
Set ws = Nothing
```

Libération de la variable « ws »

```
End Sub
```

Fin de la procédure

ANNEXE C – DÉTAIL DES PROCÉDURES DE TRI

Détection des ponts

Sub Ponts ()

Définition du nom de la procédure

Dim i As Long, j As Long

Déclaration des variables « i » et « j » (nombres)

derlig = Worksheets("Données").Cells(Rows.Count, "B").End(xlUp).Row

Attribution du nom « derlig » au numéro de la dernière ligne non-vide de la colonne B de la feuille « Données »

With Worksheets("Données")

Travail sur la feuille « Données »

For i = 5 To derlig

Itération pour « i » variant de 5 à « derlig »

j = i - 4

Attribution de la valeur i - 4 à « j »

If Cells(i, 3) = "non trouvé" Then

Si mention "non trouvé" dans la cellule [ligne i, colonne 3 (n° pont)]

Range("C" & i & ":AL" & i).Copy (Worksheets("PonceauxMursTunnels").Range("B" & j))

Copie de la plage allant de la colonne C à la colonne AL pour la ligne i vers la cellule « Bj » de la feuille « PonceauxMursTunnels »

Else

Sinon

Range("C" & i & ":AL" & i).Copy (Worksheets("Ponts").Range("B" & j))

Copie de la plage allant de la colonne C à la colonne AL pour la ligne i vers la cellule « Bj » de la feuille « Ponts »

End If

Fin de la condition Si

Next i

Passage à la ligne « i » suivante

End With

Fin du travail sur la feuille « Données »

Worksheets("Ponts").Select

Sélection de la feuille « Ponts »

Columns(2).SpecialCells(xlCellTypeBlanks).EntireRow.Delete

Suppression de la ligne entière correspondante s'il y a une cellule vide dans la 2^{ème} colonne

Worksheets("PonceauxMursTunnels").Select

Sélection de la feuille « PonceauxMursTunnels »

```
Columns(2).SpecialCells(xlCellTypeBlanks).EntireRow.Delete
```

Suppression de la ligne entière correspondante s'il y a une cellule vide dans la 2^{ème} colonne

```
End Sub
```

Fin de la procédure

Détection des ponceaux

```
Sub Ponceaux()
```

Définition du nom de la procédure

```
Dim i As Long
```

Déclaration de la variable « i » (nombre)

```
Worksheets("PonceauxMursTunnels").Select
```

Sélection de la feuille « PonceauxMursTunnels »

```
derlig = Worksheets("PonceauxMursTunnels").Cells(Rows.Count, "B").End(xlUp).Row
```

Attribution du nom « derlig » au numéro de la dernière ligne non-vide de la colonne B de la feuille « PonceauxMursTunnels »

```
For i = 1 To derlig
```

Itération pour « i » variant de 1 à « derlig »

```
If Cells(i, 28) = "non trouvé" Then
```

Si mention "non trouvé" dans la cellule [ligne i, colonne 28 (n° ponceau)]

```
Range("B" & i & ":AK" & i).Copy (Worksheets("MursTunnels").Range("B" & i))
```

Copie de la plage allant de la colonne B à la colonne AK pour la ligne i vers la cellule « Bj » de la feuille « MursTunnels »

```
Else
```

Sinon

```
Range("B" & i & ":AK" & i).Copy (Worksheets("Ponceaux").Range("B" & i))
```

Copie de la plage allant de la colonne B à la colonne AK pour la ligne i vers la cellule « Bj » de la feuille « Ponceaux »

```
End If
```

Fin de la condition Si

```
Next i
```

Passage à la ligne « i » suivante

```
Worksheets("MursTunnels").Select
```

Sélection de la feuille « MursTunnels »

```
Columns(2).SpecialCells(xlCellTypeBlanks).EntireRow.Delete
```

Suppression de la ligne entière correspondante s'il y a une cellule vide dans la 2^{ème} colonne

```
Worksheets("Ponceaux").Select
```

Sélection de la feuille « Ponceaux »

```
Columns(2).SpecialCells(xlCellTypeBlanks).EntireRow.Delete
```

Suppression de la ligne entière correspondante s'il y a une cellule vide dans la 2^{ème} colonne

```
End Sub
```

Fin de la procédure

Distinction entre les murs et les tunnels

```
Sub MursTunnels()
```

Définition du nom de la procédure

```
Dim i As Long
```

Déclaration de la variable « i » (nombre)

```
Worksheets("MursTunnels").Select
```

Sélection de la feuille « MursTunnels »

```
derlig = Worksheets("MursTunnels").Cells(Rows.Count, "B").End(xlUp).Row
```

Attribution du nom « derlig » au numéro de la dernière ligne non-vide de la colonne B de la feuille « MursTunnels »

```
For i = 1 To derlig
```

Itération pour « i » variant de 1 à « derlig »

```
If Cells(i, 36) = "non trouvé" Then
```

Si mention "non trouvé" dans la cellule [ligne i, colonne 36 (n° tunnel)]

```
Range("B" & i & ":AK" & i).Copy (Worksheets("Murs").Range("B" & i))
```

Copie de la plage allant de la colonne B à la colonne AK pour la ligne i vers la cellule « Bj » de la feuille « Murs »

```
Else
```

Sinon

```
Range("B" & i & ":AK" & i).Copy (Worksheets("Tunnels").Range("B" & i))
```

Copie de la plage allant de la colonne B à la colonne AK pour la ligne i vers la cellule « Bj » de la feuille « Tunnels »

```
End If
```

Fin de la condition Si

```
Next i
```

Passage à la ligne « i » suivante

```
Worksheets("Murs").Select
```

Sélection de la feuille « Murs »

```
Columns(2).SpecialCells(xlCellTypeBlanks).EntireRow.Delete
```

Suppression de la ligne entière correspondante s'il y a une cellule vide dans la 2^{ème} colonne

```
Worksheets("Tunnels").Select
```

Sélection de la feuille « Tunnels »

```
Columns(2).SpecialCells(xlCellTypeBlanks).EntireRow.Delete  
Suppression de la ligne entière correspondante s'il y a une cellule vide dans la 2ème colonne  
End Sub  
Fin de la procédure
```

ANNEXE D – CLASSIFICATION DE KÖPPEN-GEIGER

En 1918, le climatologue allemand Köppen se base sur des relevés de températures et de précipitations pour définir cinq types de climat principaux à travers le monde. Il s'agit de la classification primaire. Elle est ensuite perfectionnée en rajoutant des classes secondaires puis tertiaires afin de déboucher sur le modèle complet et reconnu de Köppen-Geiger.

Classification primaire

D'un point de vue notation, chacun des cinq grands climats est désigné par une lettre. Plus celle-ci est loin dans l'alphabet et plus la zone concernée est éloignée de l'équateur.

Tableau D.1: Classification primaire de Köppen-Geiger

Symbole	Désignation	Description
A	Climat tropical	Température moyenne du MPF _{année} > 18°C Pas d'hiver Fortes précipitations (PA > EA)
B	Climat sec, désertique (aride)	EA > PA Grande différence de températures jour / nuit
C	Climat tempéré, maritime (subtropical)	Températures moyennes des 3 MPF _{année} > -3°C Température moyenne du MPC _{année} > 10°C
D	Climat continental	Température moyenne du MPF _{année} < 0°C Température moyenne du MPC _{année} > 10°C Grande différence de températures été / hiver
E	Climat polaire	Température moyenne du MPC _{année} < 10°C

Classification secondaire

La classification secondaire se base sur l'étude approfondie des régimes pluviométriques. Elle débouche sur la création de plusieurs climats différents au sein d'une même classe primaire. Cela permet d'avoir une description plus précise de l'ensemble des climats mondiaux. Les nuances apportées et la liste complète des 15 "sous-climats" engendrés sont décrites dans les tableaux qui suivent (Tableau D.2 et Tableau D.3).

Remarque :

La notation $P(x)$ est utilisée pour désigner les précipitations de x .

Tableau D.2: Classification secondaire de Köppen-Geiger

Symb.	Désignation	Climats concernés	Description
w	Saison sèche en hiver	A, C et D	Pour A : $P(MPS_{hiver}) < 60 \text{ mm}$ et $100 - 0,04 \cdot PA < P(MPS_{hiver})$ Pour C et D : $P(MPS_{hiver}) < 0,1 \cdot P(MPH_{année})$
s	Saison sèche en été	A, C et D	$P(MPS_{été}) < 40 \text{ mm}$ et $0,33 \cdot (MPH_{hiver}) < P(MPS_{été})$
f	Pas de saison sèche	A, C et D	$P(MPS_{année}) > 60 \text{ mm}$
m	Mousson	A	$P(MPS_{année}) < 60 \text{ mm}$ mais $P(MPS_{année}) > 100 - 0,04 \cdot PA$
S	Steppe (pluie régulière)	B	PA entre 50 et 100% du seuil
W	Désert (sec et aride)	B	PA < 50% du seuil
T	Toundra	E	Température moyenne du MPC _{année} entre 0 et 10°C
F	Neiges éternelles	E	Température moyenne du MPC _{année} < 0°C
M	Neiges éternelles en altitude	E	Précipitations abondantes Hiver doux (Température moyenne du MPF _{année} > - 10°C)

Tableau D.3: Climats résultants de la classification secondaire

Symbol	Désignation
Af	Climat équatorial
Aw	Climat de la savane avec hiver sec
As	Climat de la savane avec été sec
Am	Climat de mousson
BS	Climat semi-aride (steppe)

Tableau D.3: Climats résultants de la classification secondaire (suite et fin)

BW	Climat aride, désertique
Cf	Climat océanique (tempéré chaud sans saison sèche)
Cw	Climat chinois (tempéré chaud avec hiver sec)
Cs	Climat méditerranéen (tempéré chaud avec été sec)
Df	Climat continental froid sans saison sèche
Dw	Climat continental froid avec hiver sec
Ds	Climat continental froid avec été sec (continental méditerranéen)
ET	Climat de la toundra
EF	Climat des inlandsis
EM	Climat subpolaire océanique

Classification tertiaire

Il s'agit de la dernière étape de la classification. Elle rajoute six nouvelles nuances basées sur les amplitudes de températures.

Tableau D.4: Classification tertiaire de Köppen-Geiger

Symb.	Désignation	Climats concernés	Description
h	Sec et chaud	B	Température moyenne annuelle > 18°C
k	Sec et froid	B	Température moyenne annuelle < 18°C
d	Hiver très froid	D	Température moyenne du MPF < - 38°C
a	Été chaud	C et D	Température moyenne du MPC > 22°C
b	Été tempéré	C et D	Température moyenne des 4 MPC > 10°C Température moyenne du MPC < 22°C
c	Été court et frais	C et D	Nb de mois avec température moy. sup. à 10°C < 4 Température moyenne du MPC < 22°C Température moyenne du MPF > -38°C

Au final, ce sont donc trente-deux zones climatologiques qui sont définies par la classification de Köppen-Geiger : Af, As, Aw, Am, BSh, BSk, BWh, BWk, Cfa, Cfb, Cfc, Csa, Csb, Csc, Cwa, Cwb, Cwc, Dfa, Dfb, Dfc, Dfd, Dsa, Dsb, Dsc, Dsd, Dwa, Dwb, Dwc, Dwd, ET, EF et EM.

Parmi ces trente-deux zones, trente sont représentées sur la Figure E.1 et trois d'entre elles couvrent le territoire québécois (Dfb, Dfc et ET).

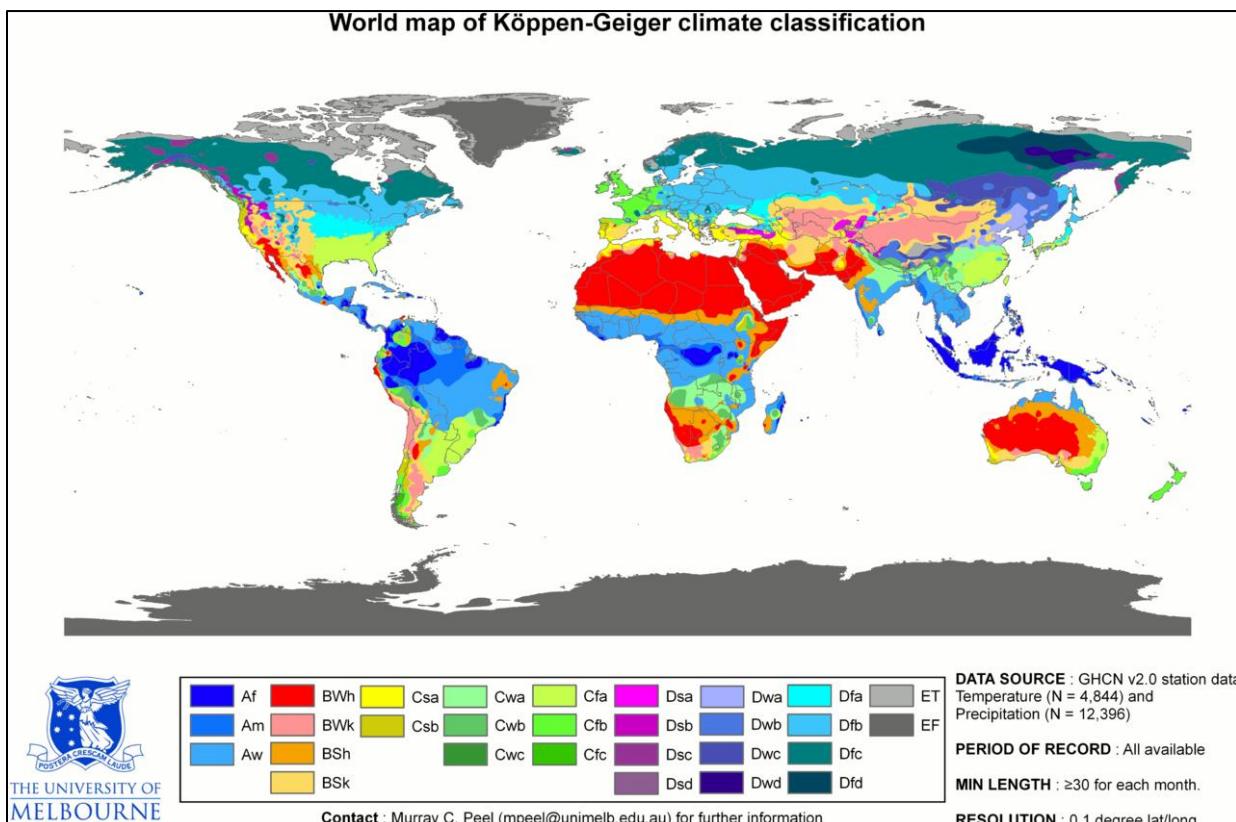


Figure D.1: Planisphère des différentes zones climatiques (*Université de Melbourne, AUS*)

ANNEXE E – CLASSIFICATION DE LITYNSKI

Principe de classification

Litynski met en place, en 1988, un modèle de classification climatique basé sur trois facteurs principaux, à savoir les températures, les précipitations et la durée de la saison de croissance. La saison de croissance correspond en fait à une période dont le début est fixé au jour, après le 1^{er} mars, pour lequel les températures quotidiennes moyennes des cinq jours suivants sont supérieures à 5°C. La fin de cette saison de croissance est définie comme la date, après le 1^{er} août, à partir de laquelle la température quotidienne moyenne devient inférieure à – 2°C. Dès lors, il est également possible de définir le nombre de degrés-jours de la saison de croissance comme étant la somme, sur une année, des températures quotidiennes moyennes supérieures à 5°C.

Application au Québec

L’obtention des différentes zones climatiques du Québec, par Gerardin et McKenney (2001), résulte d’un quadrillage de la surface du Québec (1 500 000 km²) en mailles de 2 km de côté et de l’interpolation spatiale des résultats obtenus à partir des stations météorologiques existantes.

Le principe de l’interpolation spatiale est relativement simple. Pour chaque maille du quadrillage réalisé, il est possible de connaître parfaitement trois coordonnées spatiales : la longitude, la latitude et l’altitude. À l’aide des données météorologiques fournies pendant plus de trente ans par les 396 stations du ministère et 150 stations supplémentaires installées pour l’occasion, il a été possible de définir, pour certaines mailles, neuf variables climatiques :

- ✓ Température annuelle moyenne (en °C) ;
- ✓ Amplitude journalière moyenne des températures (en °C) ;
- ✓ Température moyenne des trois mois les plus chauds (en °C) ;
- ✓ Température moyenne des trois mois les plus froids (en °C) ;
- ✓ Précipitations annuelles totales (en mm) ;
- ✓ Précipitations des trois mois les plus chauds (en mm) ;
- ✓ Précipitations des trois mois les plus froids (en mm) ;
- ✓ Durée moyenne annuelle de la saison de croissance (en jours) ;
- ✓ Nombre annuel de degrés-jours de la saison de croissance (en °C).

Des modèles mathématiques ont ensuite été mis en place pour exprimer ces neuf variables en fonction des trois coordonnées spatiales, ce qui a permis d'obtenir des valeurs pour les mailles non équipées de station météorologique.

À partir de ces valeurs et du modèle de classification, douze classes différentes de climat ont pu être définies au Québec (Tableau E.1).

Tableau E.1: Classes climatologiques du Québec (*D'après Gerardin et McKenney, 2001*)

Classe climatologique	Température	Précipitation	Saison de croissance
1	Polaire	Semi-aride	Très courte
2	Subpolaire froide	Modérée	Très courte
3	Polaire	Modérée	Courte
4	Polaire	Modérée	Très courte
5	Subpolaire froide	Modérée	Courte
6	Subpolaire froide	Subhumide	Courte
7	Subpolaire froide	Subhumide	Moyenne
8	Subpolaire	Humide	Courte
9	Subpolaire douce	Subhumide	Longue
10	Subpolaire	Humide	Moyenne
11	Modérée	Subhumide	Longue
12	Subpolaire	Subhumide	Moyenne

ANNEXE F – ÉQUATIONS PRÉDICTIVES DES RÉGRESSIONS

Pour chacune des équations de régression, les variables statistiquement significatives sont repérées par un astérisque.

ICG considéré comme continu

Régression linéaire multiple

Équation F.1 :

$$ICG = 4.563947 - 0.002247 \cdot \text{latitude} - 0.004988 \cdot \text{longitude} + 0.001515 \cdot \text{largeur HT} + 0.019025 \cdot \text{largeur carrossable}^* - 0.022095 \cdot \text{superf. tablier CR} - 0.015369 \cdot \text{âge}^* - 0.049160 \cdot \text{DJMA CR}^* + 0.165263 \cdot \% \text{ camions}$$

Méthodes de sélection des variables

Équation F.2 :

$$ICG = 4.8837103 + 0.020748 \cdot \text{largeur carrossable}^* - 0.015285 \cdot \text{âge}^* - 0.055801 \cdot \text{DJMA CR}^*$$

Régressions particulières

a) Analyse sur composantes principales ACP

Équation F.3 :

$$ICG = 4.474748 + 0.015916 \cdot F1 + 0.057811 \cdot F2 - 0.084837 \cdot F3 + 0.119073 \cdot F4 - 0.289956 \cdot F5$$

Soit

$$ICG = 4.474748 + 0.001848 \cdot \text{latitude}_{CR} + 0.006670 \cdot \text{longitude}_{CR} - 0.012643 \cdot \text{long. totale}_{CR} - 0.021285 \cdot \text{long. tablier}_{CR} + 0.005845 \cdot \text{largeur HT}_{CR} + 0.066997 \cdot \text{largeur carrossable}_{CR} -$$

$$0.008980 \cdot \text{superf. tablier CR} - 0.325937 \cdot \text{âge CR} + 0.025493 \cdot \text{DJMA CR} + 0.030735 \cdot \% \text{ camions CR} + 0.028740 \cdot \text{nombre de voies CR}^{13}$$

b) Régression ridge

Équation F.4 :

$$\begin{aligned} ICG = & 4.636132 - 0.000835 \cdot \text{latitude} - 0.001674 \cdot \text{longitude} + 0.001693 \cdot \text{largeur HT} + \\ & 0.013036 \cdot \text{largeur carrossable}^* - 0.015688 \cdot \text{superf. tablier CR} - 0.011039 \cdot \text{âge}^* - 0.022722 \cdot \\ & \text{DJMA CR} + 0.137085 \cdot \% \text{ camions} \end{aligned}$$

c) Régression PLS

Équation F.5 :

$$\begin{aligned} ICG = & 4.753045 - 0.005390 \cdot \text{latitude}^* - 0.004777 \cdot \text{longitude}^* + 0.001847 \cdot \text{largeur HT}^* + \\ & 0.016414 \cdot \text{largeur carrossable}^* - 0.025913 \cdot \text{superf. tablier CR}^* - 0.015487 \cdot \text{âge}^* - 0.037415 \cdot \\ & \text{DJMA CR}^* + 0.194842 \cdot \% \text{ camions}^* \end{aligned}$$

ICG considéré comme discret

Régression logistique binomiale

a) Détection d'un défaut de condition

Équation F.6 :

$$\begin{aligned} ICG_{\text{binomial}} = & [\exp(-7.79324 + 0.11241 \cdot \text{latitude}^* - 0.00830 \cdot \text{longitude} - 0.00814 \cdot \text{largeur HT} \\ & - 0.03677 \cdot \text{largeur carrossable}^* + 0.19787 \cdot \text{superf. tablier CR}^* + 0.04444 \cdot \text{âge}^* + 0.14548 \cdot \\ & \text{DJMA CR}^* - 0.96292 \cdot \% \text{ camions}^*)] / [1 + \exp(-7.79324 + 0.11241 \cdot \text{latitude}^* - 0.00830 \cdot \\ & \text{longitude} - 0.00814 \cdot \text{largeur HT} - 0.03677 \cdot \text{largeur carrossable}^* + 0.19787 \cdot \text{superf. tablier} \\ & \text{CR}^* + 0.04444 \cdot \text{âge}^* + 0.14548 \cdot \text{DJMA CR}^* - 0.96292 \cdot \% \text{ camions}^*)] \end{aligned}$$

¹³ Dans l'analyse ACP, toutes les variables sont prises en compte de manière "équivalente" ; il n'y a pas de variables non significatives.

b) Détection d'un bon état

Équation F.7 :

$$ICG_{binomial} = [\exp(7.79324 - 0.11241 \cdot latitude^* + 0.00830 \cdot longitude + 0.00814 \cdot largeur HT + 0.03677 \cdot largeur carrossable^* - 0.19787 \cdot superf. tablier CR^* - 0.04444 \cdot âge^* - 0.14548 \cdot DJMA CR^* + 0.96292 \cdot \% camions^*)] / [1 + \exp(7.79324 - 0.11241 \cdot latitude^* + 0.00830 \cdot longitude + 0.00814 \cdot largeur HT + 0.03677 \cdot largeur carrossable^* - 0.19787 \cdot superf. tablier CR^* - 0.04444 \cdot âge^* - 0.14548 \cdot DJMA CR^* + 0.96292 \cdot \% camions^*)]$$

Généralisation aux autres états

a) Détection d'un ICG égal à 4

Équation F.8 :

$$ICG_{binomial_4} = [\exp(-13.8184 + 0.2005 \cdot latitude^* - 0.0323 \cdot longitude^* - 0.0048 \cdot largeur HT + 0.0069 \cdot largeur carrossable + 0.1787 \cdot superf. tablier CR^* + 0.0218 \cdot âge^* + 0.0571 \cdot DJMA CR - 1.4946 \cdot \% camions^*)] / [1 + \exp(-13.8184 + 0.2005 \cdot latitude^* - 0.0323 \cdot longitude^* - 0.0048 \cdot largeur HT + 0.0069 \cdot largeur carrossable + 0.1787 \cdot superf. tablier CR^* + 0.0218 \cdot âge^* + 0.0571 \cdot DJMA CR - 1.4946 \cdot \% camions^*)]$$

b) Détection d'un ICG égal à 3

Équation F.9 :

$$ICG_{binomial_3} = [\exp(2.4686 - 0.1352 \cdot latitude^* - 0.0024 \cdot longitude - 0.0083 \cdot largeur HT - 0.0818 \cdot largeur carrossable^* + 0.0781 \cdot superf. tablier CR + 0.0362 \cdot âge^* + 0.1703 \cdot DJMA CR + 1.8312 \cdot \% camions^*)] / [1 + \exp(2.4686 - 0.1352 \cdot latitude^* - 0.0024 \cdot longitude - 0.0083 \cdot largeur HT - 0.0818 \cdot largeur carrossable^* + 0.0781 \cdot superf. tablier CR + 0.0362 \cdot âge^* + 0.1703 \cdot DJMA CR + 1.8312 \cdot \% camions^*)]$$

c) Détection d'un ICG égal à 2

Équation F.10 :

$$ICG_{binomial_2} = [exp (- 0.7505 - 0.0005 \cdot latitude + 0.0479 \cdot longitude - 0.0234 \cdot largeur HT - 0.0719 \cdot largeur carrossable^* - 0.3077 \cdot superf. tablier CR^* + 0.0557 \cdot âge^* + 0.3454 \cdot DJMA CR^* - 0.5283 \cdot \% camions)] / [1 + exp (- 0.7505 - 0.0005 \cdot latitude + 0.0479 \cdot longitude - 0.0234 \cdot largeur HT - 0.0719 \cdot largeur carrossable^* - 0.3077 \cdot superf. tablier CR^* + 0.0557 \cdot âge^* + 0.3454 \cdot DJMA CR^* - 0.5283 \cdot \% camions)]$$

Régression logistique multinomiale

a) ICG de référence pris égal à 5

Équation F.11 :

$$ICG_{multinomial_2} = [exp (- 2.8385 + 0.0348 \cdot latitude^* + 0.0436 \cdot longitude - 0.0275 \cdot largeur HT - 0.0921 \cdot largeur carrossable^* - 0.1416 \cdot superf. tablier CR^* + 0.0742 \cdot âge^* + 0.3992 \cdot DJMA CR^* - 0.8305 \cdot \% camions^*)] / [1 + exp (- 2.8385 + 0.0348 \cdot latitude^* + 0.0436 \cdot longitude - 0.0275 \cdot largeur HT - 0.0921 \cdot largeur carrossable^* - 0.1416 \cdot superf. tablier CR^* + 0.0742 \cdot âge^* + 0.3992 \cdot DJMA CR^* - 0.8305 \cdot \% camions^*) + exp (0.9534 - 0.0994 \cdot latitude^* + 0.0011 \cdot longitude - 0.0135 \cdot largeur HT - 0.1056 \cdot largeur carrossable^* + 0.2262 \cdot superf. tablier CR^* + 0.0564 \cdot âge^* + 0.2306 \cdot DJMA CR^* + 1.3908 \cdot \% camions^*) + exp (- 12.3304 + 0.1803 \cdot latitude^* - 0.0240 \cdot longitude - 0.0060 \cdot largeur HT - 0.0125 \cdot largeur carrossable^* + 0.2299 \cdot superf. tablier CR^* + 0.0345 \cdot âge^* + 0.0883 \cdot DJMA CR^* - 1.5668 \cdot \% camions^*)]$$

Équation F.12 :

$$ICG_{multinomial_3} = [exp (0.9534 - 0.0994 \cdot latitude^* + 0.0011 \cdot longitude - 0.0135 \cdot largeur HT - 0.1056 \cdot largeur carrossable^* + 0.2262 \cdot superf. tablier CR^* + 0.0564 \cdot âge^* + 0.2306 \cdot DJMA CR^* + 1.3908 \cdot \% camions^*)] / [1 + exp (- 2.8385 + 0.0348 \cdot latitude^* + 0.0436 \cdot longitude - 0.0275 \cdot largeur HT - 0.0921 \cdot largeur carrossable^* - 0.1416 \cdot superf. tablier CR^* + 0.0742 \cdot âge^* + 0.3992 \cdot DJMA CR^* - 0.8305 \cdot \% camions^*) + exp (0.9534 - 0.0994 \cdot latitude^* + 0.0011 \cdot longitude - 0.0135 \cdot largeur HT - 0.1056 \cdot largeur carrossable^* + 0.2262 \cdot superf. tablier CR^* + 0.0564 \cdot âge^* + 0.2306 \cdot DJMA CR^* + 1.3908 \cdot \% camions^*) + exp (- 12.3304 + 0.1803 \cdot$$

$$latitude^* - 0.0240 \cdot longitude - 0.0060 \cdot largeur HT - 0.0125 \cdot largeur carrossable^* + 0.2299 \cdot superf. tablier CR^* + 0.0345 \cdot \hat{age}^* + 0.0883 \cdot DJMA CR^* - 1.5668 \cdot \% camions^*)]$$

Équation F.13 :

$$ICG_{multinomial_4} = [exp (-12.3304 + 0.1803 \cdot latitude^* - 0.0240 \cdot longitude - 0.0060 \cdot largeur HT - 0.0125 \cdot largeur carrossable^* + 0.2299 \cdot superf. tablier CR^* + 0.0345 \cdot \hat{age}^* + 0.0883 \cdot DJMA CR^* - 1.5668 \cdot \% camions^*)] / [1 + exp (-2.8385 + 0.0348 \cdot latitude^* + 0.0436 \cdot longitude - 0.0275 \cdot largeur HT - 0.0921 \cdot largeur carrossable^* - 0.1416 \cdot superf. tablier CR^* + 0.0742 \cdot \hat{age}^* + 0.3992 \cdot DJMA CR^* - 0.8305 \cdot \% camions^*) + exp (0.9534 - 0.0994 \cdot latitude^* + 0.0011 \cdot longitude - 0.0135 \cdot largeur HT - 0.1056 \cdot largeur carrossable^* + 0.2262 \cdot superf. tablier CR^* + 0.0564 \cdot \hat{age}^* + 0.2306 \cdot DJMA CR^* + 1.3908 \cdot \% camions^*) + exp (-12.3304 + 0.1803 \cdot latitude^* - 0.0240 \cdot longitude - 0.0060 \cdot largeur HT - 0.0125 \cdot largeur carrossable^* + 0.2299 \cdot superf. tablier CR^* + 0.0345 \cdot \hat{age}^* + 0.0883 \cdot DJMA CR^* - 1.5668 \cdot \% camions^*)]$$

b) ICG de référence pris égal à 2

Équation F.14 :

$$ICG_{multinomial_3} = [exp (3.7868 - 0.1342 \cdot latitude^* - 0.0425 \cdot longitude + 0.0140 \cdot largeur HT - 0.0136 \cdot largeur carrossable^* + 0.3678 \cdot superf. tablier CR^* - 0.0178 \cdot \hat{age}^* - 0.1686 \cdot DJMA CR^* + 2.2213 \cdot \% camions^*)] / [1 + exp (3.7868 - 0.1342 \cdot latitude^* - 0.0425 \cdot longitude + 0.0140 \cdot largeur HT - 0.0136 \cdot largeur carrossable^* + 0.3678 \cdot superf. tablier CR^* - 0.0178 \cdot \hat{age}^* - 0.1686 \cdot DJMA CR^* + 2.2213 \cdot \% camions^*) + exp (-9.4970 + 0.1455 \cdot latitude^* - 0.0676 \cdot longitude + 0.0215 \cdot largeur HT + 0.0796 \cdot largeur carrossable^* + 0.3715 \cdot superf. tablier CR^* - 0.0397 \cdot \hat{age}^* - 0.3109 \cdot DJMA CR^* - 0.7364 \cdot \% camions^*) + exp (2.8335 - 0.0348 \cdot latitude^* - 0.0436 \cdot longitude + 0.0275 \cdot largeur HT + 0.0921 \cdot largeur carrossable^* + 0.1416 \cdot superf. tablier CR^* - 0.0742 \cdot \hat{age}^* - 0.3992 \cdot DJMA CR^* + 0.8305 \cdot \% camions^*)]$$

Équation F.15 :

$$ICG_{multinomial_4} = [exp (-9.4970 + 0.1455 \cdot latitude^* - 0.0676 \cdot longitude + 0.0215 \cdot largeur HT + 0.0796 \cdot largeur carrossable^* + 0.3715 \cdot superf. tablier CR^* - 0.0397 \cdot \hat{age}^* - 0.3109 \cdot DJMA CR^* - 0.7364 \cdot \% camions^*)] / [1 + exp (3.7868 - 0.1342 \cdot latitude^* - 0.0425 \cdot longitude +$$

$$0.0140 \cdot \text{largeur HT} - 0.0136 \cdot \text{largeur carrossable*} + 0.3678 \cdot \text{superf. tablier CR*} - 0.0178 \cdot \hat{\text{âge}}* - 0.1686 \cdot \text{DJMA CR*} + 2.2213 \cdot \% \text{ camions*}) + \exp(-9.4970 + 0.1455 \cdot \text{latitude*} - 0.0676 \cdot \text{longitude} + 0.0215 \cdot \text{largeur HT} + 0.0796 \cdot \text{largeur carrossable*} + 0.3715 \cdot \text{superf. tablier CR*} - 0.0397 \cdot \hat{\text{âge}}* - 0.3109 \cdot \text{DJMA CR*} - 0.7364 \cdot \% \text{ camions*}) + \exp(2.8335 - 0.0348 \cdot \text{latitude*} - 0.0436 \cdot \text{longitude} + 0.0275 \cdot \text{largeur HT} + 0.0921 \cdot \text{largeur carrossable*} + 0.1416 \cdot \text{superf. tablier CR*} - 0.0742 \cdot \hat{\text{âge}}* - 0.3992 \cdot \text{DJMA CR*} + 0.8305 \cdot \% \text{ camions*})]$$

Équation F.16 :

$$\text{ICG}_{\text{multinomial_5}} = [\exp(2.8335 - 0.0348 \cdot \text{latitude*} - 0.0436 \cdot \text{longitude} + 0.0275 \cdot \text{largeur HT} + 0.0921 \cdot \text{largeur carrossable*} + 0.1416 \cdot \text{superf. tablier CR*} - 0.0742 \cdot \hat{\text{âge}}* - 0.3992 \cdot \text{DJMA CR*} + 0.8305 \cdot \% \text{ camions*})] / [1 + \exp(3.7868 - 0.1342 \cdot \text{latitude*} - 0.0425 \cdot \text{longitude} + 0.0140 \cdot \text{largeur HT} - 0.0136 \cdot \text{largeur carrossable*} + 0.3678 \cdot \text{superf. tablier CR*} - 0.0178 \cdot \hat{\text{âge}}* - 0.1686 \cdot \text{DJMA CR*} + 2.2213 \cdot \% \text{ camions*}) + \exp(-9.4970 + 0.1455 \cdot \text{latitude*} - 0.0676 \cdot \text{longitude} + 0.0215 \cdot \text{largeur HT} + 0.0796 \cdot \text{largeur carrossable*} + 0.3715 \cdot \text{superf. tablier CR*} - 0.0397 \cdot \hat{\text{âge}}* - 0.3109 \cdot \text{DJMA CR*} - 0.7364 \cdot \% \text{ camions*}) + \exp(2.8335 - 0.0348 \cdot \text{latitude*} - 0.0436 \cdot \text{longitude} + 0.0275 \cdot \text{largeur HT} + 0.0921 \cdot \text{largeur carrossable*} + 0.1416 \cdot \text{superf. tablier CR*} - 0.0742 \cdot \hat{\text{âge}}* - 0.3992 \cdot \text{DJMA CR*} + 0.8305 \cdot \% \text{ camions*})]$$

Modèle général de régression

Équation F.17 :

$$\begin{aligned} \text{ICG} = & 7.103637 - 0.015769 \cdot \text{latitude} + 0.025099 \cdot \text{longitude} + 0.001712 \cdot \text{largeur HT} + \\ & 0.014375 \cdot \text{largeur carrossable*} - 0.040606 \cdot \text{superf. tablier CR*} - 0.014004 \cdot \hat{\text{âge}}* - 0.054999 \cdot \text{DJMA CR*} + 0.276220 \cdot \% \text{ camions} - 0.002061 \cdot C_{\text{type}, \text{à poutres}} - 0.286612 \cdot C_{\text{matériaux, acier-bois}*} + \\ & 0.1131655 \cdot C_{\text{matériaux, béton armé}*} + 0.226545 \cdot C_{\text{matériaux, acier}*} + 0.242247 \cdot C_{\text{matériaux, béton précontraint}*} + \\ & 0.020815 \cdot C_{\text{route, collectrice}} - 0.009141 \cdot C_{\text{route, nationale}} - 0.012263 \cdot C_{\text{route, régionale}} - 0.052644 \cdot \\ & C_{\text{route, accès ressources}} - 0.055840 \cdot C_{\text{route, local 3}} - 0.007915 \cdot C_{\text{route, autoroute}} - 0.060091 \cdot C_{\text{route, local 2}} + \\ & 0.290473 \cdot C_{\text{route, ressources}*} + 0.012288 \cdot C_{\text{région, Abitibi}} + 0.070187 \cdot C_{\text{région, Nord}} + 0.142614 \cdot \\ & C_{\text{région, Laurentides}*} - 0.437094 \cdot C_{\text{région, Centre}*} - 0.097055 \cdot C_{\text{région, Montérégie}} + 0.035903 \cdot C_{\text{région, Chaudière}} + \\ & 0.010774 \cdot C_{\text{région, Lanaudière}} - 0.220185 \cdot C_{\text{région, Gaspésie}} + 0.023758 \cdot C_{\text{région, Estrie}} + 0.132500 \cdot \\ & C_{\text{région, Mauricie}} - 0.069862 \cdot C_{\text{région, Capitale}} + 0.025718 \cdot C_{\text{région, Saguenay}} + 0.204394 \cdot C_{\text{région, Laval}} - \\ & 0.137976 \cdot C_{\text{région, Côte-Nord}} + 0.342065 \cdot C_{\text{région, Outaouais}*} + 0.108462 \cdot C_{\text{obstacle, cours d'eau}*} - 0.033990 \cdot \end{aligned}$$

$$\begin{aligned} & C_{obstacle,point\,d'eau} + 0.047732 \cdot C_{obstacle,route} + 0.027595 \cdot C_{obstacle,chemin\,de\,fer} - 0.035542 \cdot C_{obstacle,piste} - \\ & 0.000853 \cdot C_{climat,6} + 0.008662 \cdot C_{climat,7} - 0.196001 \cdot C_{climat,8} + 0.067831 \cdot C_{climat,9} + 0.097599 \cdot \\ & C_{climat,10} + 0.006607 \cdot C_{climat,11} + 0.014392 \cdot C_{parti,PLQ} - 0.056094 \cdot C_{parti,UN}^* \end{aligned}$$

ANNEXE G – APPLICATION AUX NOUVELLES OBSERVATIONS

Modèle général de régression

Tableau G.1: Application du modèle général

Référence pont	ICG _{observé}	ICG _{prédit}	ICG _{corrigé}	Résultat ?
1100	3	4,116822	3	Oui
3623	4	4,2184064	3	Non
3868	5	3,698071	2	Non
14460	5	4,494723	5	Non
14845	4	4,598617	4,562957	/
17792	5	5,076840	5	Oui
17921	5	5,019926	5	Oui
18160	5	4,975237	5	Oui
18161	5	4,975044	5	Oui
18180	5	5,260093	5	Oui
18219	5	4,917306	5	Oui
18534	5	5,312736	5	Oui
01079N	4	4,554271	4,554271	/
09760E	4	4,311726	4,311726	/
09760W	4	4,308754	4,308754	/
13167W	4	4,078530	2	Non
16858N	5	4,901441	5	Oui
16858S	5	4,914888	5	Oui
17074N	5	5,050656	5	Oui
17074S	5	5,050711	5	Oui
18219E	5	5,066490	5	Oui

Tableau G.1: Application du modèle général (suite et fin)

18219W	5	4,990643	5	Oui
18433N	5	5,103069	5	Oui
18433S	5	5,059954	5	Oui
18591N	5	5,052574	5	Oui
Prédictions exactes				17 / 21

Modèle logistique de régression

Tableau G.2: Application du modèle logistique de détection de défauts

Référence pont	Observation	Proba _{prédictive}	Proba _{corrigée}	Résultat ?
1100	1	0,585398	1	Oui
3623	1	0,574203	1	Oui
3868	0	0,809594	1	Non
14460	0	0,319233	0,319233	/
14845	1	0,204619	0	Non
17792	0	0,053357	0	Oui
17921	0	0,097958	0	Oui
18108	0	0,085884	0	Oui
18109	0	0,090474	0	Oui
18158	1	0,095564	0	Non
18159	0	0,095367	0	Oui
18160	0	0,097846	0	Oui
18161	0	0,098276	0	Oui
18180	0	0,059118	0	Oui
18219	0	0,076440	0	Oui

Tableau G.2: Application du modèle logistique de détection de défauts (suite et fin)

18534	0	0,050444	0	Oui
01079N	1	0,373887	0,364849	/
09760E	1	0,433310	1	Oui
09760W	1	0,435190	1	Oui
13167W	1	0,563024	1	Oui
16858N	0	0,080509	0	Oui
16858S	0	0,077743	0	Oui
17074N	0	0,096097	0	Oui
17074S	0	0,096074	0	Oui
18219E	0	0,052726	0	Oui
18219W	0	0,058241	0	Oui
18433N	0	0,076740	0	Oui
18433S	0	0,084977	0	Oui
18591N	0	0,061046	0	Oui
Prédictions exactes				24 / 27