POLYPUBLIE

Modélisation des temps de parcours sur un réseau routier à l'aide de données de véhicules flottants

Pierre Loustau

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

Date	2009
Type	Mémoire ou thèse / Dissertation or Thesis Mémoire de maîtrise/Master's thesis
URL de PolyPublie PolyPublie URL	https://publications.polymtl.ca/6663/
Conditions d'utilisation Terms of use	© Tous droits réservés / All rights reserved
Référence Citation	Loustau, P. (2009). Modélisation des temps de parcours sur un réseau routier à l'aide de données de véhicules flottants [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/6663/

Détails de la thèse
Thesis details

Directeurs de recherche Advisors	Catherine Morency, & Martin Trépanier
Programme Program	Génie industriel

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal
https://publications.polymtl.ca

UNIVERSITÉ DE MONTRÉAL

MODÉLISATION DES TEMPS DE PARCOURS SUR UN RÉSEAU ROUTIER À L'AIDE DE DONNÉES DE VÉHICULES FLOTTANTS

PIERRE LOUSTAU
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)
FÉVRIER 2009

© Pierre LOUSTAU, 2009.

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-53911-8
Our file Notre référence
ISBN: 978-0-494-53911-8

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire est intitulé :

MODÉLISATION DES TEMPS DE PARCOURS SUR UN RÉSEAU ROUTIER À L'AIDE DE DONNÉES DE VÉHICULES FLOTTANTS

présenté par : LOUSTAU Pierre

en vue de l'obtention du diplôme de : Maîtrise ès Sciences Appliquées

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

M. ADJENGUE Luc, Ph. D, président

Mme MORENCY Catherine, ing., Ph. D, membre et directrice de recherche

M. TRÉPANIER, Martin, ing., Ph. D, membre et codirecteur

M. GOURVIL Louis, ing., M.Sc.A., membre

À mes grands parents qui nous ont quittés,

ma famille et moi, durant ma vie montréalaise

et qui m'ont enseigné la rigueur et le sérieux.

À Aloys, ami d'enfance qui s'en est allé cet été.

REMERCIEMENTS

L'auteur du présent mémoire remercie en tout premier lieu, ses deux tuteurs, les professeurs Martin Trépanier et Catherine Morency pour leur constant soutien et leur patience. Le développement des travaux a été largement corrélé avec la qualité de leur encadrement. De plus, il remercie le Ministère des Transports du Québec pour leur aide et leur suivi des études.

L'auteur souhaite aussi remercier l'École Polytechnique et les départements de génie civil et mathématiques et de génie industriel pour la qualité des cours qu'il a suivi au cours de sa maîtrise. Il remercie de plus son école, Supélec, qui lui a donné l'opportunité d'effectuer sa troisième année de cycle ingénieur à Montréal.

Ensuite, il remercie toutes les personnes qu'il a rencontré, étudiants du laboratoire, amis d'école et amis du Rugby Club Montréal, et qui lui ont permis de profiter de la vie montréalaise et d'une ambiance de travail plus qu'appréciable.

L'auteur remercie enfin ses parents qui ont financé en particulier ses billets d'avion.

Enfin, l'auteur tient à s'excuser de parler de lui à la troisième personne.

RÉSUMÉ

Ce mémoire présente une analyse au moyen d'outils statistiques de bases de données de relevés de temps de parcours sur le réseau autoroutier de Montréal. Disposant de données odométriques recueillies par le ministère des Transports du Québec entre les années 1998 et 2004 sur diverses portions autoroutières du réseau montréalais, les objectifs de cette recherche sont en priorité : référencer et étudier les différents indicateurs de la circulation et les technologies utilisées, évaluer statistiquement l'état de la circulation autoroutière montréalaise, donner une potentielle évolution temporelle du réseau, proposer une méthode de collecte de données dans le cadre d'un futur échantillonnage, et enfin qualifier la fiabilité du réseau routier supérieur de la région de Montréal.

Les différents gestionnaires du trafic à travers le monde constatent actuellement une évolution dramatique du trafic autoroutier pour tous les types d'agglomération. Afin de pouvoir qualifier cette évolution et agissant sous la contrainte d'enjeux politiques et économiques, la notion de performance du trafic a été délaissée au profit de la notion de fiabilité. En effet, face à l'impossibilité de réduire la congestion, les gestionnaires cherchent à mieux appréhender et à savoir anticiper les causes et conséquences d'un incident sur le trafic. Afin de permettre cette nouvelle étude, de nombreuses technologies sont aujourd'hui disponibles et très souvent utilisées conjointement : technologies statiques (boucles à induction, radar, caméra), et technologies mobiles (odomètres, et plus récemment Bluetooth et téléphonie mobile).

Dans le cadre de ce projet de recherche, les données de véhicules flottants dotés de relevés odométriques recueillies par le MTQ, qui ont fait l'objet de différentes expérimentations préalables, sont exploitées de nouvelle façon ; les circuits

préalablement échantillonnés par le Ministère, ont été découpés en tronçons de longueur égale à savoir 1km. De fait, il a fallu valider les nouvelles bases de données de relevés de temps de parcours, en éliminant les observations aberrantes, et les relevés ne disposant pas de toutes les entrées. Ces dernières sont les suivantes : localisation, heure et date du relevé et valeur du temps de parcours.

Désormais disposant de bases de données cohérentes, une analyse statistique des relevés de temps de parcours du circuit 1 (autoroute 13, direction sud) a pu être effectuée. Les conclusions en ont été la mise en évidence de la signification de divers facteurs, à savoir, le mois de l'année, les conditions de la chaussée, la période durant laquelle le relevé a été effectué (AM ou PM). Cette analyse s'appuie en particulier sur les résultats d'analyse de la variance et propose de plus une interface afin de consulter et d'observer par contrôle de qualité, les relevés.

De plus, les distributions fréquentielles des relevés de temps de parcours semblent se correspondre pour plusieurs tronçons. Cette observation nous amène donc à considérer des groupes de tronçons dont les distributions fréquentielles de temps de parcours sont proches. Au moyen d'une analyse de corrélation appuyée par un groupement, plusieurs groupes ont pu être mis en évidence. De plus, nous avons choisi d'évaluer la variabilité, indicateur de fiabilité, des relevés de temps de parcours pour chaque tronçon.

Afin de satisfaire les demandes du ministère des Transports du Québec et grâce aux analyses précédentes, nous avons pu développer un échantillonnage en rapport avec la variabilité des relevés de temps de parcours. Compte-tenu de son aspect peu pratique, les résultats obtenus grâce au *clustering*, ont permis de modifier l'échantillonnage, pour réduire et mieux répartir le nombre d'échantillons nécessaires afin d'observer le trafic actuel. Après avoir appliqué l'algorithme de *clustering* sur tous les circuits, un

plan d'échantillonnage plausible a été développé prenant en compte les inégalités entre le nombre d'échantillons par circuit recueillis entre 1998 et 2004.

Puis, alors que la piste du *clustering* a été étudiée, nous avons pu constater la forme très particulière de la distribution des relevés de temps de parcours d'un circuit. En conséquence, c'est désormais une modélisation qui a été appliquée à cette même distribution. De celle-ci, nous avons pu alors simuler la moyenne et la variabilité des temps de parcours et développer plusieurs indicateurs : l'indicateur d'incidents, et l'indicateur Moyenne-Variabilité dont l'utilité est multiple. Selon des critères de bijectivité, cet indicateur peut être utilisé par le gestionnaire dans sa formulation mathématique et par l'usager de la route dans sa forme catégorielle.

Enfin, compte-tenu de ces derniers résultats, nous avons choisi d'étudier l'année 2004 face à l'analyse toutes années confondues afin de savoir s'il y a évolution critique de la circulation sur le circuit 1. Les conclusions semblent plutôt tendre vers la négative.

Les conclusions sont donc les suivantes : nous avons mis en évidence la signification de différents facteurs, proposé un plan d'échantillonnage pratique basé sur la variabilité des relevés de temps de parcours et sur l'algorithme de *clustering*, modélisé la distribution des relevés de temps de parcours pour un circuit, simulé moyenne et variabilité, puis développé deux indicateurs. Plusieurs perspectives sont alors envisageables. En effet, la modélisation a très vite montré ses limites pour certains circuits, et c'est donc vers un autre type d'analyse qu'il faut se tourner : ajouter de nouveaux facteurs tels que le type de voie, la présence ou non d'intersection, etc, puis appliquer de nouveau l'algorithme de *clustering* et ce sur l'ensemble des tronçons disponibles. Suivi d'une modélisation de la distribution fréquentielle des temps de parcours de chaque groupe, nous pourrons de nouveau simuler le trafic et évaluer les

valeurs de nos indicateurs, puis proposer un nouvel échantillonnage avec une redéfinition des circuits.

ABSTRACT

This research exposes a statistical analysis of travel time observations on critical portions of the Montréal road network. Whereas we have access to odometric data collected by the Ministère des Transports du Québec (MTQ) between 1998 and 2004 on several highway segments of the Montreal network, the required tasks are: review of existing traffic indicators and the technologies used to gather data, estimate statistically the state of the traffic on the highway network, assess the evolution of traffic on the network, propose a data collection methodology for future sampling of travel times on the road network and, finally, evaluate the reliability of travel times on the Montreal highway network.

All over the world, the different managers of the traffic have been seeing a dramatic evolution of the highway traffic, thus for all kinds of cities. Faced with increasing pressure on the network that make congestion reduction almost impossible, managers now aim at better assessing this evolution and the impacts of various types of incident to provide more reliable network with respect to travel times. In fact, managers are not trying to eradicate congestion anymore, but try to measure congestion and anticipate the causes and consequences of an incident on the traffic. To allow this new concept, several technologies are used and very often used at the same time: static technologies (induction loops, radar, camera), and mobile technologies (floating cars, and more recently, Bluetooth and Cellular).

For the current research project, floating car data collected by MTQ are processed in an innovative way. Sampled segments have been cut in equally-sized portions. Then, we had to make our data more accurate, by deleting abnormal observations, and observations with missing values. These values are spatial location, timestamp and date of the observation and travel time value.

Thanks to a reliable data collection, we applied a statistic analysis on the data of the first segment (highway 13th, direction South). The conclusions confirm the influence of several factors: month of the year, road conditions, period (AM or PM). This analysis was executed using analysis of variance, and proposes a graphical interface assisting the quality control of observed data.

Then, it appears that the frequency distribution of travel times seem similar for several network segments. That point leads us to consider groups of segments whose frequency distribution of travel times are similar. With a correlation analysis and a clustering algorithm, several groups have been exposed. Furthermore, we chose to evaluate the variability, reliability indicator, for each segment.

In order to complete the sampling requirement, and thanks to the previous analysis, we could compute a sampling frame based on the variability of travel times. Seeing the complexity of making such a sample, we used the results of the clustering algorithm and improve the future data collecting method by reducing the number of required samples. We applied this algorithm on every segment considering the differences of the number of actual samples collected per segment between 1998 and 2004.

Then, whereas the idea of clustering has been studied, we also noticed the singular form of the frequency distribution of travel times. Consequently, we applied a modelling to that distribution. From this modelling, we have been able to simulate the mean and the variability of travel times and develop several new indicators: risk of incident indicator, and Mean-Variability indicator. This last indicator can be used mathematically by the manager of the traffic or by the road user in his categorical form.

Finally, thanks to these last results, we have chosen to study the year 2004 in front of the results for all of the sampled years, in order to expose a critical evolution of the traffic on the first segment. The conclusions seem to tell us that no dramatic evolution of travel times has occurred on this segment.

The conclusions are the followings: we proved the influence of several factors, propose a sampling method based on the variability and the clustering algorithm, model the distribution of travel times, simulate mean and variability and develop new indicators. This leads to various applications. Since the modelling of the distribution is currently poor for some circuits, we will have to try a new method: add several factors to our data, like the type of way, the presence of crossing ways, etc, and apply the clustering algorithm on the entire collection of portions. Followed by a modelling of the frequency distribution of each newly created group, we will be able to simulate mean and variability, and give new values to the previously created indicators. Consequently, we will be able to propose a new sampling, with a new definition of segments.

TABLE DES MATIÈRES

DÉDICACE iv
REMERCIEMENTS v
RÉSUMÉ iv
ABSTRACT iX
TABLE DES MATIÈRES xiii
LISTE DES TABLEAUX xix
LISTE DES FIGURES xxii
LISTE DES ANNEXES xxvii
INTRODUCTION 1
CHAPITRE 1 : REVUE DE LITTÉRATURE 7
1.1 Problématique générale 7
1.2 La notion d'indicateur 9
1.3 Le suivi de la circulation 11
1.3.1 Les temps de parcours (T) 11
1.3.2 La vitesse (V) 11
1.3.3 Le débit (Q) 12
1.3.4 La densité (K) 12
1.3.5 Relation fondamentale 13
1.4 La congestion 13
1.4.1 Causes de la congestion 14
1.4.2 Indicateurs de la congestion 16

1.4.2.1 Les indicateurs de retard et de temps 17
1.4.2.2 Les indicateurs de vitesse 20
1.4.2.3 Les indicateurs de longueur et de saturation 20
1.4.2.4 Les indicateurs utilisant les débits 20
1.4.2.5 Les indicateurs de fiabilité 21
1.4.2.6 Autres travaux pertinents. 23
1.5 Technologies de collecte de données 24
1.5.1 Les systèmes de détection 25
1.5.1.1 Détecteur à tubes 25
1.5.1.2 Boucle à induction 27
1.5.1.3 Caméra et système de suivi 28
1.5.1.4 Radar 30
1.5.2 Les systèmes embarqués 31
1.5.2.1 Les véhicules taggués 31
1.5.2.2 Véhicules équipés de GPS 32
1.5.2.3 Données GPS commerciales 34
1.5.2.4 Téléphonie cellulaire 35
1.5.2.5 Relevés odométriques 37
1.5.2.6 Capteurs Bluetooth 38
1.5.3 Synthèse 41
CHAPITRE 2 : BASE DE DONNÉES 43
2.1 Synthèse de la base de données 44
2.2 Construction d'une géobase interrogative basée sur le concept de tronçons (MTQ) .46

2.3 Structure de la géobase et contenu 47
2.4 Synthèse 50
2.5 Utilisation des tronçons unitaires 53
CHAPITRE 3 : ANALYSE EXPLORATOIRE DES DONNEES BRUTES (CONTROLE DE QUALITE) ... 56
3.1 Contrôle de qualité 56
3.2 Distribution globale des temps de parcours 58
3.3 Analyse exploratoire d'un circuit 66
3.3.1 Présentation du circuit 1 67
3.3.2 Localisation des points chauds du circuit 1 68
3.3.3 Evolution temporelle des zones de congestion 70
3.3.4 Observation des temps de parcours selon le mois et l'année 73
3.3.5 Evolution d'indicateurs de fiabilité selon l'heure 75
3.3.6 Cartes de contrôle et relevés hors-limites 76
3.3.7 Observations selon les conditions de la chaussée 79
3.4 Synthèse 82
CHAPITRE 4 : *DATA MINING*, GROUPEMENT DE TRONÇONS POUR UN CIRCUIT ET ÉCHANTILLONNAGE 84
4.1 Analyse fréquentielle et *clustering* 84
4.1.1 Analyse de la corrélation 86
4.1.2 Groupement (Algorithme de *Clustering*) 90
4.1.2.1 Algorithme 90
4.1.2.2 Présentation des résultats 91
4.2 Étude de la variabilité 92

4.3 Synthèse 95

CHAPITRE 5 : ÉCHANTILLONNAGE 96

5.1 Aspects théoriques 96

5.2 Exemple pratique : cas du circuit 1 99

5.3 Échantillonnage à la semaine 101

5.4 Alternative d'échantillonnage 103

5.5 Échantillonnage sur tout le réseau montréalais 105

5.6 Synthèse 107

CHAPITRE 6 : DÉGAGEMENT DE LA CIRCULATION NORMALE 108

6.1 Présentation des distributions théoriques 108

6.1.1 La loi normale 109

6.1.2 Loi log-normale 109

6.1.3 Propriété de la loi normale 110

6.1.4 Tests sur la distribution normale 110

6.1.5 Test du χ^2 110

6.2 Démarche mathématique pour dégager la circulation normale 111

6.2.1 Principe mathématique 112

6.2.2 Optimisation 112

6.3 Résultats pour le circuit 1 113

6.3.1 Dégagement de la circulation normale 113

6.3.2 Pourcentage de relevés dans la fenêtre temporelle 115

6.4 Relevés en dehors de la fenêtre temporelle 120

6.4.1 Processus mathématique 120

6.4.2 Résultats pour le circuit 1, et problématique 120
6.5 Synthèse 122
CHAPITRE 7 : SIMULATION PAR CIRCUIT 123
7.1 Construction d'un arbre de décision 123
7.1.1 Concept mathématique 123
7.1.2 Résultats pour le circuit 1 124
7.2 Simulation 127
7.2.1 Rappel mathématique et variabilité 127
7.2.2 Simulation selon l'heure de départ 132
7.3 Synthèse 134
CHAPITRE 8 : DÉVELOPPEMENT D'INDICATEURS 135
8.1 Indicateur d'incidents 135
8.2 Indicateur Moyenne-Variabilité (IMV) 136
8.2.1 Présentation de l'indicateur 137
8.2.2 Résultats 138
8.2.2.1 Valeurs de l'IMV selon la période 138
8.2.2.2 IMV_c par créneau de 10 minutes 140
8.3 Exemple d'étude : évolution de la moyenne et de la variabilité pour l'année 2004, et indicateur d'incident 143
8.3.1 Évolution de la moyenne des temps de parcours pour 2004 144
8.3.2 Évolution de la variabilité des temps de parcours pour 2004 145
8.3.3 Pourcentage de relevés dans la fenêtre temporelle et pourcentage d'incidents graves pour 2004 146

8.3.4 Simulation de la moyenne et de la variabilité pour l'année 2004 durant la période AM 148

8.4 Synthèse 150

CHAPITRE 9 : CONCLUSIONS ET PERSPECTIVES 151

9.1 Conclusions 151

9.1.1 Synthèse 151

9.1.2 Limitations 152

9.2 Perspectives 152

9.2.1 Vers un nouveau groupement, et une vision à long terme du trafic 152

9.2.2 Vers une vision à court terme du trafic autoroutier montréalais 154

RÉFÉRENCES 150

ANNEXES 165

LISTE DES TABLEAUX

Tableau 1-1 Liste exhaustive d'indicateurs 19
Tableau 1-2 Synthèse des technologies de suivi et de détection des véhicules 41
Tableau 2-1 Nombre de fichiers disponibles pour chaque année et mois d'observation 45
Tableau 2-2 Liste des entrées du fichier "Tronçons" 48
Tableau 2-4 Liste des entrées des fichiers « Relevés-Tronçons\RelevésVF » 49
Tableau 2-6 Distribution des tronçons selon le nombre de circuits qui les impliquent 53
Tableau 3-1 Statistiques relatives aux créneaux horaires choisis 77
Tableau 3-2 Statistiques pour les tronçons 7 à 10, période du matin sans les relevés hors-limites 78
Tableau 4-1 Tableau de corrélation des fréquences AM des tronçons du circuit 1 87
Tableau 5-1 Echantillons nécessaires pour N grand 99
Tableau 6-1 Paramètres de la circulation normale du circuit 1 114
Tableau 6-2 Corrélation entre pourcentage et moyenne-écart-type-variabilité AM 119
Tableau 6-3 Corrélation entre pourcentage et moyenne-écart-type-variabilité PM 119
Tableau 6-4 Valeurs des paramètres des lois exponentielles 121
Tableau 7-1 Matrices transpositions des tronçons 1 à 4 du circuit 1 période AM 126
Tableau A-1 Statistiques descriptives 164
Tableau A-2 Fréquence des observations par année 165
Tableau A-3 Fréquence des observations par mois 166
Tableau A-4 ANOVA pour le tronçon et la période plus effet d'interaction 169
Tableau A-5 ANOVA selon l'année 170
Tableau A-6 Coefficients de la régression selon l'année 171
Tableau A-7 Moyennes et intervalles de confiances selon l'année 172
Tableau A-8 Calcul des coefficients de signification 172

Tableau A-9 Test de LSD selon l'année 172
Tableau A-10 Coefficients de corrélation 174
Tableau A-11 ANOVA selon la période et l'année avec effet d'interaction 175
Tableau A-12 ANOVA selon le mois 176
Tableau A-13 Calcul du coefficient de signification 177
Tableau A-14 Coefficients et intervalle de confiance par mois 177
Tableau A-15 Test LSD sur les moyennes par mois 178
Tableau A-16 Coefficients de corrélation 180
Tableau A-17 ANOVA selon les conditions météo 181
Tableau A-18 Coefficients de la régression et intervalles de confiance 182
Tableau A-19 Calcul des coefficients de signification 183
Tableau A-20 Test LSD 183
Tableau A-21 Coefficients de corrélation 185
Tableau A-22 Moyennes selon Mois par Année 186
Tableau A-23 Moyennes selon Mois par Année (suite et fin) 187
Tableau A-24 ANOVA Nested design 188
Tableau A-25 Moyenne et intervalle de confiance mois(année*periode) 188
Tableau A-26 Moyenne et intervalle de confiance mois(année*periode) (suite) 189
Tableau A-27 Moyenne et intervalle de confiance mois(année*periode) (suite) 190
Tableau A-28 Moyenne et intervalle de confiance mois(année*periode) (suite et fin) 191
Tableau A-29 ANOVA Mois(Année*Période) 192
Tableau A-30 Calcul du coefficient de signification 192
Tableau B-1 ANOVA des temps de parcours selon Période et tronçons 193
Tableau B-2 ANOVA Nested Design 193
Tableau B-3 ANOVA selon conditions de la chaussée 193
Tableau B-4 Groupement des tronçons du circuit 1 194
Tableau B-5 Groupement des tronçons du circuit 1 (suite et fin) 195
Tableau B-6 Variabilité des tronçons du circuit 1 196

Tableau B-7 Évolution du nombre d'échantillons nécessaires pour le circuit 1 période AM selon le nombre de relevés par tronçon initial 197
Tableau B-8 Proportions de relevés du tronçon 2 du circuit 1 période AM selon l'heure de départ 198

LISTE DES FIGURES

Figure 1 Diagramme de classe des entités rencontrées et déroulement des études (flèches)..... 5
Figure 2 Organisation du mémoire et démarche scientifique 6
Figure 1-1 : Répartition en % du nombre de causes de la congestion (FHWA, 2005), pour les agglomérations américaines 16
Figure 1-2 : Consultation en temps réel de l'état des bouchons de circulation sur le réseau parisien (http://www.sytadin.fr/) 21
Figure 1-3: Principe de la mesure par détecteur à tube 26
Figure 1-4 : Principe de la mesure par boucle à induction 27
Figure 1-5 : Principe du suivi et de la mesure par caméra 29
Figure 1-6 : Principe de la mesure par radar 30
Figure 1-7 : Principe du suivi par véhicule taggué 32
Figure 1-8 : Principe du suivi par GPS embarqué 33
Figure 1-9 : Principe du suivi par téléphone cellulaire 36
Figure 1-10 : Principe de la mesure par relevé odométrique 38
Figure 1-11 : Principe du suivi par capteurs Bluetooth 39
Figure 2-1 Carte géographique des relevés de temps de parcours (selon géobase du MTQ)..... 46
Figure 2-2 Exemple de découpage d'un circuit en tronçons de longueur variable 50
Figure 2-3 Distribution des longueurs des tronçons de longueur variable (240 tronçons uniques) 51
Figure 2-4 Exemple d'appartenance multiple du tronçon 1903 à cinq circuits différents 52
Figure 3-1 Interface de l'outil de consultation et de validation d'un parcours 57
Figure 3-2 Interface permettant le choix d'un trajet particulier à consulter 57
Figure 3-3 Diagramme espace temps du véhicule pendant tout le trajet 58
Figure 3-4 Distribution des écarts entre deux enregistrements consécutifs représentant la vitesse du véhicule entre ces deux points 58
Figure 3-5 Observations des moyennes de temps de parcours selon l'heure de départ, mise en évidence des pointes AM et PM de la circulation 60

Figure 3-6 Autres types de distribution selon l'année, le mois et la semaine 62
Figure 3-7 Temps de parcours par mois pour l'année 2000 62
Figure 3-8. Temps de parcours pour le mois de septembre à 17h 63
Figure 3-9 Temps de parcours pour le mois de septembre et la journée du vendredi 64
Figure 3-10 Temps de parcours pour le mois de janvier 2001, à 19h 65
Figure 3-11 Temps de parcours pour le mois de janvier 2001, le mercredi 65
Figure 3-12 Découpage en tronçons d'un kilomètre du circuit 1 A13 direction sud 67
Figure 3-13 Circuit 1, direction et intersections majeures 68
Figure 3-14 Evolution des temps de parcours par tronçons du circuit 1 69
Figure 3-15 Moyenne des temps de parcours du circuit 1 selon la période. 70
Figure 3-16 Evolution temporelle du temps de parcours moyen par tronçon entre 6h00 et 8h20 AM. 72
Figure 3-17 Moyenne des temps de parcours selon le mois 73
Figure 3-18 Evolution des temps de parcours pour le mois de février 2004 74
Figure 3-19 Evolution des temps de parcours pour le mois de juin 2004 74
Figure 3-20 Evolution des temps de parcours pour le mois de septembre 2004 75
Figure 3-21 Observations des *TRI* et *TTI* selon l'heure pour les tronçons 7 à 10 76
Figure 3-22 Carte de contrôle des relevés relatifs aux tronçons 7 à 10 pour les créneaux de 6h à 9h 78
Figure 3-23 Pourcentage de relevés rencontrés par condition de la chaussée par année pour le circuit 1 79
Figure 3-24 Temps de parcours moyens par conditions de la chaussée et par année pour le circuit 1. 80
Figure 3-25 temps de parcours moyen selon la condition de la route pour le circuit 1 80
Figure 4-1 Distribution de fréquences des observations du tronçon 1 du circuit 1 AM 85
Figure 4-2 Carte de corrélation des distributions fréquentielles des tronçons du circuit 1 88
Figure 4-3 Carte de corrélations et distribution des fréquences pour les tronçons 1_2, 1_12 et 1_18 89
Figure 4-4 Groupes de tronçons du circuit 1 AM 91

Figure 4-5 Distributions fréquentielles des centres des groupes 92
Figure 4-6 Variabilité AM le long du circuit 1 93
Figure 4-7 Variabilité et moyenne des tronçons du circuit 1 94
Figure 5-1 Taille de l'échantillon selon le facteur de qualité k =1 97
Figure 5-2 Taille de l'échantillon selon le facteur de qualité k =2 98
Figure 5-3 Taille de l'échantillon selon le facteur de qualité k =3 98
Figure 5-4 Nombre d'échantillons nécessaires pour le circuit 1 AM, k=2 100
Figure 5-5 Nombre d'échantillons nécessaires pour le circuit 1 PM 100
Figure 5-6 Evolution des temps de parcours par tronçon et par journée 102
Figure 5-7 Nombre d'échantillons par tronçon selon la journée 102
Figure 5-8 Répartition appropriée des observations selon le jour de la semaine sans groupement 104
Figure 5-9 Echantillonnage distributif du circuit 1 pour la période AM 106
Figure 6-1: Exemple de fonction de densité selon la loi normale 109
Figure 6-2 Fonction de densité de la loi log-normale 110
Figure 6-3 Modélisation de la circulation normale du circuit 1 114
Figure 6-4 Pourcentage de relevés dans la fenêtre temporelle par tronçon et selon la période 115
Figure 6-5 Evolution par tronçon du pourcentage de relevés dans la fenêtre par année 116
Figure 6-6 Evolution par tronçon du pourcentage de relevés dans la fenêtre par année période AM 117
Figure 6-7 Evolution par tronçon du pourcentage de relevés dans la fenêtre par année période PM 117
Figure 6-8 Evolution par tronçon du pourcentage de relevés dans la fenêtre par année septembre à décembre AM 118
Figure 6-9 Evolution par tronçon du pourcentage de relevés dans la fenêtre par année septembre à décembre PM 119
Figure 6-10 Calibrage d'une loi statistique sur la distribution observée de la partie droite 121
Figure 7-1 Corrélation des observations par tronçon du circuit 1 période AM 125

Figure 7-2 Temps de parcours moyens observé et simulé par tronçon du circuit 1 en période AM 129
Figure 7-3 Temps de parcours moyens observé et simulé par tronçon du circuit 1 en période PM 129
Figure 7-4 Temps de parcours moyens observé et simulé par tronçon du circuit 1 pour la journée entière 130
Figure 7-5 Variabilités observée et simulée par tronçon du circuit 1 en période AM 131
Figure 7-6 Variabilités observée et simulée par tronçon du circuit 1 en période PM 131
Figure 7-7 Temps de parcours moyen pour le circuit 1 selon l'heure de départ 133
Figure 8-1 Indicateurs d'incidents pour le circuit 1 136
Figure 8-2 Représentation des 4 différentes zones de l'IMV 139
Figure 8-3 Valeurs de l'IMV par tronçon du circuit 1 AM 139
Figure 8-4 Valeurs de l'IMV par tronçon du circuit 1 PM 139
Figure 8-5 Indicateur du circuit 1 période AM, 8h00 141
Figure 8-6 Indicateur du circuit 1 période AM, 8h10 142
Figure 8-7 Indicateur du circuit 1 période AM, 8h20 142
Figure 8-8 Évolution de la moyenne pour l'année 2004 période AM 144
Figure 8-9 Évolution de la variabilité pour l'année 2004 période AM 145
Figure 8-10 Pourcentage de relevés pour l'année 2004 période AM dans la fenêtre temporelle 146
Figure 8-11 Pourcentage de relevés correspondant à des incidents graves pour le circuit 1 période AM 147
Figure 8-12 Simulation de la moyenne pour AM 2004 (courbe (SAM2004) 148
Figure 8-13 Simulation de la variabilité pour AM 2004 (courbe (SAM2004) 148
Figure A-1 Histogramme des temps de parcours 165
Figure A-2 Moyennes des temps de parcours selon le tronçon 167
Figure A-3 Moyenne des temps de parcours selon la période 168
Figure A-4 Moyennes des temps de parcours selon le tronçon et la période 169
Figure A-5 Moyennes des temps de parcours selon l'année 171

Figure A-6 Résidus sur échelle normale 173
Figure A-7 Valeurs prédites vs résidus 173
Figure A-8 Résidus vs Résidus exclus 174
Figure A-9 Moyennes selon année et période 175
Figure A-10 Moyennes des temps de parcours selon le mois 176
Figure A-11 Résidus sur échelle normale 179
Figure A-12 Résidus vs résidus exclus 179
Figure A-13 Moyennes des temps de parcours selon le mois et la période 181
Figure A-14 Moyennes des temps de parcours selon les conditions météo 182
Figure A-15 Résidus sur échelle normale 184
Figure A-16 Résidus vs résidus exclus 184
Figure A-17 Moyennes selon mois par année 187
Figure A-18 Moyennes et intervalles de confiance Mois(année*période) 191

LISTE DES ANNEXES

ANNEXE A : ANALYSE STATISTIQUE DU CIRCUIT 1 AU MOYEN DE STATISTICA.....165

ANNEXE B : TABLEAUX..... 193

INTRODUCTION

Depuis plusieurs décennies, les systèmes de gestion et de contrôle de la circulation évaluent en continu l'état de la circulation sur leurs réseaux propres au moyen de nombreuses technologies et méthodologies. Les conclusions actuelles sont alarmantes, puisque, alors que le nombre et la sévérité des congestions sont en constante augmentation - quasi doublement du temps de parcours sur certaines portions autoroutières en l'espace de quelques années -, de nouveaux axes de circulation ainsi que les petites et moyennes agglomérations jusqu'alors préservés sont touchés par la congestion.

Ceci s'explique en partie par l'évolution du comportement humain : en effet, la demande a très fortement augmenté en raison de l'expansion des zones urbaines et de changements sociologiques, alors que l'offre, elle, augmente plus lentement. Il est de plus en plus difficile en zone urbaine de proposer de nouveaux axes autoroutiers, ou d'accroitre la capacité des axes existants.

La congestion, outre le fait évident qu'elle nuit aux déplacements de tout usager, est aussi un facteur important de pollution urbaine avec le dégagement important de gaz à effet de serre, et de perte financière, dans le transport des personnes et de marchandises. La congestion est par exemple un phénomène nuisible dans les 437 zones urbaines américaines étudiées dans les rapports du *Texas Transportation Institute*, ce qui provoque 4.2 milliards heures de retard et la consommation de 2.9 milliards de gallons d'essence. Le problème est qu'il n'existe pas de cause et de solution uniques aux problèmes de congestion (selon les publications du Texas Transportation Institute par Shrank et Lomax 2003, 2005, 2007). Les études comparées de 1982 et 2003 montrent une très nette augmentation du nombre d'heures de retard par voyageur ainsi que l'expansion des zones où les congestions sont observées : alors qu'en 1982, beaucoup de segments routiers échappaient à la congestion, quasiment tout le réseau semble subir des problèmes de congestion aux périodes de pointe. La congestion a donc pris une importance

considérable. Ces causes sont variées et très fréquemment intimement corrélées : goulots, incidents sur le service, travaux, système de contrôle du trafic, météorologie, fluctuations du trafic...

En raison de cette évolution critique de la circulation et dans un effort de développement durable, les centres de gestion ont dû abandonner l'objectif de réduire la congestion et d'améliorer la performance du réseau routier. En effet, les politiques actuelles ont amené un nouveau type de contrôle de la circulation : assurer la fiabilité des temps de parcours sur le réseau pour permettre à l'usager de mieux prévoir ses déplacements. Les décisions ne visent donc plus la minimisation des congestions et autres problèmes de la circulation mais plutôt l'appréhension des temps de parcours d'un trajet prédéfini. Cette idée s'intègre plus généralement dans le concept politico-économique de la livraison de biens. Il est certain que de la précision du moment de livraison d'un bien, découlent des économies conséquentes, alors que le retard ou même l'avance dans la livraison correspondent à des pertes financières.

Deux concepts subséquents à la performance et à la fiabilité sont à différencier : la variabilité et la fiabilité. Bien que les notions soient très proches l'une de l'autre, la fiabilité est une référence à la consistance du service, alors que la variabilité relève justement de l'inconsistance du service autoroutier. Alors que les causes de la congestion peuvent être énoncées (incidents, zones de travaux, système de contrôle, météorologie, évènements spéciaux, fluctuations de la demande, et manque d'ajustement de la capacité de circuits), l'objectif de la fiabilité est de donner une mesure consistante : cela passe donc par la prédiction et l'anticipation. Par exemple, l'impact de la météo sur la circulation peut être anticipé, la présence d'incident sur des portions de route à risque appréhendée, ainsi que la planification des travaux sur voies routières.

Les concepts statistiques employés ont donc évolué en même temps que la fiabilité était développée. Alors que les études actuelles se concentrent de plus en plus sur la variable « temps de parcours », les centres de gestion utilisent de nouveaux modèles mathématiques : *data mining* devant le nombre souvent trop important de données recueillies, réseaux de

neurones, analyse logistique, analyse stochastique et optimisation, corrélation entre circuits, etc.

C'est donc dans ce principe que s'inscrit ce mémoire. Disposant de données odométriques (relevés distances-temps) sur les principaux axes autoroutiers de la région montréalaise de 1998 à 2004, on cherchera à mettre en évidence plusieurs points de la circulation et de la congestion.

- Premièrement, la recherche d'une confirmation montréalaise d'une évolution critique de la circulation mondiale des métropoles au cours des années d'observation au moyen d'une analyse exploratoire simple des circuits qui permettra aussi de dégager les zones très congestionnées du réseau autoroutier montréalais.
- Ensuite, en considérant non plus l'entité « observation » mais l'entité « tronçon d'un circuit», une analyse de la distribution fréquentielle des relevés de temps de parcours sur un circuit donné permettra de comparer par corrélation et *clustering*, les différents tronçons de 1 km de ce même circuit. Ceci permettra par la suite de proposer un plan d'échantillonnage : celui-ci reposera alors sur une méthodologie de recueil de données systématique.
- Dans le but ultime alors de développer des indicateurs, en choisissant désormais l'entité « circuit », il sera nécessaire de qualifier statistiquement l'ensemble des relevés de temps de parcours de celui-ci. Par conséquent, toujours au moyen de la distribution fréquentielle des temps de parcours du circuit, sa modélisation statistique permettra par la suite de simuler le trafic sur le réseau, et d'observer les points chauds de la circulation. Plusieurs approches en découleront : simulation de la moyenne, simulation de la variabilité et courbes espace-temps d'un trajet sur un circuit.
- S'ensuivra la conception de plusieurs indicateurs, dérivés d'indicateurs existants, tels que la moyenne, la variabilité, le TRI (*Travel Rate Index*) et TTI (*Travel Time Index*), et un indicateur exposant le risque d'incidents sur un circuit durant une période donnée.

Pour souligner le développement d'indicateurs de fiabilité spécifiques à la région montréalaise, plusieurs approches statistiques seront simultanément exploitées. Tout d'abord, une analyse statistique descriptive complète de la base de données, avec mise en évidence des tendances lourdes, et confirmation géographique de points chauds temporels et spatiaux du réseau autoroutier. Cette première étude sera la base d'une étape de *data mining* de la base de données. Par la suite, alors que nous emploierons plusieurs facteurs dont nous espérons prouver la signification (facteurs catégoriels : types de tronçon, année, mois ou saison, jour de la semaine conditions météorologiques, périodes d'observation que nous allierons aux données des enquêtes O-D ; facteurs continus : heure de relevé), nous confirmerons l'utilité de plusieurs indicateurs dont ceux cités plus haut, et développerons des indicateurs conséquents aux analyses statistique et spatio-temporelle, et aux indicateurs précédents.

Comme dit dans la liste précédente, plusieurs entités sont considérées :

- L'entité « ensemble des observations » : c'est-à-dire l'ensemble des relevés sur un ou des circuits, sur un ou des tronçons ;
- L'élément « ensemble des observations d'un tronçon d'un circuit» qui est un élément de l'ensemble des tronçons sur **un** circuit ;
- L'entité « ensemble d'observations sur un circuit » qui est un élément de l'ensemble des circuits proposés dans les bases de données ;
- L'entité « ensemble des observations d'un tronçon » qui correspond à un élément de l'ensemble des tronçons de tous les circuits.

En clair, on se concentre dans un premier temps sur des généralités, puis après une réduction certaine des observations (nous passons de la totalité des observations à seulement quelques observations sur un tronçon), nous cherchons à agrandir notre ensemble d'étude.

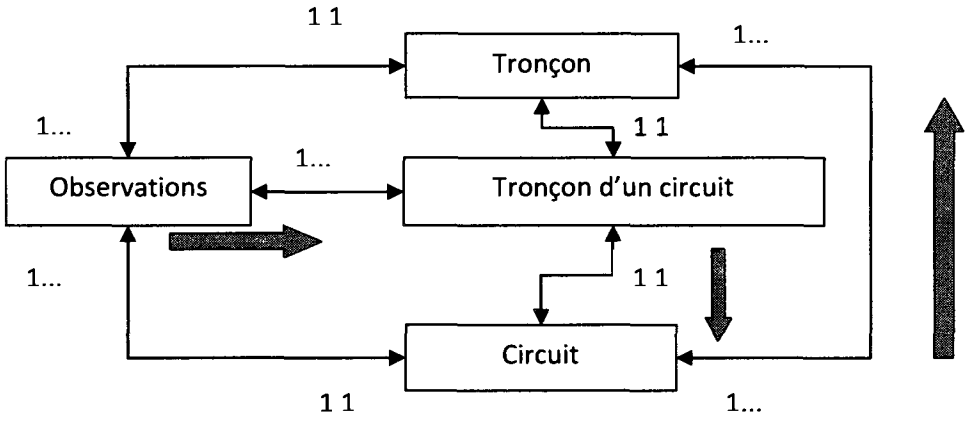


Figure 1 Diagramme de classe des entités rencontrées et déroulement des études (flèches)

Le but suivant sera donc de fournir un outil de consultation robuste et interactif pour analyser les données et donnant les informations nécessaires à l'utilisateur usager ou gestionnaire de service. Il sera donc possible d'anticiper le temps de déplacement avec un intervalle de variabilité, et d'appréhender « stochastiquement » pour le gestionnaire, les risques potentiels non prévisibles (incidents, fluctuations soudaine) de congestion.

Nous présentons dans la Figure 2 Organisation du mémoire et démarche scientifique, l'évolution des travaux ainsi que ses diverses branches. Ce graphe reprend donc en détail les différents chapitres de ce mémoire (cercles numérotés). Il demeure important de noter que cette analyse est un déroulement continu : compte -tenu de la nature des résultats que l'on peut obtenir à chaque partie de ce mémoire, il faut sans cesse chercher à confronter ou conforter les conclusions des diverses méthodes mathématiques. Notre démarche a été, hormis l'analyse exploratoire, d'augmenter l'ensemble de données que l'on analyse au fur et à mesure des analyses.

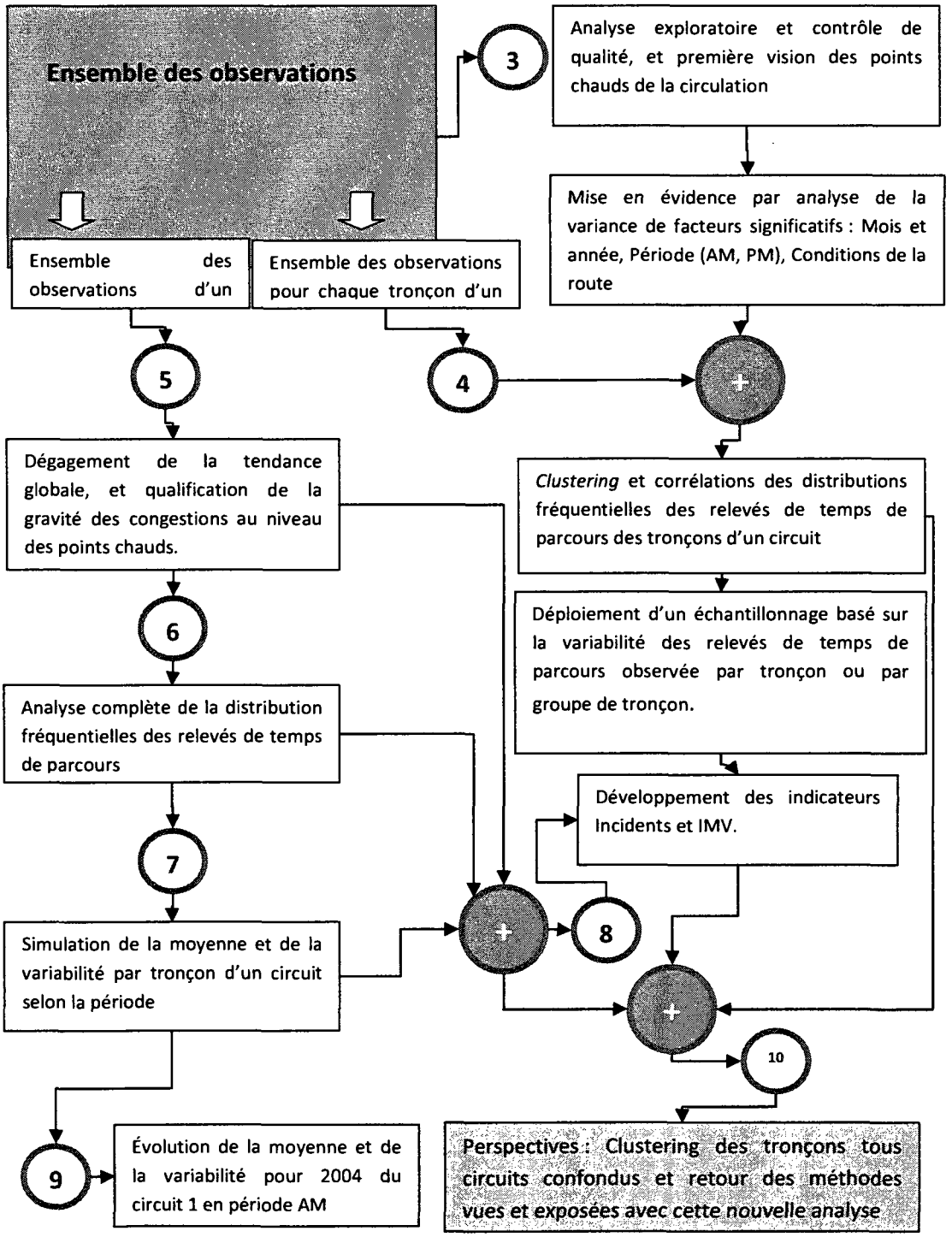


Figure 2 Organisation du mémoire et démarche scientifique

CHAPITRE 1 : REVUE DE LITTÉRATURE

Cette section vise d'abord à présenter les principaux concepts liés à l'étude de la circulation et, plus spécifiquement, de la congestion. Par la suite, en s'inspirant de différents travaux publiés sur des thématiques connexes, les principaux indicateurs actuellement utilisés pour mesurer ce phénomène ainsi que ses impacts sur la circulation sont présentés.

1.1 Problématique générale

> ***"Congestion is worse in areas of every size "***. (Shrank et Lomax 2007)

Mieux comprendre l'utilisation des réseaux routiers afin de mieux modéliser la congestion est un enjeu critique pour les planificateurs et professionnels responsables des réseaux de transport. Ceux du Ministère des transports (MTQ) ne sont pas en reste.

Des centaines d'études ont été publiées sur les impacts négatifs de la congestion et de nombreux travaux sont en cours. En outre, les préoccupations actuelles de développement durable font de plus en plus migrer les recherches vers une meilleure compréhension et modélisation des impacts indirects de la congestion, notamment les aspects environnementaux (Nesamani et al. 2007), les notions d'équité et d'accessibilité, voire même les modifications aux rythmes d'activités des personnes et ménages notamment la baisse des déplacements de type magasinage (Schmöcker et al. 2006). En parallèle, de nombreux travaux sont réalisés afin de clarifier le concept de développement durable lors de son application particulière au transport et à la mobilité. Litman (2008) fournit à cet égard une documentation très intéressante. Il propose en effet une série d'indicateurs de transport durable permettant d'apprécier, objectivement, les projets de transport et d'orienter le processus de planification des

comme indicateurs économiques de transport durable où une réduction est perçue comme une contribution au transport durable.

La nécessité de bien comprendre les problèmes de circulation, couplée à la disponibilité croissante de technologies, se traduit par une disponibilité grandissante de données diverses sur les mouvements des personnes et véhicules, certaines recueillies spécifiquement à cette fin, d'autres provenant de systèmes d'observation. Plusieurs recherches récentes portent par conséquent sur les méthodologies d'exploitation de ces ensembles de données pour modéliser les temps de parcours ainsi que sur la définition d'indicateurs permettant de décrire justement les conditions de circulation. Aux États-Unis, les travaux du Texas Transportation Institute (Schrank et Lomax 2003, 2005, 2007), proposent depuis une vingtaine d'années, différents indicateurs de l'état de la congestion dans les principales agglomérations du pays.

D'ailleurs, dans leur plus récent rapport, ces auteurs reconfirment certains éléments. La congestion est un phénomène nuisible dans les 437 zones urbaines américaines étudiées, ce qui provoque 4.2 milliards d'heures de retard et la consommation de 2.9 milliards de gallons d'essence. Le problème est qu'il n'existe pas de cause et de solution uniques aux problèmes de congestion. Pour ce plus récent rapport, les méthodes de collecte de données ont été revisitées, et, contrairement aux rapports précédents les zones urbaines de faible taille ont aussi été examinées. Selon ces mêmes auteurs, les causes de la congestion sont diverses : les voyageurs eux-mêmes (transport des personnes), le lent développement du réseau routier urbain, les incidents. Ces causes ont eu des impacts de plus en plus importants sur la circulation. Les études comparées de 1982 et 2003 montrent une très nette augmentation du nombre d'heures de retard par voyageur ainsi que l'expansion des zones où on observe de la congestion : alors qu'en 1982, beaucoup de segments routiers échappaient à la congestion, la donne diffère désormais, alors que tout le réseau semble subir des problèmes de congestion aux périodes de pointe. La congestion a donc pris une importance considérable.

Face à un tel problème ainsi qu'aux préoccupations grandissantes de développement durable, il semble que l'enjeu ne soit plus nécessairement de réduire les temps de parcours mais bien d'assurer leur fiabilité. Cette préoccupation de fiabilité du réseau se traduit par le

développement d'indicateurs tels que le « *Planning Time Index* » (PTI), qui évalue le temps nécessaire pour effectuer un trajet considéré dans les temps, 19 fois sur 20. Aux États-Unis, cet indicateur, tout comme le « *Travel Time Index* » (TTI), ont très fortement augmenté au cours des années en périodes de pointe (ces indicateurs sont définis dans les sections suivantes). Un usager doit donc prévoir plus de temps pour effectuer son trajet.

Dans le même esprit, le Département des transports des États-Unis (FHWA 2004) poursuit un effort continu de développement et de mesure d'indicateurs de la congestion à travers son programme de suivi, en continu, de l'état de la mobilité (« *Mobility Monitoring Program* »), lequel se base sur les capteurs électroniques existant dans une trentaine d'agglomérations pour observer l'évolution de la congestion. Une autre contribution importante est celle du Département des transports de la Pennsylvanie (Szekeres et Heckman 2005) qui s'est attaqué à la mise en place d'un système de gestion de la congestion (Congestion Management System – CMS) comme outil de planification des transports à l'échelle de l'État.

1.2 La notion d'indicateur

Selon l'*OCDE (2001), « un indicateur est un paramètre ou une valeur dérivée de paramètres donnant une information sur un phénomène »*. Litman (2008) propose, quant à lui, la définition suivante *: « Indicators are things we measure to evaluate progress toward goals and objectives ».*

Un indicateur est donc une donnée obéissant à des contraintes établies par le gestionnaire et visant à transmettre un maximum d'information en une seule valeur (cette valeur peut être de toute forme, un réel, un entier, un mot, ou bien encore une couleur).

Les indicateurs peuvent être séparés en deux selon le type de leur utilisateur (Ladier, 2005). Dans un contexte de suivi de la circulation, ceux-ci s'expriment comme suit:

- les indicateurs destinés à l'usager, qui ont généralement une qualité plus informationnelle;
- les indicateurs destinés au gestionnaire qui offrent des possibilités de gestion du trafic, permettant de prendre action.

Dans les deux cas, ces indicateurs doivent présenter certaines caractéristiques afin de les rendre valides et utilisables. La méthode de calcul des indicateurs doit correspondre aux critères suivants (Ladier, 2005):

- *robustesse :* comportement des indicateurs face au manque de données, ou à des données fausses;
- *précision* : marge d'erreur acceptable (en fonction de la précision des mesures sur les grandeurs observables);
- *fiabilité* : qualité qualifiant la procédure de calcul de l'indicateur.

Les données utilisées pour l'estimation des indicateurs doivent elles aussi présenter les propriétés suivantes :

- *actualité :* qui définit le temps qui sépare la mesure physique faite de l'utilisation effective, de l'observation;
- *disponibilité :* permettant de distinguer les indicateurs suivant le moment où l'indicateur est disponible;
- *période de calcul :* elle correspond à la période de variation du phénomène représenté dans l'indicateur;
- *durée de mise à jour :* soit la durée de calcul entre deux calculs de l'indicateur.

De ces qualités découlent alors les qualités propres à l'utilisation correcte d'un indicateur. Cette utilisation doit s'inscrire dans plusieurs idées :

- *praticité du recueil,* ou facilité de collecte de données;
- *injectivité* de l'indicateur, toute valeur de l'indicateur ne correspond qu'à un seul état de la circulation;
- *représentativité* et *la sensibilité* i.e. la valeur d'un indicateur ne doit pas dépendre du domaine de variation du phénomène;

- *praticité d'analyse,* à savoir le fait que l'indicateur doit être facile à comprendre et permettre de synthétiser un maximum d'information.

Lorsque tous ces critères semblent s'accorder, l'indicateur peut alors être validé. La validation d'un indicateur par un usager de la route prend souvent la forme d'un questionnaire où celui-ci expose son évaluation de la qualité de l'information fournie par l'indicateur; de la part du gestionnaire, c'est une comparaison avec la réalité qui est effectuée, ce qui exige souvent la collecte de nouvelles données.

1.3 Le suivi de la circulation

Plusieurs concepts sont impliqués dans le suivi et la mesure de la circulation notamment : le temps de parcours, la vitesse, le débit, la densité.

1.3.1 Les temps de parcours (T)

Les temps de parcours représentent les temps que met un usager pour se rendre d'un point initial à un point final. Les temps de parcours sont très informatifs ; en effet, l'observation de leur distribution sur un segment routier permet de mettre en évidence des phénomènes réguliers ou des incidents. Toutefois, ne juger la qualité du trafic que sur l'observation d'un temps de parcours moyen semble limité. Des indicateurs statistiques développés à partir de distributions de temps de parcours (ensemble d'observations) sont privilégiés.

1.3.2 La vitesse (V)

Tout comme les temps de parcours, la vitesse est une donnée très importante relatant de la qualité du trafic. De la même manière, l'observation des vitesses sur un circuit permet d'obtenir des informations pertinentes quant à la qualité du trafic.

Rappelons que nous distinguons plusieurs types de vitesse :

La vitesse instantanée, V_t qui correspond à la vitesse mesurée en un endroit précis pendant un moment donné, c'est une vitesse mesurée dans le temps. On déduit la vitesse moyenne dans le temps: $\overline{V_t} = \frac{\sum V_i}{n}$.

La vitesse de parcours, V_s qui est une vitesse dans l'espace : $v_s = \frac{\sum distances\ parcourues}{\sum temps\ de\ parcours} = \frac{\sum dis\tan ce(véh)}{\sum temps(véh)}$, la moyenne des vitesses de parcours de plusieurs véhicules est alors la **moyenne harmonique** des vitesses de parcours de ceux-ci.

1.3.3 Le débit (Q)

Le débit est un indicateur déduit directement par des mesures. Il est calculé à un endroit donné sur le réseau, pendant un temps donné (généralement une heure) ; il s'agit du nombre de véhicules traversant la portion considérée, pendant une heure.

On peut souligner l'existence d'une technique de calcul du débit basée sur la répétition du passage d'une voiture flottante dans un sens donné, qui compte le nombre de voitures qu'elle dépasse, qui la dépassent, et qu'elle croise lorsqu'elle effectue le retour. Cette méthode a été employée par Moran Toledo (2008), dans la région de Stockholm. Q est alors défini ainsi :

$$Q_1 = \frac{3600\left(N_{veh\ croisés\ sens\ 2} + N_{véh\ doublantes\ sens\ 1} - N_{véh\ doublées\ sens\ 1}\right)}{t_{trajet\ sens\ 1} + t_{trajet\ sens\ 2}}$$

Cette formule et cette méthodologie de comptage sont donc simples d'utilisation et confèrent des résultats pertinents.

1.3.4 La densité (K)

La densité est mesurée à un instant donné sur une longueur donnée (généralement 1 km, sinon on parle de concentration), et correspond au nombre de véhicules présents sur une portion à un moment donné(unité : véhicules par km).

1.3.5 Relation fondamentale

De par la définition de ces différents indicateurs, il vient une relation fondamentale :

$$Q = K \cdot V_s .$$

Ces premiers éléments donnent de très importantes indications quant à la qualité de la circulation. Cette circulation est sujette à des variations importantes au cours de la journée, au cours de la semaine ou bien du mois. Les divers travaux effectués sur l'observation, la simulation ou le maintien de la circulation urbaine, ont démontré que malheureusement ces indicateurs ne fournissent pas suffisamment d'information. Très souvent, des indicateurs statistiques plus complexes seront préférés.

1.4 La congestion

La congestion est un phénomène récurrent du trafic automobile. Un rapport de Cambridge Systematics, Inc. (2004) cite une étude de Schrank et Lomax (2003) qui confirme une augmentation de la congestion dans 75 régions américaines importantes entre 1983 et 2003. Les études de 2005 (Schrank et Lomax, 2005) et 2007 (Schrank et Lomax, 2007) viennent aussi confirmer ce fait. En fait, en 2007, ces auteurs indiquent que depuis 1982, « *congestion is worse in areas of every size* ».

L'usager de la route est donc confronté à des situations de conduite désagréables, dont la conséquence directe est le retard. La congestion est aussi une contrainte importante pour le transport de marchandises et peut mener à des répercussions néfastes sur l'économie.

Définition : La définition de la congestion a constamment évolué, d'une part parce que l'on a fait appel à l'usager pour la définir, d'autre part, parce que les systèmes de contrôle de la circulation ont évolué. On définit de manière générale la congestion selon deux seuils de vitesse : dans l'exemple tiré des travaux de Ladier (2005), un véhicule, quelque soit le type, est en situation de congestion sur une autoroute à vitesse limite égale à 100km/h, lorsque sa

vitesse descend en dessous de 25 ou 30 km/h et le demeure tant que sa vitesse n'atteint pas à nouveau 60 km/h.

La congestion peut elle-même être classée selon deux catégories (Robitaille et Nguyen, 2003) :

- La **congestion acceptable**, lorsque le temps de parcours est supérieur à celui encouru dans des conditions normales de trafic, et reste en deçà d'un maximum qui est défini par la collectivité.
- La **congestion inacceptable** qui correspond simplement au cas où le temps de parcours dépasse cette limite.

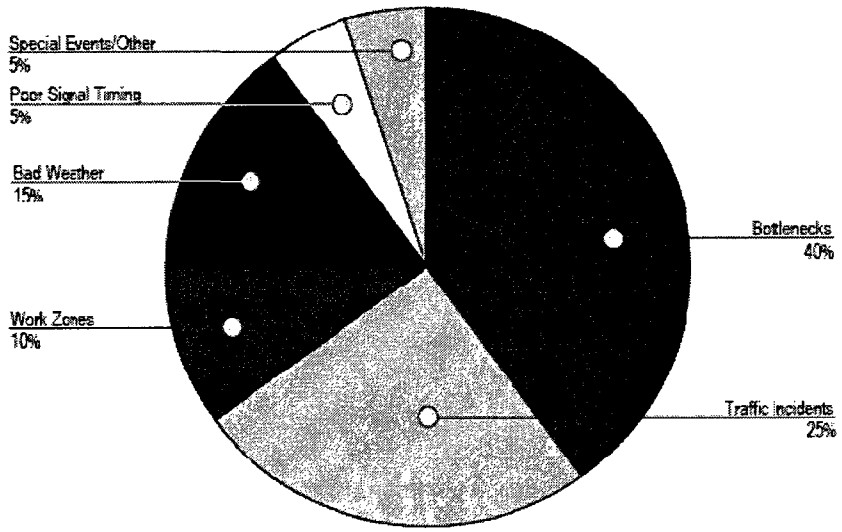
1.4.1 Causes de la congestion

Quelles sont alors les causes essentielles de la congestion? On peut référencer sept grandes causes principales des conditions de congestion (Laider, 2005, FHWA, 2005), dont la FHWA propose une répartition dans le cas américain (voir Figure 1-1).

- *Les goulots d'étranglements* (lorsque la demande est supérieure à la capacité du circuit). La capacité d'un circuit est le débit maximal possible sur une portion considérée. Celle-ci est calculée de manière théorique ou de manière empirique (observations) selon plusieurs facteurs : nombre de voies, entrecroisements contenus sur la portion, et sa topographie (lignes droites, virages, ...).
- *Les incidents du trafic* (évènements qui interrompent le trafic normal). Ce sont par exemple, les accidents, les pannes de véhicules, ou la présence d'éléments étrangers sur la route.
- *Les zones de travaux* (reliées aux activités d'entretien du réseau). Celles-ci ont un impact important sur la circulation, puisque des travaux peuvent conduire à la fermeture d'une ou plusieurs voies, à des bifurcations, la fermeture d'échangeurs, ou parfois la fermeture totale d'une portion.

- *La météorologie* (conditions climatiques et de la chaussée). Il est évident que des conditions climatiques critiques comme la chute de neige, la pluie voire le soleil, peuvent changer de manière brutale la conduite de tout usager. Schneider et al. (2007) proposent une étude de l'influence des conditions météorologiques sur la circulation. Ils utilisent des méthodes de *data mining* et les réseaux neuronaux, notamment les SOM (*Self-Organising Map*) pour modéliser cette relation.
- *Les systèmes de contrôle du trafic* (comme par exemple les sections à péage, ou en milieu urbain les feux de circulation). Ils ont un impact important sur la congestion et les variations des temps de parcours.
- *Les évènements spéciaux* qui impliquent des changements brutaux de la circulation sur des périodes et zones données.
- *Les fluctuations dans le trafic normal* basées sur les variations de la demande d'un jour à l'autre, ou d'une heure à une autre.

Figure ES.2 The Sources of Congestion
National Summary



Source: http://www.ops.fhwa.dot.gov/aboutus/opstory.htm.

Figure 1-1 : Répartition en % du nombre de causes de la congestion (FHWA, 2005), pour les agglomérations américaines

Nous venons donc de donner une définition possible de la congestion et d'en identifier les principales causes. Il est donc nécessaire désormais de développer des outils et des indicateurs permettant de quantifier et qualifier la congestion.

1.4.2 Indicateurs de la congestion

Plusieurs notions ont été sous-entendues dans l'énoncé des concepts précédents, comme le retard provoqué par une situation congestionnée, le temps de parcours dans une situation de trafic normal ou congestionné. Il est important de noter de plus que les différents centres de gestion de la circulation et professionnels du domaine développent des indicateurs ainsi que leurs propres outils pour décrire la congestion ; il en existe donc plusieurs.

Les sections qui suivent dénombrent et définissent plusieurs indicateurs recensés dans la littérature. Ils sont classés selon le concept-clé depuis lequel ils sont estimés. Ces indicateurs sont présents dans plusieurs sources notamment Ladier (2005) qui offre déjà un inventaire

intéressant, constitué depuis plusieurs sources notamment la HA (Highways Agency, Angleterre : http://www.trafficengland.com/TCC/), VicRoads (Australia : http://www.vicroads.vic.gov.au/Home), Texas Transport Institute (USA : http://tti.tamu.edu/), le rapport du NCHRP : Synthesis 311 (2003), le groupe de travail sur la LOLF (loi organique relative aux lois de finances, France : http://www.performance-publique.gouv.fr/). Pour le moment, il est impossible de dire si les données disponibles à Montréal permettront d'estimer certains ou l'ensemble des indicateurs présentés.

1.4.2.1 Les indicateurs de retard et de temps

- Retard : minutes perdues par rapport au temps en écoulement libre; Temps (minutes) supplémentaire requis pour parcourir un kilomètre par rapport à une situation normale (écoulement libre);
- Temps de parcours : temps requis pour parcourir un tronçon;

Van Lint et Van Zuylen (2005) estiment les temps de parcours selon l'heure de la journée (TOD, *time of day*), le jour de la semaine (DOW, *day of week*) ou bien le mois de l'année (MOY, *month of the year*). Ceux-ci deviennent des outils essentiels lors de la modélisation du trafic et de la congestion par reconnaissance de forme au moyen de réseaux de neurones. Les auteurs utilisent alors les 90^{e}, 50^{e} et 10^{e} percentiles, et considèrent deux indicateurs :

$$\lambda^{skew} = \frac{T_{90} - T_{50}}{T_{50} - T_{10}}, \lambda^{var} = \frac{T_{90} - T_{10}}{T_{50}}$$

Sont dégagées alors trois conclusions selon les valeurs de ces variables :

- pour $\lambda^{skew} = 1$ et $\lambda^{var} < 0.1$, le temps de parcours est fiable, et le trafic suit des conditions normales;
- pour $\lambda^{skew} << 1$ et $\lambda^{var} >> 0.1$, le trafic est congestionné, et plus λ^{var} est élevé plus les temps de parcours sont peu fiables;

- pour $\lambda^{skew} >> 1$ et $\lambda^{var} > 0.1$, la congestion s'installe ou se décompose, et donc des conditions de trafic normales et de grands temps de parcours peuvent être attendus; par contre, plus λ^{skew} est grand plus les temps de parcours sont peu fiables.

Des cartes de fiabilité des temps de parcours sont ensuite construites.

D'autres indicateurs recensés sont représentés dans le tableau suivant :

Tableau 1-1 Liste exhaustive d'indicateurs

« *Travel Delay* »	temps de parcours additionnel créé par la congestion
Taux de déplacement (*Travel Rate*)	inverse de la vitesse (temps de déplacement / longueur du tronçon)
Taux de retard	taux de déplacement réel – taux de déplacement acceptable défini comme le taux de déplacement en écoulement libre
Retard total	taux de retard * débit; nombre total d'heures perdues dans la congestion
Taux relatif de retard	Taux de retard (min / km) / taux de déplacement en écoulement libre(min / km)
Ratio de retard	taux de retard / taux de déplacement
TRI (*Travel Rate Index*)	rapport entre le temps additionnel et le temps requis en situation normale (écoulement libre)
TTI (*Travel Time Index*)	rapport entre le temps de parcours observé et le temps de parcours en situation normale
PTI (*Planning Time Index*)	Temps de parcours observé au 95^{e} percentile
Retard par véhicule	retard total / nombre de véhicules
Pourcentage du temps passé en situation de congestion	
Hkm	indicateur en heures*km correspondant au temps perdu dans les encombrements

1.4.2.2 Les indicateurs de vitesse

La vitesse des véhicules étant aussi une information pertinente de l'état du trafic, plusieurs indicateurs existent à partir de cette information.

- Vitesse observée : par classe de vitesses ou par rapport à une valeur moyenne annuelle;
- Vitesse moyenne générale ou pour différentes périodes critiques.

1.4.2.3 Les indicateurs de longueur et de saturation

Cette autre classe d'indicateurs correspond aux indicateurs utilisant l'information de la capacité et la topographie des voies.

- Taux de saturation, calculé à partir d'une capacité échelon de 2000 véhicules par voie et par heure;
- Route congestionnée et itinéraire congestionné (km): le premier indicateur correspond à la somme des portions en situation de congestion, le deuxième à la somme des portions congestionnées multipliées par leur débit respectif;
- Indice de mobilité par corridor : nombre de personne multiplié par la vitesse moyenne divisé par une valeur normalisée dépendant de la topographie du corridor étudié;
- Taux d'occupation des voies : indicateur selon quatre classes de condition de circulation : Intermittente (*Stop and Go*), Lourde (*Heavy*), Intermédiaire (*Moderate*) et fluide (*Wide-Open*).
- Facteur de blocage (*Jam Factor*) : indicateur calculé à partir de la vitesse moyenne et du temps de parcours ; il s'apparente à une échelle de Richter pour les conditions de circulation.

1.4.2.4 Les indicateurs utilisant les débits

Enfin, en utilisant toujours la relation fondamentale (1.3.5), plusieurs indicateurs utilisent l'information contenue par le débit de la circulation.

- Le cumul de bouchons : kilométrage de voies où l'on constate des bouchons dont la vitesse est descendue en dessous de 30 km/h (http://www.sytadin.fr/, Paris, France)

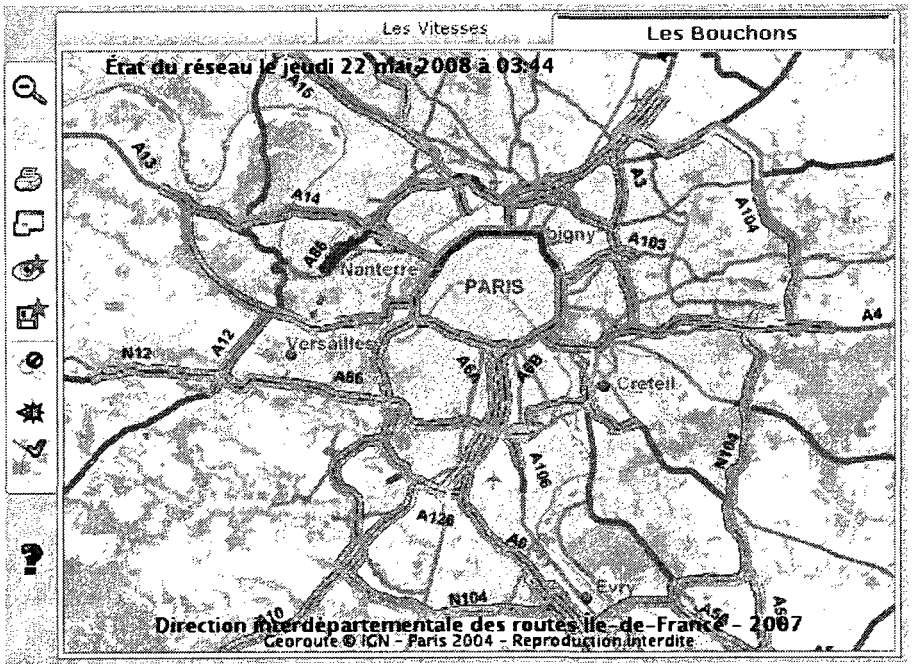


Figure 1-2 : Consultation en temps réel de l'état des bouchons de circulation sur le réseau parisien (http://www.sytadin.fr/)

- Indice de congestion des routes (RCI : *Roadway Congestion Index*) : rapport entre le débit total et l'offre du réseau routier;
- Véhicules-kilomètres parcourus (VMT) : nombre de véhicules dans le réseau multiplié par la longueur qu'ils parcourent.

1.4.2.5 Les indicateurs de fiabilité

Selon Lomax et al. (2003), deux concepts sont à différencier : la fiabilité et la variabilité. La fiabilité est une référence à la consistance du service, alors que la variabilité relève justement de l'inconsistance du service autoroutier. Nombreuses sont les causes de la congestion : incidents, zones de travaux, système de contrôle, météorologie, évènements spéciaux, fluctuations de la demande. La finalité des indicateurs de fiabilité est, dans ce contexte, de donner une mesure consistante du temps possible de déplacement étant donnée certaines

probabilités que la circulation soit affectée par certains phénomènes, récurrents ou non, qui exigent prévision ou anticipation. Par exemple, l'impact de la météo sur la circulation peut être anticipé, la présence d'incidents sur des portions de route à risque appréhendée, tout comme la planification des travaux sur voies routières.

Pour décrire la fiabilité du service, il est intéressant d'utiliser plusieurs indices : des mesures statistiques (fenêtre de temps de parcours, moyenne, écart-type, pourcentage de variation, indice de variabilité, indice de temps de parcours), des tampons (buffer time – différence entre 95^{e} percentile et moyenne des temps de parcours, *Buffer Time Index, Planning Time Index*) et des indicateurs de nombre de voyages retardés (par exemple la statistique de fiabilité du réseau en Floride égale à 100% - pourcentage de temps de parcours supérieurs à un seuil fixé, ou l'indice de malchance (MI), Shaw, T. et D. McLeod (1998)) .

Quelques éléments de discussion sont présentés ci-dessous :

- Tu et al. (2007) discutent du temps de parcours comme indice de fiabilité du réseau autoroutier. Dans cette étude le temps de parcours s'écrit par hypothèse

$$\widehat{TT} = t_f(1 + \alpha \left(\frac{q}{c}\right)^{\beta})$$

(il s'agit de la forme générale des courbes volume-délais), où t_f , q, c désignent respectivement le temps de parcours en écoulement libre, le débit, et la capacité du circuit. Les paramètres α, β diffèrent entre types de routes et selon l'état de la circulation (libre, état transitoire, congestionnée). Le temps de parcours est ici considéré comme une fonction du débit. Ces équations amènent donc deux nouvelles variables λ_1 et λ_2 – d'après les travaux de Brilon (1999) - qui caractériseront les états cités auparavant. L'estimation de ces paramètres avec des relations de continuité de la fonction TT s'effectue grâce à la méthode des moindres carrés. Dans un second temps, des études empiriques sur les villes de Delft et de Pékin exposent les relevés des mesures statistiques du temps de parcours: temps de parcours moyen, 10^{e} et 90^{e} percentiles. Les graphes mettent en évidence la supposition d'existence des paramètres λ1 et λ2.

- Mazloumi et al. (2007) proposent différents indicateurs de fiabilité du réseau, notamment : l'indice de **fiabilité du temps de parcours** (*TTR* : *travel time reliability*) : indice qui correspond au pourcentage des déplacements avec un temps de parcours (*TT*) supérieur au temps de parcours moyen en conditions normales (TT_{moy})

$$TTR = 100 - pourcentag\ e\ de\ voyage\ avec\ TT > TT^{moy}$$

- Selon Schrank et Lomax (2007), le manque de fiabilité des temps de parcours est un des principaux problèmes causés par la congestion. Ils proposent à cet effet d'utiliser le *PTI* (planning time index) qui est similaire au « Travel time index » mais qui fait plutôt référence au temps de parcours requis pour atteindre la destination à temps (selon le temps souhaité d'arrivée) 19 fois sur 20.

- **L'indice de malchance** (*MI*) et la **variabilité du temps de parcours** (*TTV* : Travel time variability) sont aussi utilisés comme indicateur de fiabilité des réseaux :

$$MI = \frac{TT_{moy}\ des\ 20\ plus\ longs\ voyages - TT_{moy}Global}{TT_{moy}Global}$$

$$TTV = \sqrt{\frac{\sum (TT_i - TT_{moy})^2}{n}}$$ où n est le nombre de déplacements

La liste d'indicateurs présentée dans cette section n'est pas exhaustive mais demeure suffisante. Cet inventaire alimente les réflexions tenues autour de l'exploitation des bases de données provenant des véhicules flottants à Montréal.

1.4.2.6 Autres travaux pertinents.

Il existe plusieurs travaux périphériques impliqués dans le processus de modélisation de la congestion :

Dans un contexte de collecte automatique de données de flot de circulation grâce à la disponibilité croissante de systèmes de cueillette, Qu et al. (2007) appliquent la méthode des

composantes principales pour réduire la quantité de données disponibles et permettre d'estimer des modèles robustes avec une quantité moindre de données.

Bernard et al. (2006) étudient l'autocorrélation entre les vitesses observées sur des tronçons en fonction de leur proximité et connexité. Leur étude s'appuie sur des données de temps de parcours recueillies à l'aide de véhicules flottants équipés de GPS. Les relevés ont été segmentés en tronçons de 500 m.

La corrélation entre les N vitesses observées $(v_{1i})_i$ et $(v_{1i})_i$ entre deux liens s'estime comme suit :

$$\rho = \frac{1}{N} \frac{\sum_{i=1}^{N} (v_{1i} - \overline{v_1})(v_{2i} - \overline{v_2})}{\sigma_{v1}\sigma_{v2}}$$

Où $\overline{v_1}$ *et* $\overline{v_2}$ désignent la moyenne des vitesses respectivement aux liens 1 et 2, et σ_{v1} et σ_{v2} sont les écarts-types des vitesses aux liens 1 et 2. Emam et Al-Deeik (2006) expérimentent et proposent des définitions et méthodes de calcul d'indicateurs de fiabilité des temps de parcours à partir d'une banque de relevés de détecteur double-boucles, dans un corridor autoroutier floridien.

Quiroga (2000) rapporte par ailleurs qu'un système de gestion de la congestion efficace ne peut se contenter de tels capteurs ponctuels de vitesse, mais doit être complété de lectures dynamiques effectuées sur le réseau. Il propose l'utilisation des traces GPS, tout comme Taylor et al. (2000). Les méthodes de correction des données GPS permettent aujourd'hui d'en arriver à un haut niveau de précision (Du et Aultman-Hall 2007).

1.5 Technologies de collecte de données

Durant les trois dernières décennies, un grand nombre de nouvelles technologies ont été développées, certaines liées directement à l'étude de la circulation. La disponibilité et la démocratisation de ces technologies se traduisent par une augmentation de la quantité de

données, de leur niveau de résolution ainsi que de leur précision. Ceci contribue favorablement au raffinement des mesures visant à décrire plus précisément les conditions de circulation.

Dans une perspective de mesure des temps de parcours sur un tronçon routier, Robinson et Polak (2004) identifient trois critères principaux dans le choix d'une technologie :

- Capacité à mesurer la vitesse de chaque véhicule avec une précision et une validité connues;
- Capacité d'identifier si un véhicule s'est comporté normalement (pas d'arrêt, etc.) entre le début et la fin d'un tronçon (identification des véhicules valides);
- Capacité de capturer un échantillon représentatif de la population totale de véhicules valides.

Ces critères devraient alimenter les réflexions quant au choix d'une technologie pour la collecte de données à Montréal.

L'objectif des sections qui suivent est d'exposer différentes technologies, actuellement utilisées ou en attente d'utilisation car encore au plan expérimental. Nous différencions deux grandes classes de technologies :

- les systèmes de détection : technologies utilisant des capteurs sur le bord des portions étudiées, et qui donnent en général des données macroscopiques sur le trafic;
- les systèmes embarqués : technologies localisées à bord d'un véhicule (ou sur un objet mobile) qui stockent et/ou transmettent des détails sur les traces spatio-temporelles des véhicules équipés.

1.5.1 Les systèmes de détection

1.5.1.1 Détecteur à tubes

Les compteurs-détecteurs à tubes permettent de compter le nombre d'essieux de véhicules qui passent à un endroit donné.

Principe de fonctionnement

Un ou plusieurs tubes de caoutchouc ou d'autre matière flexible sont placés en travers de la chaussée et reliés à un dispositif d'enregistrement. Lorsque qu'un pneu écrase le tube, l'afflux de pression est enregistré. Dans le cas de tubes dotés de capteurs piezzo-électriques (convertissant une pression en un signal électrique) des tables de classification peuvent être construites et utilisées pour identifier le nombre de véhicules qui ont traversé les tubes, car on doit tenir compte des camions et autres véhicules à essieux multiples. Certains détecteurs peuvent déterminer des vitesses de passage à l'aide de deux tubes disposés en parallèle.

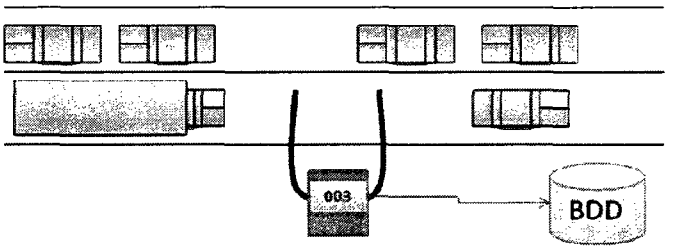


Figure 1-3: Principe de la mesure par détecteur à tube

Données récoltées

Bien que le principe soit ancien, les capteurs de pressions sont de plus en plus perfectionnés. Ils permettent entre autres de tenir compte des différences de pression dans les pneus et des erreurs dues aux tubes qui rebondissent. À la base, les capteurs enregistraient le nombre de passages, qu'il fallait ensuite diviser par un facteur (souvent 2, faute de données plus justes) pour déterminer le nombre de véhicules. Les nouveaux capteurs enregistrent chaque passage avec une lecture temporelle extrêmement fine et il est donc possible, avec un logiciel sophistiqué, d'estimer a posteriori le nombre de véhicules, pour différentes classes.

Limitations

L'utilisation de tubes comporte plusieurs limitations. Il arrive que la pression des pneus soit insuffisante pour que le passage soit enregistré. La trajectoire des véhicules doit être perpendiculaire afin que les roues ne soient pas comptées en double. Les tubes risquent de

rebondir sur la chaussée, ce qui fausse les données. La durée de vie des tubes étant limitée, le dispositif à tube est destiné aux prises de mesure temporaires.

1.5.1.2 Boucle à induction

La boucle de détection est un mécanisme qui permet de comptabiliser les passages de véhicules; la boucle à induction est un système simple, robuste et bien connu (Cohen et al., 2002).

Principe de fonctionnement

Une boucle à induction est un circuit électrique situé en dessous de la chaussée qui, au passage d'une voiture (donc, d'un corps métallique imposant), transmet un courant à un enregistreur. Certaines installations comptent deux boucles placées à courte distance l'une de l'autre. La Figure 1-4 illustre le concept.

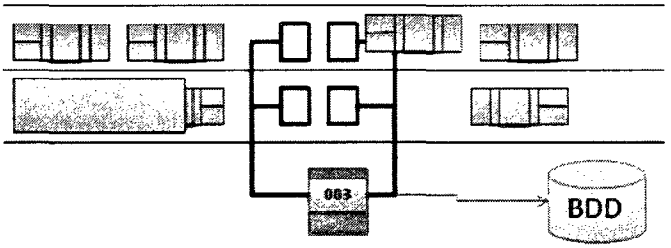


Figure 1-4 : Principe de la mesure par boucle à induction

Données récoltées

C'est une collecte continue de données qui a lieu. Deux passages successifs (donc deux véhicules différents) sur la même boucle peuvent être considérés comme le passage d'une seule et même voiture, et la vitesse doit alors être estimée à partir d'une valeur moyenne d'empattement. Il faut également tenir compte de la présence de camions pour calibrer les résultats. Avec une double boucle, il est plus facile d'estimer la vitesse car il suffit alors de calculer la différence de temps entre les deux lectures. Cependant, en circulation très dense, les lectures peuvent être faussées par la trop grande proximité des véhicules. L'étude consécutive des résultats permet ensuite d'évaluer les distributions de ces paramètres selon

l'heure, le jour et la position géographique. L'évolution d'une boucle à induction simple à une boucle à induction double, permet en particulier de différencier le type de véhicule qui traverse le circuit en mesurant « électriquement » la taille de l'essieu de celui-ci, en autant que le système soit calibré.

Limitations

Bien que les derniers modèles de détecteurs puissent mieux détecter la présence de camions, le calibrage et la classification des résultats demeure un problème majeur pour ces systèmes. Des mesures supplémentaires doivent être faites pour valider les résultats. L'installation des boucles requiert des travaux imposants, nécessitant souvent la fermeture de plusieurs voies de circulation lors de l'installation, sous la chaussée. D'autre part, les boucles sont sensibles aux fortes variations climatiques et aux conditions extrêmes de températures qui risquent de fausser les résultats. Les détecteurs doivent prendre des mesures à des fréquences très élevées sinon ils risquent d'échapper des lectures.

Exemples d'applications

Robinson and Polak (2005) modélisent le temps de parcours sur un lien à l'aide de données issues de boucles de détection. Ils utilisent une méthode de classification par proximité (k-NN) pour estimer le temps de parcours. La méthode est testée sur des données réelles de Londres (UK). Ils suggèrent de combiner ces données à des données de véhicules flottants équipés de GPS pour obtenir des estimations plus précises.

1.5.1.3 Caméra et système de suivi

Les caméras sont utilisées principalement pour surveiller l'état des routes et de la circulation. Elles peuvent cependant être utilisées pour le suivi des véhicules.

Principe de fonctionnement

La caméra est, en général, installée en hauteur. Elle permet d'obtenir une collecte quasi-continue de données. On peut alors effectuer un suivi de véhicules par reconnaissance de plaques d'immatriculation ou par analyse d'images avec des boucles virtuelles (Moran Toledo, 2008).

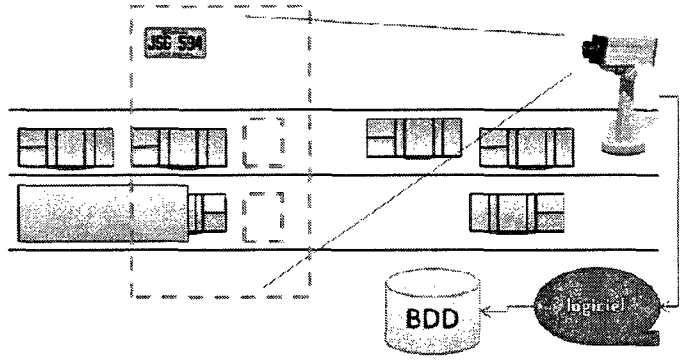


Figure 1-5 : Principe du suivi et de la mesure par caméra

Données récoltées

Le suivi de véhicules par reconnaissance de plaque minéralogique sur un tronçon donné, permet d'estimer les temps de parcours de ceux-ci, après avoir mesuré et enregistré les heures de passage et la direction de parcours à au moins deux endroits du réseau. D'autre part, il existe des logiciels qui permettent de "dessiner" des boucles virtuelles sur la chaussée. Par reconnaissance d'image, les véhicules sont détectés dès qu'ils traversent la boucle virtuelle, ce qui permet de récolter des données similaires à celles obtenues des boucles de détection. On obtient alors la vitesse instantanée (avec deux boucles), la densité et le nombre de véhicules.

Défauts du système

Cette technologie présente plusieurs limitations. La plus importante est que les systèmes vidéos ne peuvent mesurer que ce qu'ils voient. En d'autres termes, dans des conditions de faible visibilité (brouillard, manque de clarté, la nuit par exemple), les données recueillies sont peu fiables. De plus, de par leur situation, une caméra est plus soumise aux conditions météorologiques, ce qui nécessite des interventions d'entretien plus fréquentes, l'eau et la pollution dégradant considérablement la qualité des données. Les systèmes de comptage par boucles virtuelles nécessitent de plus une habile calibration, mais qui peut être faite par l'œil humain par observation.

1.5.1.4 Radar

La technologie radar a été développée toujours dans le but de compenser les défauts majeurs rencontrés par l'utilisation des technologies citées précédemment.

Principe de fonctionnement

Un détecteur radar est généralement un petit boîtier électronique. Ce dernier émet des ondes électromagnétiques qui seront réfléchies par les véhicules en mouvement. La variation des longueurs d'ondes entre onde émise et onde réfléchie (effet Doppler), ainsi que le temps mesuré entre émission et réception du signal sont directement recueillis.

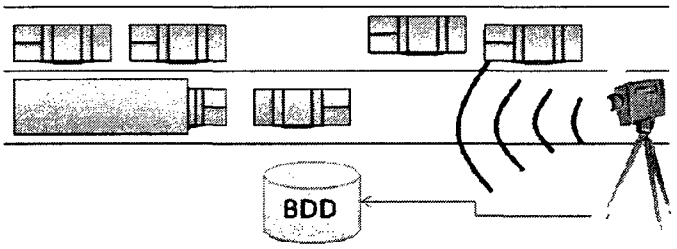


Figure 1-6 : Principe de la mesure par radar

Données récoltées

Nous pouvons constater que le radar permet de recueillir de nombreuses données : vitesse, débit (fiabilité moindre que dans le cas des technologies précédentes), distance.

De manière générale, les systèmes de détection ne fournissent des informations que sur une zone précise, malheureusement cette même zone peut ne pas être représentative de la qualité du trafic sur le reste du réseau. Il est clair, cependant, que la multiplication de points de détection, ainsi qu'une fréquence plus élevée de rencontre de ceux-ci sur le réseau, permettront de remédier de manière probante à ce problème. Toutefois, ceci nécessite un investissement important ainsi qu'une surveillance de l'état de fonctionnement des technologies plus accrue.

Limitations

Le radar possède cependant des défauts mineurs. D'une part, il recueillera des données moins fiables quant au volume de véhicules. D'autre part, le radar ne permet pas encore de différencier le type de véhicule observé. Toutefois, les recherches actuelles amènent à penser par reconnaissance de forme spectrale que le suivi d'une voiture est technologiquement et mathématiquement faisable. Cet outil est aussi sensible aux conditions climatiques, ce qui limite son utilisation.

1.5.2 Les systèmes embarqués

Alors que nous avons vu dans un premier temps les technologies de détection, il peut être intéressant d'utiliser des systèmes embarqués et, par exemple, de lancer dans la circulation des véhicules flottants qui recueilleront les données. On peut citer alors de nombreux travaux, d'autant plus que c'est un domaine, qui comme nous allons le voir, évolue fortement.

1.5.2.1 Les véhicules taggués

Les véhicules taggués agissent comme des observatoires locaux, ou des sondes de la circulation.

Principe de fonctionnement

Un capteur utilisant des communications dédiées à courte portée (DSRC), enregistre les passages de véhicules spécialement équipés d'émetteurs-récepteurs. En d'autres termes, dès qu'un véhicule passe devant le capteur, l'émetteur inclus dans le véhicule dialogue avec le capteur, en lui remettant les informations qu'il a recueillies.

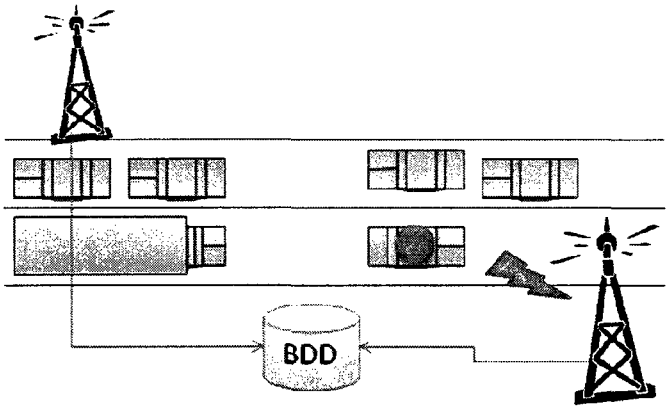


Figure 1-7 : Principe du suivi par véhicule taggué

Données récoltées

Cette technologie permet le calcul des temps de parcours entre deux stations fixes. Il s'agit généralement d'un échantillon de la population de véhicules spécifiquement dotés de la technologie, soulevant encore des préoccupations de représentativité de l'échantillon.

Limitations

Ce système présente quelques inconvénients. D'une part, la qualité des données dépend fortement de la situation géographique des capteurs, ce qui pourrait nuire par exemple à l'observation et à la description d'un phénomène de congestion. De plus, ce système ne se base que sur un échantillon de véhicules.

1.5.2.2 Véhicules équipés de GPS

Le GPS, dans l'utilisation faite pour le trafic, permet de calculer les mêmes paramètres que la technologie précédente ; cependant, ils permettent plus facilement la couverture d'une plus grande région géographique sans les coûts liés à l'installation de capteurs sur le réseau. Les données GPS ont été employées par exemple, sur les agglomérations de Montréal, dans les travaux de Robitaille et Nguyen (2003).

Principe de fonctionnement

Un GPS (*Global Positioning System*) est un système de positionnement géographique en temps réel, *i.e.,* lorsqu'un véhicule est équipé d'un récepteur GPS, sa position est calculée au moyen de plusieurs satellites - une précision fiable, de l'ordre de quelques mètres, est possible avec au moins quatre satellites -. En clair, la technologie GPS permet de suivre les déplacements d'un véhicule sur le réseau routier, l'enjeu étant d'assurer l'échange de cette information avec les contraintes du monde extérieur (tunnels, zone urbaine,...).

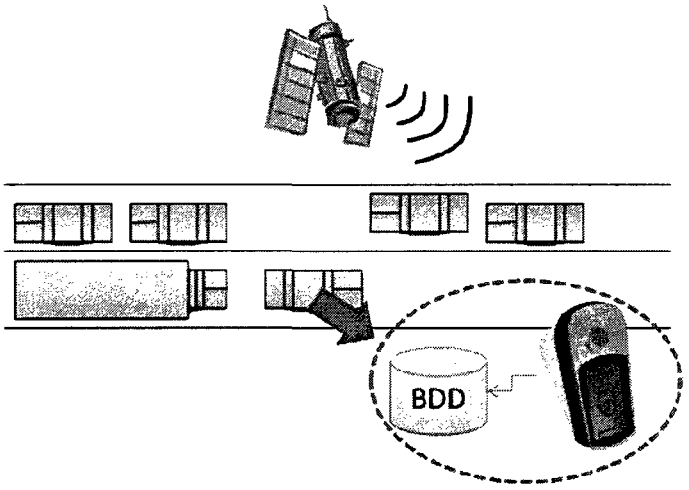


Figure 1-8 : Principe du suivi par GPS embarqué

Données récoltées

Le récepteur GPS récupère à intervalles réguliers les date et heure de lecture, la longitude, la latitude (et possiblement l'altitude du véhicule). Ces données brutes permettent alors d'estimer les temps de parcours sur des tronçons donnés et des itinéraires de déplacements, en autant que les localisations soient appariées sur un réseau routier analytique spatialisé dans un système d'information géographique.

Limitations

Il existe une forte contrainte liée à l'investissement. En effet, malgré la réduction constante des prix et l'augmentation du nombre de GPS, l'acquisition de telles technologiques représente

toujours un coût certain. Toutefois, dans l'optique d'un projet à long terme, cela peut s'avérer être très intéressant, car la collecte de données est alors peu coûteuse.

Deuxièmement, le récepteur GPS a besoin de 4 satellites pour déterminer une position fiable. En plus de pertes GPS, produites en raison de l'architecture du réseau (canyons urbain, tunnel,...), les données peuvent apparaître aberrantes. Des outils ont toutefois été développés pour remédier à ce genre de problèmes. On peut citer par exemple le « map-matching » (appariement) (Scott, 1994), où l'on modélise les erreurs de mesures comme un bruit blanc, que l'on cherche donc à filtrer.

1.5.2.3 Données GPS commerciales

Certains fournisseurs de services GPS embarqués vendent des ensembles de données, épurées des informations qui permettraient de reconnaître un individu ou un véhicule particuliers (retrait des données nominatives afin d'assurer le respect de la confidentialité).

Principe de fonctionnement

Plusieurs véhicules commerciaux disposent de capteurs GPS embarqués qui transmettent leurs données de façon synchrone ou non à des serveurs centralisés qui permettent aux gestionnaires des entreprises de faire la gestion de leurs flottes de véhicules (taxis, ambulances, flottes de camions, véhicules de services publics, autobus, etc). Les données récoltées sont des traces GPS plus ou moins détaillées spatialement, ce qui permet de déterminer des temps de parcours sur le réseau routier, une fois celles-ci appariées.

Limitations

Généralement, les données commerciales touchent des flottes de camions ou de véhicules de service ayant des comportements de vitesse et de temps de parcours possiblement différents des automobiles conduites par des particuliers. De plus, les itinéraires des camions ne couvrent pas nécessairement tous les axes du réseau avec suffisamment de représentativité. Enfin, les données achetées risquent d'être diluées ou réduites par le vendeur pour conserver l'anonymat de ses clients.

1.5.2.4 Téléphonie cellulaire

Gonzalez et al. (2008) utilisent les traces de 100 000 usagers anonymes de téléphones cellulaires (suivi pendant 6 mois) pour étudier les patrons de mobilité des individus. En outre, l'utilisation de traces cellulaires pour produire un portrait de la circulation fait l'objet de différentes études. Selon Wunnava et al. (2007), la technologie des traces cellulaires est viable et mature pour faire le suivi de la circulation en situation normale. Cependant, ce n'est pas le cas lors de conditions congestionnées alors que les données ne sont pas précises. En fait, la précision des données diminue avec l'augmentation du niveau de congestion alors que la criticité des mesures (indicateurs), elle, augmente.

Principe de fonctionnement

Un téléphone cellulaire comporte un composant émetteur-récepteur. Pour qu'une communication puisse s'effectuer, le téléphone doit être en dialogue avec l'antenne qui couvre une certaine aire, ou cellule. Cette même antenne est connectée en haut débit avec le reste du réseau téléphonique. Ainsi, une région géographique est morcelée en cellules et un téléphone portable appartient à la cellule dans laquelle le signal de l'appareil est le plus fort.

C'est donc au moyen de ce découpage que l'on peut effectuer un repérage géographique de l'appareil avec une précision entre 100 et 600 mètres. En effet, on peut visualiser le déplacement d'un véhicule, lorsque que celui change de cellules. Par extension, on peut également procéder à une triangulation des signaux pour déterminer une position plus précise.

Il apparaît que la quantité d'informations récupérable par les téléphones portables est très importante. Toutefois, cette technologie demeure encore à l'étape de validation - expérimentée par exemple sur les autoroutes du Sud de l'Allemagne, avec le réseau téléphonique *Vodafone* (Alger, 2004), quant à son utilisation pour décrire le trafic autoroutier. Elle a l'avantage d'être démocratisée : un très grand nombre de véhicules, ou plutôt de personnes localisées dans un véhicule, possèdent un téléphone cellulaire. En conséquence, une quantité très importante de données peut éventuellement être mise à disposition pour faire des analyses.

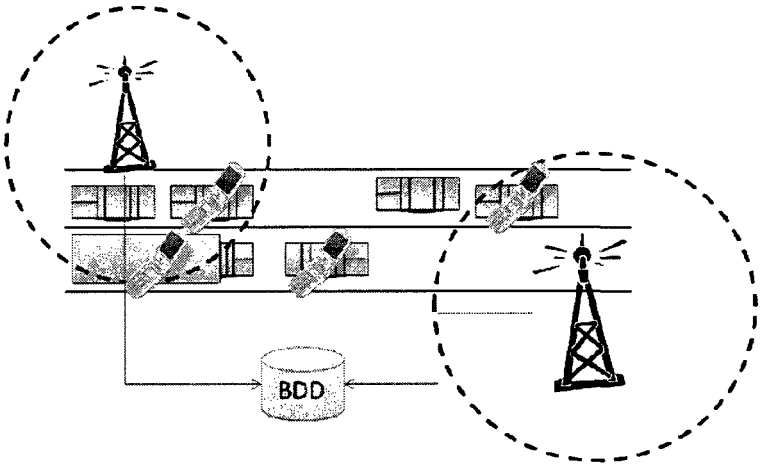


Figure 1-9 : Principe du suivi par téléphone cellulaire

Données récoltées

Similairement aux technologies présentées auparavant, notamment les véhicules tagués et les GPS, la téléphonie permet d'estimer les mêmes indicateurs. On peut noter les travaux d'Alger (2004) sur le calcul de la vitesse moyenne avec des fonctions de poids :

$\bar{v}(t) = \frac{\sum I_{(t-\delta t,t)}(t_i) \cdot v_i}{\sum I_{(t-\delta t,t)}(t_i)}$, où I désigne la fonction échelon, et t_i et v_i sont respectivement le temps et la vitesse estimée d'un véhicule à une cellule donnée.

$\bar{v}(t) = \frac{\sum e^{-(t-t_i)/v_i} \cdot v_i}{\sum e^{-(t-t_i)/v_i}}$, vitesse avec poids.

Limitations

D'un point de vue technologique, il semble que cette technologie produise un certain nombre de données aberrantes. Par exemple, il s'est avéré impossible de trouver sur le réseau allemand (Alger, 2004), trois cellules successives donnant des mesures corrélées. Il est donc nécessaire d'effectuer un traitement conséquent des données recueillies. De plus, l'imprécision est aussi un des inconvénients majeurs de la téléphonie cellulaire. En effet, dans des zones urbaines, la transmission d'un signal est compromise par la topographie du lieu : présence

d'immeubles, tunnels... ; le téléphone cellulaire est aussi très peu fiable en situation de congestion, phénomène que justement nous étudions. C'est la raison pour laquelle on préfère utiliser ce type de données uniquement pour des corridors routiers longue distance.

L'impossibilité de savoir précisément si l'usager est dans un véhicule privé, un véhicule de transport en commun ou simplement à pied ou à vélo à côté de la voie congestionnée est aussi un frein majeur à l'analyse.

Wunnava et al. (2007) identifient quatre enjeux principaux liés à l'utilisation de traces cellulaires :

- Le respect de la confidentialité des usagers des téléphones cellulaires dont les transmissions sont utilisées par les compagnies téléphoniques et vendues à différentes fins (question éthique);
- L'irrégularité des données (temps de parcours et vitesse), particulièrement en situation de congestion ou lorsque les conditions météorologiques sont mauvaises;
- Les capacités limitées des fournisseurs de données de temps de parcours d'être en phase avec les compagnies cellulaires en ce concerne le format et la structure des données (évolution);

L'incompatibilité des données entre fournisseurs potentiels.

1.5.2.5 Relevés odométriques

Les odomètres, technologie bien connue, sont employés actuellement dans la région montréalaise (Robitaille et Nguyen, 2003).

Principe de fonctionnement

Le capteur de données est mis en fonction manuellement, le système vérifie date et heure, ensuite il récupère les informations à différents rythmes (10 lectures par seconde par exemple). A chaque passage devant un repère, un son est émis; au repère final, un son différent est émis, moment auquel l'utilisateur suspens l'enregistrement des données.

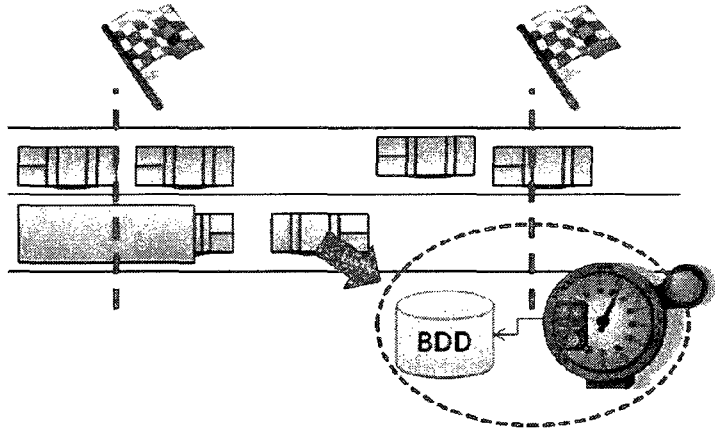


Figure 1-10 : Principe de la mesure par relevé odométrique

Données récoltées

Les circonstances spatio-temporelles (date, heure et localisation) sur le tronçon sont directement mesurées. Ces données permettent alors d'estimer les temps de parcours, les vitesses, ainsi que les délais occasionnés par la congestion.

Limitations

Se rapprochant du système du véhicule taggué, ce système présente le désavantage d'être semi-automatique. En effet, sa mise ne marche et son extinction dépendent implicitement de l'attention que le conducteur lui porte. Il est clair que par exemple, dans le cas d'une situation de trafic dense, le conducteur du véhicule flottant s'attachera plus à surveiller sa conduite qu'à mettre en route le système. Ceci implique donc des erreurs de mesure, non négligeables, dans la base de données.

1.5.2.6 Capteurs Bluetooth

Technologie nouvelle, l'emploi du capteur Bluetooth est donc récent. De premiers travaux commencent donc à émerger, et peuvent relater des possibilités de l'application de cette technologie. Des études de faisabilité sont désormais en cours. Nous pouvons citer par exemple les travaux de Philip J. Tarnoff, où la technologie Bluetooth® est utilisée dans la ville d'Indiana.

Principe de fonctionnement

Cette approche consiste à installer des capteurs de signaux Bluetooth® (émis par tout appareil disposant de cette technologie et dont la fonction est activée) aux abords des infrastructures routières. Le couplage des numéros d'identification des appareils ayant émis les signaux permet de déduire leur cheminement plausible sur le réseau et de dériver des temps de parcours et éventuellement des itinéraires cordons (d'un capteur à l'autre).

Les capteurs Bluetooth® échangent par des communications courte-portée sans fil des informations diverses dans une bande de fréquence de 2.45GHz. La portée d'un capteur peut varier entre 10 et 100 mètres. De plus un capteur Bluetooth® émet de manière continue.

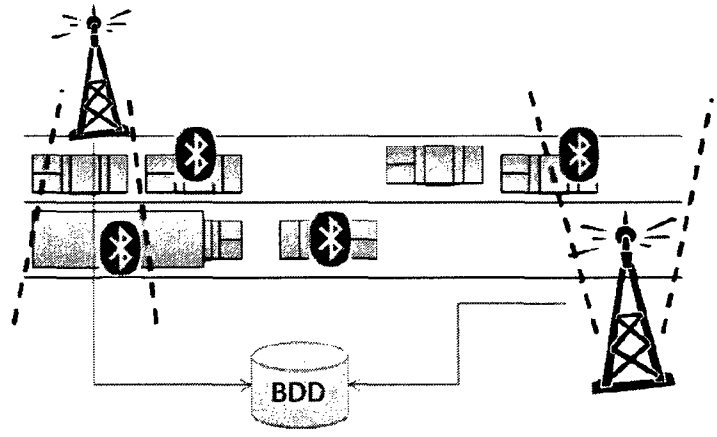


Figure 1-11 : Principe du suivi par capteurs Bluetooth

Données récoltées

Chaque lecture par un capteur permet de recueillir un numéro d'identification d'appareil ainsi qu'une heure de passage à un endroit précis (où pointe le capteur). Les lectures prises par tous les capteurs sont ensuite intégrées afin de reconstituer des données différentielles (vitesses, temps de parcours), connaissant la localisation spatiale des différents capteurs. Avec l'interdiction du combiné cellulaire au volant, l'utilisation des dispositifs Bluetooth dans les véhicules routiers devient commune.

Avantages

Les avantages issus de l'utilisation sont nombreux et compensent très largement les défauts de toutes les technologies citées auparavant (Philip J. Tarnoff) :

- C'est une technologie dont le prix est peu élevé :
- De très nombreux appareils utilisent cette technologie (ordinateur, gps, kits mains-libres, téléphones portables,...) ;
- Il n'y a pas de problèmes de confidentialité dans le cas où tous les échanges entre capteurs sont recueillis, les adresses des appareils étant alors inconnues ;
- Meilleure précision de mesure que tous les autres dispositifs ;
- Collecte de très nombreuses données, d'où implicitement, une meilleure vision des requêtes Origine-Destination.

Limitations

Les principales limitations sont liées à la rapidité des capteurs par rapport aux vitesses de passage des véhicules. De plus, les dispositifs Bluetooth risquent d'être captés plusieurs fois par seconde; il faut alors établir des règles pour déterminer l'heure exacte de détection. L'expansion rapide de ce type de technologie (ainsi que du WiFi) pourrait se traduire par une augmentation de l'interférence (utilisation d'une bande publique à 2.4 GHz) et une impossibilité de transmettre efficacement des données (Brookes, 2005).

Il existe de plus une erreur expérimentale mais qui reste bien moindre que celles que l'on peut trouver avec une autre technologie.

1.5.3 Synthèse

Le **Erreur ! Source du renvoi introuvable.** résume les technologies et les catégorise selon les critères décrits ci-après.

Tableau 1-2 Synthèse des technologies de suivi et de détection des véhicules

Technologie	**Indicateurs mesurés**	**Coût**	**Nb d'obs.**	**Représ. spatiale**	**Calibration**	**Analyse des données**
Tubes (au moins deux)	C (D) V	$	Élevé	Faible	Nécessaire	Légère
Boucles de détection (au moins deux)	C D V	$$	Élevé	Faible	Nécessaire	Légère
Caméras + suivi (au moins deux)	C D IC T V	$$$	Élevé	Faible	Nécessaire (boucles virtuelles)	Lourde
Radar	(C) (D) V	$	Moyen	Faible	Mineure	Légère
Véhicules taggués	IC T	$$	Peu	Faible	Non	Moyenne
Véhicules équipés de GPS	I T V	$$	Peu	Forte	Non	Lourde
Données GPS commerciales	I T V	$*	Moyen	Forte	Non	Lourde
Téléphonie cellulaire	I T	$*	Élevé	Forte	Non	Lourde
Relevés odométriques	T	$$$	Peu	Moyenne	Mineure	Moyenne
Capteurs Bluetooth (au moins deux)	IC T	$$	Moyen	Faible	Mineure	Moyenne

Indicateurs mesurés: C=compte, D=densité, I=itinéraire complet, IC=itinéraire cordon (à un point et un autre, sans connaître quels tronçons le véhicule a emprunté), V=vitesse instantanée, T=temps de parcours.

- Coût: bien que ce soit difficile à établir, donne une indication du coût nécessaire à l'installation et à l'opération de la technologie (* ne comprend pas les coûts d'acquisition des données).
- Nb obs.: nombre d'observations de mesure et de suivi possibles.
- Calibration: indique si la technologie demande une calibration afin d'établir des mesures. Les indicateurs récoltés dépendent alors des paramètres de calibration.
- Analyse des données: indique le niveau d'effort requis pour analyser les données.

CHAPITRE 2 : BASE DE DONNÉES

Propos recueillis depuis le devis de recherche émis par le Ministère des transports du Québec « évaluation des enjeux analytiques et des potentialités des données de temps de parcours pour mesurer la fiabilité du réseau routier montréalais », Projet R.612, 2007 :

Les préoccupations de mesure de la congestion se sont aussi, au Québec, traduites par la mise en place de processus de collecte de données. En effet, le MTQ a mis en place dans la région de Montréal, depuis 1998, un programme de relevés de temps de parcours et de mesure spatio-temporelle des files d'attente qui s'est poursuivi, bon an mal an, jusqu'à l'automne 2004. Un effort identique s'est déroulé dans la région de Québec de mars 2003 à février 2006. Les relevés, exécutés à l'aide de la méthode des voitures flottantes, ont chaque fois été remis au MTQ sous forme de base de données sur CD-ROM. Certains travaux ont été effectués à l'aide de ces données par la firme MIRO. Une restructuration de ces données, sous forme d'une géobase regroupant plus de 200,000 observations sur métasegments du réseau autoroutier montréalais, a été réalisée par le Ministère (Babin et al., 2007) afin d'en faciliter l'exploitation intégrée.

Parallèlement à cela, le MTQ s'interroge depuis plusieurs années sur les façons de mesurer systématiquement la congestion routière et son évolution dans la région de Montréal. Ainsi, des évaluations expérimentales des données de stations permanentes de mesure de la circulation sur les autoroutes Métropolitaine et Décarie ont été conduites (Whitehead et Fournier, 2000) et des propositions d'indicateurs ont ensuite été avancées (Tremblay et Fournier, 2001). Mais le manque de données réellement exploitables, autant en termes de qualité que de quantité, a empêché à l'époque tout développement concret d'indicateurs.

Le Ministère s'est alors tourné vers une autre approche complémentaire d'évaluation de la congestion, basée sur son modèle régional de simulation des transports pour la région de Montréal, pour produire deux études estimant les coûts de la congestion urbaine (ADEC, 1997,

Joubert, 2000, Gourvil et Joubert 2004) à partir des situations respectivement dérivées des enquêtes Origine-Destination de 1993 et de 1998. Au moment d'écrire ces lignes, le Ministère est encore en réflexion sur l'opportunité de republier un tel ouvrage sur la base des données associées à la plus récente enquête Origine-Destination régionale (2003), en raison notamment des sérieuses limitations constatées dans la précision et la représentativité de la méthode.

La problématique de la mesure de la congestion et de son évolution, de même que celle de la diffusion (en différé ou en temps réel) de cette information interpellent donc clairement le Ministère et requièrent de façon pressante de sa part un positionnement technique et stratégique.

Pour développer les outils et les indicateurs que nous voulons utiliser par la suite, il est nécessaire de revenir en détail sur la base de données de relevés de temps de parcours développée par le MTQ. Tout d'abord, celle-ci est le résultat de 6 différentes années d'observation (1998, 1999, 2000, 2001, 2002, 2004) odométriques (cf 1.5.2.5).

2.1 Synthèse de la base de données

Entre 1998 et 2004, le MTQ a procédé à une opération de collecte de données de temps de parcours sur certains tronçons du réseau routier montréalais à l'aide de la technique des voitures flottantes référencées. Dans la région de Montréal, 29 229 itinéraires ont été ainsi enregistrés sur une période de 7 ans (51 mois d'observation répartis entre 1998 à 2004, aucune observation en 2003). Le tableau qui suit résume la distribution de l'information disponible, dans le temps, en termes de fichiers de relevés.

Tableau 2-1 Nombre de fichiers disponibles pour chaque année et mois d'observation

Nb. Obs.	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre	Total
1998		528	561	544	564	546	559	563	532	557	556	548	6058
1999	525								555	567	573	557	2777
2000	545	533	543	533	556	556	562	563	556	565	563	310	6385
2001							553	620	562	579	567	598	3479
2002	594	602	580	530	521	642	595						4064
2003													
2004	451	1065	868	577	568	524	467	436	799	712			6467
Total	**2115**	**2728**	**2552**	**2184**	**2209**	**2268**	**2736**	**2182**	**3004**	**2980**	**2259**	**2013**	**29230**

Il existe 48 extensions de fichier différentes qui correspondent, selon notre compréhension, grosso modo à des numéros de circuits différents. Ces différentes informations seront validées ultérieurement.

La Figure 2-1 présente l'ensemble des circuits pour lesquels des relevés de temps de parcours ont été recueillis par le MTQ. Selon cette représentation géographique, approximativement 800 km de routes ont été échantillonnés.

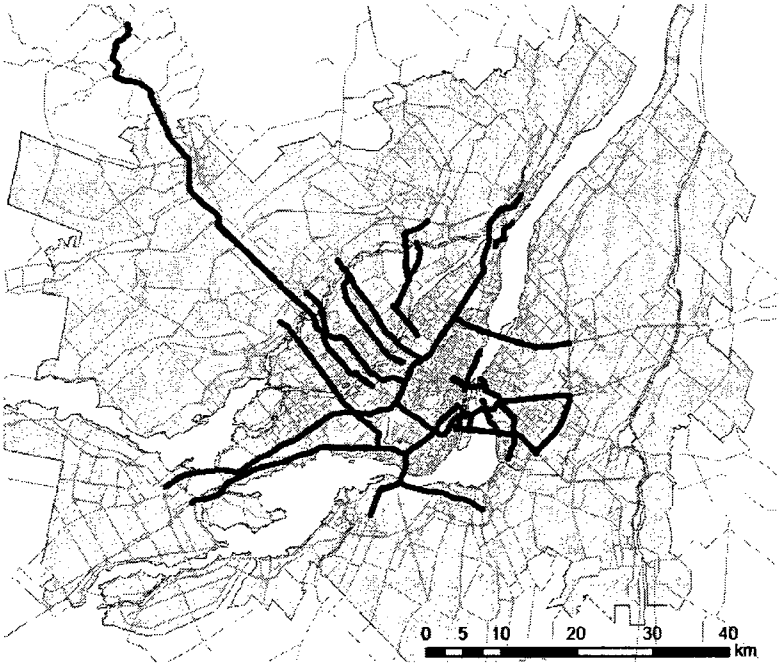


Figure 2-1 Carte géographique des relevés de temps de parcours (selon géobase du MTQ)

2.2 Construction d'une géobase interrogative basée sur le concept de tronçons (MTQ)

Afin de valoriser l'information contenue dans les fichiers de relevés, le MTQ a développé une géobase interrogative, visant à faciliter l'exploitation des données de voitures flottantes recueillies entre les années 1998 et 2004, et d'analyser statistiquement les observations réparties sur un total de 51 mois d'échantillonnage. Sur cet ensemble de relevés, 51 circuits directionnels ont été identifiés – en effet, certains trajets sont parcourus dans les deux sens, et sont donc considérés comme deux circuits distincts -. Toutefois, en raison de la modification de plusieurs circuits entre 1998 et 2004, 55 circuits directionnels sont plutôt identifiés, chacun doté d'un identifiant unique. En outre, suite à un processus de filtrage basé sur différentes règles, un total de 29 048 relevés sur les 29 229 sont retenus pour fins d'analyse (les 181 délaissés représentent des fichiers avec données manquantes).

Pour des fins d'analyse, le MTQ s'est appuyé sur le concept de tronçon unique, chaque circuit étant constitué d'un ou plusieurs tronçons. Ainsi, chaque circuit, hormis deux d'entre eux, a été découpé en tronçons. Chaque tronçon est délimité par deux bornes. Cette procédure a permis d'identifier un total de **296 tronçons**. Seulement, puisqu'un tronçon appartient parfois à plusieurs circuits, **240 tronçons uniques** sont requis pour reconstituer les **55 circuits**, auxquels on donne aussi un identifiant unique.

Enfin, une dernière étape de filtrage de la base de données est effectuée : seuls 27 193 trajets sont retenus, 1846 relevés ne vérifiant pas le critère suivant :

$$|\text{DistanceCircuit} - \text{DistanceReleve}| / \text{DistanceCircuit} < 2\%$$

ou

$$|\text{DistanceCircuit} - \text{DistanceReleve}| < 250 \text{ mètres.}$$

Ce filtrage est justifié, simplement pour obéir aux contraintes géométriques des circuits afin d'obtenir une base de données sans relevés aberrants.

2.3 Structure de la géobase et contenu

La géobase interrogative vise à faciliter la consultation des données de relevés. Chaque tronçon est identifié à par un numéro d'identification unique et peut être interrogé.

D'abord, les différents tronçons sont eux-mêmes définis selon les entrées suivantes dans le fichier « Tronçons » :

Tableau 2-2 Liste des entrées du fichier "Tronçons"

un numéro d'identification unique : 1 à 240
un identifiant numérique (NoCircuit*100+SeqTroncon), si le tronçon appartient à plusieurs circuits, on considère le plus petit numéro des circuits incluant le tronçon
un nom décrivant la situation géographique du tronçon
le nom du circuit considéré
la longueur du tronçon
le type de route du le tronçon
le nombre de voies
la liste des numéros de circuits empruntant ce tronçon
les identifiants du tronçon pour chaque circuit le contenant

Un fichier de données par relevé est développé : chaque fichier « Relevés-Tronçons\RelevésVF » contient 240 sous-fichiers de données relatifs par conséquent à chaque tronçon. Dans un des sous-fichiers, les entrées seront les suivantes :

Tableau 2-3 Liste des entrées des fichiers « Relevés-Tronçons\RelevésVF »

un numéro unique de relevé
identifiant du relevé
numéro séquentiel du tronçon
numéro du tronçon
date du relevé
jour de la semaine
heure du relevé
identifiant du tronçon
modèle de voiture flottante
conditions de la chaussée
distance de la partie du relevé associée au tronçon en mètres
durée du trajet
vitesse moyenne observée sur le tronçon

L'outil développé permet de spécifier le type de relevés à étudier, selon divers types de filtrage : date du relevé et heures de passage au tronçon à retenir, jours de la semaine, type de chaussée, type de voitures. En conséquence, sous forme d'un tableau, et selon les filtres, l'utilisateur de la base de données peut obtenir les relevés relatifs à sa requête. Enfin, sur ce tableau, sont effectués des calculs statistiques : longueur moyenne du tronçon, vitesses moyennes et variabilité.

2.4 Synthèse

La géobase interrogative fait partie des données fournies pour les fins de la recherche. La Figure 2-2 Exemple de découpage d'un circuit en tronçons de longueur variableprésente un exemple de segmentation d'un circuit en tronçons de longueur variable (circuit : a13sud).

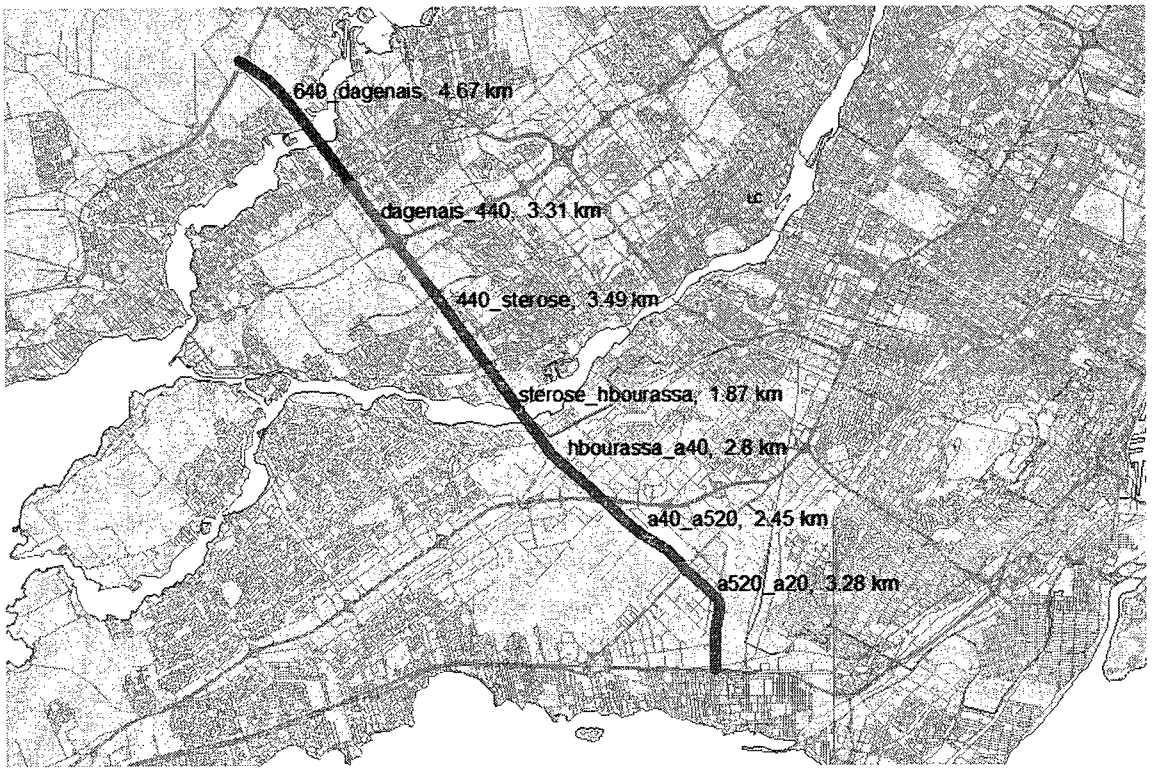


Figure 2-2 Exemple de découpage d'un circuit en tronçons de longueur variable

La longueur des tronçons varie entre 100 mètres et 15.41 kilomètres. 29 tronçons ont moins d'un kilomètre alors que 48 en ont plus de 5. La longueur moyenne d'un tronçon d'analyse est de 3.33 km mais cette longueur varie beaucoup (voir Figure 2-3).

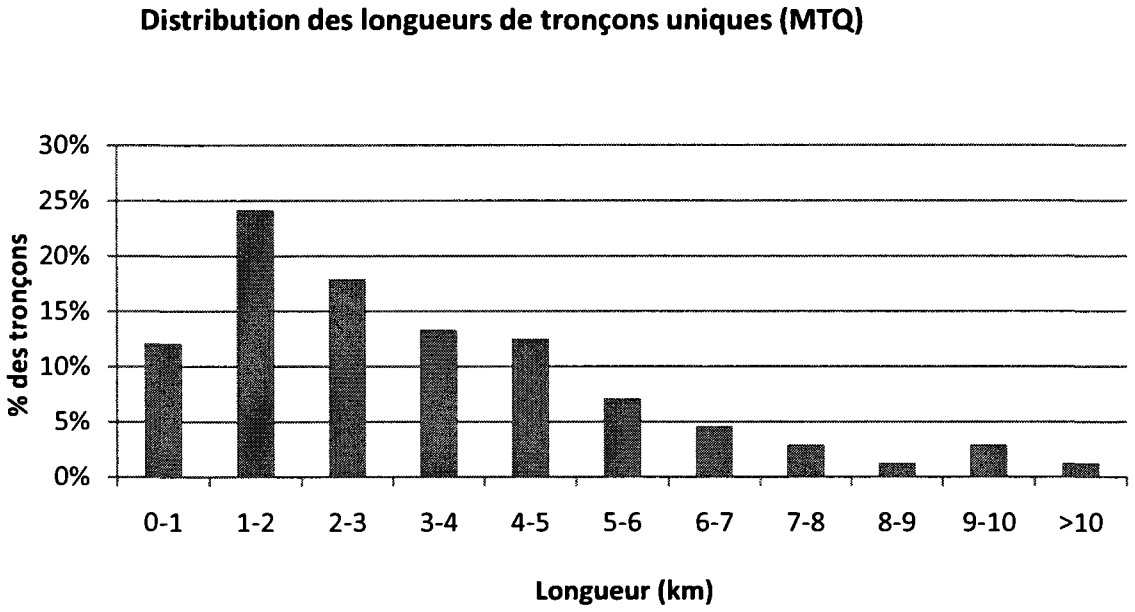


Figure 2-3 Distribution des longueurs des tronçons de longueur variable (240 tronçons uniques)

En outre, tel que mentionné précédemment, certains circuits se chevauchent ce qui signifie que des tronçons appartiennent à plusieurs circuits. En fait, l'étude de la base de données liée à la géobase permet d'identifier ces chevauchements. Il ressort de cette analyse qu'un peu moins de 50% des tronçons appartiennent à plusieurs circuits, certains appartenant même à 5 circuits. Le Tableau 2-4 indique le nombre et le pourcentage des tronçons qui sont impliqués dans un ou plusieurs circuits différents alors que la Figure 2-4 présente un exemple d'appartenance multiple d'un tronçon à plusieurs circuits.

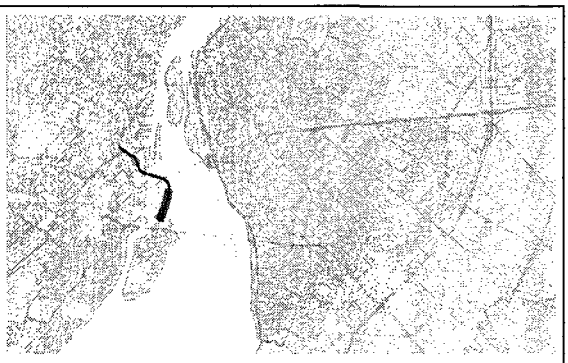
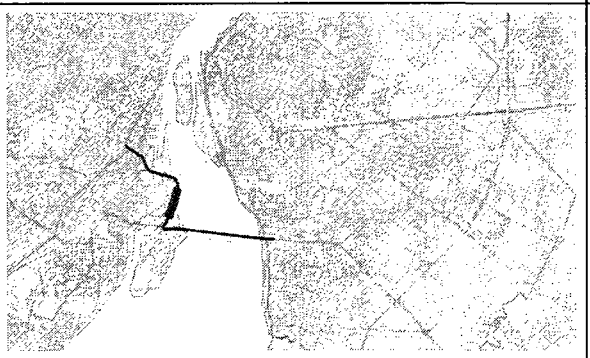
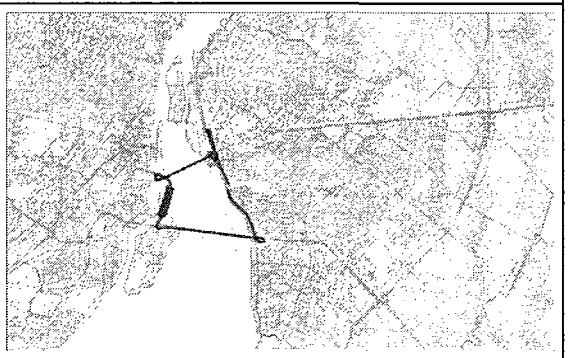
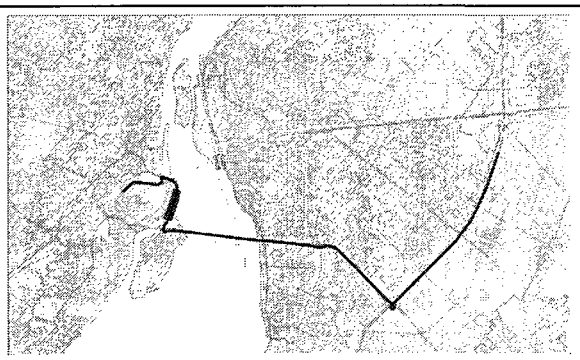
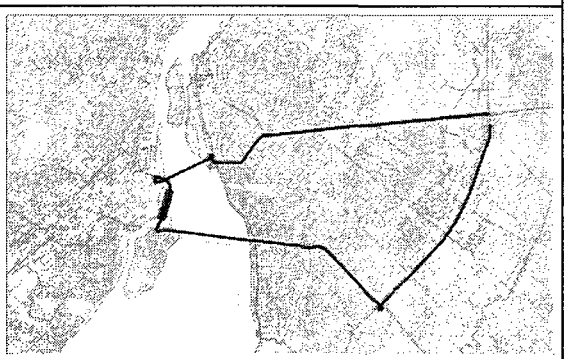
Circuits impliquant le tronçon 1903	 Circuit 19 – bonavsud
 Circuit 21 - bonav_champ_sud	 Circuit 34 - 116nord
 Circuit 32 - bona_30_sud	 Circuit 33 - victochamp_antih

Figure 2-4 Exemple d'appartenance multiple du tronçon 1903 à cinq circuits différents

Tableau 2-4 Distribution des tronçons selon le nombre de circuits qui les impliquent

Nombre de circuits	Tronçons	
	Nombre	%
1	128	53.33%
2	86	35.83%
3	17	7.08%
4	6	2.50%
5	3	1.25%
Total	240	

2.5 Utilisation des tronçons unitaires

Dans le but d'uniformiser notre base de données de manière à obtenir une analyse systématique, a été choisie dans cette recherche, l'idée d'une segmentation des circuits proposés précédemment en tronçons unitaires de 1 km.

- D'abord, les relevés de temps de parcours ont été intégrés dans une base de données intégrées sous *Visual Foxpro*. Un premier contrôle de qualité, assisté par un outil de consultation interactif (XLS), a permis de détecter certains problèmes liés aux données brutes :

- Certains relevés comptaient peu d'observations, non ordonnées chronologiquement, et dont les lectures odométriques semblaient erronées ou désynchronisées (102 relevés).

- La présence de points de contrôle dans les relevés venait « polluer » la séquence spatiale (des distances) et temporelles des lectures odométriques. Bien que ces points de contrôle aient eu une utilité lors de la collecte, ils ne sont d'aucun secours ici car la précision de

marquage de ces points par les conducteurs (fractions de seconde à 100 km/h) n'est pas suffisante.

Cette étape a permis de constituer une liste de circuits uniques (possédant des propriétés de longueur similaires) pour l'ensemble des relevés. Chaque relevé a par la suite été associé à un de ces circuits unique. En outre, une couche géographique a été crée pour représenter chacun de ces circuits.

Les relevés ont ensuite été segmentés en tronçons unitaires de 1000 mètres, tout comme les circuits. Cette segmentation a été effectuée par un programme (*Visual Foxpro*) qui a fractionné les relevés de façon continue depuis leur point de départ. Ce découpage a permis d'identifier des problèmes supplémentaires quant aux temps de parcours. Par exemple, pour un certain nombre de segments, le temps de parcours calculé est inférieur à 10 secondes, ce qui est très court pour parcourir 1 km (vitesse de 360 km/h). Ainsi, une validation supplémentaire a permis de filtrer environ 1000 segments sur les 600 000, ce qui est très peu.

Les segments unitaires ont fait l'objet de quatre traitements majeurs :

- Analyse exploratoire des temps de parcours des segments orientée sur certains attributs de base (heure, circuit impliqué, etc.). Cette analyse s'appuie de plus sur des analyses de variance ANOVA.
- Analyse statistique avancée à l'aide de techniques de *data mining*. Cette analyse a pour but de caractériser les tronçons ayant des distributions des temps de parcours semblables au sein d'un même circuit. La finalité est alors en s'appuyant sur un indicateur de fiabilité à savoir la variabilité des temps de parcours, de concevoir un plan d'échantillonnage.
- Dégagement de la circulation normale d'un circuit dans l'espoir de différencier la congestion récurrente de la congestion non-récurrente. Cette analyse, complétée par une analyse de la corrélation des temps de parcours des tronçons, permettra par la suite de modéliser le trafic et de le simuler.

- Simulation de la moyenne et de la variabilité des relevés de temps de parcours pour un tronçon dans le but de développer des indicateurs.

CHAPITRE 3 : ANALYSE EXPLORATOIRE DES DONNEES BRUTES (CONTROLE DE QUALITE)

Dans un premier temps, il est important, alors que nous venons de créer notre base de données, d'observer celles-ci. Quelles en sont les caractéristiques ? Où sont les points à priori chauds de la circulation ? Combien de temps dure la congestion ? Les données sont-elles vraiment intéressantes ?

3.1 Contrôle de qualité

Dans un souci de valider les données brutes de temps de parcours et d'identifier les procédures automatiques à mettre en place pour améliorer leur contenu, un outil de consultation interactive des relevés bruts a été développé au moyen de *Microsoft Excel 2007*. L'outil permet de choisir manuellement un parcours pour une année et un mois choisi au préalable (voir Figure 3-1). Quelques caractéristiques de base du parcours sont énumérées dont le nom et la description du circuit, le nombre d'enregistrements le temps du trajet et la distance parcourue. Les enregistrements associés permettent ensuite d'illustrer graphiquement le parcours dans le temps et l'espace (voir Figure 3-3 : diagramme espace-temps). L'outil permet aussi de repérer les points de contrôle impliqués dans le relevé actif et de valider leur position sur un tronçon. Les trajets qui sont jugés suspects peuvent, par la suite, être examinés plus précisément. L'objectif principal de cet outil est d'assurer la qualité des données avant d'approfondir les analyses. Les objectifs secondaires sont notamment:

- Représenter graphiquement le cheminement d'un relevé;
- Valider les enregistrements d'un relevé;
- Valider le trajet parcouru en vérifiant la position des points de contrôle;
- Écrire un commentaire dans la base de données sur la validité de l'information;

- Repérer les erreurs systémiques qui se retrouvent dans la base de données pour alimenter le programme de traitement des données.

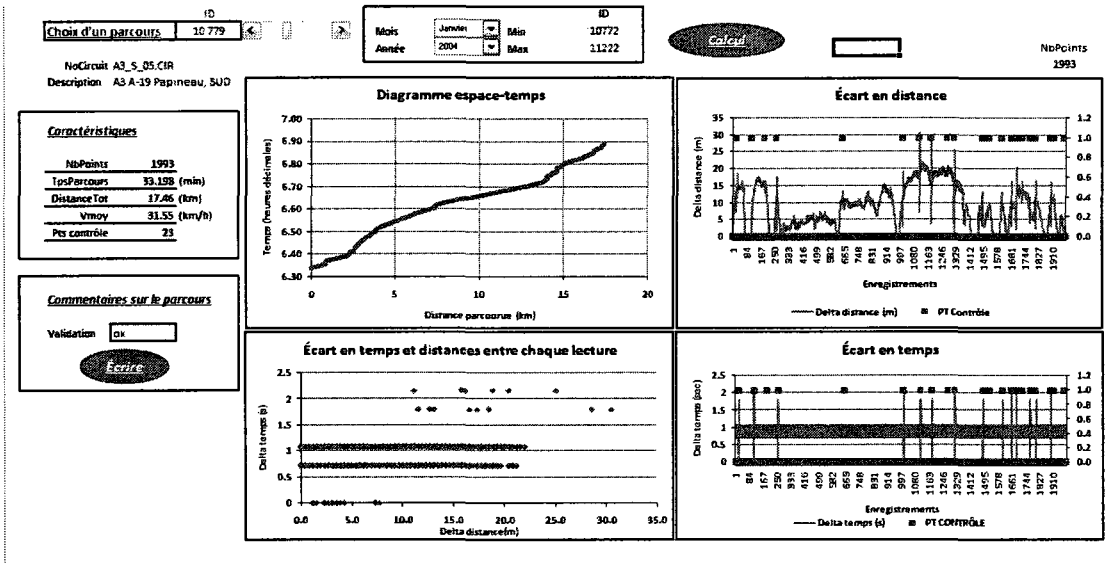


Figure 3-1 Interface de l'outil de consultation et de validation d'un parcours

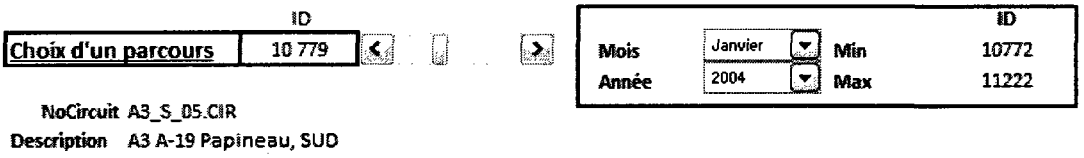


Figure 3-2 Interface permettant le choix d'un trajet particulier à consulter

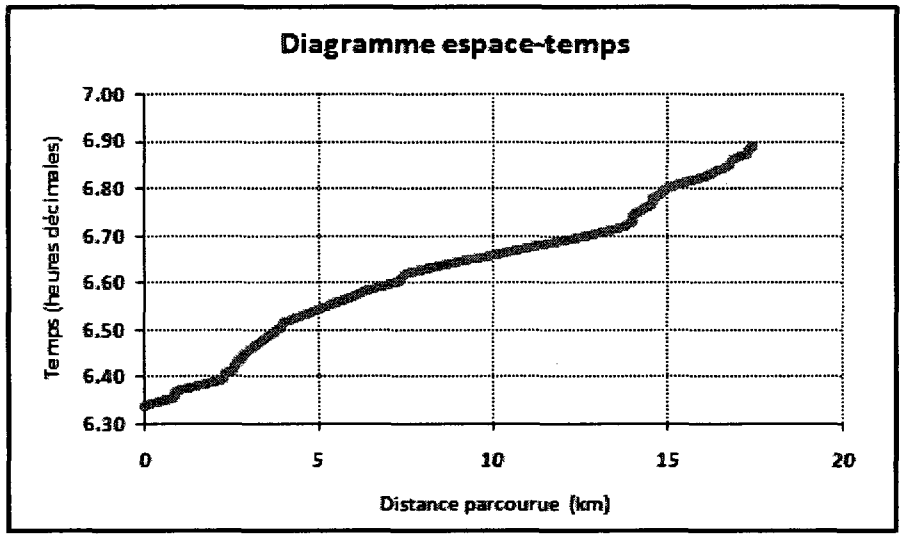


Figure 3-3 Diagramme espace temps du véhicule pendant tout le trajet

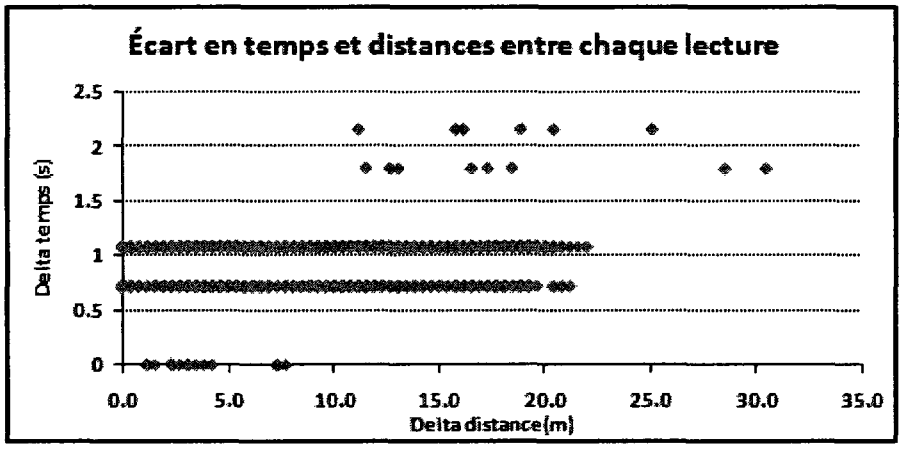


Figure 3-4 Distribution des écarts entre deux enregistrements consécutifs représentant la vitesse du véhicule entre ces deux points

3.2 Distribution globale des temps de parcours

Un second outil a été développé afin d'observer la distribution des temps de parcours par kilomètre (tronçon unitaire) selon différentes variables de segmentation. L'outil s'appuie sur des statistiques a priori, calculées à l'aide d'un logiciel de statistiques (STATA) et vise à mieux comprendre l'ampleur de la variabilité contenue dans les données empiriques.

Plusieurs observations peuvent être faites. En effet, on peut analyser les temps de parcours par heure, jour de la semaine, mois, ou bien année :

- La Figure 3-5 Observations des moyennes de temps de parcours selon l'heure de départ, présente l'évolution du temps de parcours moyen, par kilomètre, selon l'heure de départ (heure d'arrivée en tête de tronçon unitaire). On confirme d'après ce graphe, que seulement les périodes de pointe ont été étudiées, de plus, que malgré l'absence de traitement de données, les deux périodes de pointes (matin et après-midi) sont nettement mises en exergue.
- La Figure 3-6 Autres types de distribution selon l'année, le mois et la semaine illustre d'autres distributions de temps moyens de parcours par kilomètre selon différentes variables de segmentation : l'année, le mois, la semaine. Évidemment, il s'agit là de distributions unidimensionnelles qui ne tiennent pas compte de la structure de l'échantillon. Celles-ci ne permettent pas de se prononcer sur des tendances mais visent surtout à informer sur le contenu des données. Ainsi, à la vue de ces graphes, il ne faut toutefois pas tirer de conclusions hâtives. En effet, on pourrait penser naïvement que l'année 1999 fut une mauvaise année quant à la circulation, mais premièrement tous les mois de l'année n'ont pas été échantillonnés (ce sont en fait les mois de septembre à décembre qui l'ont été, mois durant lesquels la moyenne des temps de parcours ainsi que leurs écarts-types sont critiques), deuxièmement, le nombre d'observations pour l'année 1999 est bien inférieur à ceux des autres années. Par contre, l'observation du temps de parcours moyen selon le mois, en attente d'une confirmation statistique (ANOVA), nous amènera à tester les hypothèses suivantes : la moyenne des temps de parcours durant les mois d'été est inférieure aux autres mois, le mois de septembre est atypique (phénomène de rentrée ?), etc.

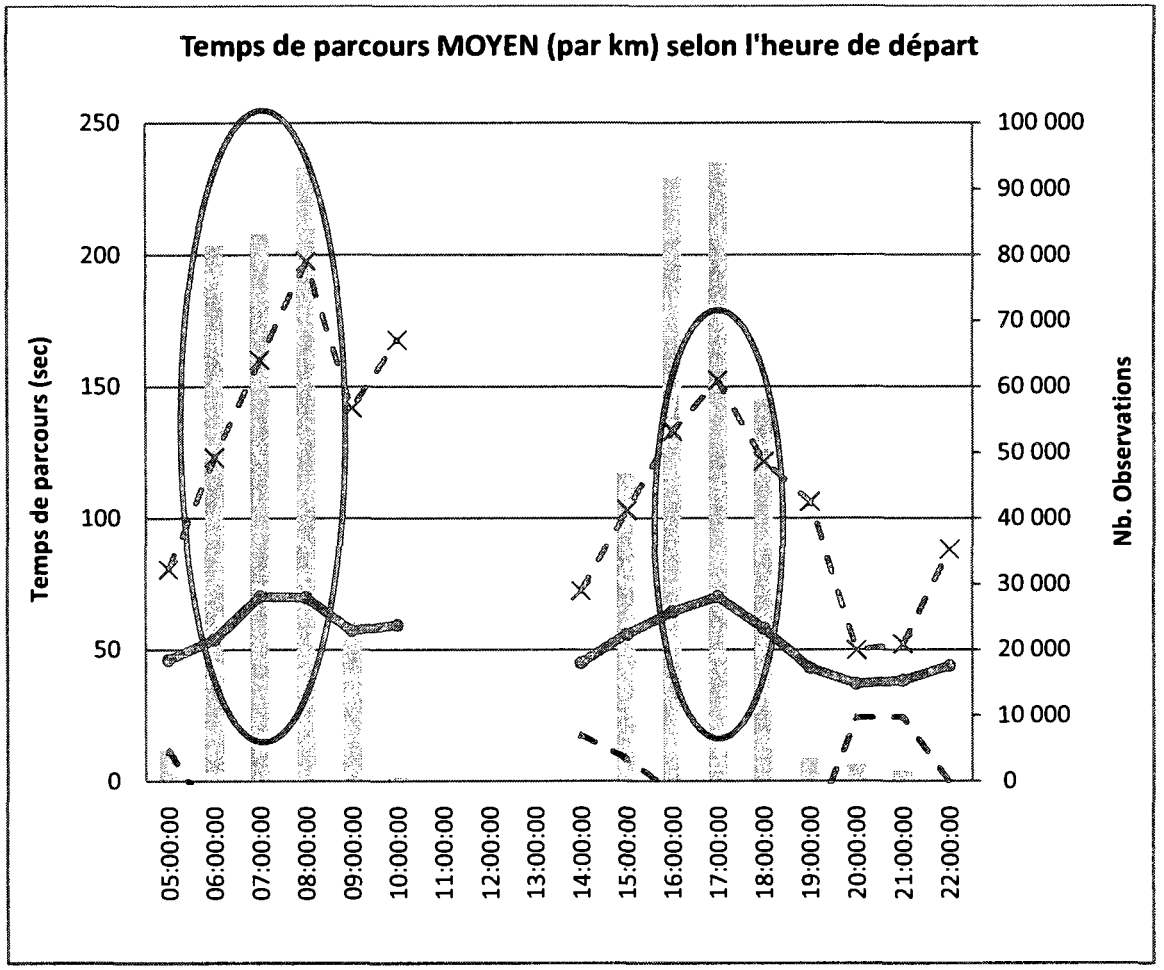


Figure 3-5 Observations des moyennes de temps de parcours selon l'heure de départ, mise en évidence des pointes AM et PM de la circulation

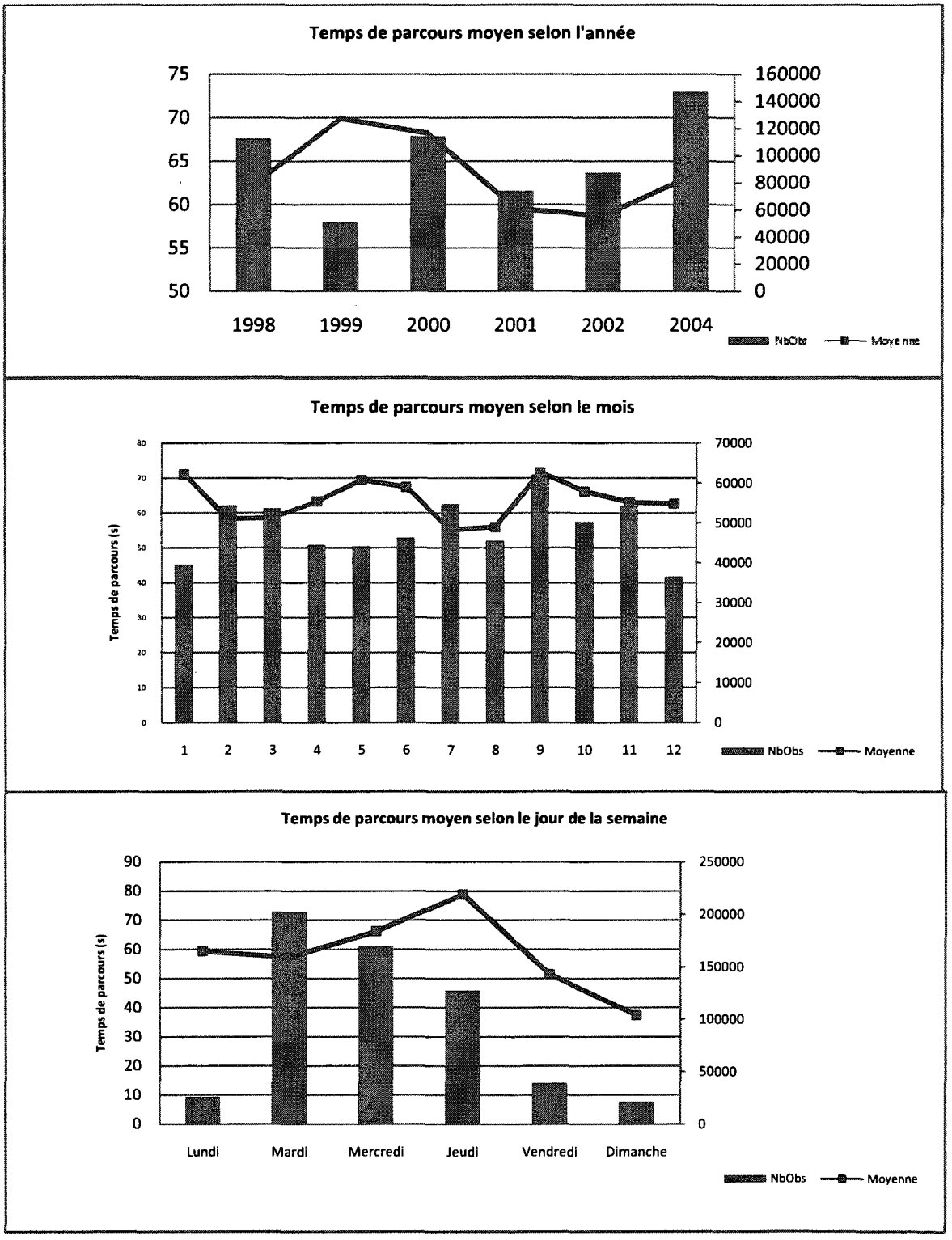

Temps de parcours moyen selon l'année
75
70
65
60
55
50
160000
140000
120000
100000
80000
60000
40000
20000
0
1998
1999
2000
2001
2002
2004
NbObs
Moyenne
Temps de parcours moyen selon le mois
Temps de parcours (s)
80
70
60
50
40
30
20
10
0
70000
60000
50000
40000
30000
20000
10000
0
1
2
3
4
5
6
7
8
9
10
11
12
NbObs
Moyenne
Temps de parcours moyen selon le jour de la semaine
Temps de parcours (s)
90
80
70
60
50
40
30
20
10
0
250000
200000
150000
100000
50000
0
Lundi
Mardi
Mercredi
Jeudi
Vendredi
Dimanche
NbObs
Moyenne

Figure 3-6 Autres types de distribution selon l'année, le mois et la semaine

Les résultats statistiques concernant la base de données permettent de croiser les facteurs heure, jour de la semaine, mois et année. Par conséquent plusieurs outils de visualisations ont été développés (Figure 3-7 à Figure 3-11) :

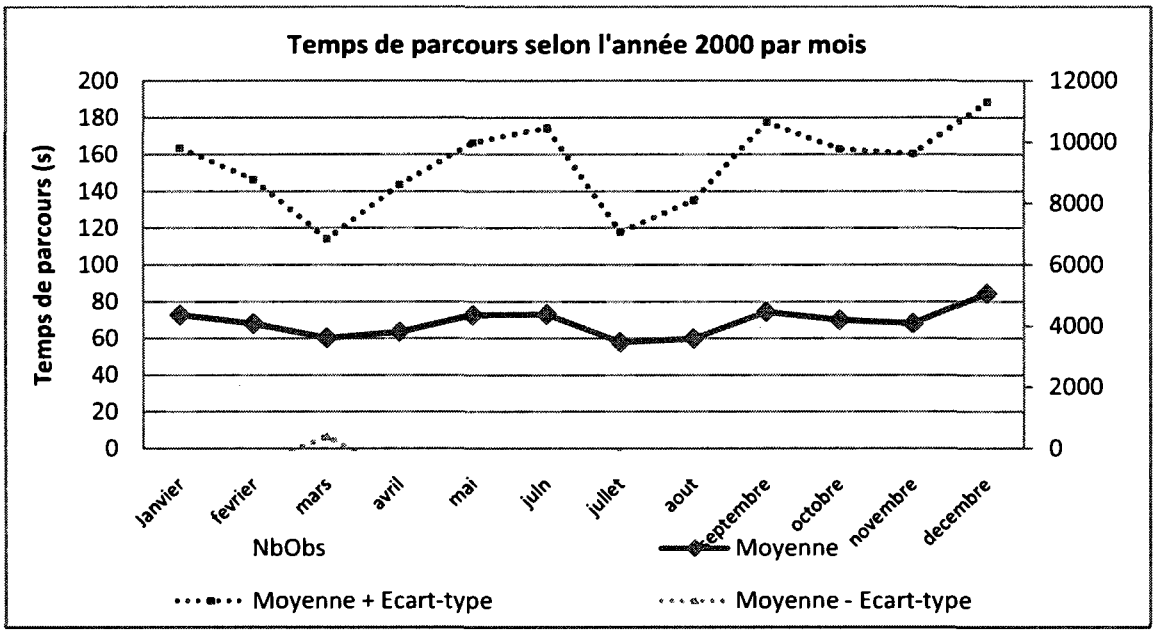


Figure 3-7 Temps de parcours par mois pour l'année 2000

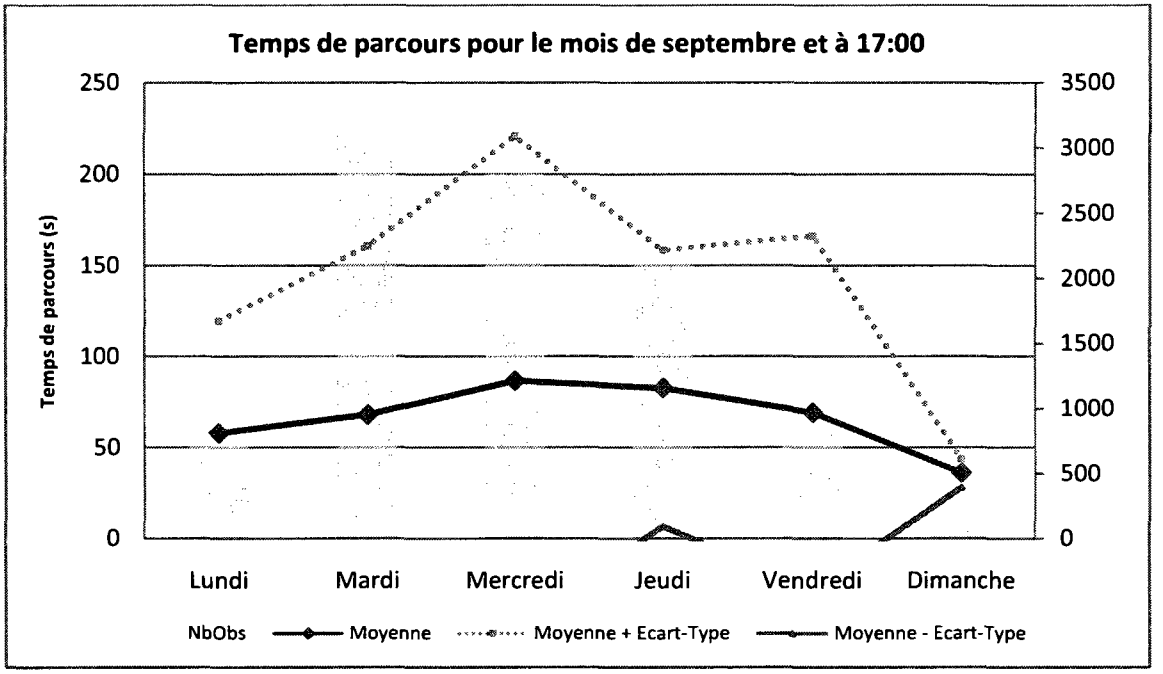


Figure 3-8. Temps de parcours pour le mois de septembre à 17h

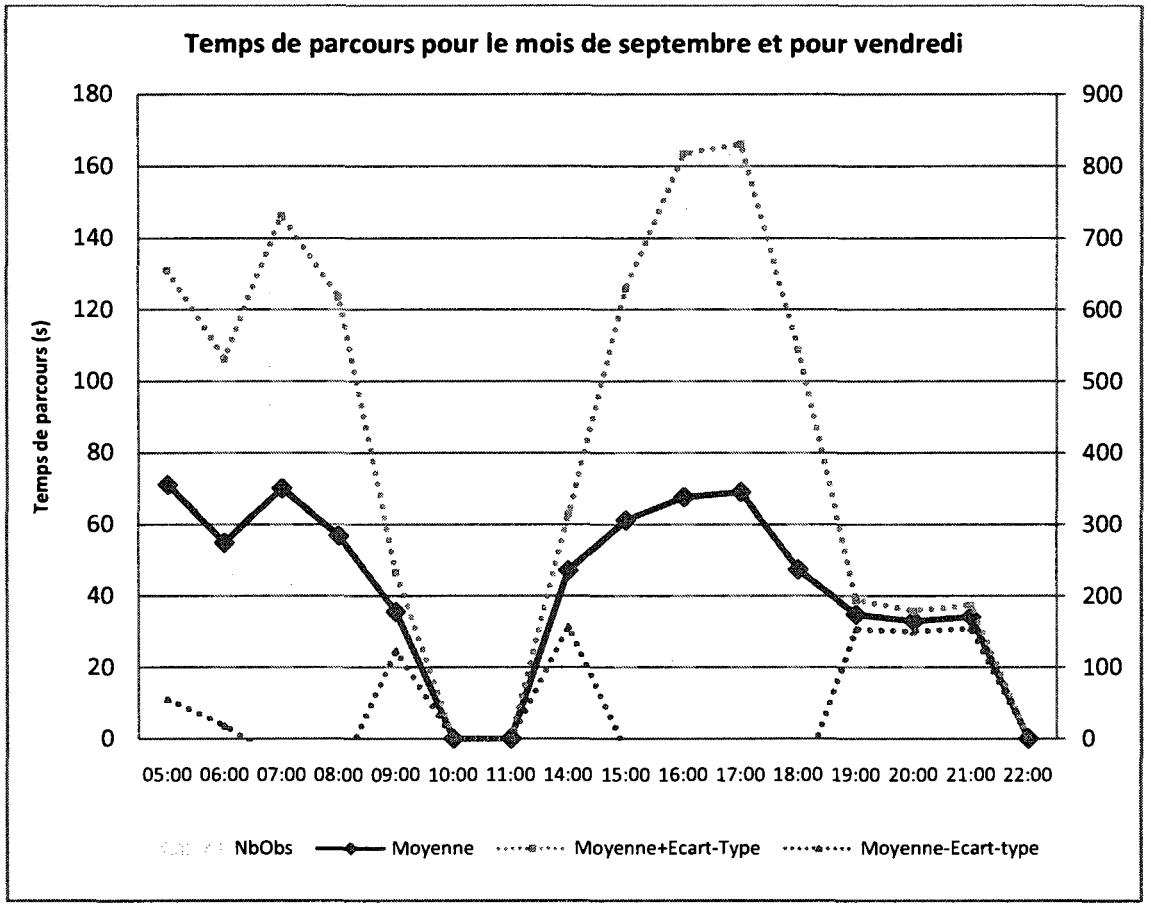


Figure 3-9 Temps de parcours pour le mois de septembre et la journée du vendredi

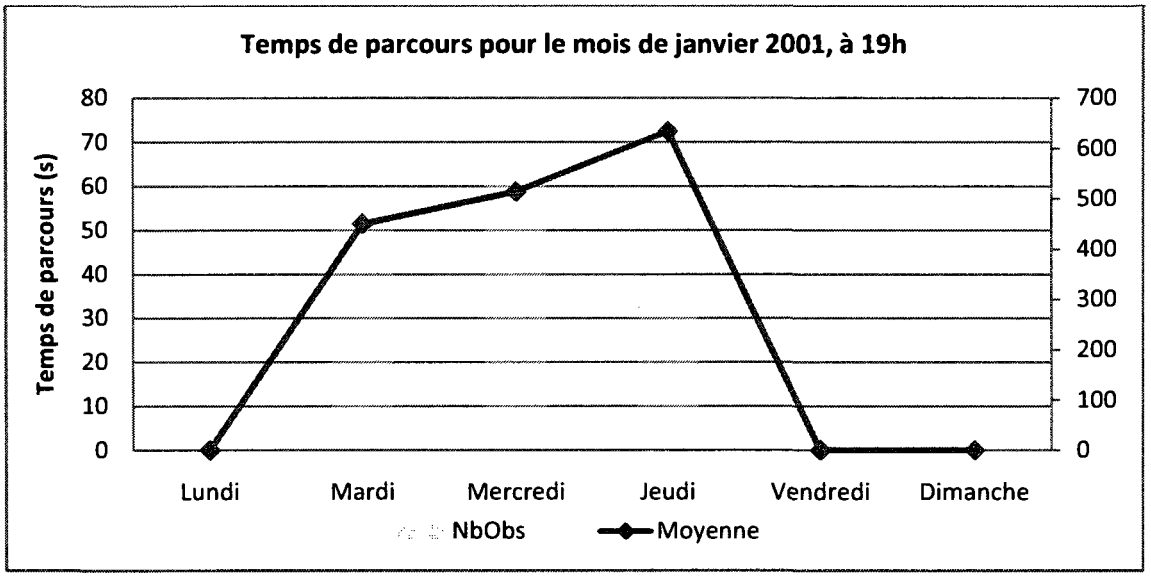


Figure 3-10 Temps de parcours pour le mois de janvier 2001, à 19h

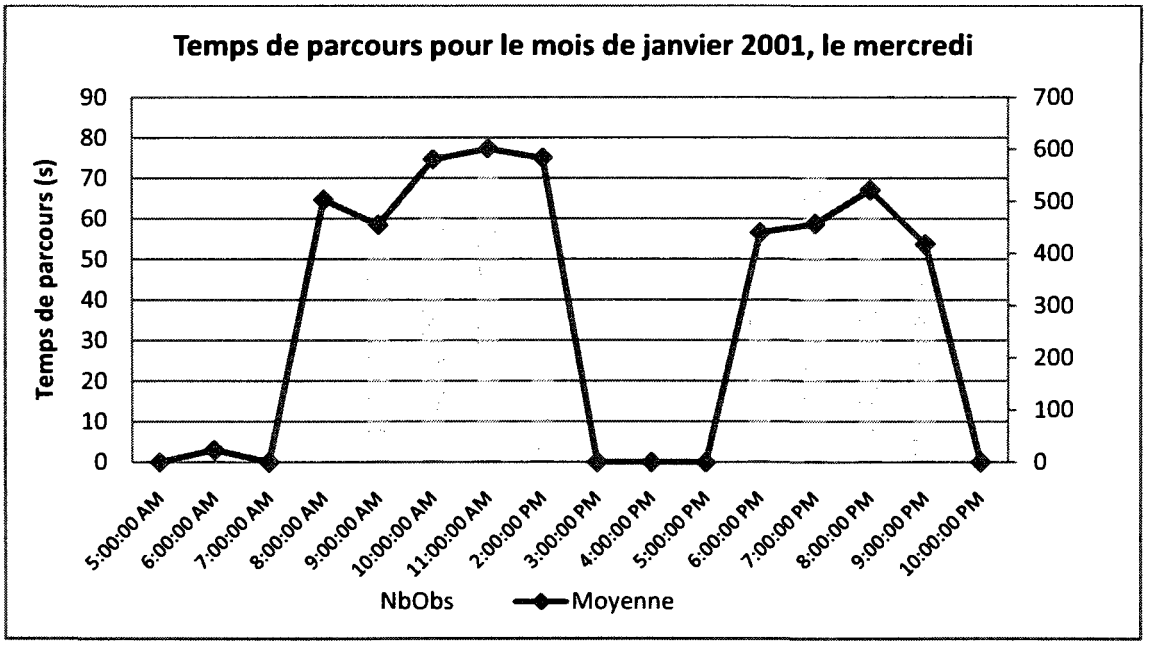


Figure 3-11 Temps de parcours pour le mois de janvier 2001, le mercredi

Cette partie expose donc les différents outils de visualisation des premières valeurs statistiques calculées. Ce fichier Excel n'est donc pas destiné à la visualisation d'UN relevé mais bien à la visualisation de TOUS les relevés de TOUS les circuits. Par conséquent, il sera malencontreux de

tirer toutes formes de conclusion hâtives par l'observation des graphes présentés ci-dessus. Il y a seulement une justification a priori de phénomènes propres à la circulation déjà connus.

3.3 Analyse exploratoire d'un circuit

Dans la partie précédente, nous avons présenté les valeurs statistiques usuelles de la distribution des temps de parcours sur le réseau autoroutier montréalais. Toutefois, cette analyse utilisait la base de données dans sa globalité, sans distinction aucune du circuit, du tronçon ou bien de facteurs particuliers. Or, il pourrait apparaître normal que le comportement de la distribution statistique des relevés de temps de parcours dépende statistiquement de manière significative de la forme du circuit, de la période (matin ou après-midi), du mois, ou bien encore des tronçons où se trouve une zone d'entrecroisement. En conséquence, c'est sur les relevés de temps de parcours propres à un circuit particulier que va s'effectuer cette nouvelle analyse ; nous en donnerons les caractéristiques statistiques, sans toutefois se contenter simplement de la moyenne et de l'écart-type des relevés qui ne permettent pas de qualifier correctement les temps de parcours sur un circuit.

Dans cette partie, nous nous intéresserons tout particulièrement à l'étude du circuit 1, c'est-à-dire, celui empruntant l'autoroute 13, direction sud.

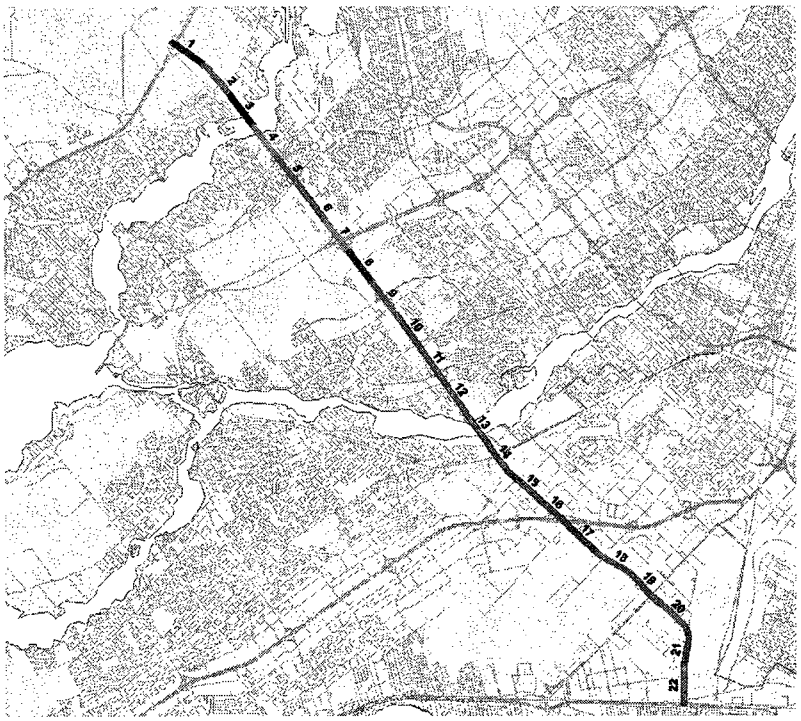


Figure 3-12 Découpage en tronçons d'un kilomètre du circuit 1 A13 direction sud

La signification statistique des facteurs période, mois de l'année, conditions de la chaussée est exposée en ANNEXE A : ANALYSE STATISTIQUE DU CIRCUIT 1 AU MOYEN DE STATISTICA.

3.3.1 Présentation du circuit 1

Le circuit 1 emprunte l'autoroute 13, direction sud. Celui-ci commence donc au croisement en l'A640, coupe ensuite l'A440 au tronçon 7, l'A40 au 16, l'A520 au 19 et enfin l'A20 à la fin du circuit. Par conséquent, une voiture empruntant le circuit 1, se dirige vers Montréal. L'intérêt de ce circuit repose dans le fait que celui-ci dispose d'une propriété simple mais affreusement pratique : la vitesse limite y est constante et égale à 100 km/h. Cette remarque, qui peut sembler hors de propos pour l'instant, aura son importance dans les futures parties de ce mémoire.

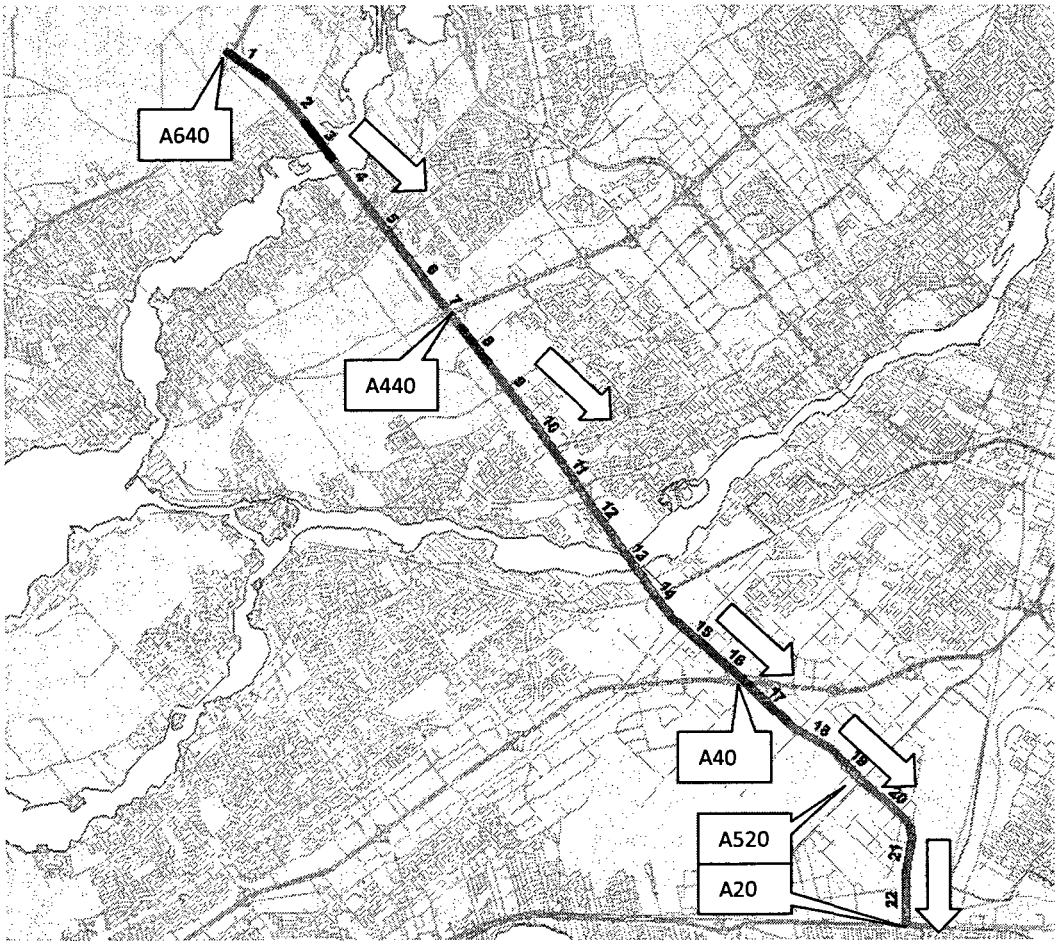


Figure 3-13 Circuit 1, direction et intersections majeures

3.3.2 Localisation des points chauds du circuit 1

Compte-tenu de la configuration spatiale et du sens de déplacement sur le circuit 1, il devient pertinent alors de considérer simplement la moyenne des temps de parcours tout au long de notre circuit, selon la période d'étude (AM ou PM). En effet dans l'ANALYSE STATISTIQUE DU CIRCUIT 1 AU MOYEN DE STATISTICA, nous avons précisé la signification de ce facteur catégoriel à 2 niveaux (AM/PM) par analyse de la variance.

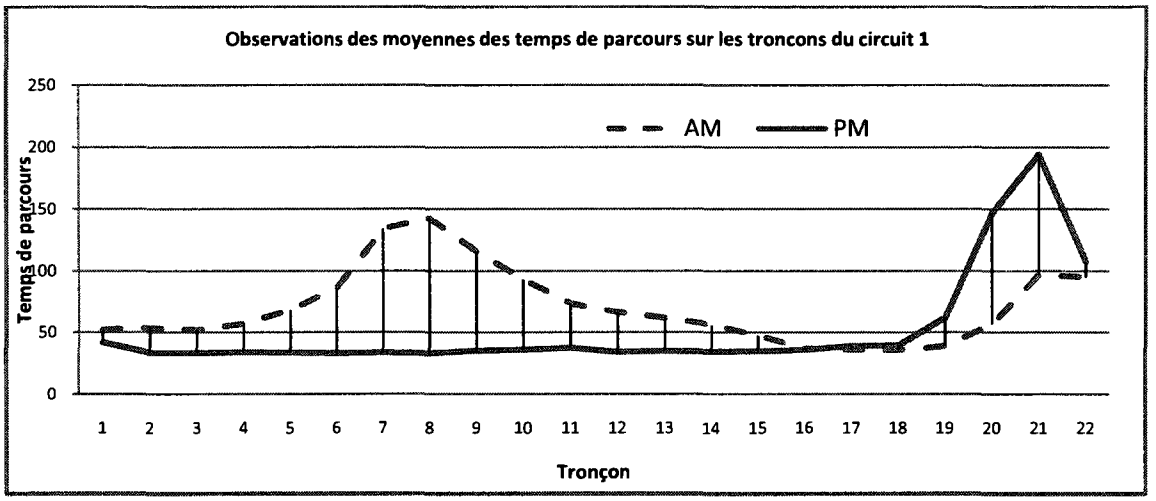


Figure 3-14 Evolution des temps de parcours par tronçons du circuit 1

La Figure 3-14 démontre donc une différence notoire entre le comportement du temps de parcours moyen du matin et celui de l'après-midi. Durant la matinée, la moyenne des temps de parcours tout au long du circuit 1 est donc plus importante que pour l'après-midi. Ce constat est conforme à notre connaissance des patrons de mobilité typiques des montréalais à savoir que ceux-ci convergent vers les parties centrales de Montréal en pointe du matin afin d'accéder aux pôles principaux d'emplois. Cette conclusion est statistiquement confirmée grâce à une analyse de variance avec la période d'étude comme facteur catégoriel à 2 niveaux (AM et PM).

D'un point de vue observationnel, durant le matin, il y a deux zones de congestion. La première a lieu au tronçon 8 et s'étend entre le 7 et le 10, ce qui signifie que la congestion s'étend sur 4 km. La deuxième zone s'étend autour du tronçon 21 et est commun aux périodes matin et après midi. Rappelons que par rapport aux zones d'entrecroisement, nous pouvons émettre l'hypothèse que durant la matinée, de nombreuses voitures s'insèrent dans le circuit 1 au niveau de l'entrecroisement A13-A440 et semblent quitter ce circuit au niveau de l'entrecroisement A13-A40. En effet, la moyenne des relevés de temps de parcours est décroissante du tronçon 8 au tronçon 16. Par ailleurs la valeur de la moyenne des relevés des temps de parcours du tronçon 16 pour la période matinale, tous autres facteurs confondus, est

de 37 secondes ce qui correspond à une vitesse 97 km/h. En clair, la vitesse limite étant constante et égale à 100km/h sur la totalité du circuit 1, on peut supposer que le trafic n'est plus congestionné dès que l'échangeur de l'A40 est passé. Enfin, on retrouve un pic en fin de circuit qui correspond simplement à une fluctuation normale de sortie d'autoroute.

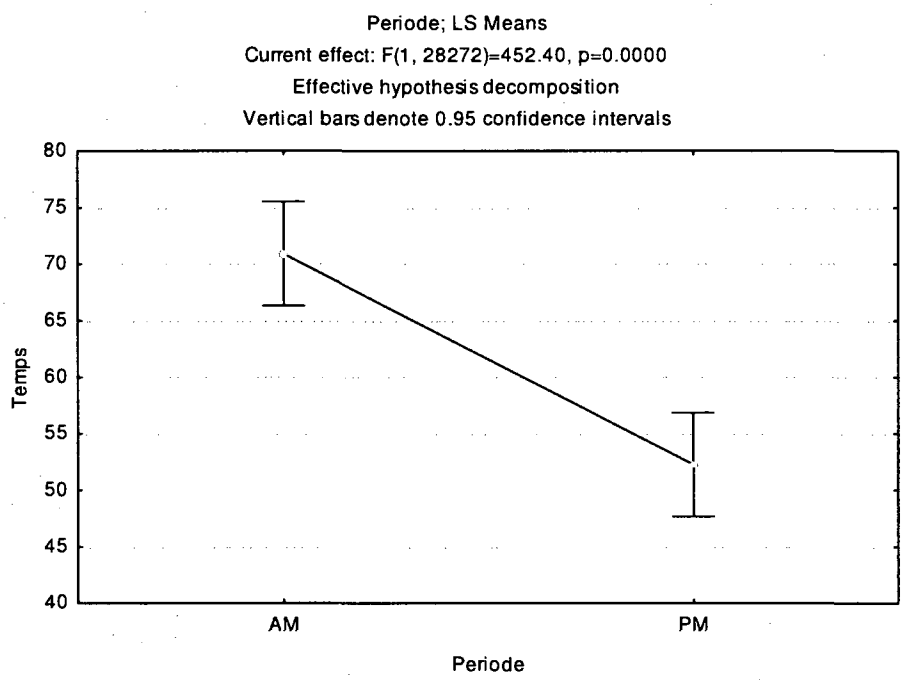


Figure 3-15 Moyenne des temps de parcours du circuit 1 selon la période.

En conclusion, nous venons de confirmer une différence significative entre le matin et l'après-midi avec une localisation spatiale pertinente des points chauds, puisque, effectivement, les pics de la courbe matinale des moyennes de temps de parcours (cf. Figure 3-13 et Figure 3-14 Evolution des temps de parcours par tronçons du circuit 1), correspondent très justement aux zones d'intersection avec d'autres axes autoroutiers.

3.3.3 Evolution temporelle des zones de congestion

Au moyen des relevés dont nous disposons, comment caractériser temporellement ces zones de trafic congestionnées ? En effet, au cours de la matinée, un pic de congestion peut

apparaitre et par la suite disparaitre, sa période de « vie » est forcément délimitée dans le temps mais nous pouvons retrouver ce pic plus tard, sur le même circuit.

Ainsi, pour observer l'évolution temporelle des moyennes des relevés de temps de parcours tout au long du circuit 1 nous choisissons de considérer des créneaux de 20 minutes soit 1200 secondes ou bien encore une vitesse de parcours de 5 km/h (!!!), ce qui permet d'éviter des chevauchements temporels (en effet très peu de relevés de temps de parcours sont supérieurs à cette valeur) : en clair, nous nous assurons que si une voiture est prise dans une congestion, vingt minutes plus tard, il est quasi certain que le véhicule a parcouru plus d'un kilomètre. Au niveau de la base de données, cela revient simplement à dire que si l'heure d'un relevé de temps de parcours est égale à 8h23, alors la valeur du créneau horaire auquel appartient ce relevé sera 8h20.

La Figure 3-16 expose donc le fait de l'évolution temporelle de la zone congestionnée. Les courbes noires correspondent aux moyennes des relevés de temps de parcours par tronçon sur le circuit 1 durant le créneau horaire considéré, la zone grise correspond aux moyennes des relevés de temps de parcours sur la journée. On constate qu'entre 6h40 et 8h20, la circulation est très congestionnée au niveau des tronçons 7 à 10 (intersection avec l'A440). Par conséquent, nous pouvons réitérer ce questionnement entre problème de congestion de la circulation sur les grands axes routiers et autoroutiers et le concept de déplacement OD (lignes de désir). De plus, le pic justement observé au niveau des tronçons au cours du temps, signifie que la situation est congestionnée de manière uniforme au cours du temps. Une des questions futures sera donc de qualifier la variabilité de cette congestion.

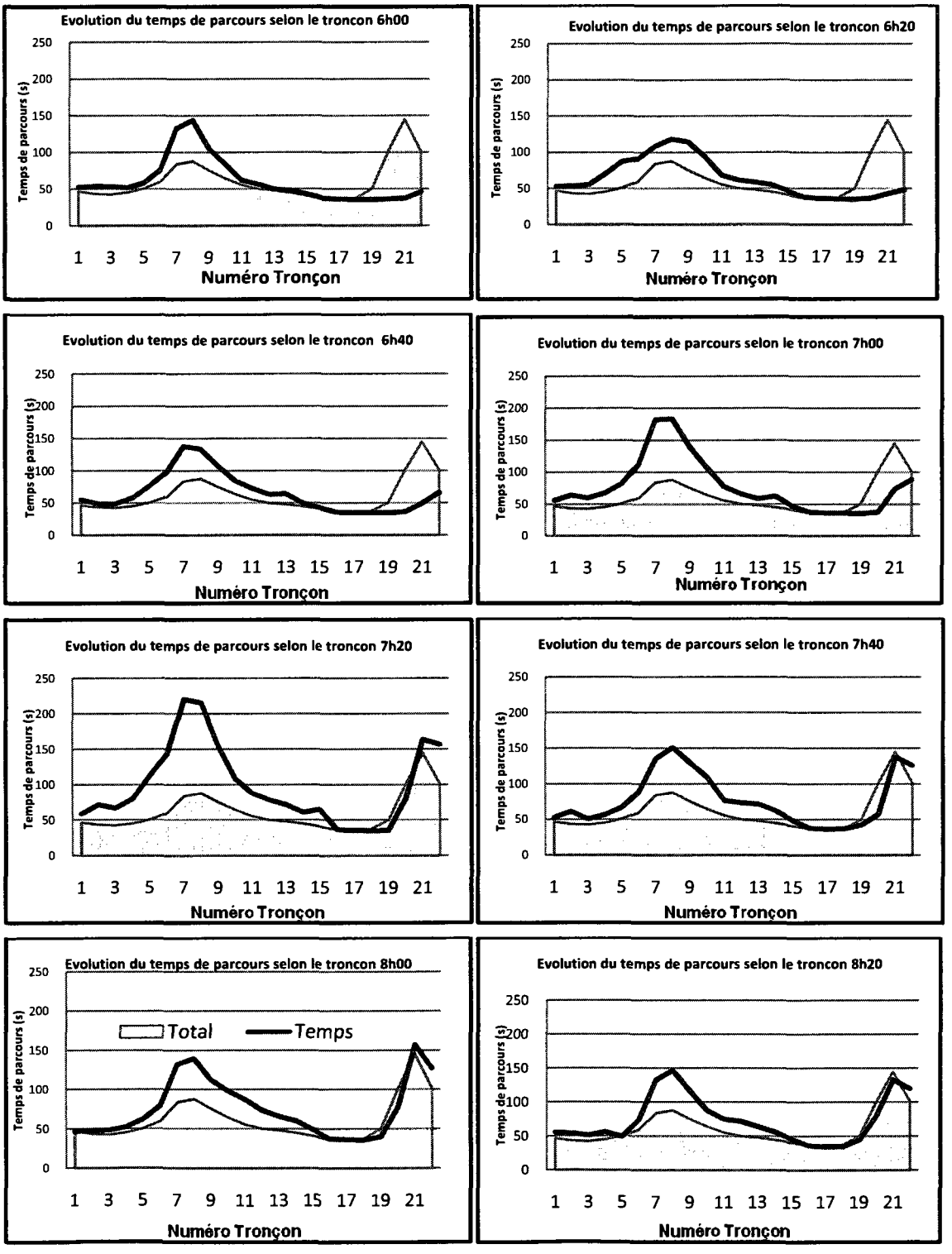


Figure 3-16 Evolution temporelle du temps de parcours moyen par tronçon entre 6h00 et 8h20 AM

3.3.4 Observation des temps de parcours selon le mois et l'année

Comme cité dans la partie 3.2, tous les mois d'une année n'ont pas forcément été échantillonnés. Ce constat est considérablement malheureux : en effet, premièrement, nous intuitons la différence de comportement de la circulation durant le mois de septembre et les autres mois, et cette différence est elle-même confortée par une analyse ANOVA (Figure 3-17). Par conséquent, il serait alors malencontreux par exemple, de comparer la moyenne des temps de parcours par année et de conclure que l'année 1999 fut une année critique pour la circulation autoroutière. En effet, ce sont seulement les mois de septembre à décembre qui ont été échantillonnés pour cette année, mois justement critiques.

Il devient par contre très intéressant de concevoir un autre type d'analyse de la variance qui prendra justement en compte le fait que tous les mois d'une même année ne sont pas échantillonnés (ce type d'analyse statistique correspond à la régression *ANOVA Nested Design*).

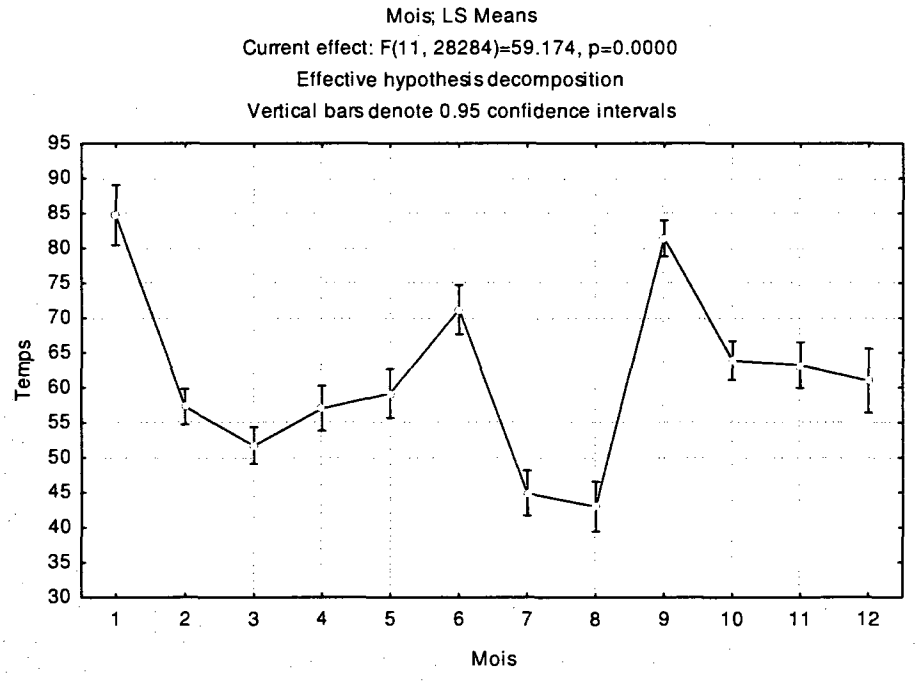


Figure 3-17 Moyenne des temps de parcours selon le mois

Ainsi, le Tableau A-24 se trouvant en Annexe A explique que, effectivement la fonction Mois(Année) est significative, les p-values étant très proches de 0. On peut alors observer le comportement moyen des temps de parcours, par exemple de l'année 2004, selon le mois de l'année. En clair, il existe une différence significative entre plusieurs mois d'une même année. Nous confirmons de plus que le mois de septembre est le mois le plus critique, et ce pour toutes les années, grâce à l'interface Excel proposée.

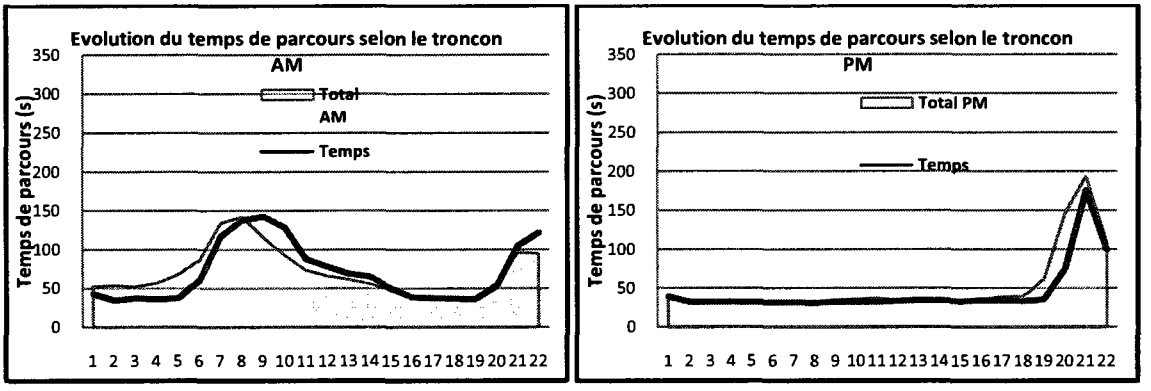


Figure 3-18 Evolution des temps de parcours pour le mois de février 2004

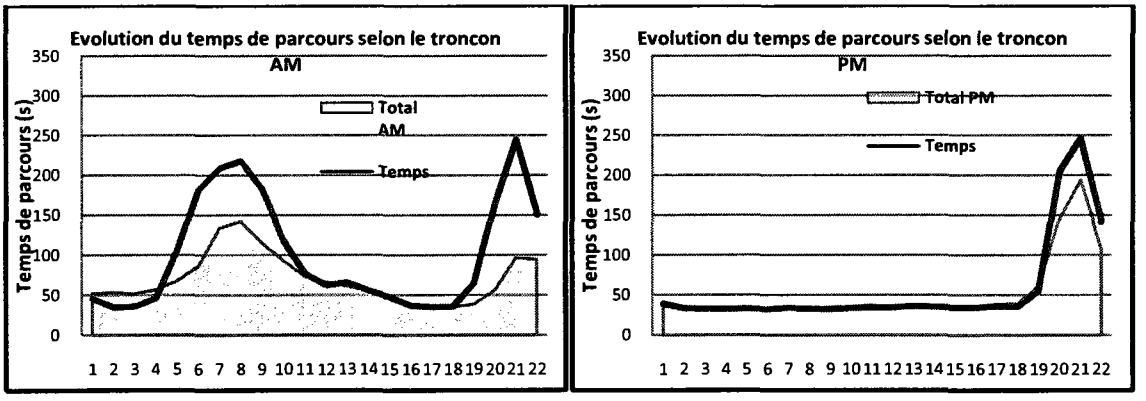


Figure 3-19 Evolution des temps de parcours pour le mois de juin 2004

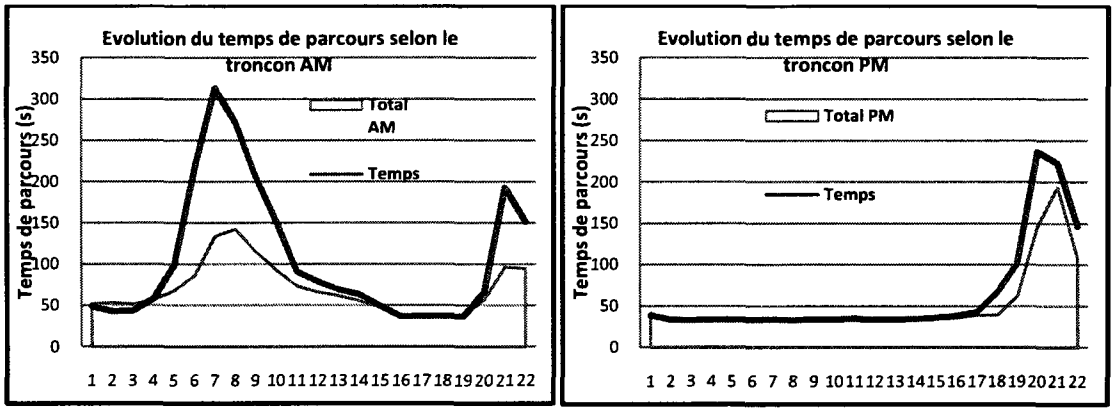


Figure 3-20 Evolution des temps de parcours pour le mois de septembre 2004

L'observation de cette liste de graphes permet en outre de caractériser le comportement moyen des temps de parcours. En effet, pour le mois de septembre 2004, le pic caractéristique des tronçons 7 et 8 est bien plus grand que la tendance totale (moyenne confondue de tous les relevés, zone grisâtre sur les graphes). Toujours dans un contexte de localisation de la congestion, ceci signifie que la circulation est plus congestionnée au niveau des tronçons cités précédemment. Ainsi, il peut apparaître logique de déclarer que, au niveau de cette zone d'entrecroisement, plus de voitures s'intègrent, ou que, en raison du phénomène de rentrée, les conducteurs appréhendent mal cette zone.

3.3.5 Evolution d'indicateurs de fiabilité selon l'heure

Sachant que l'objectif principal de cette étude est de qualifier la fiabilité du réseau autoroutier montréalais, il faut désormais appliquer quelques indicateurs de fiabilité propres à la circulation, que nous avons évoqués durant la revue de littérature. Observons par exemple le *TTI* et le *TRI* pour les relevés réunis des tronçons 7 à 10 qui sont a priori de bons indicateurs de fiabilité.

Nous rappelons les formules suivantes,

$TRI = \frac{T_{90}-T_{10}}{T_{50}}$ *et* $TTI = \frac{T_{90}-T_{50}}{T_{50}}$, où *T(i)* désigne le ième centile.

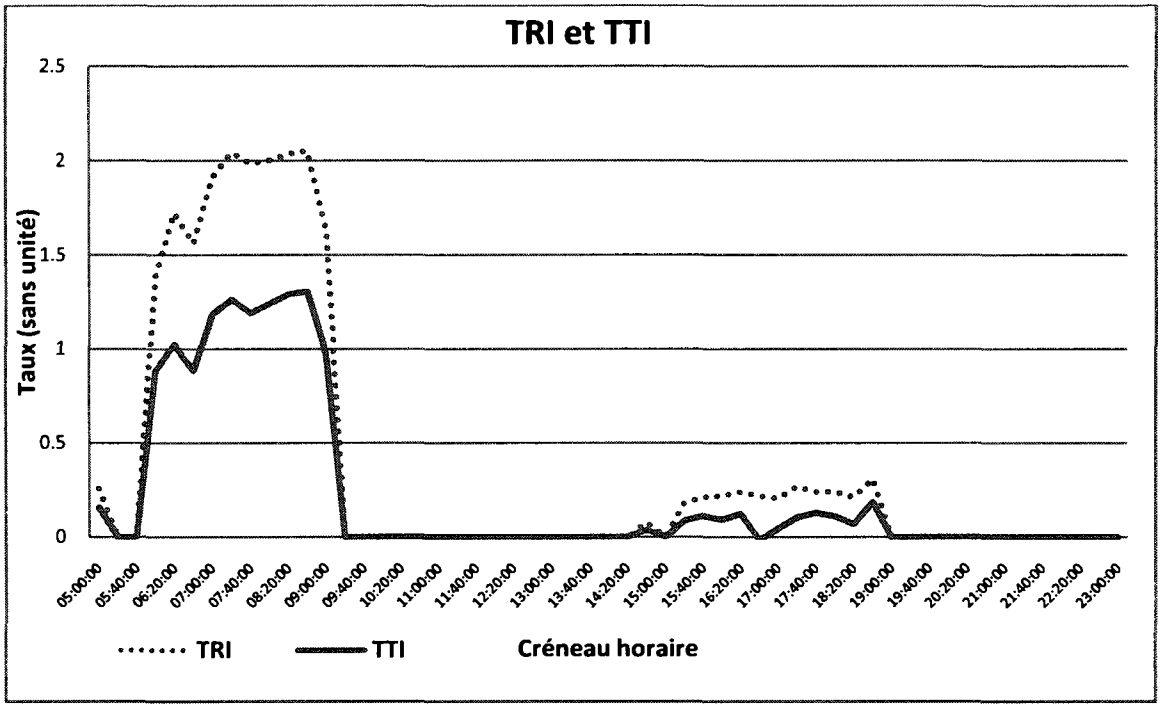


Figure 3-21 Observations des *TRI* et *TTI* selon l'heure pour les tronçons 7 à 10

La Figure 3-21 expose de nouveau le manque de fiabilité des tronçons 7 à 10. Il faut compter deux fois le temps de parcours moyen pour parcourir ne serait-ce, qu'un seul de ces tronçons durant le pic matinal. Autrement dit, un conducteur empruntant ces tronçons doit multiplier son temps utile de parcours au niveau de ces tronçons entre 6h20 et 9h20 par 2 ! Signalons toutefois que nous ne connaissons pas l'état de ces indicateurs entre 9h40 et 15h00 d'où les valeurs nulles du graphique ci-dessus.

3.3.6 Cartes de contrôle et relevés hors-limites

Malgré un nombre certain de relevés de temps de parcours élevés, la base de données comporte beaucoup de relevés autour d'une valeur moyenne de temps de parcours que l'on pourrait identifier comme temps de parcours en circulation fluide ou hors congestion. L'intérêt d'une carte de contrôle est que celle-ci permet de mettre alors en avant les relevés hors limites, c'est-à-dire les temps de parcours dont la valeur est élevée.

a. Principe d'une carte de contrôle

Une carte de contrôle permet, comme son nom l'indique, un contrôle pertinent de la valeur des relevés. Ceux-ci correspondent à l'axe des x sur la carte, on observe alors plusieurs droites parallèles à ce même axe :

- la moyenne ;
- la moyenne plus et moins l'écart-type ;
- la moyenne plus et moins trois écarts-types.

L'axe des y représente alors les valeurs des relevés.

Ce procédé s'appuie très fortement sur les propriétés de la loi normale et sur le théorème de la limite centrale. Ces derniers impliquent a priori que 70% des relevés doivent être compris entre la moyenne plus ou moins un écart-type et 95% des relevés doivent être compris entre la moyenne plus ou moins trois écart-type.

Constatons que les limites sont peu importantes dans notre cas ; en effet nous recherchons seulement à exposer les valeurs importantes.

b. Carte de contrôle pour les tronçons 7 à 10 durant la période matinale

Dans le cas de l'étude faite pour les tronçons 7 à 10, nous obtenons la carte de contrôle suivante pour la période du matin :

Tableau 3-1 Statistiques relatives aux créneaux horaires choisis

Statistique des créneaux horaires choisis

Nombre de relevés	**2584**	10ème percentile	**32.123**
Moyenne	**121.30**	90ème percentile	**269.95**
Ecart-type	**114.65**		
Minimum	**26.09**	TRI	**1.96**
Maximum	**1397.7**	TTI	**1.22**

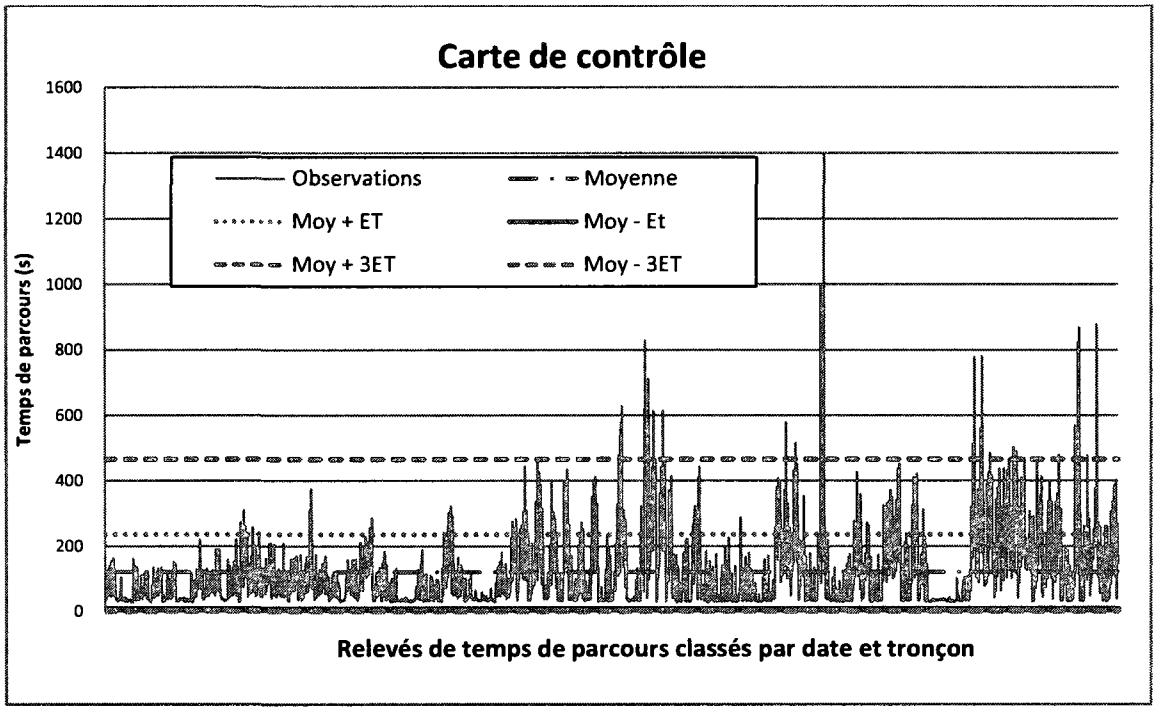


Figure 3-22 Carte de contrôle des relevés relatifs aux tronçons 7 à 10 pour les créneaux de 6h à 9h

L'observation de la carte de contrôle où les données ont été classées chronologiquement (Figure 3-22), confirme les très grandes fluctuations observées grâce au *TRI* et au *TTI*. On constate tout de même que ces fluctuations sont devenues de plus en plus importantes au cours des années. Il y a donc un grand nombre de relevés hors limites, c'est-à-dire dépassant la moyenne des temps de parcours plus un écart-type.

Tableau 3-2 Statistiques pour les tronçons 7 à 10, période du matin sans les relevés hors-limites

Statistique des créneaux horaires choisis sans les relevés hors limites			
Nombre de relevés	**2242**	10ème percentile	**31.88**
Moyenne	**85.98**	90ème percentile	**173.054**
Ecart-type	**55.84**		
Minimum	**26.09**	TRI	**1.64**
Maximum	**235.93**	TTI	**1.01**

Le Tableau 3-2 indique que 88% (2242/2584) des relevés sont compris dans un intervalle entre la moyenne plus ou moins un écart-type. L'hypothèse d'une ***circulation normale*** sur ces tronçons durant cette période semble alors se confirmer. Un des objectifs futurs sera donc de la caractériser, d'en voir et d'en peser les conséquences sur ce même circuit.

3.3.7 Observations selon les conditions de la chaussée

Une autre problématique importante de la circulation et de sa fiabilité est la signification des conditions de la chaussée sur le temps de parcours moyen sur un tronçon. Il a alors été développé après avoir confirmé cette signification (analyse de la variance), une interface permettant de comparer les temps de parcours moyens selon les conditions de la chaussée.

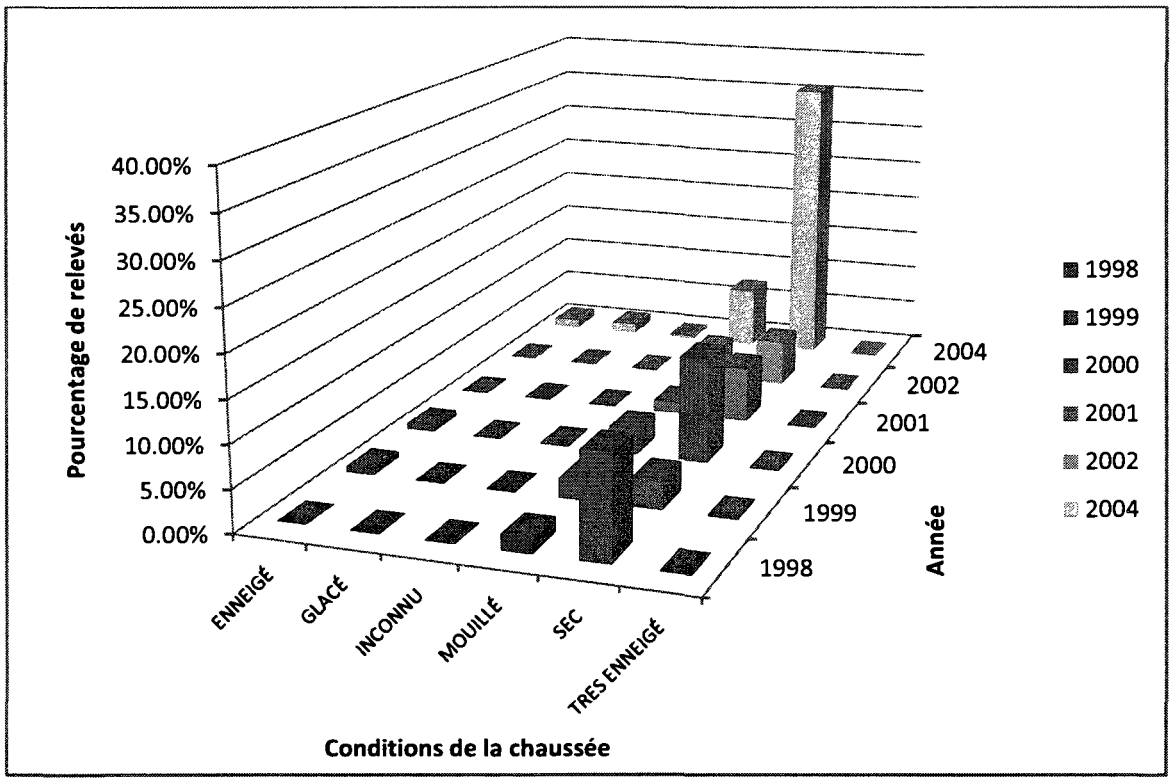


Figure 3-23 Pourcentage de relevés rencontrés par condition de la chaussée par année pour le circuit 1

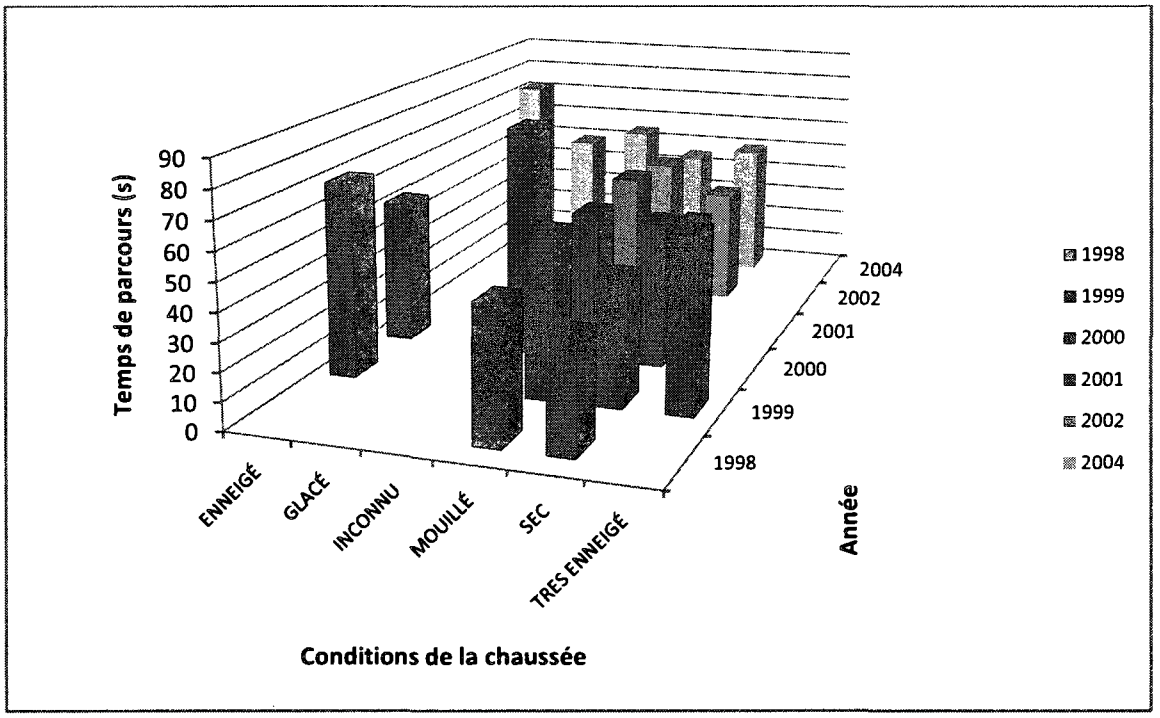


Figure 3-24 Temps de parcours moyens par conditions de la chaussée et par année pour le circuit 1

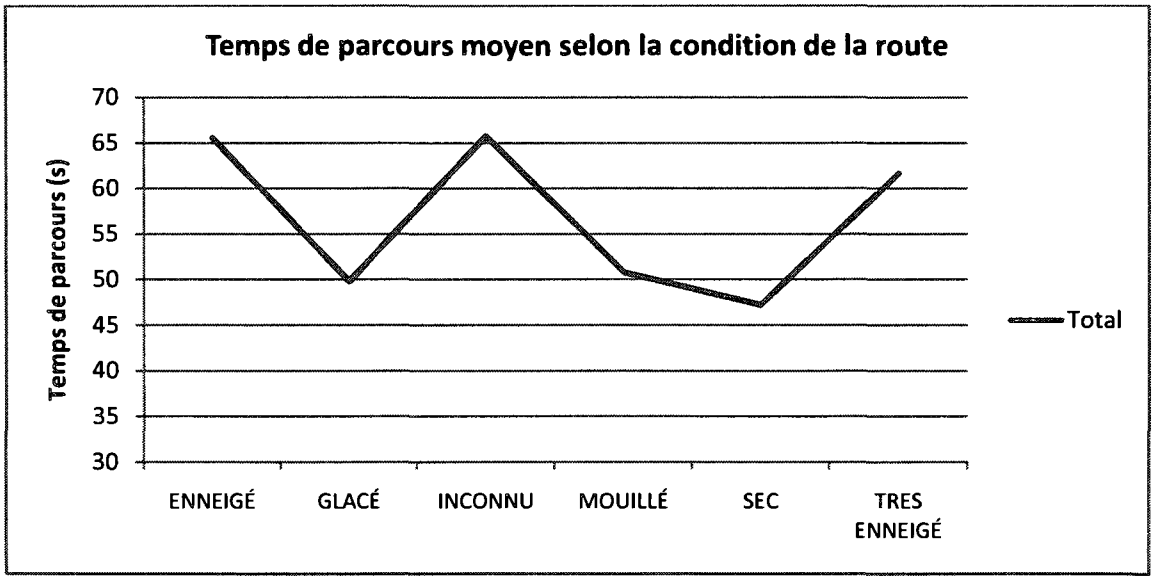


Figure 3-25 temps de parcours moyen selon la condition de la route pour le circuit 1

Nous constatons alors, que seulement lorsque le temps est enneigé, le temps de parcours moyen est alors plus élevé. Exposons tout de même que cette analyse est très restrictive. Premièrement, la qualité du déneigement implique que la chaussée est rarement enneigée (Figure 3-23). De plus, les valeurs des temps de parcours moyens par condition météorologique sont calculées en considérant tous les relevés de temps de parcours du circuit 1.

En conclusion de cette partie, nous venons de mettre en évidence la signification statistique de nombreux facteurs. De plus l'analyse exploratoire a permis de dégager en particulier l'influence plus importante sur le comportement des temps de parcours du circuit 1, de la période d'observation, influence liée à un problème de convergence de plusieurs paires origine-destination. Enfin, après observations des points chauds, et des cartes de contrôles, il devient intéressant de se poser les questions suivantes : les tronçons congestionnés n'ont-ils pas la même distribution fréquentielle de relevés de temps de parcours, peut-on dégager une circulation normale des temps de parcours et extraire la congestion récurrente de la congestion non-récurrente ?

3.4 Synthèse

Conclusions : Au cours de cette analyse exploratoire, plusieurs concepts statistiques ont été appliqués. Après une étude de l'ensemble des relevés de temps de parcours sur tout le réseau montréalais au moyen de moyenne et variabilité, l'ensemble des observations a été décomposé en sous-groupes correspondant aux relevés de temps de parcours par circuit. Une analyse statistique est alors nécessaire pour chaque circuit. Le cas du circuit 1 (Autoroute A13 Direction Sud) expose plusieurs conclusions propres à ce circuit : les différentes analyses ANOVA ont mis en exergue la signification de la différence entre niveaux de plusieurs facteurs tels que le mois de l'année, la période d'observation, ou bien encore les conditions de la chaussée. Attention tout de même à ne pas en conclure que les facteurs cités précédemment sont significatifs pour tous les circuits.

Enfin, les cartes de contrôle construites pour des tronçons du circuit 1 démontrent le contrôle de qualité et implicitement l'évolution de la variabilité des relevés de temps de parcours sur le circuit.

Perspectives : au moyen des cartes de contrôle ainsi que de l'analyse ANOVA, certaines tendances ont été exposées. Premièrement, on constate, dans le cas du circuit 1 qu'un très grand nombre de relevés de temps de parcours se situent dans un intervalle de temps de parcours. Deuxièmement, les distributions statistiques des relevés de temps de parcours entre différents tronçons semblent se correspondre.

En conclusion, la suite des travaux se divisera en deux. Le premier point devra relater de l'observation de similarité entre les distributions fréquentielles de relevés de temps de parcours de différents tronçons. Cette analyse permettra de déboucher sur une possibilité d'échantillonnage.

Le deuxième point devra développer les outils nécessaires à la mise en évidence de cette fenêtre temporelle – *ie* l'intervalle de temps où se situe une majeur partie des relevés de temps de parcours-. Cette nouvelle analyse passera par la modélisation de la distribution

fréquentielle des relevés de temps de parcours pour un circuit. De cette dernière, une simulation et le développement de plusieurs indicateurs deviendront réalisables.

CHAPITRE 4 : *DATA MINING*, GROUPEMENT DE TRONÇONS POUR UN CIRCUIT ET ÉCHANTILLONNAGE

Dans le cadre d'un échantillonnage des circuits autoroutiers montréalais pour la saison automnale et hivernale (mois de septembre à décembre) au moyen de nouvelles technologies telle que la technologie Bluetooth, il devient important d'observer les points chauds de la circulation autoroutière et d'en mesurer la portée, ainsi que de grouper les tronçons par circuit ayant le même comportement. En effet rappelons que dans l'exemple du circuit 1, certains tronçons ne semblent jamais congestionnés alors que d'autres le sont constamment.

L'étude menée dans cette partie concerne donc les observations des mois cités précédemment, selon la période d'observation (AM ou PM).

En considérant l'entité circuit (55 circuits sont donc disponibles à l'étude), il est intéressant d'observer les similarités comportementales entre les différents tronçons du circuit étudié. En effet, plusieurs tronçons semblent congestionnés durant une même période. Les temps de parcours y sont-ils alors proches, ou bien encore les distributions fréquentielles des relevés de temps de parcours sont-elles semblables ? Par conséquent, au lieu de s'orienter sur l'analyse statistique des observations, nous préfèrerons une analyse mathématique basée sur les fréquences des observations dont nous donnons les détails dans la partie suivante.

4.1 Analyse fréquentielle et *clustering*

Choisissons $tr_{j,c}^{p}$ le tronçon j du circuit c, pour la période d'observation $p = AM$ ou PM, et soit un relevé de temps de parcours t sur ce tronçon.

On définit le vecteur $T = [t_1, t_2, \ldots, t_m] = [20, 25, \ldots, 360]$. Par conséquent, il existe t_i et t_{i+1} telle que t appartienne à $[t_i ; t_{i+1})$, où $|t_i - t_{i+1}| = 5s$ si $t < 360s$ (vitesse de 10km/h) ou $[360 ; \infty[$ si $t > 360s$.

Soit $n_{j,c}{}^{p}$ le nombre total d'observations pour le tronçon $tr_{j,c}{}^{p}$, on désignera alors par

$$f^{p}_{j,c,t1,t2} = \frac{n^{p}_{j,c,t1,t2}}{n^{p}_{j,c}},$$

où $n_{j,c,t1,t2}{}^{p}$ désigne le nombre d'observations entre $[t_1 ; t_2)$, pour le tronçon j du circuit c durant la période p, la fréquence d'observations du tronçon $tr_{j,c}{}^{p}$, comprises entre $[t_1 ; t_2]$. Par conséquent, il vient que $\sum_{k\in[\![1;m]\!]} f^{p}_{j,c,t(k),t(k+1)} = 1$.

Par exemple dans le cas du tronçon 1 du circuit 1 nous observons la distribution suivante :

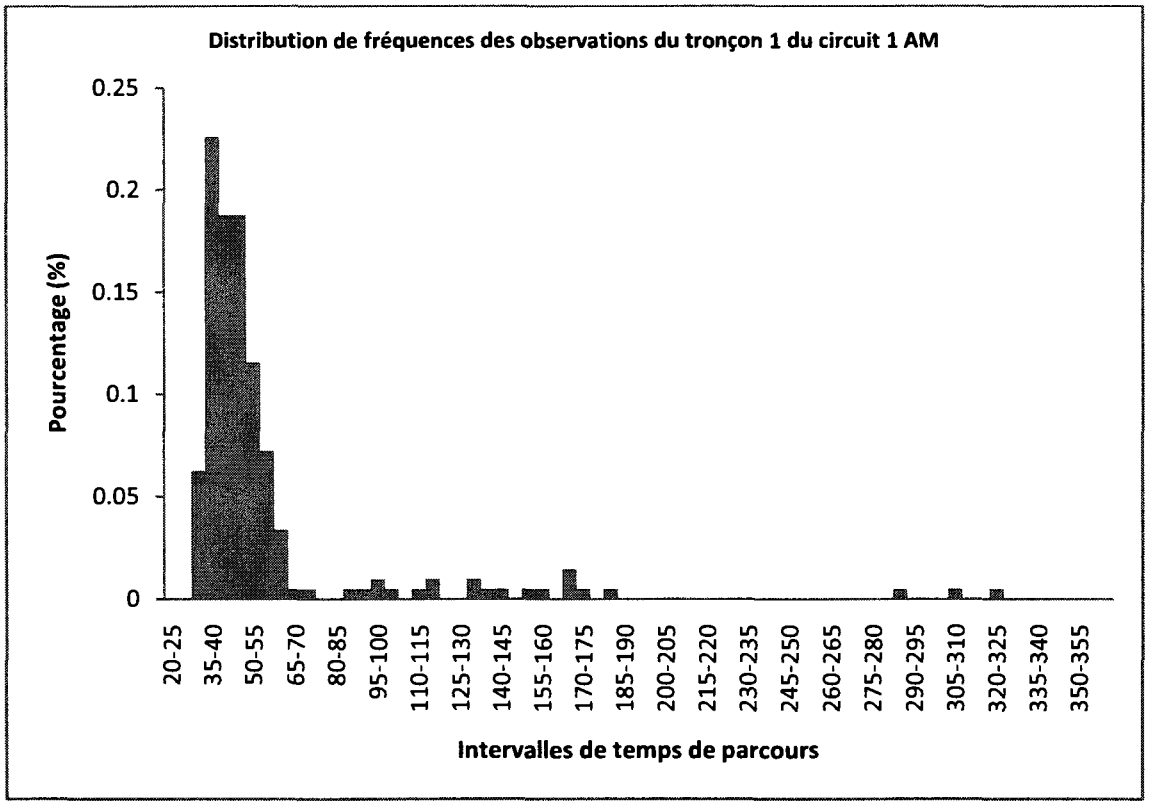


Figure 4-1 Distribution de fréquences des observations du tronçon 1 du circuit 1 AM

Ont été choisis des intervalles de 5 secondes pour obtenir une bonne précision, en effet, une différence de 5 secondes entre deux observations correspond à une différence de vitesse de *5/3.6 = 1.4 km/h.*

4.1.1 Analyse de la corrélation

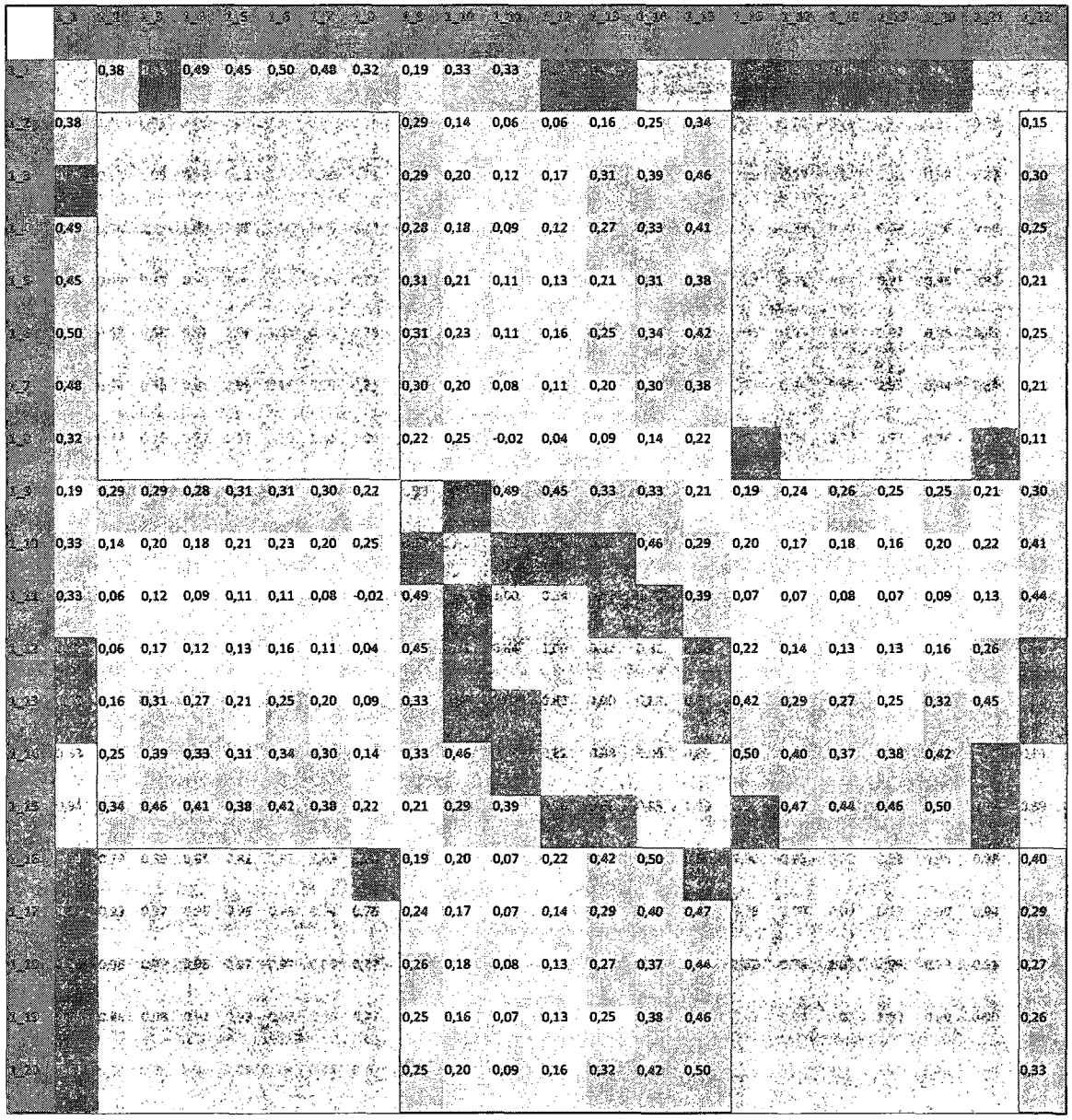
Dans un but exploratoire, l'analyse de la corrélation de la distribution des fréquences entre les tronçons d'un même circuit peut apparaître très pertinente. Cette analyse ne permet pas par contre de dégager ni un groupement statistiquement viable des tronçons du circuit, ni une idée du comportement de ces tronçons.

On rappelle la formule générale de la corrélation entre deux séries de valeurs x_1 et x_2. Si (x_1^i) et (x_2^i) désignent les éléments de chaque série, alors la corrélation entre ces deux séries est :

$$corr(x_1, x_2) = \frac{\sum_i (x_1^i - \bar{x}_1).(x_2^i - \bar{x}_2)}{\sqrt{\sum_i (x_1^i - \bar{x}_1)^2 \sum_i (x_2^i - \bar{x}_2)^2}}$$, où $\overline{x_k}$ désigne la moyenne des (x_k).

Ainsi, plus les distributions de deux séries sont proches ou antonymes, plus ce coefficient en valeur absolue sera proche de 1. L'indépendance entre ces deux séries signifiera une valeur nulle du coefficient de corrélation. Dans le cas du circuit 1, voici le tableau de corrélation des distributions fréquentielles des divers tronçons du circuit 1.

Tableau 4-1 Tableau de corrélation des fréquences AM des tronçons du circuit 1



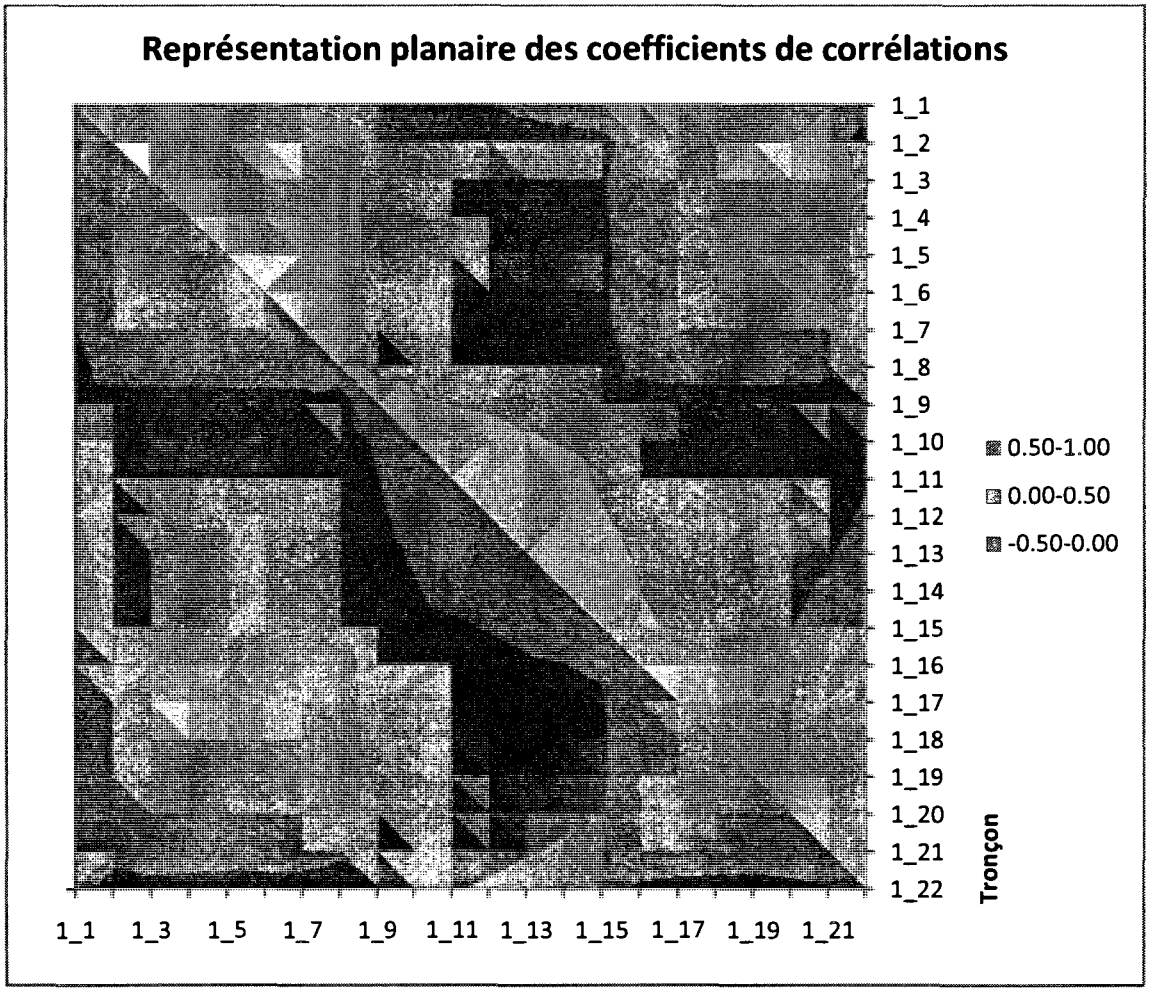


Figure 4-2 Carte de corrélation des distributions fréquentielles des tronçons du circuit 1

Nous constatons ainsi que certains des tronçons du circuit 1 présentent de très fortes corrélations (souvent voisines de 1) ce qui signifie que justement leurs distributions fréquentielles sont très proches l'une de l'autre. Déjà nous pouvons voir plusieurs groupes se dessiner sur la représentation planaire des coefficients de corrélations. Les tronçons 1 à 6 sont très corrélés, de même pour les tronçons 10 à 15 et pour les tronçons 17 à 20. Il semble donc se dégager des comportements particuliers propres des groupes identifiés a priori. En effet, certains comportements avaient été observés pour ces tronçons lors de la précédente partie, en particulier, le caractère de congestion ou de fluidité qui ressort de cette analyse.

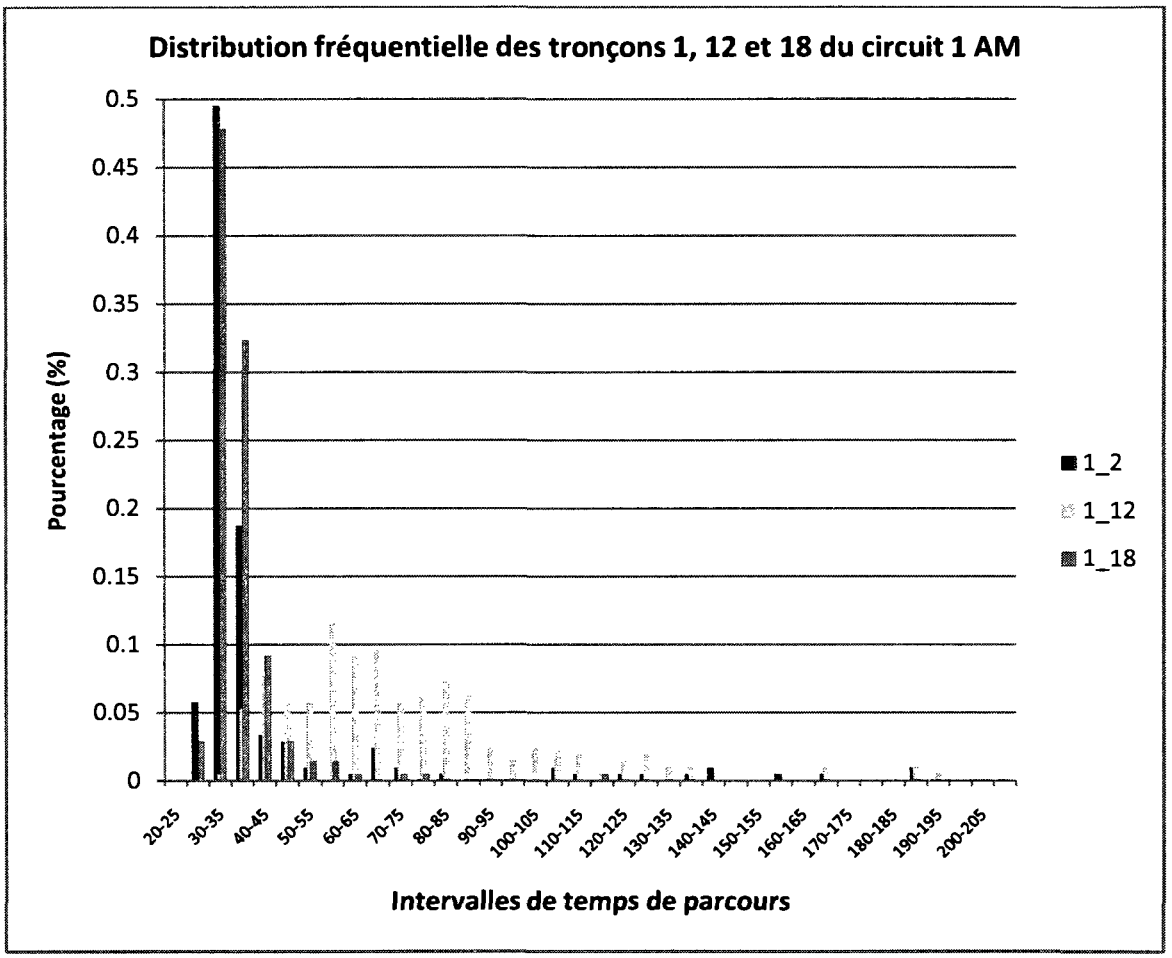


Figure 4-3 Carte de corrélations et distribution des fréquences pour les tronçons 1_2, 1_12 et 1_18

Dans cet exemple, nous confirmons bien l'idée que les distributions fréquentielles des relevés de temps de parcours des tronçons 1 et 18 sont très voisines, alors que celle du tronçon 12 est bien plus étalée.

Par conséquent, on peut donc mettre en avant une possible forme de groupement des tronçons pour chaque circuit. Mais il est alors nécessaire de la justifier d'un point de vue mathématique.

4.1.2 Groupement (Algorithme de *Clustering*)

Cette partie s'inscrit dans la perspective de groupement des tronçons d'un même circuit. Il est basé sur un algorithme de groupement issu du *data mining*.

4.1.2.1 Algorithme

Soit $(x_1, x_2, \dots, x_n)$ un ensemble de vecteurs-observations. On désignera alors $x_i^T = (x_i^1, x_i^2, \dots, x_i^m)$, le vecteur de dimension m correspondant aux valeurs définissant le vecteur-observation x_i.

On rappelle que la distance euclidienne entre deux vecteurs x et y est égal à :

$$\|x - y\| = \sqrt{\sum_{k=1}^{m} (x_k - y_k)^2}$$

On définit alors une valeur réelle strictement positive δ.

La construction des différents groupes s'effectue alors ainsi :

- Soit Ω_1, le groupe singleton comprenant le vecteur x_1 et de centre $C_1 = x_1$.
- Comparons la distance euclidienne entre $\|x_1 - x_2\|$:
    - Soit $\|x_1 - x_2\| \leq \delta$, dans ce cas alors on inclut dans le groupe Ω_1 le vecteur x_2 et le centre C_1 possède alors les nouvelles coordonnées $\left[\frac{1}{2}\left(x_1^i + x_2^i\right)\right] = \frac{1}{2}(x_1 + x_2)$.
    - Soit $\|x_1 - x_2\| > \delta$, alors on définit le nouveau groupe Ω_2 de centre $C_2 = x_2$,
- Ainsi de suite, pour les *n-2* autres vecteurs.

Le résultat conséquent est la création d'un nombre de groupe distincts (Ω_k). Le groupe k-ième comprend alors l observations $(x_{\Omega(1)}, x_{\Omega(2)}, \dots, x_{\Omega(n)})$ de centre C_k de coordonnées $C_k^i =$

$\frac{1}{l}\sum x^i_{\Omega(s)}$ et est tel que quelque soit le vecteur-observation x élément de Ω_k, alors $\|x - C_k\| = \sqrt{\sum(x^i - C^i_k)^2} \leq \delta$.

4.1.2.2 Présentation des résultats

L'application de l'algorithme a donc été effectuée pour chacun des 55 circuits selon la période d'observation (AM/PM). Cette analyse permet donc de légitimer les possibilités de groupements observés par l'analyse de la corrélation. Dans le cas du circuit 1 nous trouvons les résultats suivants présentés en annexe, où les colonnes portant le titre *essaim(i)* sont les valeurs des centres des essaims, et les colonnes *éléments(i)* désignent les tronçons du groupe *i*. La Figure 4-4 expose les groupes créés par l'algorithme. Nous retrouvons donc bien les conclusions spéculatives de la partie précédente basées sur l'étude de la corrélation de la distribution fréquentielle des relevés de temps de parcours. Les tronçons 10à 14 sont en particuliers ensemble, de même pour 16 à 20 et enfin 2 à 6.

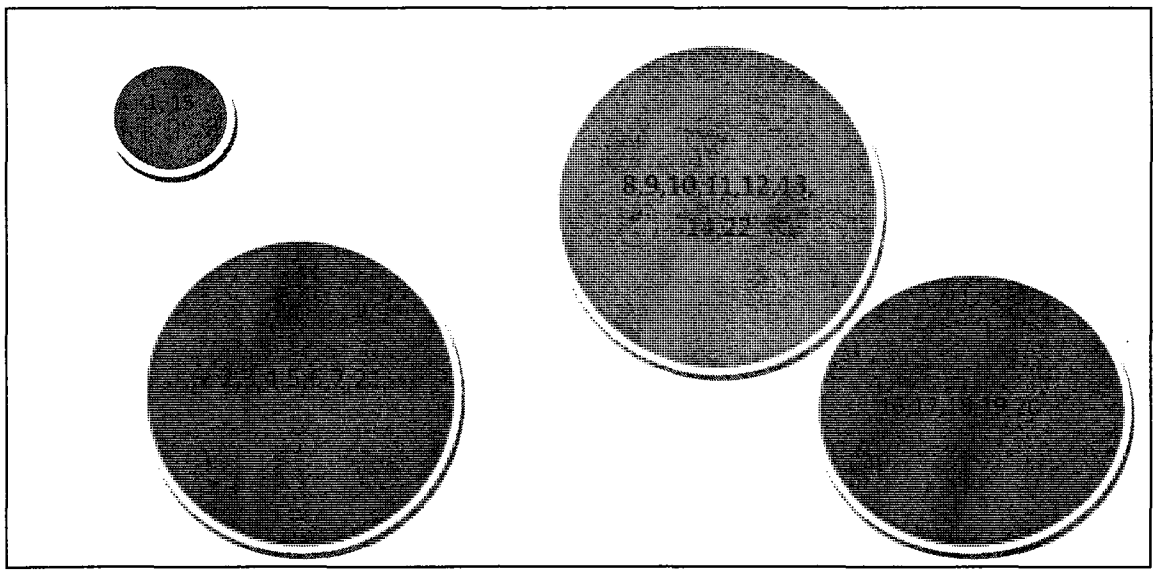


Figure 4-4 Groupes de tronçons du circuit 1 AM

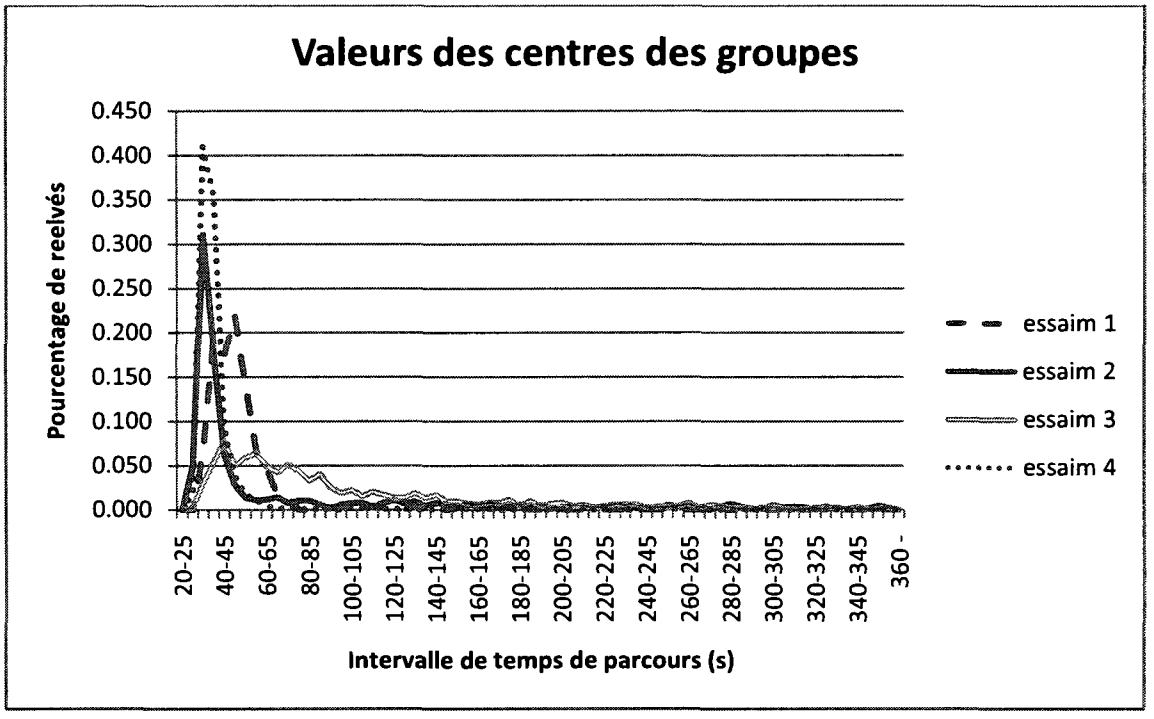


Figure 4-5 Distributions fréquentielles des centres des groupes

D'après la Figure 4-5, nous constatons donc que les quatre groupes présentent bien des centres de natures différentes. On observe en particulier concernant le groupe 3, un plus grand étalement des fréquences. Ceci signifie que les tronçons de ce groupe sont sujets à de fortes congestions et à une variabilité importante, alors que les tronçons de l'essaim 1 sont sujets à une congestion sans variabilité puisque l'étalement y est moindre. D'un autre côté les essaims 2 et 4 correspondent à des groupes bénéficiant d'une importante fluidité vis-à-vis des deux autres groupes.

4.2 Étude de la variabilité

Dans un second temps, compte-tenu des impératifs de la recherche, il est important de qualifier les tronçons de chacun des circuits. Le coefficient de variation v est défini comme le rapport entre l'écart-type et la moyenne et semble être un bon indicateur des points chauds et donc de la fiabilité de la circulation. On peut constater de plus que cet indicateur se rapproche

des deux indicateurs (*TRI* et *TTI*) que nous avons utilisés dans la partie précédente. Nous ne considérons, en perspective de la prochaine partie concernant l'échantillonnage, de ne considérer que les relevés des mois de septembre à décembre afin de calculer les valeurs de cet indicateur de fiabilité.

Autrement écrit, si $X = (x_1, x_2, \ldots, x_n)$, alors la variabilité v (CV) de cette série d'observations vaut

$$v = \frac{S}{m} = \frac{n}{\sqrt{n-1}} \cdot \frac{\sqrt{\sum_{i=1}^{n}(x_i - m)^2}}{\sum_{i=1}^{n} x_i}$$

Où m et S désignent respectivement la moyenne et l'écart-type des observations.

Les deux figures suivantes (Figure 4-6 et Figure 4-7) présentent la variabilité observée sur le circuit 1.

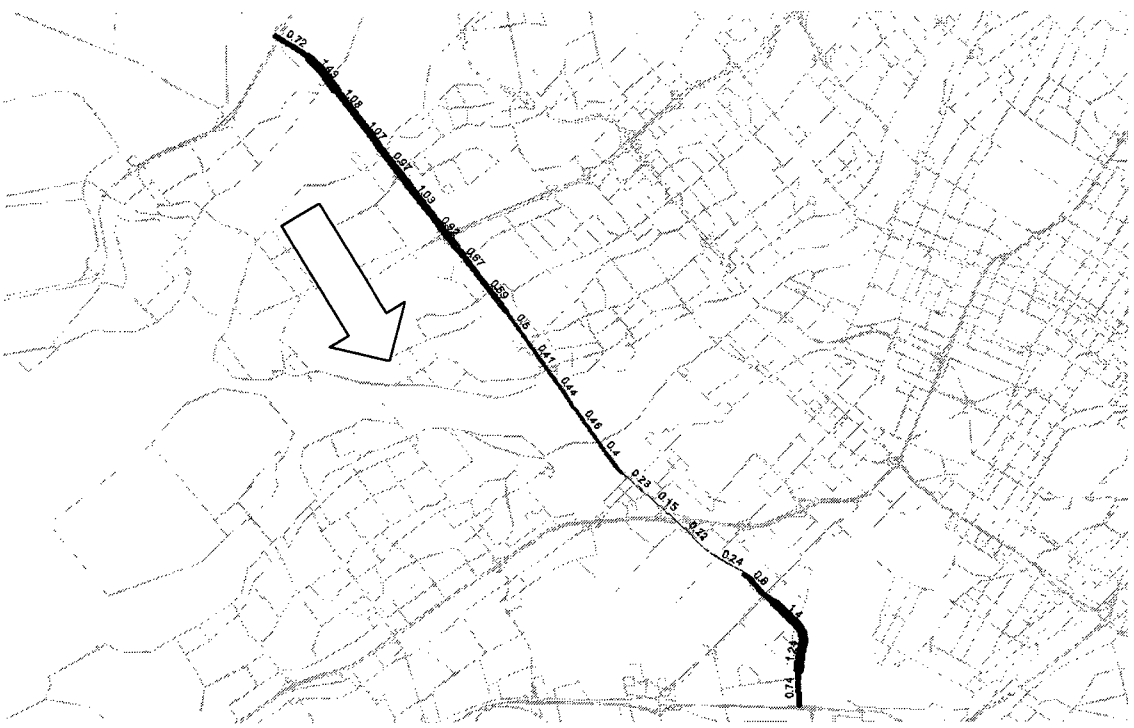


Figure 4-6 Variabilité AM le long du circuit 1

Le coefficient de variation sera nommé dans la suite du mémoire « variabilité » des relevés de temps de parcours

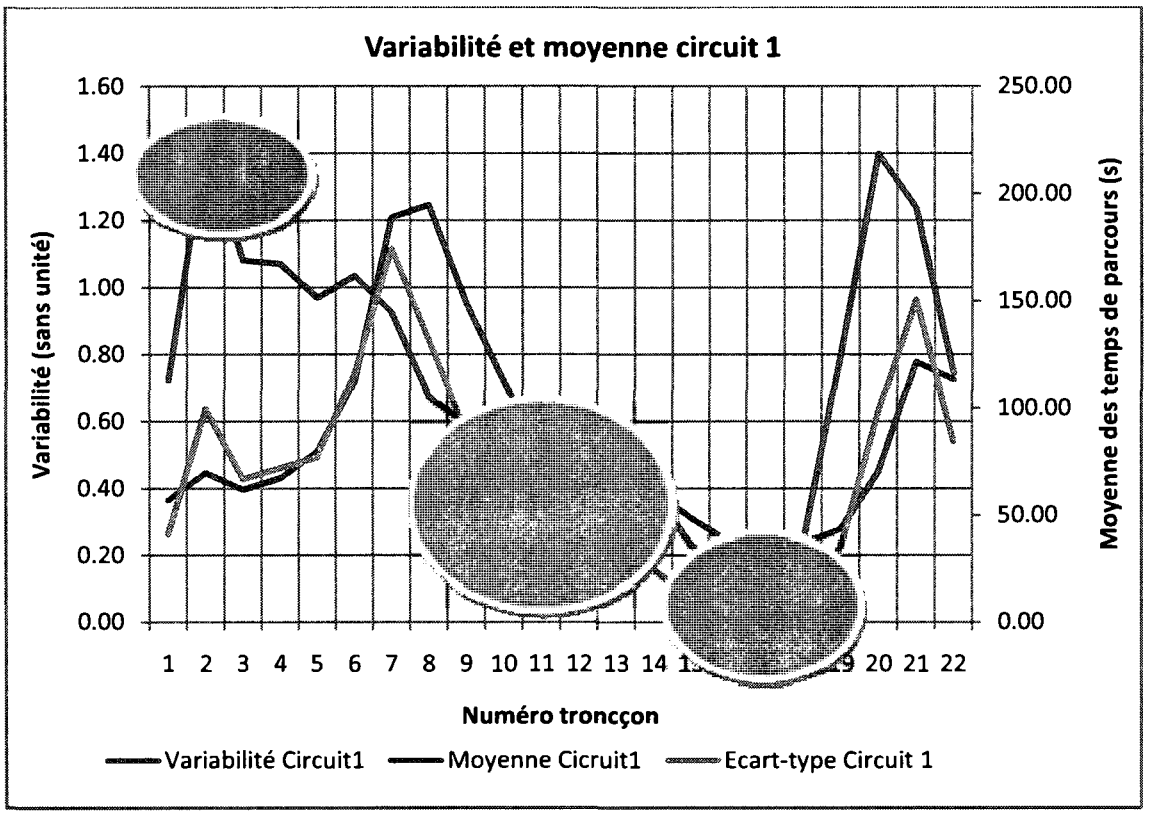


Figure 4-7 Variabilité et moyenne des tronçons du circuit 1

En conclusion, les différents groupes mis en avant dans la partie précédente sont donc visibles dans la Figure 4-7. On retrouve aussi les conclusions faisables dès le groupement des tronçons. En clair, nous venons de conforter les résultats.

Cette analyse complète est d'un point de vue pratique particulièrement intéressante. On entend par là que de nouveaux relevés peuvent être ajoutés aux bases de données et donc que l'on peut réactualiser les résultats pour surveiller l'état et l'évolution de la circulation sur les circuits autoroutiers montréalais et donc effectuer de nouveau un *clustering*.

En effet, les données recueillies par la technologie Bluetooth par exemple, demeurent très proches de celles prises grâce à la méthode actuelle à savoir la technique de relevés odométriques.

4.3 Synthèse

Conclusions : Au moyen de l'analyse du tableau de corrélation et du *clustering* des distributions fréquentielles de relevés de temps de parcours par tronçon du circuit 1, certaines parties du circuit se comportent de la même manière. Il a été mis en avant plusieurs groupes : des groupes de tronçons sans congestion, un peu congestionnés, complètement congestionnés.

La variabilité par tronçon rend compte alors de ce l'étalement de chacune des distributions de relevés de temps de parcours, et expose donc la présence ou non de fiabilité du temps de parcours d'un tronçon et a posteriori, du circuit.

Perspectives : Grâce aux résultats du *clustering* et l'étude de la variabilité, il devient désormais possible de concevoir un échantillonnage cohérent ; la partie suivante portera sur ce propos. Enfin, le *clustering* et donc l'observation de la distribution des relevés de temps de parcours de chaque groupe de tronçons exposent encore des pics d'observations dans une plage de temps bien définie.

Nous parlerons alors désormais de circulation normale pour définir cette plage temporelle non encore mathématiquement définie.

CHAPITRE 5 : ÉCHANTILLONNAGE

Connaissant désormais les variabilités et moyennes des relevés de temps de parcours des tronçons ou des groupes de tronçons de chaque circuit, un échantillonnage devient réalisable. Celui-ci sera donc dépendant de ces deux indicateurs.

D'un point de vue théorique supposons que nous disposions d'une base de données initiale contenant N^i observations pour un tronçon i. L'objectif de cette partie est de donner une idée de l'échantillonnage à produire selon la variabilité observée. Chaque tronçon est caractérisé par une moyenne, un écart-type et par conséquent par une variabilité. Ces trois données bien qu'expérimentales, seront toutefois considérées comme l'initialisation d'un futur échantillonnage.

On se contente dans notre cas de n'étudier que les mois de septembre à décembre.

5.1 Aspects théoriques

Soit v la variabilité observée d'un tronçon i.

Soit $(x)_n$ n nouveaux échantillons, définissons par $f = n/N^i$ le ratio entre le nombre de nouveaux échantillons et le nombre d'observations actuellement disponibles pour ce tronçon.

Soit D la différence relative entre moyenne observée et moyenne des nouveaux échantillons. On désire que la variabilité de notre nouvel échantillon vérifie $kV' = D$, où k correspond à un facteur de qualité. Il équivaut donc à l'écart maximal de variabilité acceptable.

$$\text{Sachant que } V'^2 = (1-f)\cdot\frac{v^2}{n}$$

$$\text{Par conséquent} \quad \frac{D^2}{k^2} = (1-f)\frac{v^2}{n} = \frac{N^i-n}{N^i}\frac{v^2}{n}$$

$$\text{Ainsi} \quad n = \frac{N^i k^2 v^2}{k^2 v^2 + N^i D^2}.$$

Dans le cas du tronçon 1 du circuit 1, nous disposons de 208 relevés, et d'une variabilité égale à *0.72*. Pour $k = 3$, et $D = 0.05$, il vient alors que la taille de l'échantillon nécessaire sera : $n = 187$ échantillons. Dans le cas où l'on ne désire qu'une précision de 10% alors n = 143 relevés.

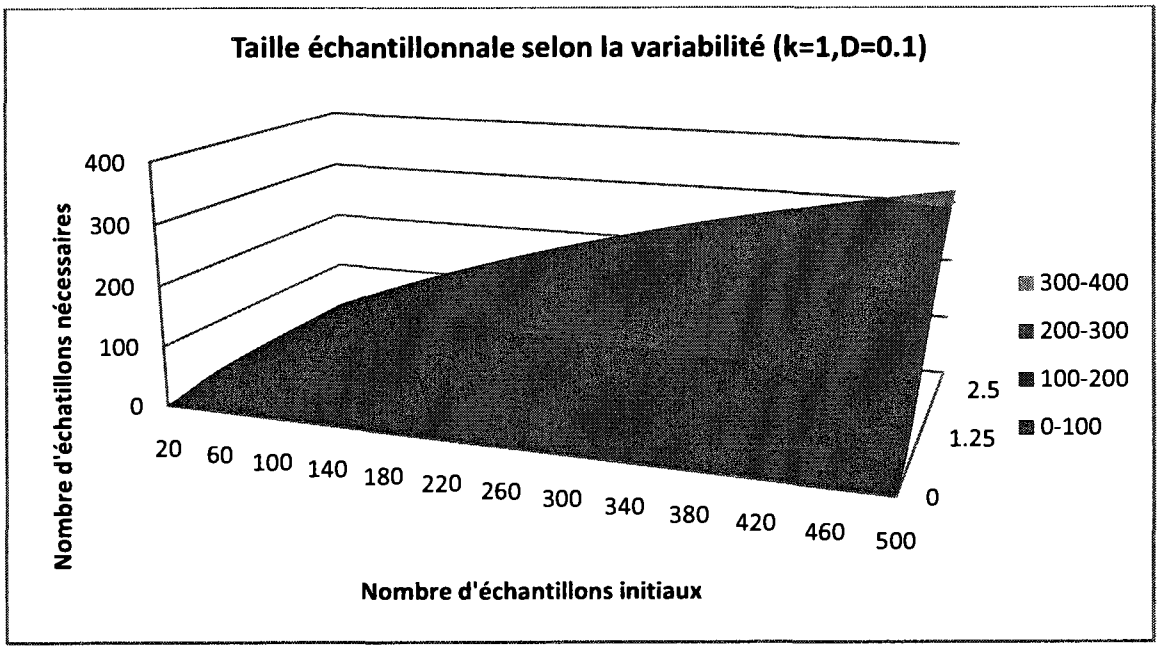


Figure 5-1 Taille de l'échantillon selon le facteur de qualité k =1

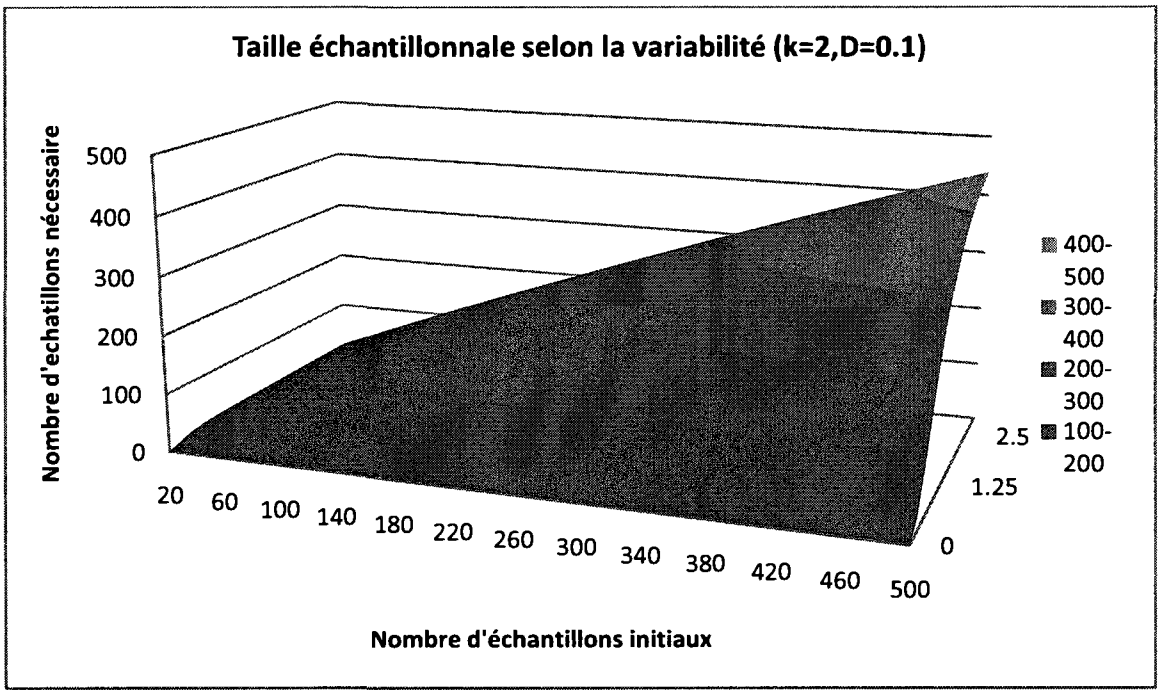


Figure 5-2 Taille de l'échantillon selon le facteur de qualité k =2

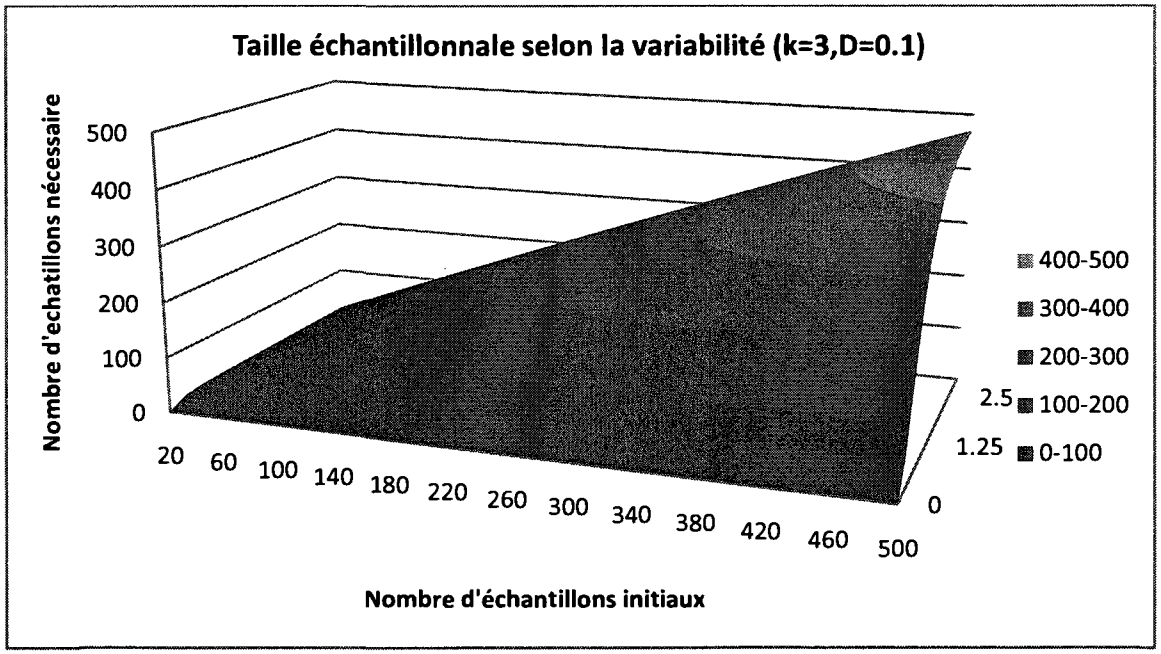


Figure 5-3 Taille de l'échantillon selon le facteur de qualité k =3

On constate donc que dès que les trois facteurs k, N, *et* v augmentent alors la taille de l'échantillon requis augmente et tend vers N. Il serait donc assez utile de relier certains tronçons dont les comportements sont similaires.

En effet, dans le cas où alors le nombre d'observations N devient très grand (agrégation de tronçons), il vient par passage à la limite : $n = \frac{k^2V^2}{D^2} + O(\frac{1}{N})$ (cf. Tableau 5-1)

Pour $k = 1$, on obtient alors les possibilités suivantes si $D=0.1$:

Tableau 5-1 Echantillons nécessaires pour N grand

Variabilité	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75	3
Echantillons nécessaires	7	25	57	100	156	225	307	400	507	625	757	900

5.2 Exemple pratique : cas du circuit 1

Nous allons observer les types de résultats obtenus lors de l'expérimentation concernant le circuit 1. On distinguera alors, les résultats avec groupement (ag) et les résultats sans groupement (sg). En effet, comme cité dans la partie précédente, si le nombre d'échantillons N augmente, alors le nombre de nouveaux échantillons à produire décroît.

En considérant les résultats du *clustering* (cf *DATA MINING*, GROUPEMENT DE TRONÇONS POUR UN CIRCUIT ET ÉCHANTILLONNAGE), les tronçons d'un même groupe suivent approximativement la même distribution de fréquences. Donc au lieu de donner un échantillon de chaque tronçon, il semble judicieux de répartir de manière équivalente le nombre d'échantillons à effectuer sur le centre du groupe selon sa variabilité sur les différents tronçons de ce groupe.

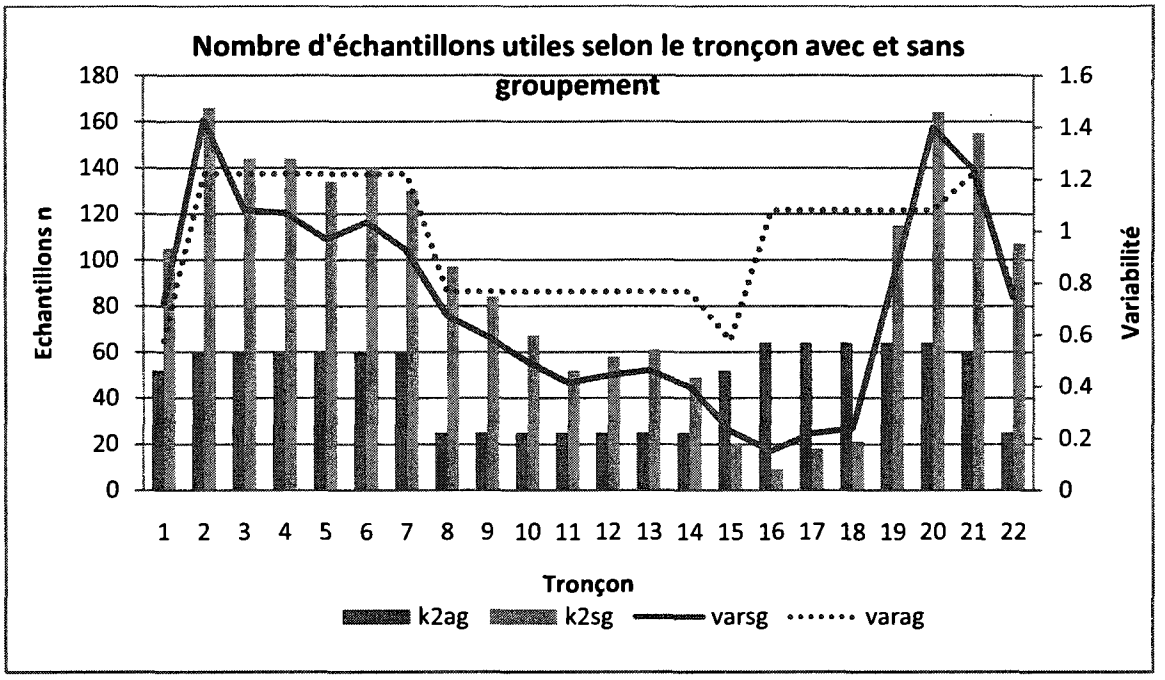


Figure 5-4 Nombre d'échantillons nécessaires pour le circuit 1 AM, k=2

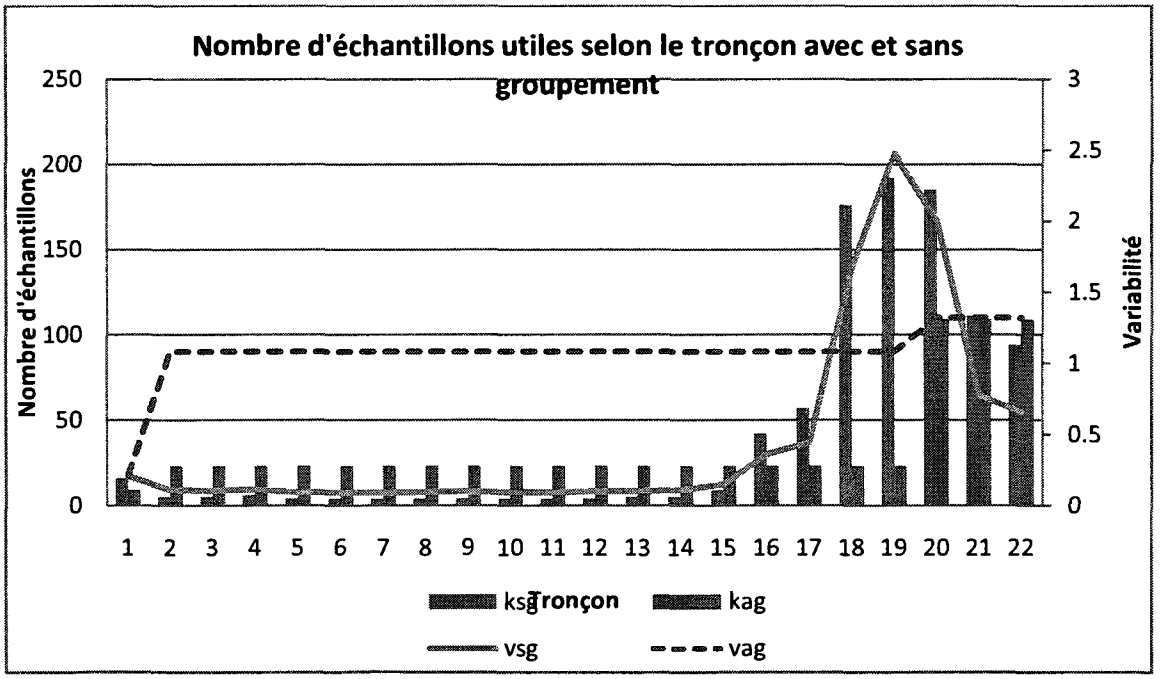


Figure 5-5 Nombre d'échantillons nécessaires pour le circuit 1 PM

Nous constatons que le principe permet alors de largement diminuer le nombre de relevés à effectuer. Tous tronçons confondus pour la période *AM*, pour $k = 2$, si nous ne considérons pas le groupement nous avons 2040 relevés à effectuer, alors qu'avec groupement seulement 1044 relevés sont nécessaires. On constate que cette analyse apporte de plus un côté pratique non négligeable en raison de la méthode de collecte de données (voiture flottante). En effet, il sera plus évident d'envoyer le véhicule sur tout le circuit plutôt que de lui faire effectuer des trajets de seulement 1 km afin de surveiller la variabilité d'un seul tronçon. Pour la période PM par exemple (Figure 5-5), seulement 23 passages seront nécessaires pour les tronçons 1 à 19, alors que si nous avions effectué un échantillonnage pour chaque tronçon, la voiture flottante aurait dû passer 192 fois sur le tronçon 19 !

5.3 Échantillonnage à la semaine

Dans l'analyse exploratoire, nous avions mis statistiquement en avant la signification du facteur journée de la semaine (Figure 5-6). En effet, le comportement des gens semblent différer entre un mercredi normal et un vendredi normal. Il devient donc intéressant de choisir l'échantillonnage nécessaire possible selon la journée.

Dans la méthode précédente, il est important de considérer que le nombre d'observations nécessaires par journée et par tronçon dépend fortement des variabilités observées par tronçon et par journée, et aussi du **nombre initial d'observations** par tronçon et par jour. Ce dernier paramètre peut être responsable d'un effet pervers qui mènerait à un échantillonnage faussé. Nous proposerons alors une méthode dans la partie suivante pour pallier ce problème.

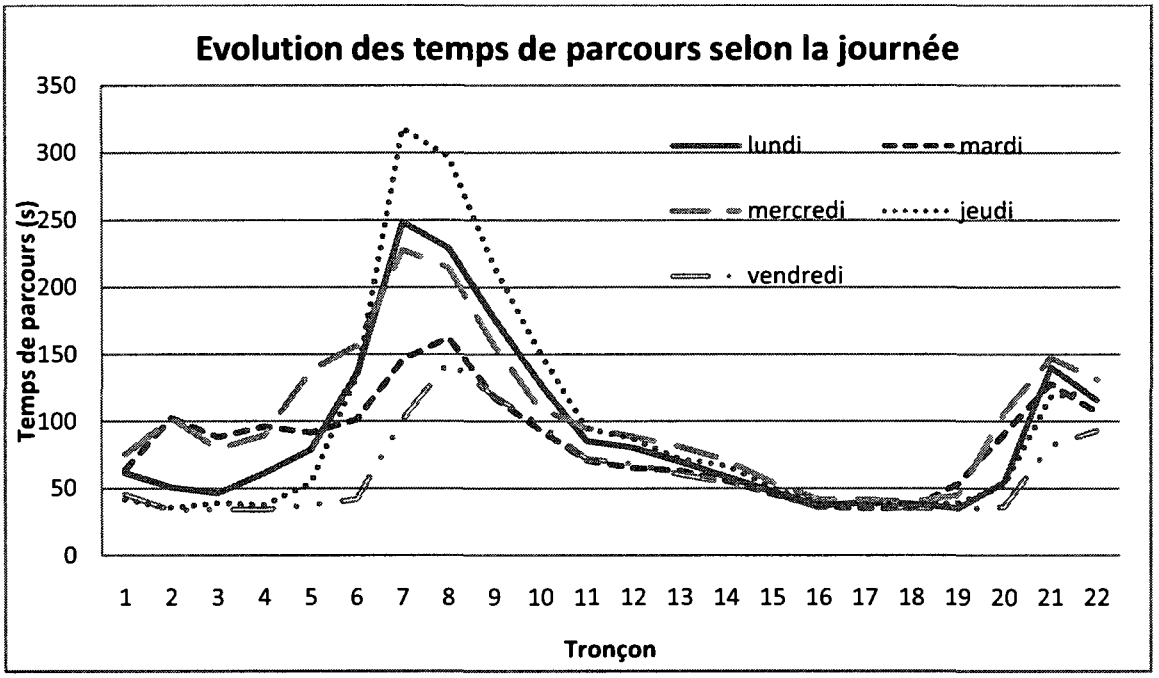


Figure 5-6 Evolution des temps de parcours par tronçon et par journée

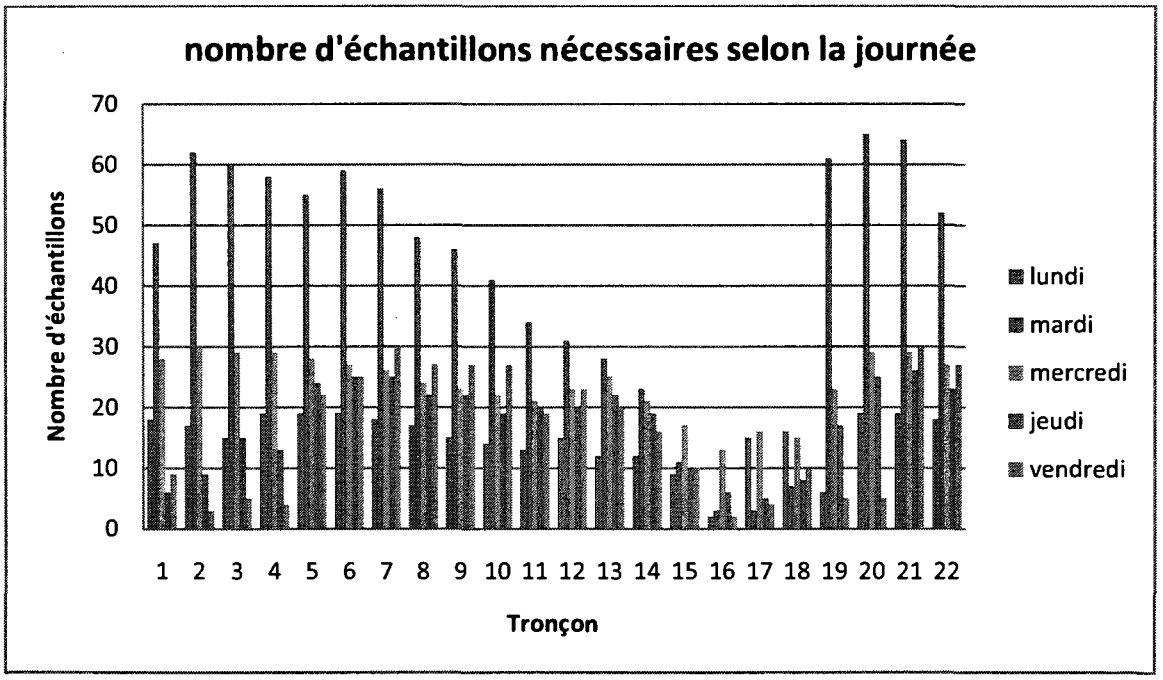


Figure 5-7 Nombre d'échantillons par tronçon selon la journée

On confirme bien cette idée par le fait que le mardi doit être plus échantillonné simplement parce qu'il a été initialement plus échantillonné que les autres journées (nous disposons de plus d'observations pour cette journée que pour les autres). La méthode ne peut donc s'appliquer directement.

5.4 Alternative d'échantillonnage

On peut voir alors le problème différemment. Nous disposons par exemple d'un « portefeuille » de n'= 1044 observations à réaliser, il faut alors les répartir de manière logique, selon la variabilité de la circulation journalière. Nous allons donc reprendre l'équation de la taille de l'échantillon :

$$n_j^i = \frac{N_j^i k^2 v_j^2}{k^2 v_j^2 + N_j^i D^2}$$, où j désigne le jour de la semaine

Nous supposerons alors dans ce cas, que chaque tronçon a été échantillonné autant de fois tous les jours de la semaine, et autant de fois que tous les autres tronçons. En conséquence nous ne connaissons pas le $N_j^i = N'$ mais nous savons par contre que $\sum_i \sum_j n_j^i = n' = 1044$.

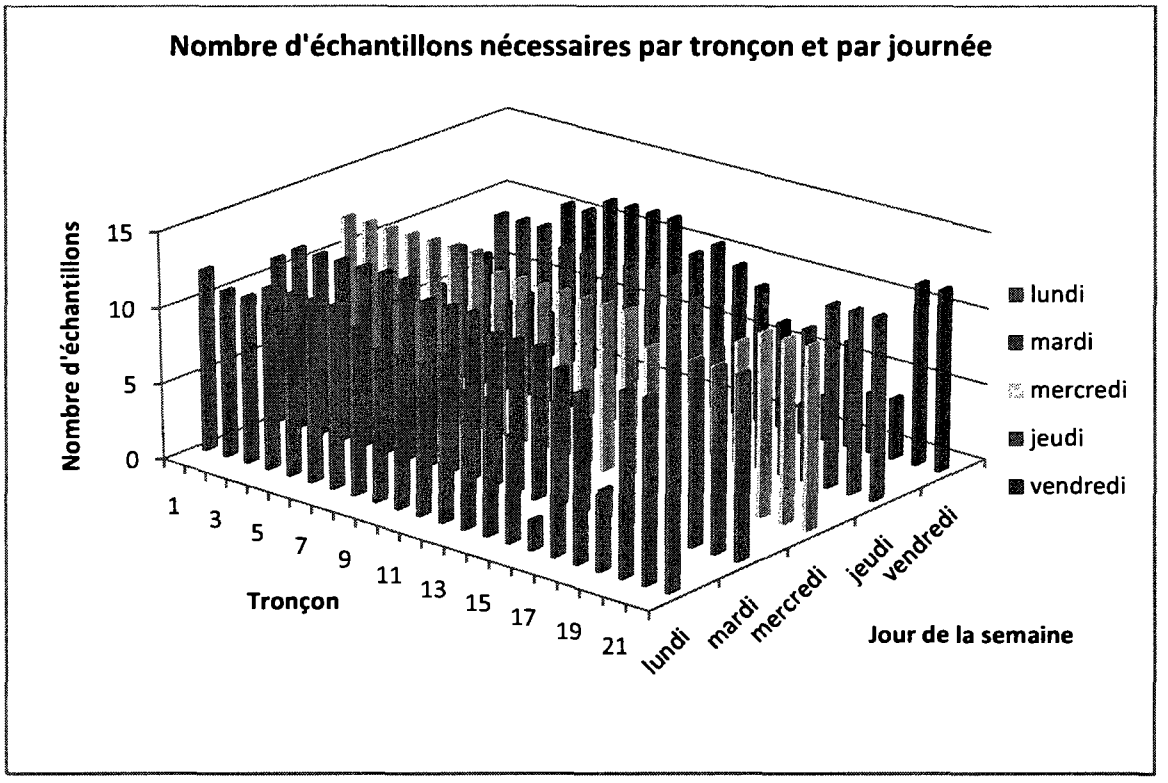


Figure 5-8 Répartition appropriée des observations selon le jour de la semaine sans groupement

Cette méthode alternative permet donc d'uniformiser l'échantillonnage et donc de ne pas léser certaines journées. Mais dans ce cas, on pourrait perdre de l'information concernant la congestion spécifique à une journée.

Toutefois, selon le gestionnaire, il est tout à fait concevable de préférer conserver un échantillonnage plus important pour les jours de fortes congestions mais alors l'évolution de la circulation sur le circuit sera moins représentative.

Il en est alors de la responsabilité du gestionnaire de décider de quelles mesures il a besoin. Précision ou comportement global...

5.5 Échantillonnage sur tout le réseau montréalais

On peut aussi généraliser la méthode d'échantillonnage et l'appliquer aux 55 circuits étudiés. Toutefois, il se repose le même problème que celui rencontré pour un échantillonnage selon le jour de la semaine. En effet, tous les circuits n'ont pas été échantillonnés le même nombre de fois.

Par conséquent l'application brute de la méthode lèserait un circuit où le nombre de relevés est faible face à un circuit très échantillonné. En effet, ceci impliquerait un faible nombre d'échantillons à produire pour le premier circuit face à un grand nombre d'échantillons pour le deuxième.

Il faut donc encore une fois répartir de manière équitable les échantillons dont on dispose. Dans ce cas, c'est encore une fois le nombre initial d'observations qui sera considéré comme constant et dont on cherche la valeur pour qu'alors, si n' est égal au nombre total d'échantillons que l'on peut produire,

$\sum_{circuit\ j=1}^{55} \sum_{troncon\ i=1}^{k_j} n_j^i = n'$, où k_j désigne le nombre de tronçon du circuit j.

Voici alors l'échantillonnage possible, dans le cas du circuit 1 pour la période du matin, si nous disposons d'un portefeuille de 25000 relevés soit un nombre de relevés N_j^i = 2000.

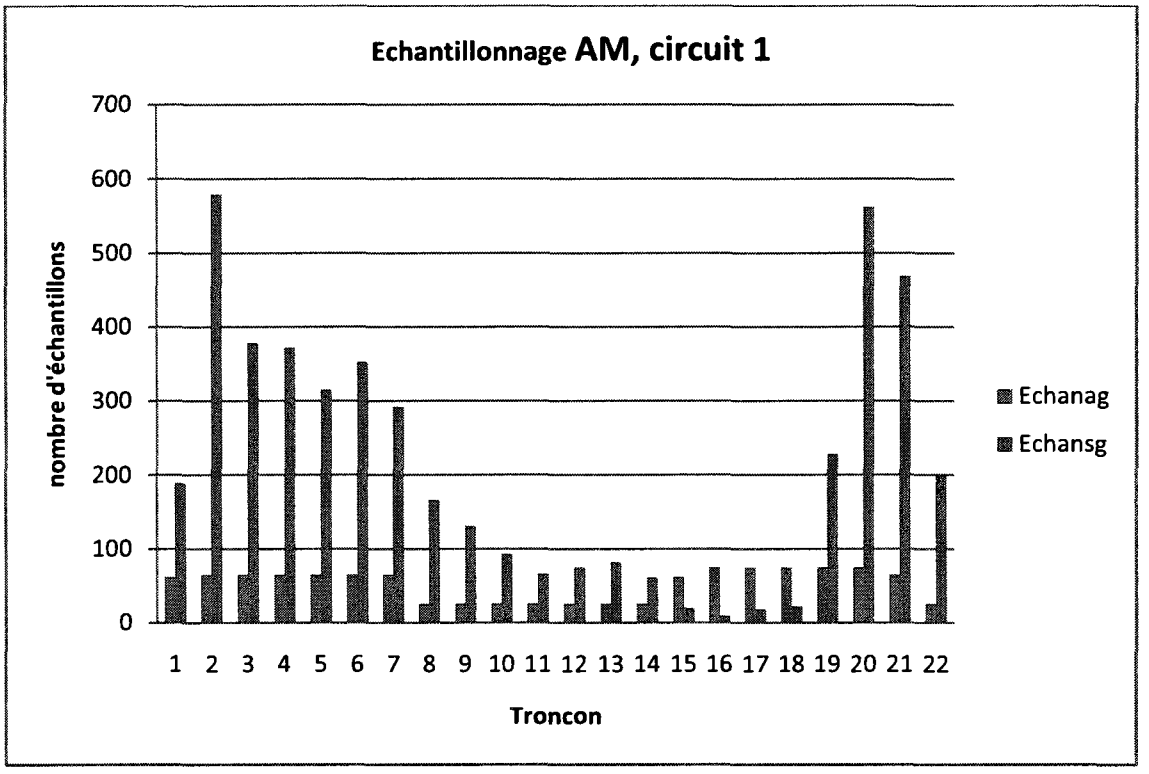


Figure 5-9 Echantillonnage distributif du circuit 1 pour la période AM

La valeur 2000 n'a pas été choisie au hasard. En effet, nous constatons qu'à partir de 2000, le nombre d'échantillons nécessaires augmente beaucoup plus lentement (cf 5). Nous pouvons aussi rajouter que la répartition dépend aussi du nombre de tronçons que contient chaque circuit.

Pour conclure cette partie, nous venons de constater qu'un échantillonnage est tout à fait réalisable, que la méthode de *clustering* permet de réduire significativement le nombre d'échantillons à produire et que celui-ci peut être équitable pour tous les circuits. Par conséquent, c'est au tour du gestionnaire de définir son cahier des charges et donc de définir ses objectifs d'études.

5.6 Synthèse

Conclusions : L'échantillonnage du réseau autoroutier montréalais fait partie des préoccupations énoncées par le ministère des Transports du Québec. La question est donc d'effectuer un échantillonnage facilement applicable et rendant (on l'espère) compte de l'état de la circulation sur le réseau autoroutier montréalais.

Nous avons constaté que l'application brute de la formule mathématique pour définir l'échantillonnage ne permet pas de satisfaire la condition de faisabilité de l'expérience. C'est pourquoi le *clustering*, et ses résultats sont finalement la véritable alternative. Au moyen de l'étude faite dans la partie précédente, on peut effectuer un échantillonnage cohérent qui, a priori, rendra compte de l'état de la circulation mais qui surtout, réduit de manière considérable le nombre d'échantillons à réaliser et qui satisfait la condition de faisabilité.

Perspectives : Le problème de la démonstration de cette méthode ne permet pas d'effectuer un échantillonnage cohérent sur tout le réseau montréalais. En effet, la méthode pourrait léser un circuit vis-à-vis d'un autre. C'est pour cela que nous avons proposé une alternative « économique » qui pallie justement ce problème.

La perspective est donc que ce type d'analyse peut être reproduit pour tous les autres facteurs tels que conditions de la chaussée, jour de la semaine, mois de l'année, etc.

CHAPITRE 6 : DÉGAGEMENT DE LA CIRCULATION NORMALE

L'analyse fréquentielle effectuée dans la partie précédente a mis en exergue une circulation normale existant sur les différents circuits, tendance que nous avions commencé à pressentir lors de l'utilisation des cartes de contrôle dans l'analyse exploratoire. En effet, nous constatons que les distributions fréquentielles des essaims 2 et 4 du circuit 1 par exemple étaient très asymétriques et présentaient un pic (ou un extremum) autour d'un temps de parcours de 36 secondes, ce qui correspond finalement à une vitesse de 100 km/h soit la vitesse légale sur le circuit 1.

L'objectif de cette partie est donc de dégager de la base de données du circuit 1 un maximum de relevés tels qu'ils soient dans un intervalle de temps, et que leur distribution soit la plus proche d'une distribution théorique connue.

On parle donc dans cette partie de modélisation, et d'explication d'un phénomène aléatoire. Dans un premier temps nous rappellerons quelques définitions et propriétés de lois connues, puis exposerons notre démarche mathématique et les résultats et conclusions qui en découlent.

6.1 Présentation des distributions théoriques

Il est fréquent, lors de l'observation d'un phénomène aléatoire, de chercher à le modéliser selon une distribution statistique connue, généralement la loi de distribution normale par application du théorème centrale limite. Typiquement, la question qui se pose est celle de la normalité de la distribution des données de temps de parcours sur le réseau montréalais (distributions normales, ou log-normales) telles que citées dans la revue de littérature.

6.1.1 La loi normale

Une variable aléatoire X possède une distribution normale de paramètres μ, σ, avec les propriétés suivantes : $E(X) = \mu$, $Var(X) = \sigma^2$.

La fonction de densité est donnée par $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} . e^{\frac{(x-\mu)^2}{2\sigma^2}}$.

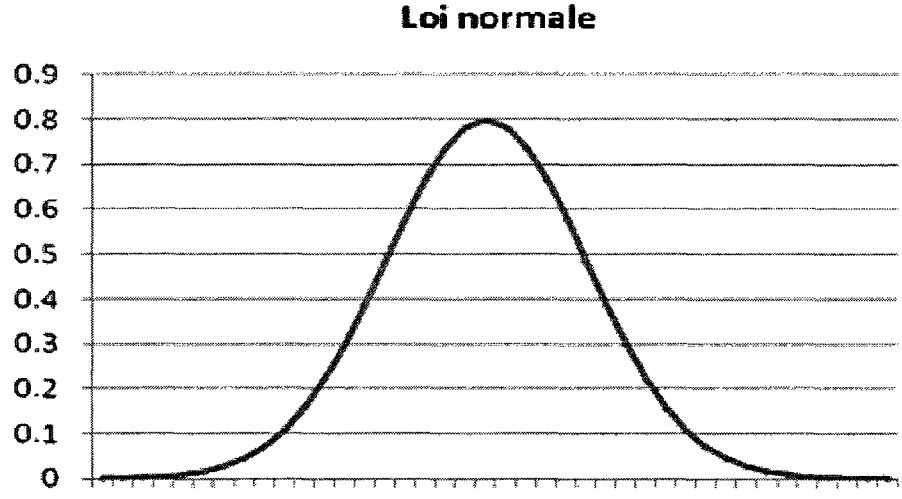


Figure 6-1: Exemple de fonction de densité selon la loi normale

6.1.2 Loi log-normale

On définit la loi log-normale comme une transformation bijective de la loi normale.

Soit X une variable aléatoire suivant une loi log-normale de paramètres μ, σ. Alors la variable $Y = ln(X)$ suit une loi normale dont les paramètres α, β, respectivement moyenne et écart-type, sont reliés aux paramètres μ, σ par :

$$\mu = e^{\alpha+\beta^2/2}$$

$$\sigma^2 = e^{2\alpha+\beta^2}(e^{\beta^2} - 1)$$

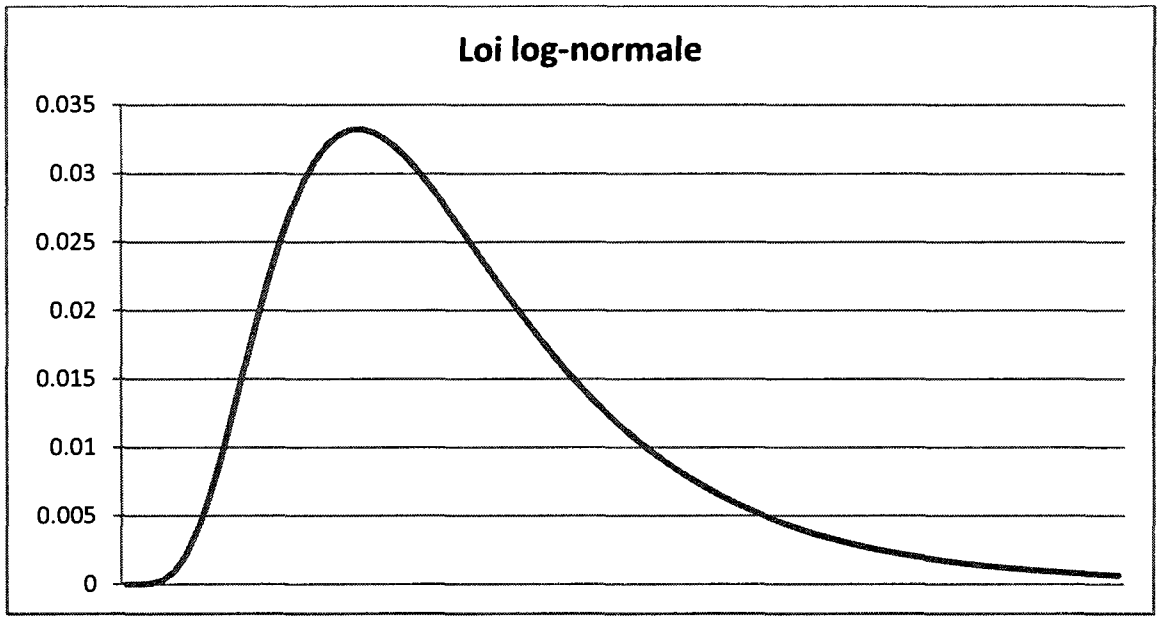


Figure 6-2 Fonction de densité de la loi log-normale

6.1.3 Propriété de la loi normale

Propriété: Soit $(X_i)_{i\in J}$, où $J = [\![1; N]\!]$, N variables aléatoires indépendantes suivent des lois normales de paramètres μ_i , σ_i alors la variable $X = \sum_{i\in J} \gamma_i X_i$ suivra une loi normale de paramètres moyenne et écart-type respectivement égaux à :

$$\sum_{i\in J} \gamma_i \mu_i \ , \ \sqrt{\sum_{i\in J} \gamma_i^2 \sigma_i^2}$$

6.1.4 Tests sur la distribution normale

Il existe plusieurs tests permettant d'accepter ou refuser l'hypothèse de normalité de la distribution de *X*, X représentant dans notre cas le temps de parcours (ou autre variable estimée à l'aide des mêmes données).

6.1.5 Test du χ^2

Après avoir classé les valeurs observées selon m intervalles et estimé la moyenne et la variance expérimentales, μ_0, σ_0, nous posons les hypothèses :

$H_0 : X \sim N(\mu_0, \sigma_0)$,
$H_1 : X$ ne suit pas une distribution normale de paramètres μ_0, σ_0.

Soit alors $f_{i,0}$ la fréquence des valeurs observées dans la i-ème classe et $f_{i,t}$ la fréquence théorique d'une loi normale de paramètres μ_0, σ_0, alors la variable $Y = \sum_{i=1}^{m} \frac{(f_{i,0} - f_{i,t})^2}{f_{i,t}}$ suit alors une loi $\chi^2(m-3)$.

Si la valeur Y_0 calculée s'avère être **inférieure** à $\chi^2_{m-3,\alpha}$ où α est une valeur dans $[0,1]$ correspondant au seuil préalablement choisi, alors **nous accepterons H_0**. Dans le cas contraire, nous la rejetterons.

6.2 Démarche mathématique pour dégager la circulation normale

Il a été précisé dans la partie précédente la méthode du χ^2 afin de vérifier qu'une distribution expérimentale s'approche d'une distribution théorique connue (on choisira ici de considérer des distributions log-normales en raison de la forte asymétrie précédemment observée).

C'est en partant de cette méthode que nous allons étudier les relevés d'un circuit.

Afin de retirer des observations, les observations correspondant à la circulation normale d'un circuit, c'est-à-dire les relevés que l'on pourrait qualifier de relevés en situation normale, il devient alors pertinent de chercher à comparer la distribution que l'on possède avec une distribution statistique classique. Selon la revue de littérature, nous savons que la distribution devrait se rapprocher d'une loi log-normale (May, A., Traffic Flow Fundamentals). Toutefois, la décroissance du nombre de relevés de temps de parcours croissants est trop lente pour obtenir des résultats corrects, il devient préférable de considérer la somme de deux lois log-normales pour définir le phénomène.

Soit $\Omega = (x_i)_i$ l'ensemble des observations d'un circuit.

6.2.1 Principe mathématique

Soit Ω' un sous-ensemble fini de Ω, il existe alors x_{min} et x_{max} tels que $\forall x \in \Omega', x \in [x_{min}, x_{max}]$.

On définit de plus l'application bijective $\varphi : \Omega' \rightarrow \Omega''$

$$x \rightarrow \ln(x)$$

On notera par μ et σ la moyenne et l'écart-type des éléments de Ω''

On supposera en particulier (déformation du principe de Pareto), que $\frac{card(\Omega')}{card(\Omega)} \geq 0.70$, où $card(\Omega')$ est définit comme le nombre d'éléments de Ω'. En clair on essaie de confirmer qu'au moins 70% des relevés explique le comportement des temps de parcours sur le circuit.

On choisit alors de décomposer Ω' en n intervalles de tailles égales. Notons $\Omega'_1, \ldots, \Omega'_n$ ces sous-ensembles de Ω' d'extrema respectifs $(x^1_{min}, x^1_{max}), \ldots, (x^n_{min}, x^n_{max})$, avec comme conditions aux frontières $x^1_{min} = x_{min}$ et $x^n_{max} = x_{max}$.

On note alors $f_i^{exp} = card(\Omega'_i)$, la fréquence expérimentale de relevés de Ω'_i. Par conséquent $\sum_i f_i^{exp} = card(\Omega')$.

On note de plus $f_i^{théorique} = card(\Omega') * \{\theta(x^i_{max}) - \theta(x^i_{min})\}$, où la fonction θ est la cumulative de deux lois log-normales de moyenne et d'écart-type respectifs, μ_1 et σ_1, μ_2 et σ_2.

On définit alors par $\pi_{\Omega'} = \sum_i \frac{(f_i^{exp} - f_i^{théorique})^2}{f_i^{théorique}}$.

6.2.2 Optimisation

Par conséquent l'étape suivante est donc de trouver l'intervalle ou ***fenêtre temporelle*** *décrivant la circulation normale*, tel que le sous ensemble soit optimal. Mathématiquement, cela revient à rechercher l'optimum :

$$\Pi = \min_{\Omega'} \pi_{\Omega'}.$$

Cette optimisation dépend donc de 6 paramètres, les bornes de l'intervalle correspondant à la fenêtre temporelle sous la contrainte d'avoir au moins 70% des relevés, enfin les paramètres des deux lois log-normales. Afin d'obtenir une précision correcte et de trouver un minimum pertinent, nous utiliserons la méthode des gradients pour s'approcher des valeurs optimales des paramètres.

6.3 Résultats pour le circuit 1

L'algorithme a été appliqué aux données du circuit 1, et nous en donnons les résultats principaux. Une étude peut être effectuée évidemment pour tous les circuits du réseau autoroutier montréalais. Il est important toutefois de noter que plusieurs circuits ne peuvent être étudiés. En effet, deux raisons particulières sont à prendre en compte : soit le nombre de relevés est insuffisant (n'oublions pas la condition des 70% de relevés dans la fenêtre temporelle), soit la vitesse maximale autorisée peut être inférieure à 100km/h sur trop de tronçons du circuit.

6.3.1 Dégagement de la circulation normale

Afin d'étayer le concept mathématique étudié, nous choisissons de donner les résultats du circuit 1 résumés dans le Tableau 6-1. La Figure 6-3 présente les deux distributions observée et théoriques. On remarquera que la valeur minimale $\Pi = \min_{\Omega'} \pi_{\Omega'}$ de la méthode exposée est 54, et que cette fenêtre concerne 71% des relevés.

Tableau 6-1 Paramètres de la circulation normale du circuit 1

Lognormale	log1	log2	Intervalles
Moyenne	3.49	3.586	26.1
Ecart-type	0.074	0.12	43.5

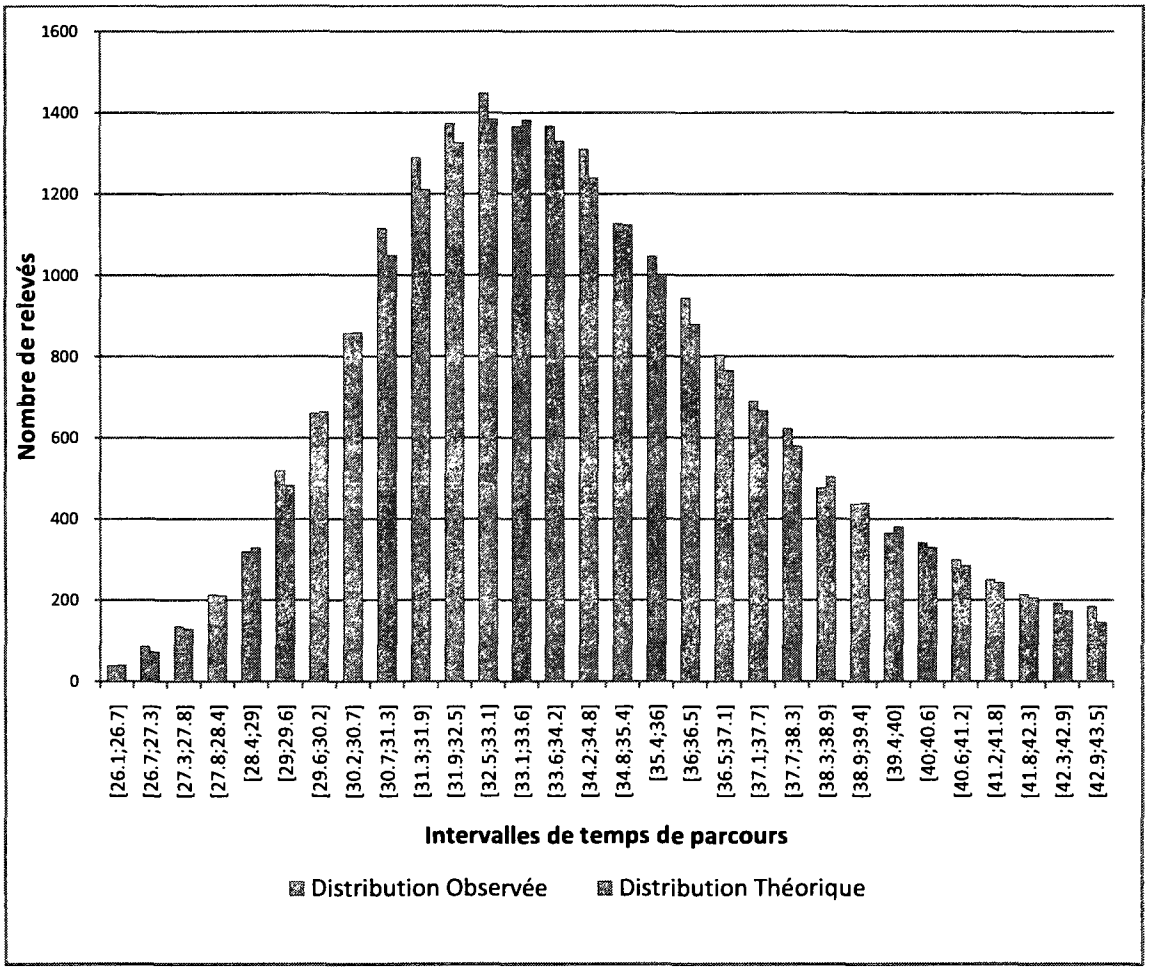


Figure 6-3 Modélisation de la circulation normale du circuit 1

Cette analyse permet alors de dégager les points chauds tout au long de notre circuit. Nous rappelons que tous les circuits ont été découpés en tronçons de 1 km, ce qui permettra juste

par observation du pourcentage de relevés dans la fenêtre temporelle (Figure 6-4) de mettre en avant les zones du circuit congestionnées selon la période d'observation (AM ou PM).

6.3.2 Pourcentage de relevés dans la fenêtre temporelle

Cette partie est en fait un argument d'autorité des différentes méthodes appliquées précédemment, localisation des points chauds, différences entre année, différence de comportement entre les périodes, *data mining* et étude de la variabilité du circuit.

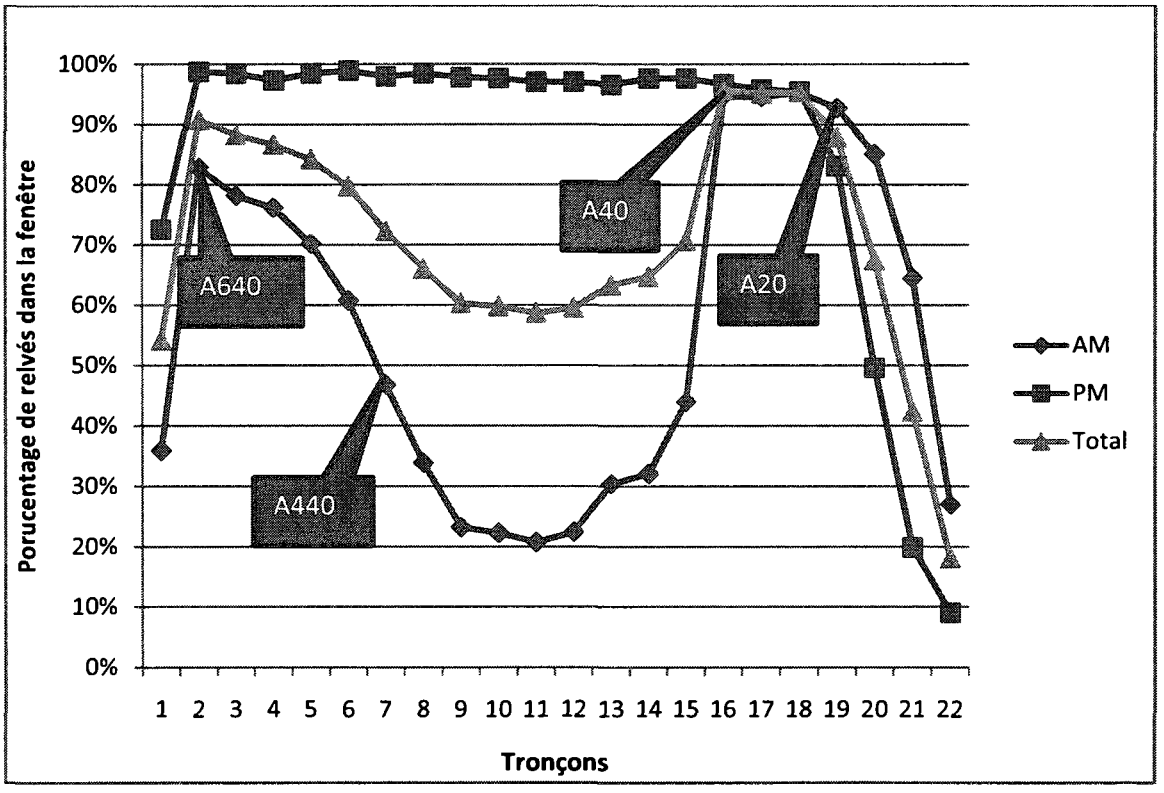


Figure 6-4 Pourcentage de relevés dans la fenêtre temporelle par tronçon et selon la période

Dans le cas du circuit 1 correspondant à l'A13 direction sud, on retrouve de manière très claire deux phénomènes : premièrement le problème de convergence des origines et destinations des voyageurs. En effet, le circuit est plus congestionné le matin ce qui signifie que les conducteurs semblent utiliser cette autoroute le matin pour se rendre à Montréal ; deuxièmement, on retrouve de même toutes les intersections entre l'A13 et les autres

autoroutes, nous constatons donc que trois intersections sur quatre sont des facteurs de congestion. Par contre, dès le passage de l'entrecroisement entre l'A40 et l'A13, la circulation redevient fluide, ce qui signifie que une grosse partie des véhicules initialement sur l'A13 abandonnent cette autoroute pour l'A40 et se rendre à Montréal.

Une autre idée serait donc, puisque l'effet significatif du facteur année a été démontré (ANOVA), est de comparer l'évolution des pourcentages par tronçon, par année. On constate donc à la vue de la Figure 6-5, que l'année 1999 semble plus congestionnée que les autres pour les tronçons 1 à 7. Toutefois, nous avions pu remarquer que l'année 1999 n'avait été échantillonnée que pour les mois de septembre à décembre.



Figure 6-5 Evolution par tronçon du pourcentage de relevés dans la fenêtre par année



Figure 6-6 Evolution par tronçon du pourcentage de relevés dans la fenêtre par année période AM

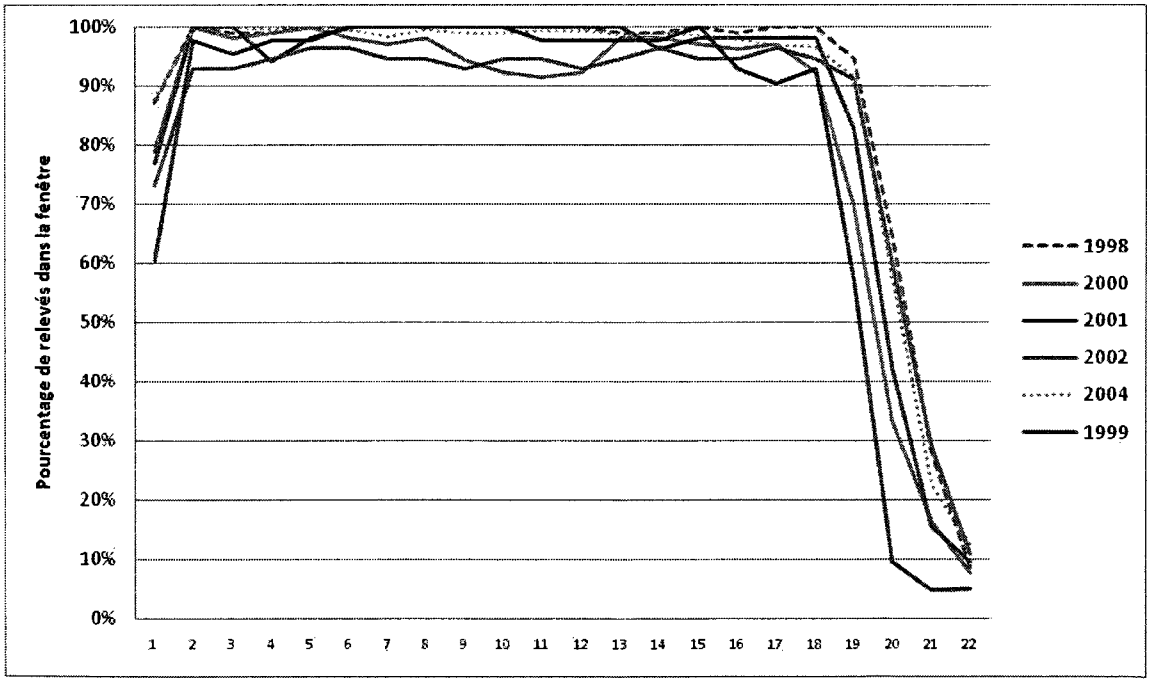


Figure 6-7 Evolution par tronçon du pourcentage de relevés dans la fenêtre par année période PM

Pour confirmer alors l'idée que l'année 1999 est bien une année critique, nous allons alors ne conserver que les valeurs des mois de septembre à décembre pour toutes les années. Malgré une réduction alors importante du nombre d'échantillons disponibles, le résultat est tout de même probant. En effet, nous constatons, d'après la Figure 6-8, que l'année 1999 est bien plus critique pour les tronçons 1 à 7.

Faut-il justement y voir une circonstance particulière de longue durée, tel que des travaux ?

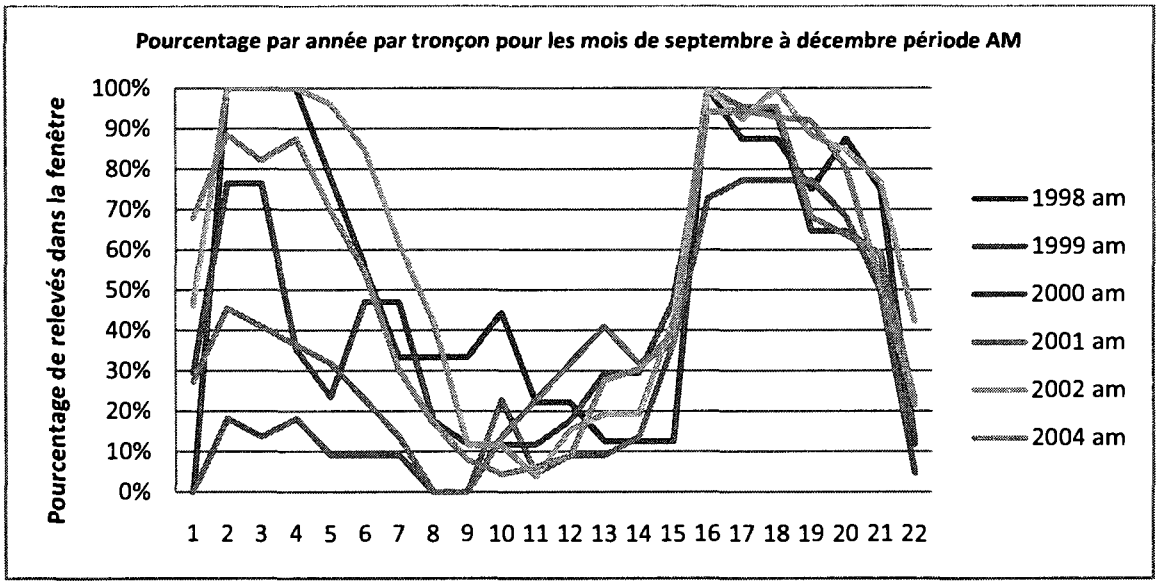


Figure 6-8 Evolution par tronçon du pourcentage de relevés dans la fenêtre par année septembre à décembre AM

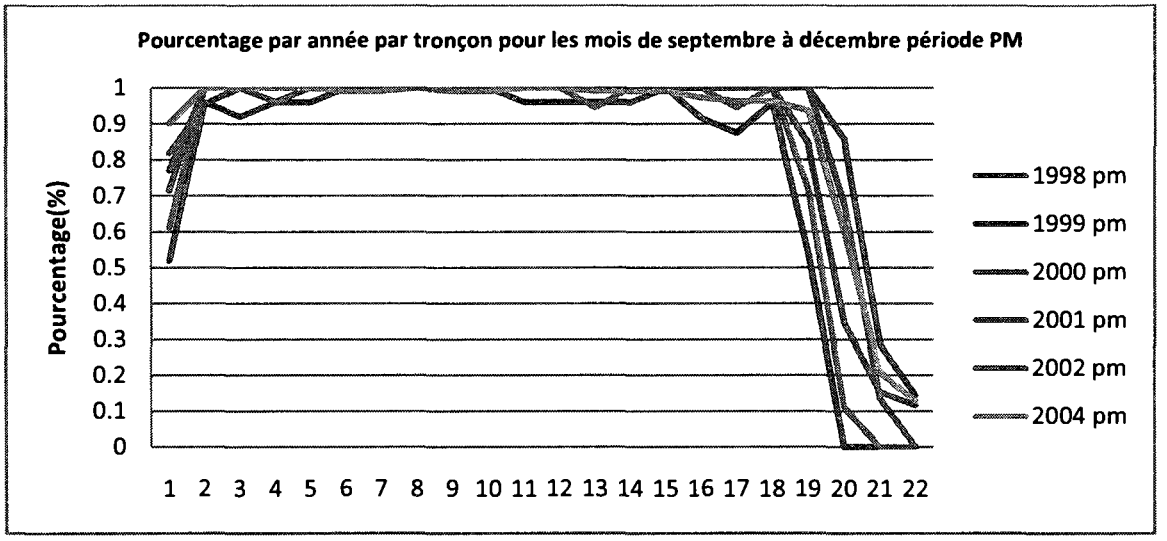


Figure 6-9 Evolution par tronçon du pourcentage de relevés dans la fenêtre par année septembre à décembre PM

Tableau 6-2 Corrélation entre pourcentage et moyenne-écart-type-variabilité AM

Corrélation	Corrélation Moyenne	Corrélation Ecart-type	Corrélation Variabilité
Fréquences AM	-0.53	-0.14	0.23

Tableau 6-3 Corrélation entre pourcentage et moyenne-écart-type-variabilité PM

Corrélation	Corrélation Moyenne	Corrélation Ecart-type	Corrélation Variabilité
Fréquences PM	-0.90	-0.66	-0.38

L'analyse de la corrélation présente alors les résultats escomptés, puisque la corrélation entre les pourcentages de relevés dans la fenêtre et la série des moyennes des temps de parcours par tronçon est négative et inférieure à -0.5.

En conclusion de cette analyse, on peut donc constater que nous venons de légitimer le phénomène de circulation normale, que nous reconfirmons encore une fois le problème lié au facteur période, et que la circulation normale permet de dégager les congestions de la base de données.

6.4 Relevés en dehors de la fenêtre temporelle

Nous avons dans la partie précédente exposé la modélisation statistique de plus de 70% des relevés, mais que dire donc des relevés restants. Une première idée, plutôt logique serait de déclarer que plus le temps de parcours augmente, moins il y aura de relevés. En clair nous parlons d'une décroissance exponentielle.

6.4.1 Processus mathématique

Comme nous l'avons pressenti dans l'introduction de cette partie, la modélisation de la « partie droite », ou ensemble des relevés dont la valeur est supérieure à la borne supérieure de la fenêtre temporelle, doit être faite selon des lois exponentielles. Toutefois, compte-tenu du fait que le nombre de relevés dont la valeur est proche de la borne supérieure de la fenêtre temporelle demeure élevé, il devient donc nécessaire de ne considérer pas une loi exponentielle mais deux, agrémentée chacune d'un coefficient. Ce choix de modèle est propre aux données dont nous disposons et est confirmé sur d'autres circuits mais ne dépend pas de travaux antérieurs.

En résumé, voici la formule définissant la fonction de densité de cette distribution statistique :

$$f(x) = \begin{cases} 0 \; si \; x < x_{max} \\ \lambda_1 . e^{-\alpha_1 (x - x_{max})} + \lambda_2 . e^{-\alpha_2 (x - x_{max})}, \; \lambda_1, \lambda_2, \alpha_1, \alpha_2 > 0, si \; x > x_{\max} \end{cases},$$

il vient par conséquent que par intégration sur l'ensemble des réels, $\frac{\lambda_1}{\alpha_1} + \frac{\lambda_2}{\alpha_2} = 1$. La valeur x_{max} est la valeur maximale de la circulation normale. Le test est ensuite appliqué sur l'intervalle de temps de parcours $[x_{max}; 360]$.

6.4.2 Résultats pour le circuit 1, et problématique

Dans cette partie nous présentons succinctement les résultats obtenus par application d'un algorithme basé sur la méthode des gradients afin de trouver un la fonction de densité optimale. La valeur pivot du test minimale trouvée est 94.5.

Tableau 6-4 Valeurs des paramètres des lois exponentielles

exp	e1	e2	Intervalle
lambda	0.0311	0.007	43.5
alpha	0.1155	0.00957	360

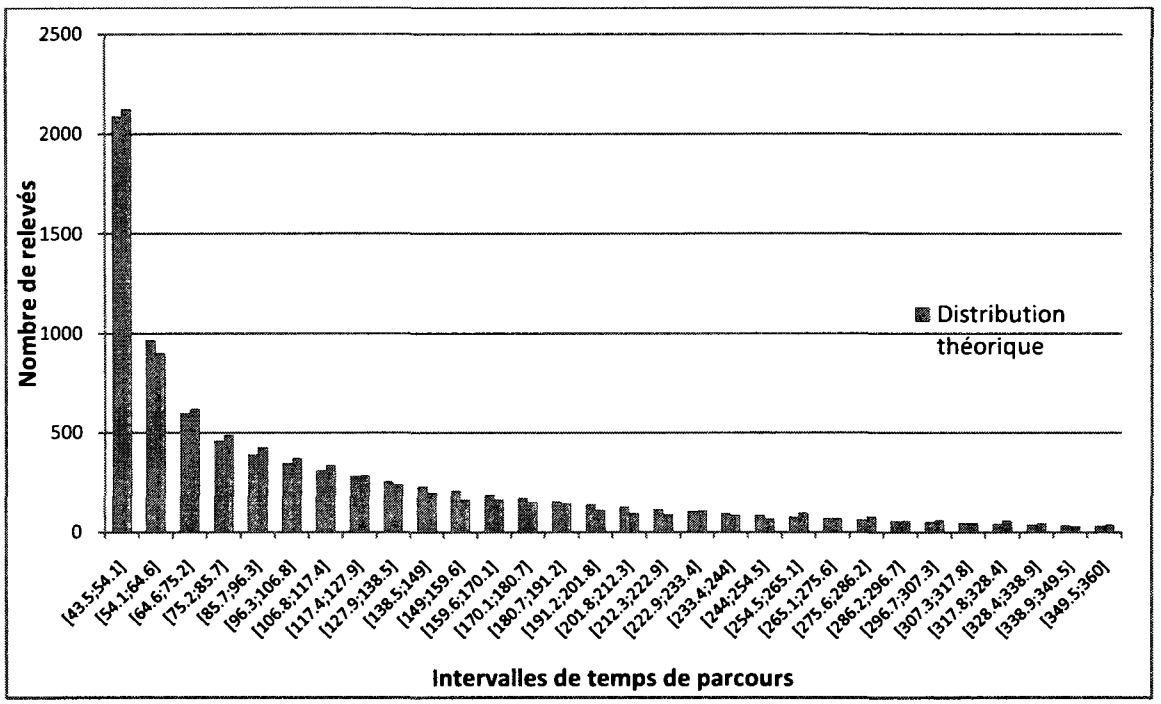


Figure 6-10 Calibrage d'une loi statistique sur la distribution observée de la partie droite

En conclusion, nous en déduisons que désormais, nous avons pu modéliser la distribution fréquentielle des relevés des temps de parcours avec une précision correcte, et que dans le cas du circuit nous venons d'expliquer 98,5% de la totalité des relevés ce qui au premier abord ne semblait pas une chose facile. Désormais, ayant connaissance des distributions statistiques, une nouvelle approche de la circulation autoroutière est donc la simulation et le développement issu de celle-ci. Ainsi, il sera alors possible de distinguer des segments potentiellement dangereux de segments présentant une congestion récurrente et donc appréhendable.

6.5 Synthèse

Conclusions : L'analyse de la distribution des relevés de temps de parcours pour un circuit a démontré l'existence dans le cas du circuit 1, d'une concentration importante des relevés de temps de parcours dans une fenêtre temporelle que nous avons nommée « circulation normale » d'un circuit. La modélisation statistique, au moyen d'une minimisation d'une distance se rapprochant du test du khi-carré, a donc été réalisable sur cette circulation normale, et les résultats sont assez probants.

L'observation du pourcentage de relevés de temps de temps de parcours compris dans cette circulation normale par année, par période, confirme les points chauds de la circulation est peut aussi donner une explication de déplacement des flots sur le circuit.

Enfin, afin d'avoir une modélisation complète, il a fallu se soucier de la distribution des relevés de temps de parcours supérieurs à la borne supérieure de la circulation normale. Cette modélisation finie, nous aboutissons à une conclusion certaine : la quasi-totalité des relevés a été considérée.

Perspectives : Les perspectives deviennent alors évidentes. Nous venons de modéliser la distribution des temps de parcours. L'étape suivante est donc de simuler les temps de parcours sur le circuit. Cette simulation débouchera aussi sur la création de différents indicateurs, et relatera de la récurrence ou non récurrence de la congestion.

CHAPITRE 7 : SIMULATION PAR CIRCUIT

Grâce à l'analyse précédente, nous venons de caractériser ni plus ni moins l'évolution des temps de parcours dans des conditions de circulation libres.

La congestion, comme nous l'avons vu dans la revue de littérature, se dissocie en deux types, la congestion récurrente dont nous pouvons mesurer les effets, et une congestion non-récurrente plus contraignante à identifier et à caractériser. De plus, une phase de congestion est précédée par une phase de début de congestion et de fin de congestion.

L'objectif de cette partie est de construire un arbre de décision des temps de parcours, et de plus de simuler le trafic sur une portion de tronçon d'un circuit donné.

7.1 Construction d'un arbre de décision

7.1.1 Concept mathématique

Soit un ensemble d'observations Ω d'un circuit comprenant N tronçons. Soit alors le vecteur-observation, $T = [t_1, t_2, ..., t_N]$, où t_i désigne le temps de parcours observé sur le tronçon i.

Nous disposons alors d'un ensemble de cardinal m de vecteurs-observations $(T^j)_{j=[1,m]}$. Nous pouvons donc désormais calculer la corrélation entre l'ensemble des observations de deux tronçons différents, c'est-à-dire par exemple dans le cas des tronçons 1 et 2 $\frac{\sum_i (t_1^i - \bar{t}_1).(t_2^i - \bar{t}_2)}{\sqrt{\sum_i (t_1^i - \bar{t}_1)^2 \sum_i (t_2^i - \bar{t}_2)^2}}$, où $\bar{t}_1$ désigne la moyenne des temps de parcours du tronçon 1.

On choisit ensuite selon la fenêtre temporelle I que nous avons calculée dans la partie précédente, de coder les valeurs de la manière suivante :

$$x = \begin{cases} 0 \; si \; t \in I \\ 1 \; si \; t \in J \\ 2 \; si \; t \notin I \; et \; t \notin J \end{cases}.$$

On définit J comme un nouvel intervalle, correspondant à un « ventre mou » c'est-à-dire aux observations dont les temps de parcours sont supérieurs à la borne supérieure de I, mais qui ne dénotent pas d'un immobilisme de la voiture flottante.

Ainsi, nous pouvons construire une matrice de transposition $P_{i,i+1}$ d'un tronçon i à son suivant $i+1$ dont les éléments sont calculés selon le maximum de vraisemblance :

$P_{i,i+1} = \begin{bmatrix} p_{i,i+1}^{0,0} & p_{i,i+1}^{0,1} & p_{i,i+1}^{0,2} \\ p_{i,i+1}^{1,0} & p_{i,i+1}^{1,1} & p_{i,i+1}^{1,2} \\ p_{i,i+1}^{2,0} & p_{i,i+1}^{2,1} & p_{i,i+1}^{2,2} \end{bmatrix}$, où $p_{i,i+1}^{r,s}$ désigne la probabilité de passer de l'état r à l'état s.

Par conséquent $\sum_{s=0}^{2} p_{i,i+1}^{r,s} = 1$.

Le modèle est donc un modèle trinomial. L'intérêt désormais est de revenir à des variables aléatoires discrètes en sortie.

7.1.2 Résultats pour le circuit 1

L'étude de la corrélation des observations pour la période AM donne le graphe suivant (Figure 7-1) :

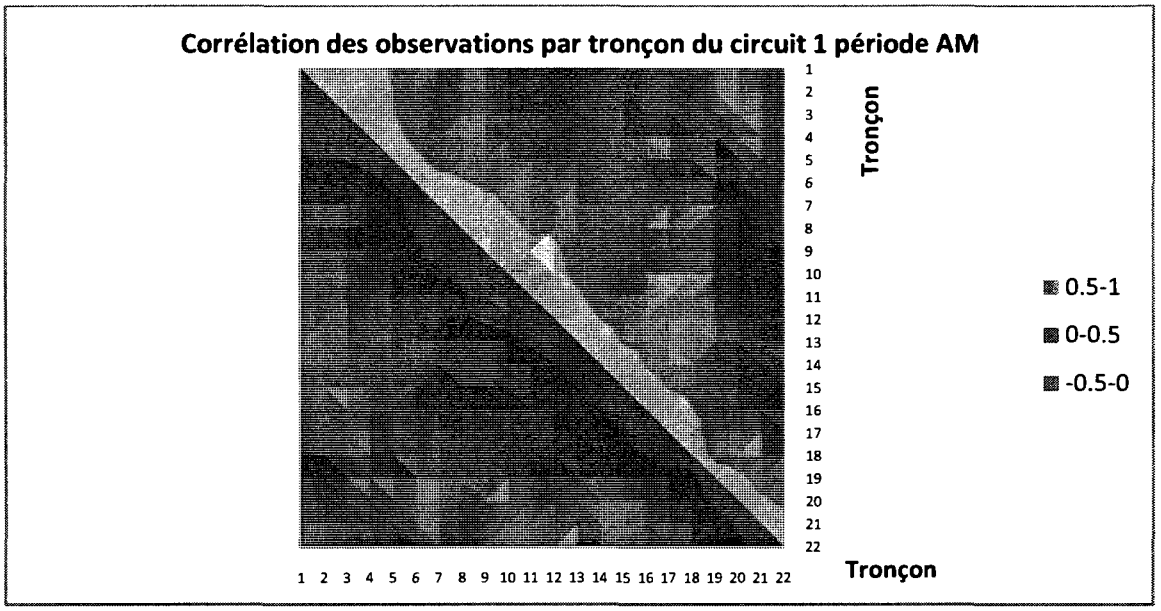


Figure 7-1 Corrélation des observations par tronçon du circuit 1 période AM

Les résultats sont probants ; en effet, nous constatons que très vite la corrélation décroit dès que les tronçons sont loin l'un de l'autre. Ceci permet donc de dire que pour pouvoir concevoir un modèle, nous pouvons faire l'hypothèse d'indépendance entre deux tronçons dès lors que leur éloignement grandit.

On constate de plus une réduction certaine au niveau du tronçon 18, ce ci signifie donc que le que ce tronçon (entre deux intersections ou deux tronçons à congestion alors que lui-même ne semble pas congestionné) est atypique de ses voisins.

Dans le cas du circuit 1 période AM, pour les quatre premiers tronçons nous trouvons alors les matrices de transition suivantes, où le ventre mou est défini comme l'intervalle, en secondes, [43.5,70]:

Tableau 7-1 Matrices transpositions des tronçons 1 à 4 du circuit 1 période AM

Calcul de la matrice de probabilité de 1 vers 2			
Etat Tronçon1/ Etat Tronçon 2	0	1	2
0	0.96	0.03	0.02
1	0.86	0.06	0.08
2	0.13	0.00	0.88

Calcul de la matrice de probabilité de 2 vers 3			
Etat Tronçon 2/ Etat Tronçon 3	0	1	2
0	0.94	0.04	0.02
1	0.20	0.56	0.24
2	0.00	0.06	0.94

Calcul de la matrice de probabilité de 3 vers 4			
Etat Tronçon 3/ Etat Tronçon 4	0	1	2
0	0.94	0.03	0.03
1	0.45	0.33	0.23
2	0.01	0.15	0.83

Dans le cas du passage du tronçon 3 au tronçon 4, on constate donc que si le temps de parcours observé au tronçon 3 est dans la fenêtre temporelle (non-congestion), alors il y a 94% de chances que le temps de parcours du tronçon 4 soit aussi dans la dans la fenêtre temporelle. De même si le tronçon 3 est congestionné alors il y a 83% de chances que le tronçon 4 le soit aussi.

L'intérêt de cette construction est donc que cela permet pour l'usager de la route de savoir à quoi s'attendre. Dans cette optique, si l'usager est pris dans une congestion à un tronçon donné, et connaissant les matrices de transposition pour les tronçons suivants, peut-être lui sera-t-il bénéfique de choisir un circuit alternatif.

7.2 Simulation

L'étape de simulation est un pilier des conclusions que nous pourrons tirer de ce document. En effet, ayant connaissance des distributions statistiques acceptées dans les parties précédentes, nous allons pouvoir modéliser le trafic autoroutier. Mais avant de rentrer dans le vif du sujet, il est important désormais de concevoir une catégorisation des relevés de temps de parcours. Ce débat plus pragmatique que scientifique, correspond seulement à la définition que l'on donne aux congestions acceptables et inacceptables. Dans le cadre de ce document, nous choisirons de considérer des relevés de temps de parcours comme appartenant à une congestion inacceptable si leur valeur est supérieure à 120 secondes (soit une vitesse de 30 km/h) ; le nombre de relevés de temps de parcours en dessous de cette limite correspond à plus de 70% des relevés de temps de parcours de la partie droite que nous avions définie auparavant ; nous nommerons l'intervalle contenant ces résultats comme le « ventre mou du circuit 1 ». La valeur limite entre congestion acceptable et inacceptable variera selon le type de route que l'on considère.

Deux approches seront envisagées. Premièrement, la simulation sera effectuée en omettant la conception de l'heure de départ, puis par la suite, c'est justement ce facteur significatif qui servira à concevoir une simulation selon l'heure de départ.

7.2.1 Rappel mathématique et variabilité

La simulation est le résultat de calculs probabilistes simples. Dans le cas de l'espérance du temps de parcours moyen par tronçon, nous pouvons écrire alors si Y est la variable aléatoire à trois niveaux 0, 1, 2 :

$$E(X) = E(X|Y = 0)P(Y = 0) + E(X|Y = 1)P(Y = 1) + E(X|Y = 2)P(Y = 2)$$

Donc connaissant les fonctions de distributions de chaque intervalle,

$$E(X) = E(X|X \in [20, x_{max}])P(Y = 0) + E\big(X|X \in [x_{max}, x_{acceptable}]\big)P(Y = 1) + E\big(X|X \in [x_{acceptable}, 360]\big)P(Y = 2)$$

Dans le cas du circuit 1, nous choisissons de définir la congestion inacceptable à partir du moment où la valeur d'un relevé de temps de parcours pour un tronçon quelconque devient supérieure à 120 secondes soit une vitesse de 30 km/h, ce qui s'accorde avec les considérations liées aux définitions actuelles de la congestion.

Les trois figures suivantes confirment l'idée que la modélisation choisie est plus que valable ; en effet, les courbes sont très corrélées, et nous avons peu de pertes d'informations. On notera tout de même que la précision est bien plus grande lorsque que les tronçons sont peu congestionnés, ce qui semble viable, puisque dans ce cas, la quasi-totalité des relevés est contenu dans la circulation normale.

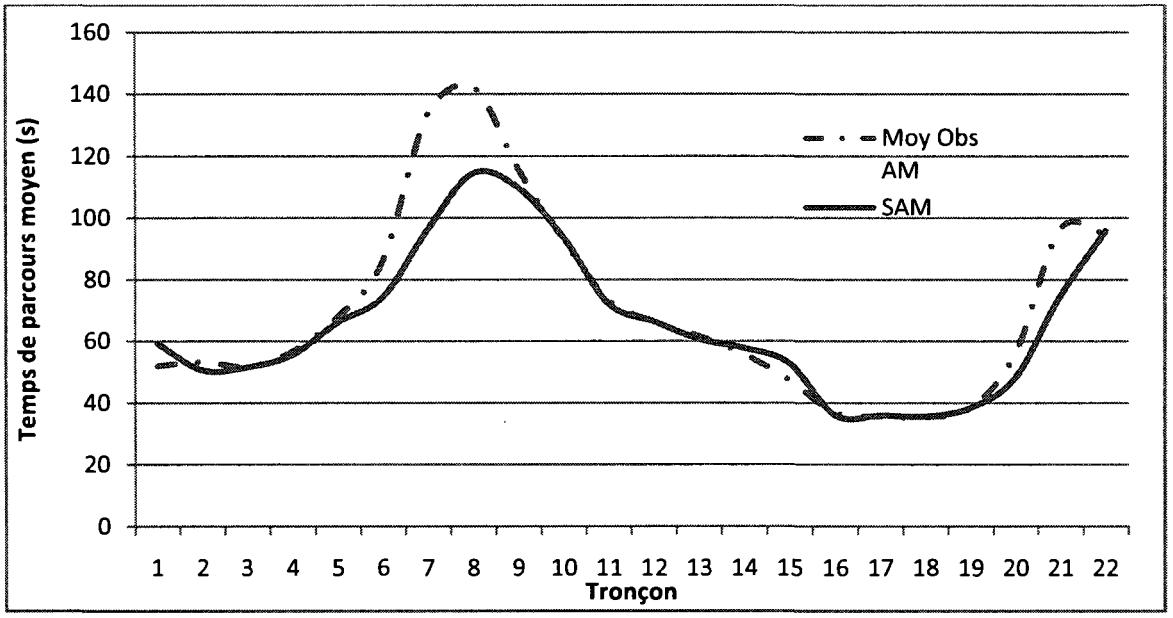


Figure 7-2 Temps de parcours moyens observé et simulé par tronçon du circuit 1 en période AM

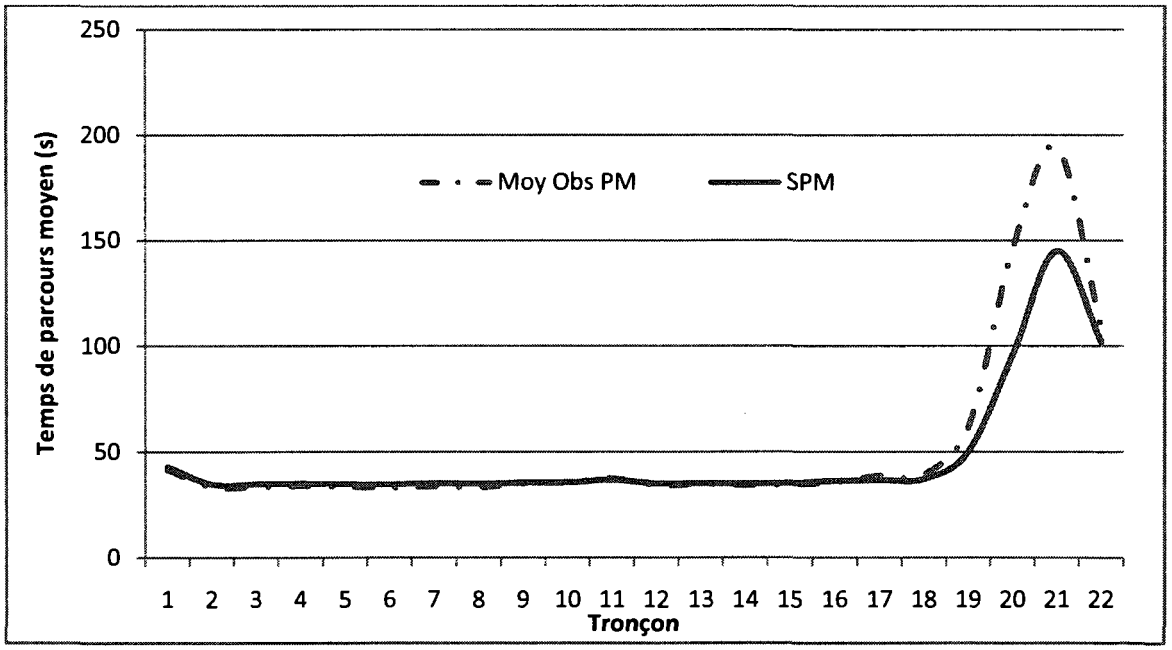


Figure 7-3 Temps de parcours moyens observé et simulé par tronçon du circuit 1 en période PM

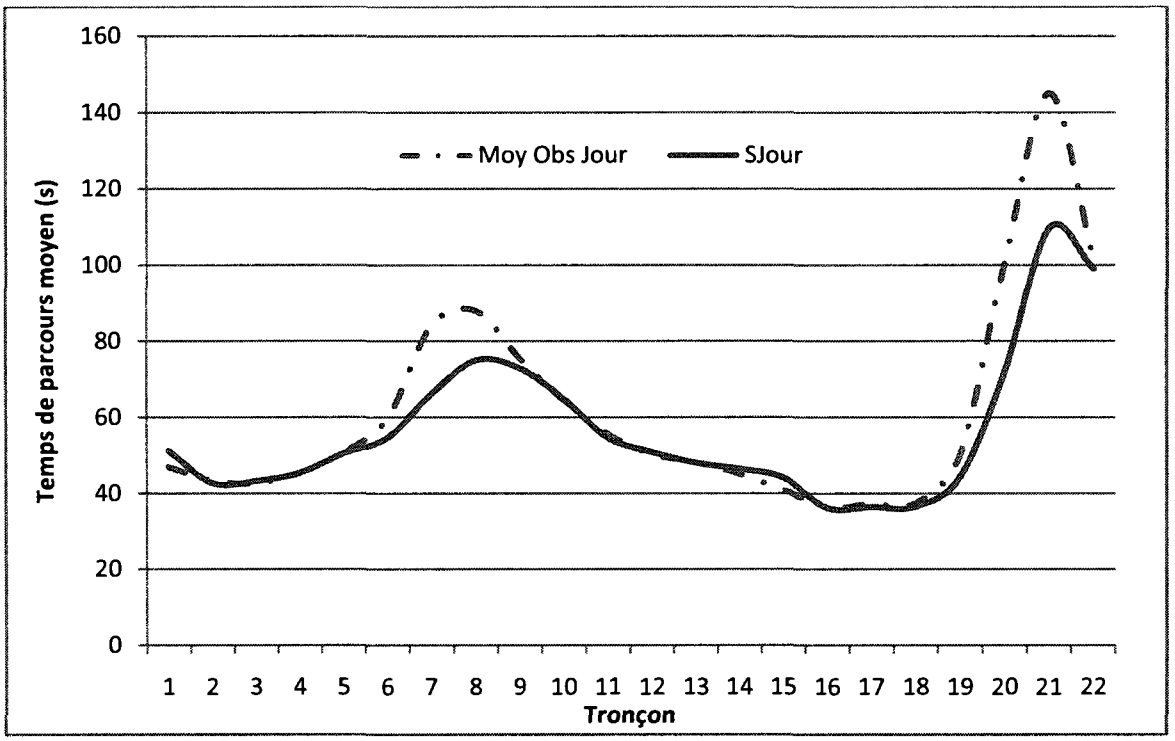


Figure 7-4 Temps de parcours moyens observé et simulé par tronçon du circuit 1 pour la journée entière

Observons désormais les résultats concernant la confrontation entre variabilités (AM ou PM) observée et simulée.

La Figure 7-5 et la Figure 7-6 présentent les variabilités observées ainsi que les variabilités simulées. Il apparait donc de nouveau que l'allure générale est conservée grâce à la modélisation. On constate tout de même plusieurs pics importants qui n'existent pas dans la modélisation. Ceci vient en effet, de la non-considération des relevés donc la valeur est supérieure à 360 secondes et qui influent énormément sur la variabilité (rappelons que la variabilité est le rapport entre écart-type et moyenne).

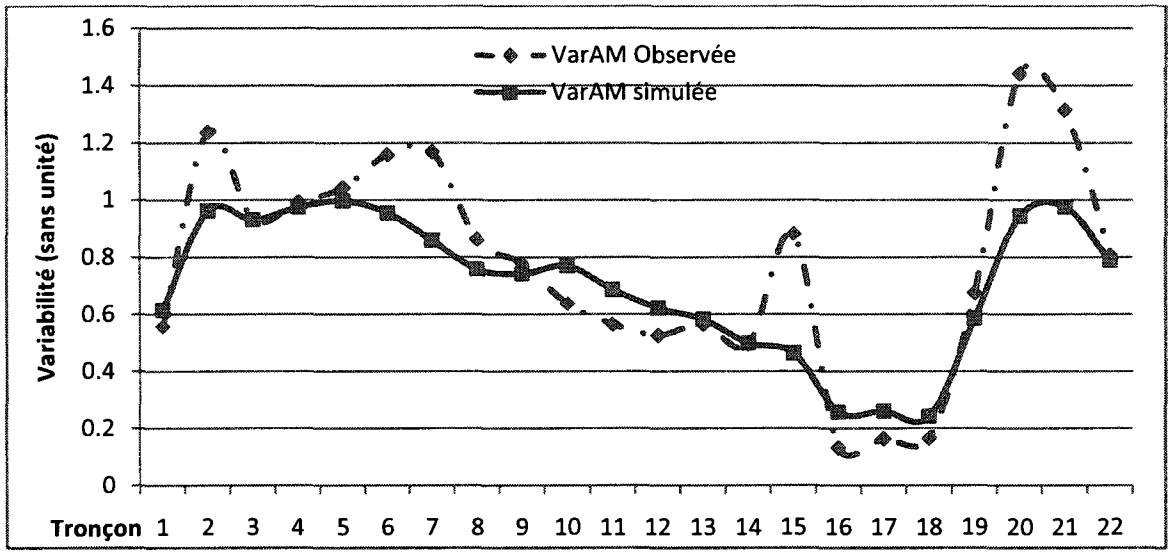


Figure 7-5 Variabilités observée et simulée par tronçon du circuit 1 en période AM

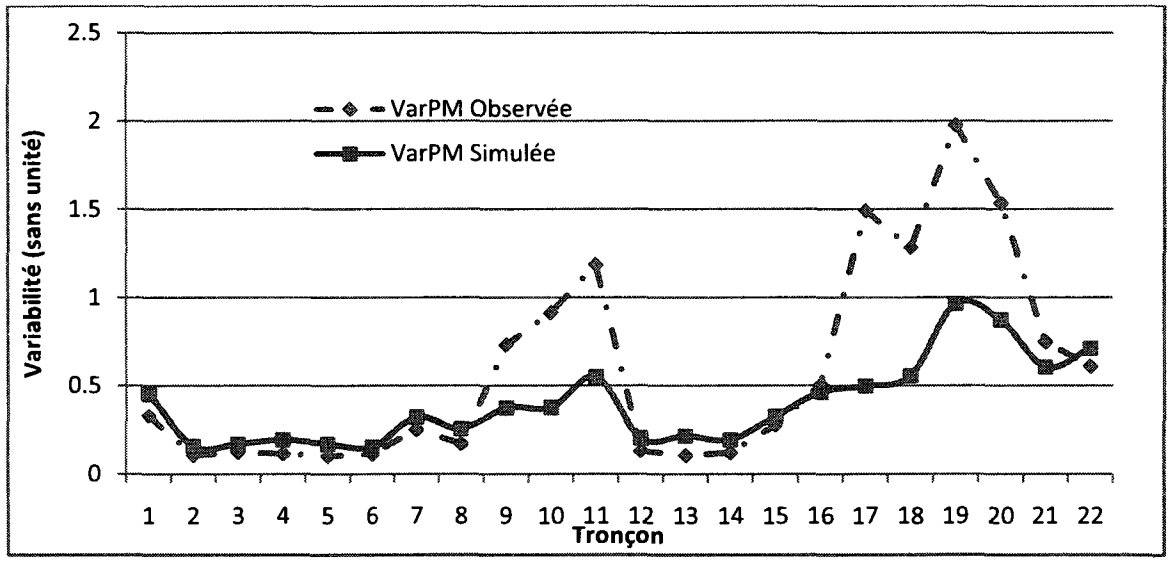


Figure 7-6 Variabilités observée et simulée par tronçon du circuit 1 en période PM

Cette constatation permet d'émettre plusieurs suppositions concluantes. Premièrement, la modélisation relate donc de la congestion récurrente, en effet les valeurs des moyennes et variabilités des tronçons peuvent être considérées comme des valeurs types, correspondant à une circulation normale (congestionnée ou non). Ensuite, les pics observés dans la Figure 7-6,

relèveraient justement de la non récurrence d'évènements « catastrophiques », c'est-à-dire des incidents hors congestions.

En conséquence, il devient alors intéressant de concevoir deux indicateurs permettant de séparer la circulation normale sur le circuit 1 et la probabilité d'incidents graves tout au long du circuit.

7.2.2 Simulation selon l'heure de départ

Comme nous l'avons présenté dans les parties précédentes, l'heure de départ sur un tronçon est un facteur significatif important. En effet, dans le cas des tronçons 7 à 10, la période de congestion matinale est limitée dans le temps. Hors la modélisation choisie ne permet pas de clarifier ce comportement : nous ne possédons ici qu'une moyenne et une variabilité globale des temps de parcours pour chaque tronçon durant la période matinale.

Par conséquent, toujours en considérant la codification, il est intéressant de calculer sur des créneaux de 10 minutes, la proportion de relevés de type 0, 1 ou 2.

Le Tableau B-8 donne les résultats pour le tronçon 2 durant la période matinale, tronçon choisi comme étant le tronçon initial du circuit puisque le tronçon 1 est sujet à des questionnements concernant la fiabilité expérimentale de ses relevés (précision géographique du début du circuit).

Ainsi nous pouvons pour chaque tronçon, selon l'heure de départ, évaluer le temps de parcours moyen et ainsi reconstituer un temps de parcours moyen du circuit.

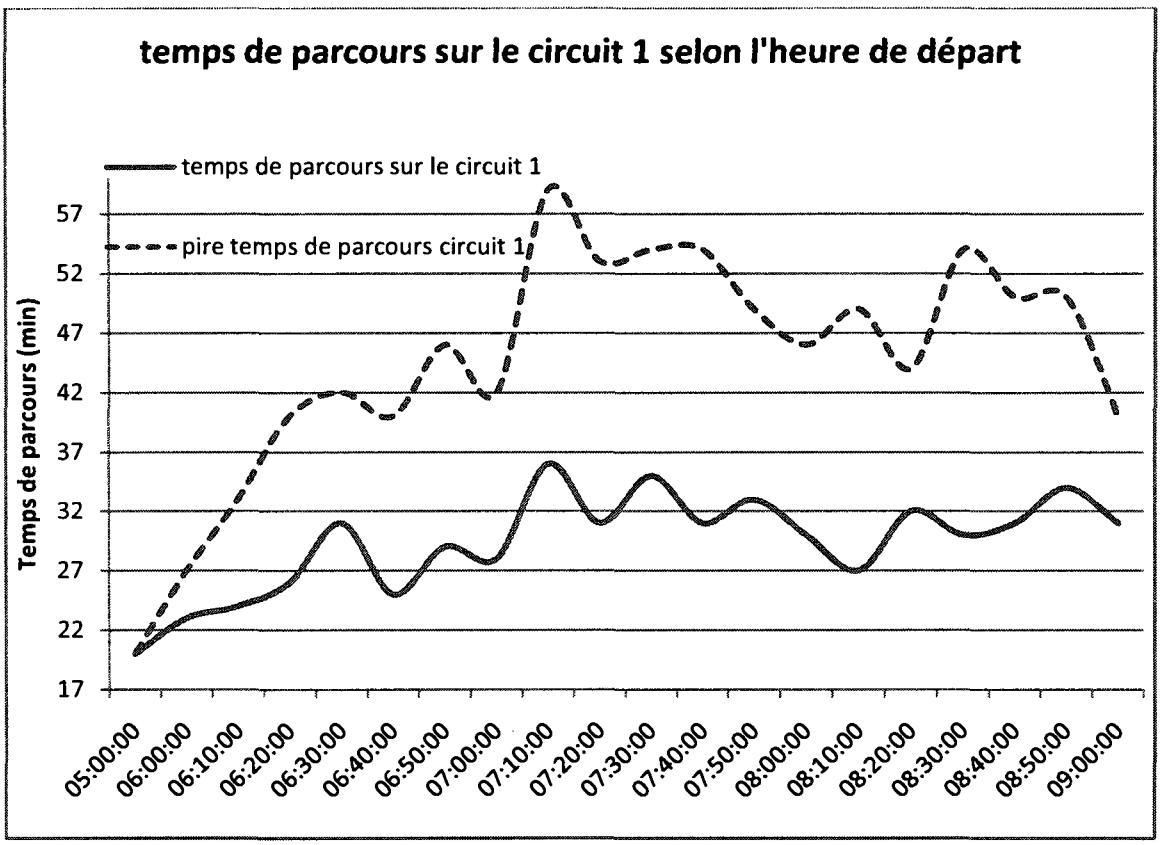


Figure 7-7 Temps de parcours moyen pour le circuit 1 selon l'heure de départ

Le pire temps de parcours correspond au temps de parcours moyen ajouté du temps de parcours multiplié par la variabilité pour chaque tronçon. Par conséquent, on observe sur la Figure 7-7 Temps de parcours moyen pour le circuit 1 selon l'heure de départ un pic correspondant à une heure de départ sur le circuit 1 au niveau du tronçon 2 à 7h10.

La simulation a donc permis de donner une représentation statistique du circuit 1, mais celle-ci ne peut être utilisée directement par le gestionnaire ou bien par l'usager de la route.

7.3 Synthèse

Conclusions : La simulation issue de du calcul simple de la moyenne et de la variabilité des temps de parcours, semble relativement cohérente. Nous avons aussi auparavant choisi de codifier les temps de parcours. Cette codification permet de savoir pour l'usager de la route pris dans une congestion la probabilité d'être toujours dans une congestion un kilomètre plus loin.

Perspectives : Munis désormais de nouvelles valeurs descriptives issues de la simulation, nous pouvons développer les indicateurs découlant de toutes les idées que nous avons préalablement exposées.

On constate aussi que la simulation et l'étude de la circulation normale incluent les très grands relevés et les considèrent comme étant inférieurs à 360 secondes (borne supérieure partialement choisie). Ont pourrait donc extrapoler nos résultats et donc signifier que si la variabilité observée est supérieure à la variabilité simulée, c'est qu'il y a présence de congestion non-récurrente.

CHAPITRE 8 : DÉVELOPPEMENT D'INDICATEURS

Un indicateur est une valeur permettant de donner une information pertinente pour son utilisateur. Celui-ci doit donc être bijectif et fiable. Sachant qu'une modélisation a été effectuée, et que la simulation de la variabilité a mis en exergue les problèmes liées aux relevés « catastrophiques » de temps de parcours, nous allons développer deux indicateurs, le premier donnant une information sur la proportion d'incidents « graves » et le second unissant moyenne et variabilité simulées.

8.1 Indicateur d'incidents

L'indicateur d'incidents est la mesure du pourcentage de relevés « catastrophiques ». Dans l'exemple du circuit 1, il correspond aux relevés de temps de parcours dont la valeur est supérieure à 360 secondes, soit une vitesse inférieure à 10 km/h. Cet indicateur est donc directement issu des données dont nous avons connaissance.

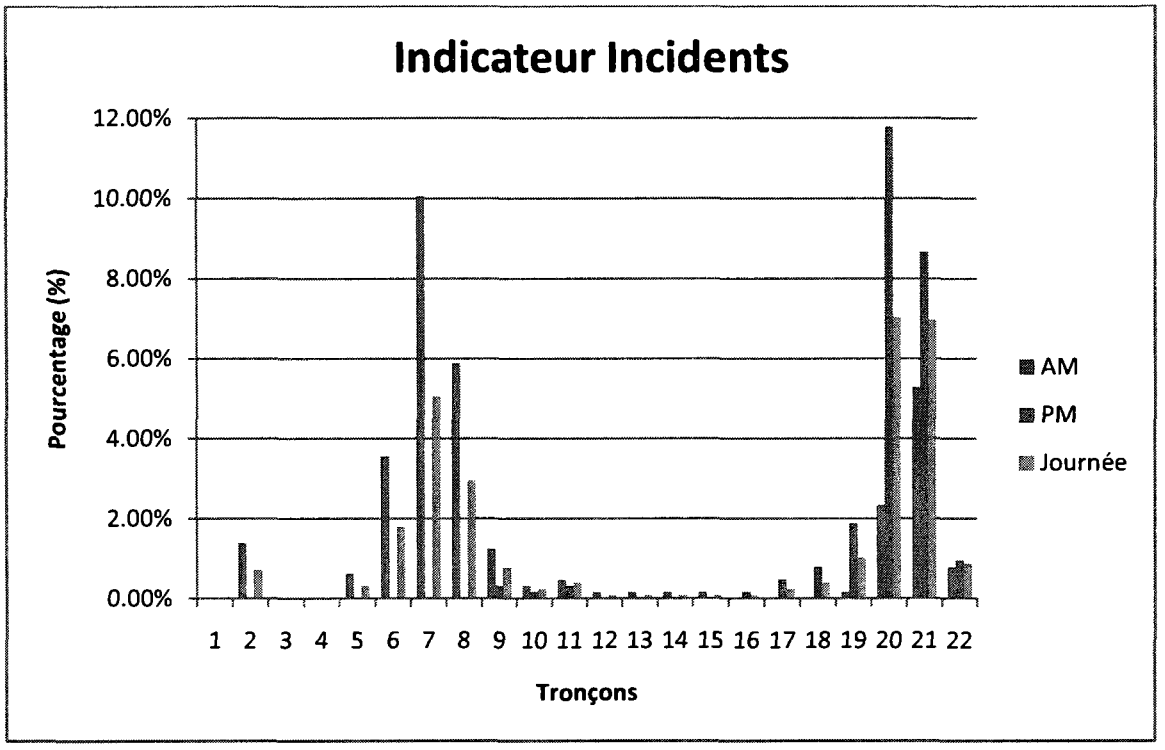


Figure 8-1 Indicateurs d'incidents pour le circuit 1

La Figure 8-1 expose donc le fait que les tronçons 2, 6, 7, 8, 19,20, 21, ont de gros pourcentages d'incidents « graves ». Effectivement, nous retrouvons les pics observés dans la Figure 7-5 et la Figure 7-6. En conséquence de quoi, cet indicateur donne une idée correcte des tronçons qui sont des zones propices à des incidents « catastrophiques ».

8.2 Indicateur Moyenne-Variabilité (IMV)

Comme dit dans la partie 7, à la vue des résultats donnés, et à partir de la simulation, il devient possible de construire un indicateur. En faisant écho à la revue de littérature, un indicateur est d'autant plus pertinent, qu'il porte de l'information et qu'il présente une forme de bijection.

Quelles sont les deux valeurs que nous avons utilisées tout au long de cette analyse ? La moyenne et la variabilité des temps de parcours durant une période donnée. Il devient donc

dans une idée toujours permanente de fiabilité, de créer un indicateur pouvant à la fois porter une information de performance (moyenne) et une information de fiabilité (variabilité).

L'objectif majeur, toujours dans un but de fiabilité, est donc de différencier quatre états de circulation sur un tronçon :

- Petite variabilité et petite moyenne ;
- Petite variabilité et grande moyenne ;
- Grande variabilité et petite moyenne ;
- Forte variabilité et forte moyenne.

Ces quatre états semblent donc convenir complètement à toutes les études précédemment présentées au cours de ce mémoire.

8.2.1 Présentation de l'indicateur

Pour différencier ces quatre états au moyen d'un simple graphe, nous choisissons de construire l'indicateur IMV, pour « indicateur de moyenne et de variabilité », comme un nombre complexe dont la formule suit :

$IVM = (m - m_{seuil}) - i.\frac{(v_{seuil}-v)}{v_{seuil}}$, où m désigne la moyenne, v la variabilité et i le nombre tel que $i^2 = -1$. Les valeurs seuils sont choisies de manière partiale ; on choisira pour la valeur seuil de la moyenne la borne supérieure de la circulation normale dans le cas de circuits du groupe 1 et 2, et la valeur seuil de la variabilité sera définie par l'utilisateur. Notons que la formulation complexe de cet indicateur est complètement partiale, elle a été simplement choisie afin de réduire en une seule information la moyenne et la variabilité.

Dans le but d'une utilisation simple pour l'usager de la route en particulier, l'indicateur peut être catégorisé. Effectivement, la formule mathématique définissant l'indicateur IVM permet de caractériser un nouvel indicateur catégorique IVM_c, à 4 états selon les signes des parties réelle et imaginaire.

En voici la codification :

- Si Re(IVM) < 0, et Im(IVM)<0, alors IVM_c = 0 (petite moyenne, petite variabilité).
- Si Re(IVM) > 0, et Im(IVM)<0, alors IVM_c = 1 (grande moyenne, petite variabilité).
- Si Re(IVM) < 0, et Im(IVM)>0, alors IVM_c = 2 (petite moyenne, grande variabilité).
- Si Re(IVM) > 0, et Im(IVM)<0, alors IVM_c = 3 (grande moyenne, grande variabilité).

8.2.2 Résultats

8.2.2.1 Valeurs de l'IMV selon la période

Nous présentons dans cette partie les valeurs de l'IMV pour le circuit 1 par tronçon. On constate donc dans le cas de la Figure 8-4 que presque tous les tronçons (1 à 18) ont une partie imaginaire presque nulle et négative et une partie réelle est elle aussi négative et proche de 0 : il n'y a pas de congestion forte, les temps de parcours moyens sur ces tronçons seront donc faibles, et la variabilité est aussi faible. Nous retrouvons par ailleurs que ce soit en matinée ou en après-midi, quelques groupes sur les graphes similaires à ceux trouvés lors du *clustering* des tronçons du circuit 1.

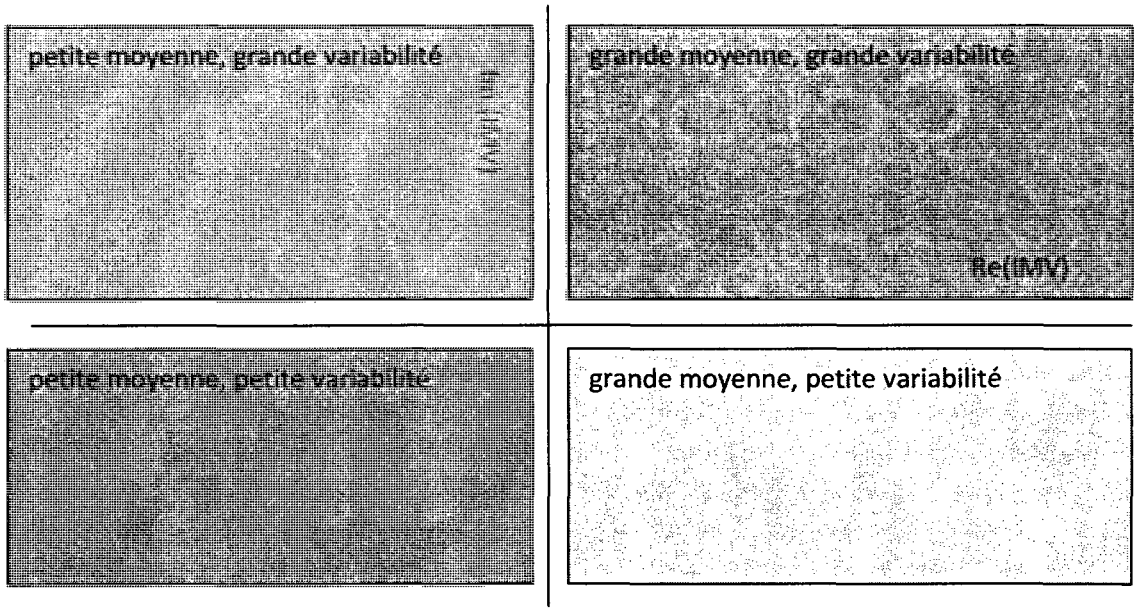


Figure 8-2 Représentation des 4 différentes zones de l'IMV

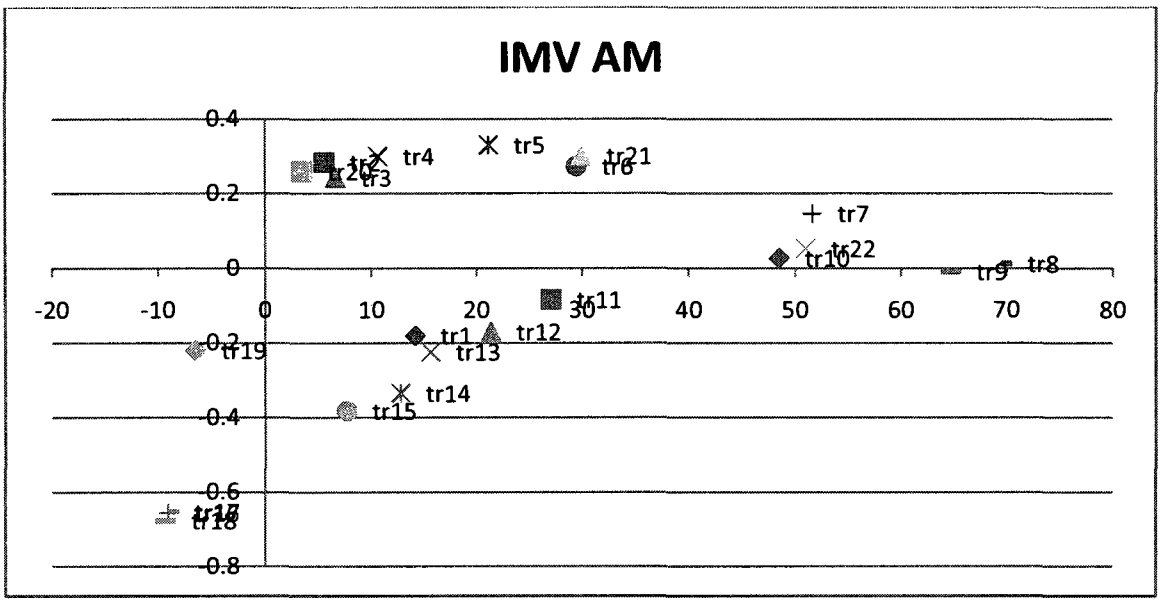


Figure 8-3 Valeurs de l'IMV par tronçon du circuit 1 AM

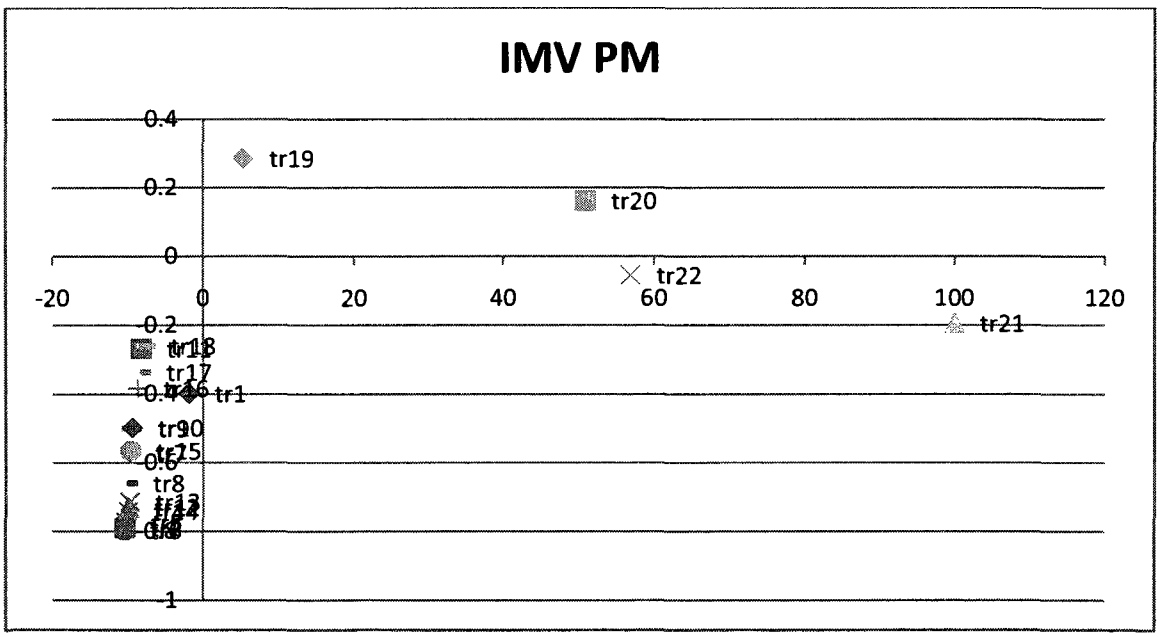


Figure 8-4 Valeurs de l'IMV par tronçon du circuit 1 PM

8.2.2.2 IMV_c par créneau de 10 minutes

Les résultats concernent toujours le circuit mais dans la période matinale. Les valeurs de l'indicateur sont calculées à partir de la modélisation statistique de la distribution des temps de parcours avec la codification précédemment utilisée.

Dans la mesure où les tronçons ont été échantillonnés, l'indicateur est évalué sur tous les tronçons du circuit 1 en période matinale, toutes les dix minutes.

On rajoute de plus un code de couleur pour la représentation sur Arc Gis :

- $IVM_c = 0$, couleur verte (petite moyenne, petite variabilité).
- $IVM_c = 1$, couleur jaune (grande moyenne, petite variabilité).
- $IVM_c = 2$, couleur orange (petite moyenne, grande variabilité).
- $IVM_c = 3$, couleur rouge (grande moyenne, grande variabilité).

Dans le but ensuite de valider le modèle statistique ainsi que la pertinence de l'indicateur, on cherchera alors à comparer les valeurs de l'indicateur issu de la modélisation et les valeurs de l'indicateur selon les observations.



Figure 8-5 Indicateur du circuit 1 période AM, 8h00



Figure 8-6 Indicateur du circuit 1 période AM, 8h10



Figure 8-7 Indicateur du circuit 1 période AM, 8h20

Les valeurs de l'indicateur indiquent donc que les tronçons 2 à 15 sont sujets à de fortes variabilités et/ou sujets à de fortes congestions. Toutefois on constate que les tronçons 16 à 19 ne sont pas congestionnés. Effectivement une analyse complémentaire démontre que, au niveau de l'échangeur A13S – A40E, la situation est très congestionnée sur l'A40. De plus, la valeur de l'indicateur pour tous les créneaux de 10 minutes en matinée des 6 tronçons les plus proches de l'échangeur est 1, ce qui rappelons-le, équivaut à une faible variabilité et une forte moyenne des temps de parcours.

Dans le but ensuite de comparer les valeurs de l'indicateur obtenues par simulation et les valeurs directement issues des observations, nous observons sur les 400 valeurs de l'indicateur, 313 concordances directes, et 73 valeurs acceptables. Les valeurs acceptables correspondent à des valeurs limites. Par exemple, la simulation donne une variabilité de 0.90 alors que les observations donnent une variabilité de 1.05, ceci correspond donc à un changement d'état de notre indicateur.

C'est ainsi que l'on pourrait optimiser la formule de l'indicateur en jouant sur les valeurs seuils, afin soit de prendre en compte la topographie et les caractéristiques d'un tronçon (présence d'un échangeur, limitation de vitesse...), soit de trouver des valeurs seuils afin de maximiser le nombre de concordances directes entre valeurs de l'indicateur simulé, et observé.

8.3 Exemple d'étude : évolution de la moyenne et de la variabilité pour l'année 2004, et indicateur d'incident

Compte-tenu du mandat, l'année 2004 étant l'année la plus récente de notre base de données, il faudrait au moyen des analyses précédentes évaluer l'évolution de la congestion pour cette année, face à la congestion toutes années confondues.

8.3.1 Évolution de la moyenne des temps de parcours pour 2004

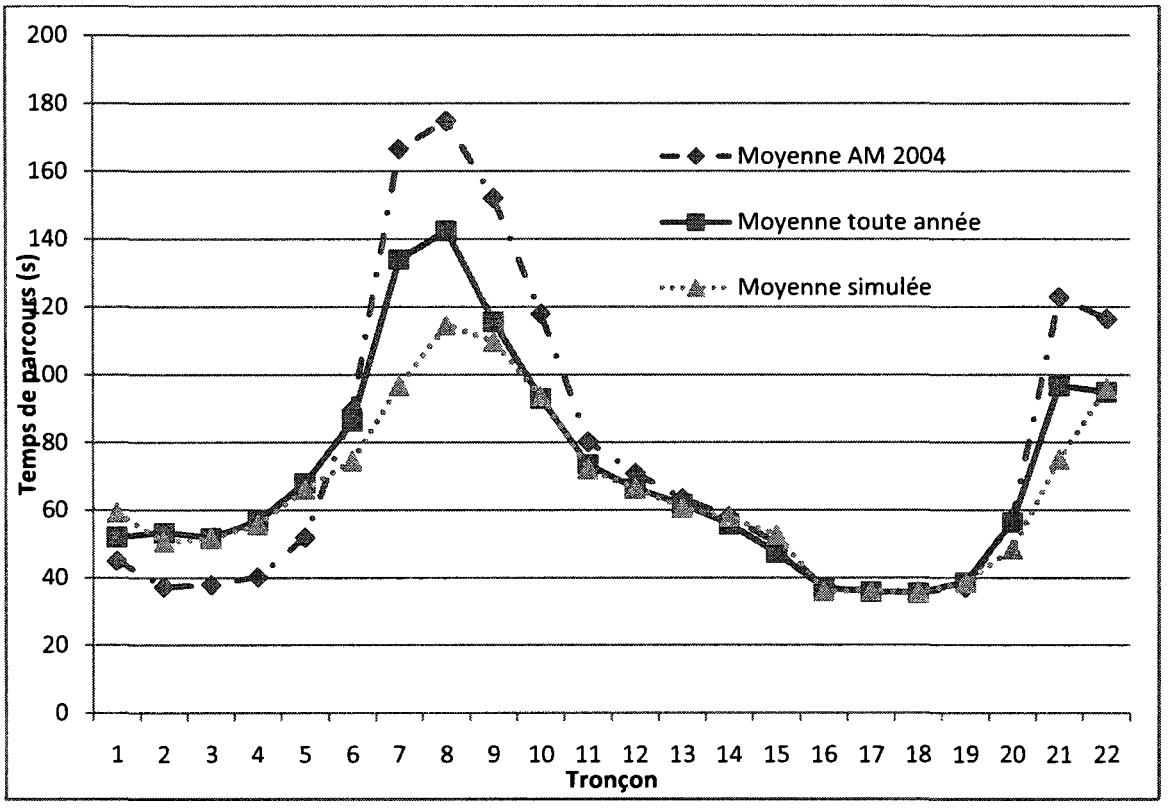


Figure 8-8 Évolution de la moyenne pour l'année 2004 période AM

A la vue de la figure précédente, la congestion a donc évolué en particulier au niveau des tronçons 6 à 9. Plusieurs hypothèses peuvent en découler : est-ce une élévation de la variabilité, du nombre d'incidents graves, ... ?

8.3.2 Évolution de la variabilité des temps de parcours pour 2004

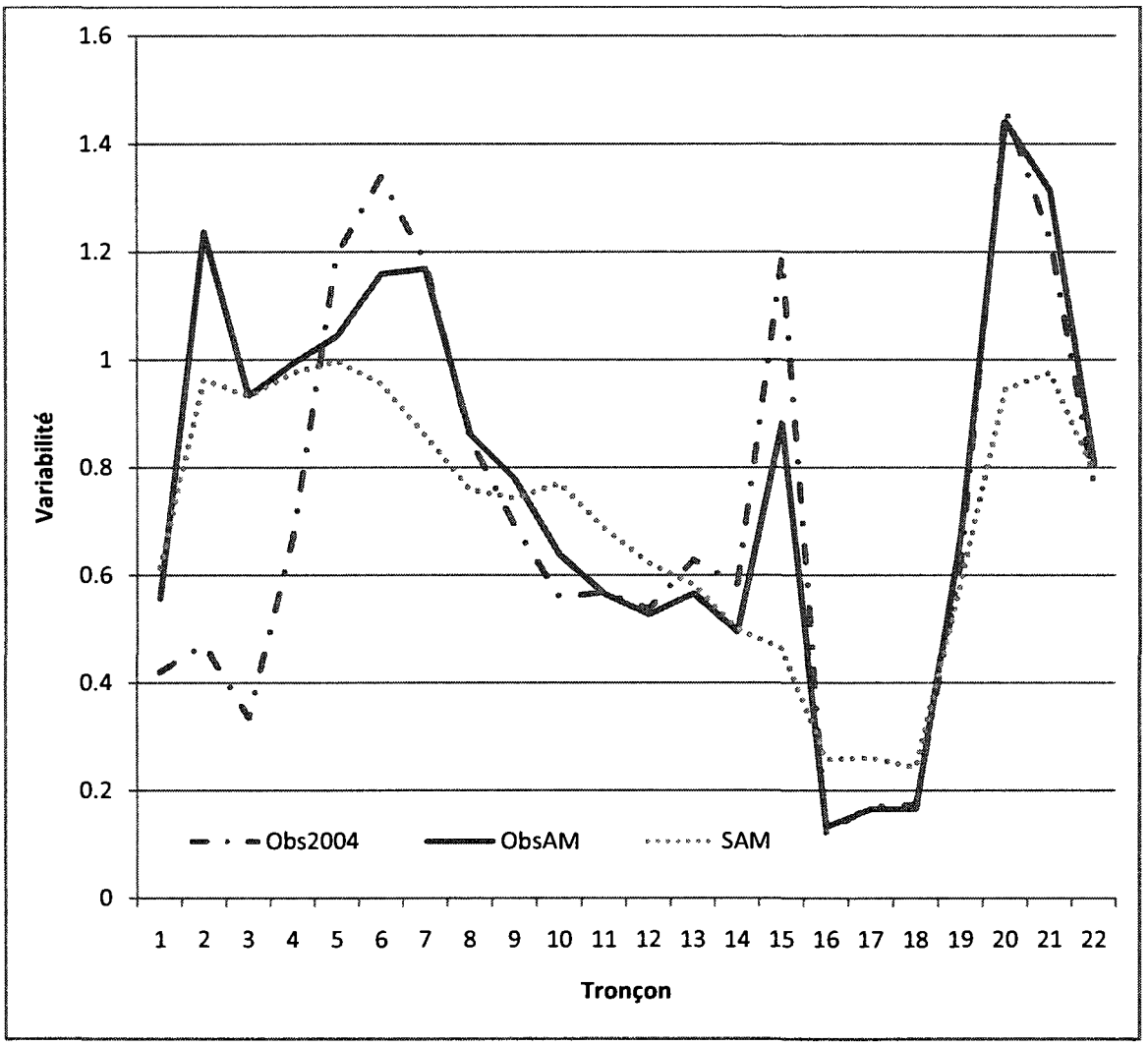


Figure 8-9 Évolution de la variabilité pour l'année 2004 période AM

D'après le graphe précédent, nous constatons donc seulement une augmentation de la variabilité au niveau du tronçon 6. En clair, si la moyenne des relevés de temps de parcours a largement augmenté, alors l'écart-type a encore plus grandi. Pour le reste des tronçons, l'évolution ne semble pas si critique.

8.3.3 Pourcentage de relevés dans la fenêtre temporelle et pourcentage d'incidents graves pour 2004

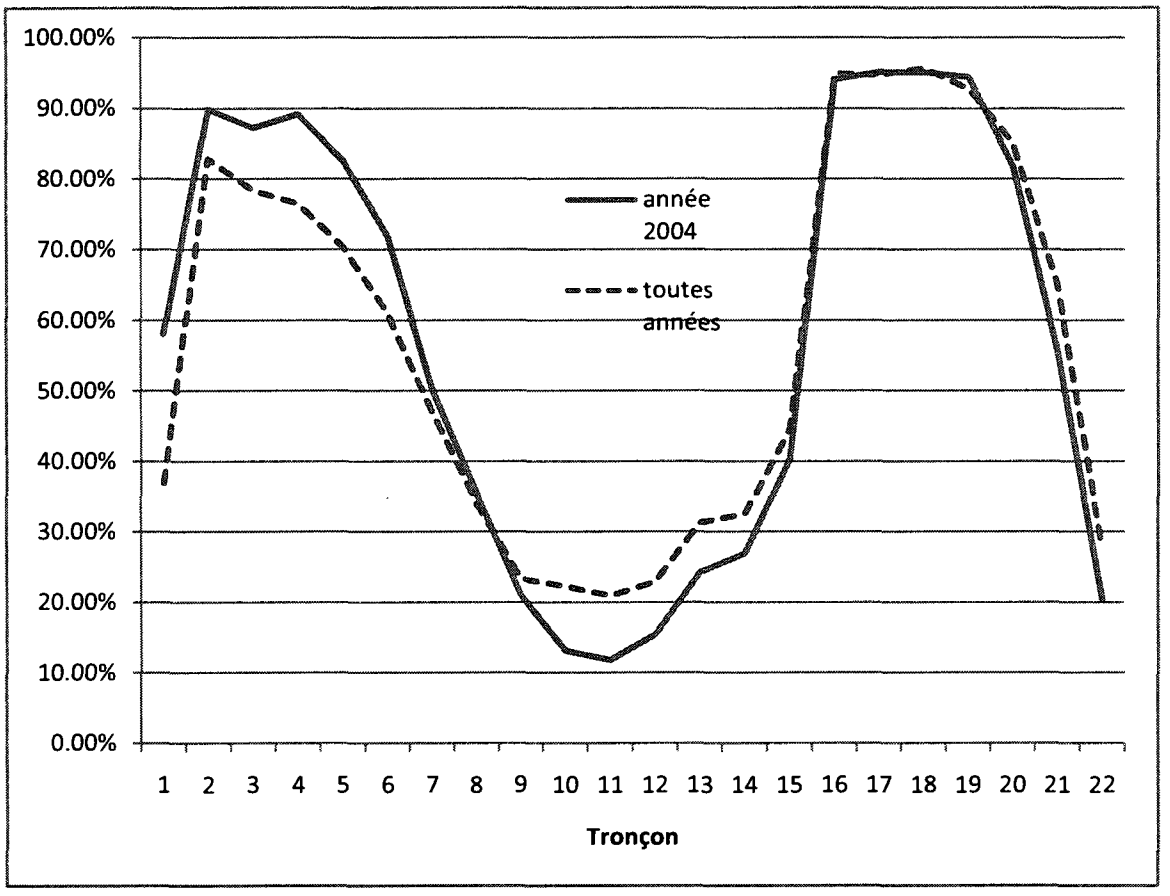


Figure 8-10 Pourcentage de relevés pour l'année 2004 période AM dans la fenêtre temporelle

La Figure 8-10 Pourcentage de relevés pour l'année 2004 période AM dans la fenêtre temporelle montre donc que mis à part les tronçons 10 et 11, le nombre de relevés dans la fenêtre temporelle a augmenté, ce qui signifierait une meilleure circulation. Toutefois la moyenne et la variabilité pour les tronçons 6 à 9 ont augmenté. Observons alors le pourcentage de relevés correspondant, selon notre définition, à des incidents graves.

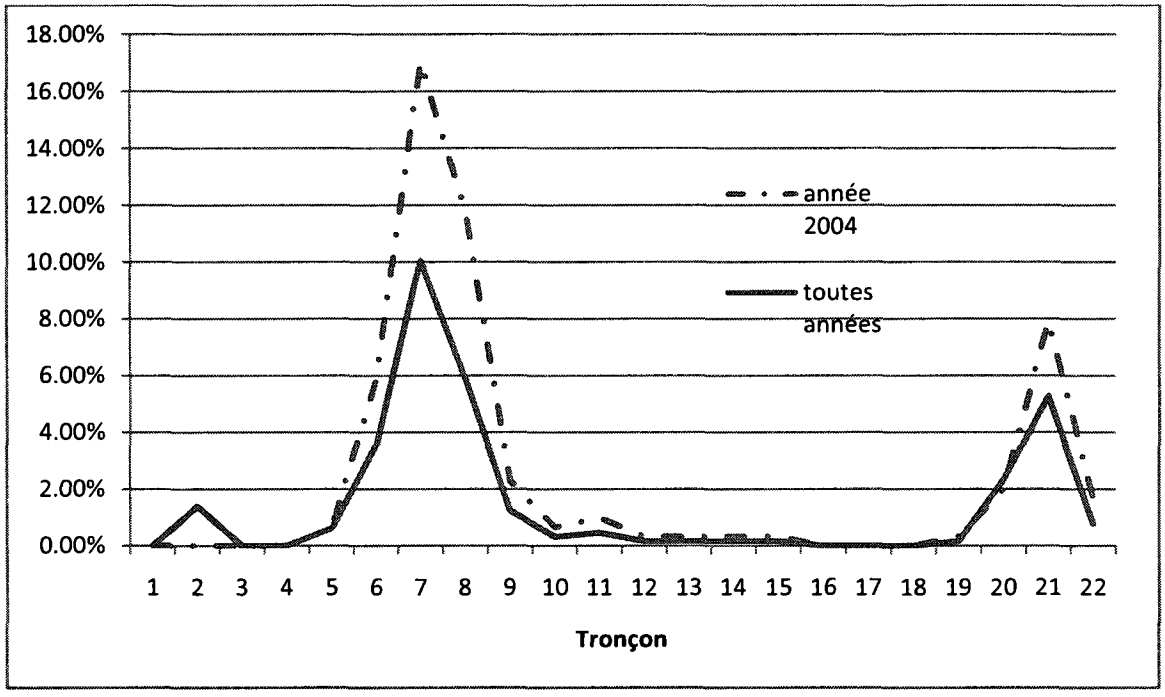


Figure 8-11 Pourcentage de relevés correspondant à des incidents graves pour le circuit 1 période AM

Nous constatons directement que c'est exactement en raison de l'augmentation du nombre de relevés au-dessus de 360 secondes, qui est la raison de ces évolutions catastrophiques. On pourrait par exemple justifier cela par un accroissement d'incidents non-récurrents, puisque le pourcentage de relevés dans la fenêtre temporelle pour ces tronçons a augmenté.

8.3.4 Simulation de la moyenne et de la variabilité pour l'année 2004 durant la période AM

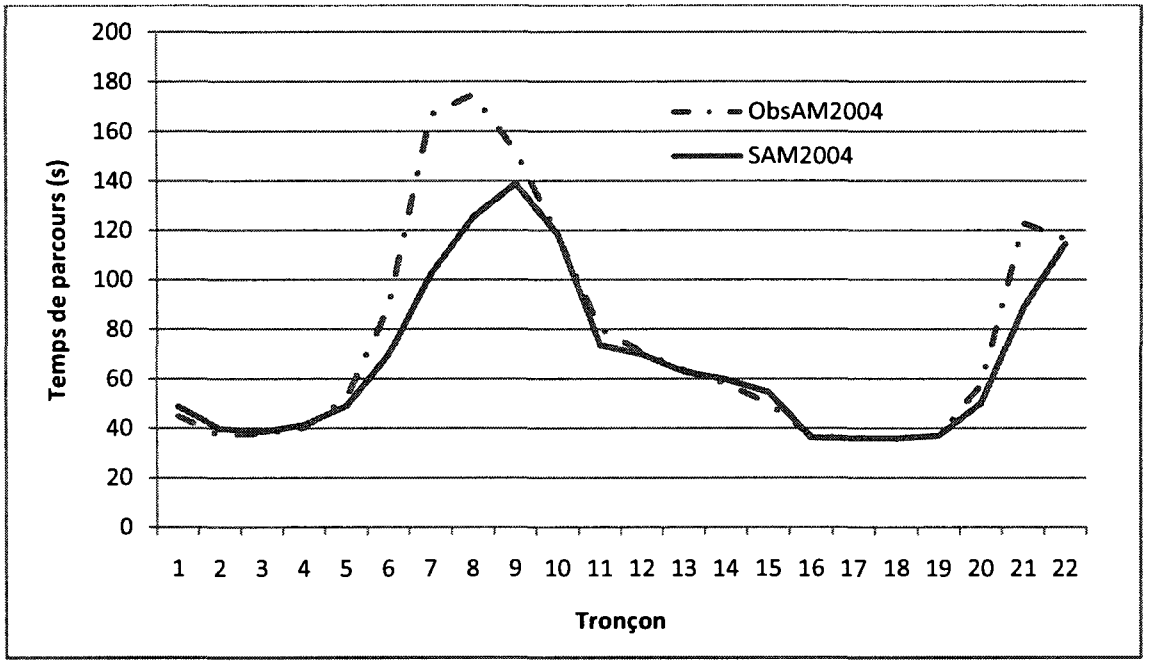


Figure 8-12 Simulation de la moyenne pour AM 2004 (courbe (SAM2004)

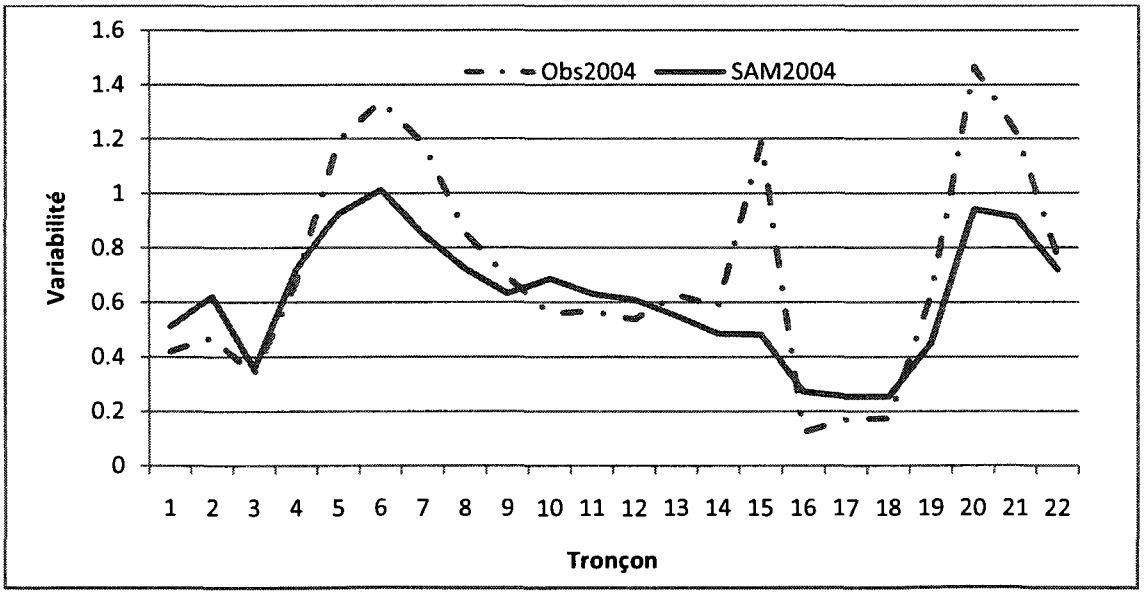


Figure 8-13 Simulation de la variabilité pour AM 2004 (courbe (SAM2004)

Les simulations de la moyenne et de la variabilité pour l'année 2004 durant la période matinale, confortent toutes les conclusions tirées dans les différentes parties de ce chapitre. En effet, on observe bien une variabilité augmentant pour le tronçon 15 seulement causée par des incidents non récurrents, puisque la différence entre les valeurs de moyenne simulées et observées est quasi-nulle. Enfin, pour les tronçons 6 à 9 on retrouve aussi une augmentation de ce pourcentage de relevés correspondant à des incidents graves.

En conclusion, la congestion récurrente n'a pas augmenté spatialement et en force durant l'année 2004, mais la proportion d'incidents importants a augmenté. Doit-on y voir un accroissement du nombre d'incidents graves tels que des accidents ?

8.4 Synthèse

Conclusions : Les données issues de la simulation ont permis de créer deux indicateurs. Le premier, indicateur d'incidents correspond au pourcentage de relevés dont la valeur est supérieure à un seuil (360 secondes, dans notre cas). Le deuxième indicateur est un indicateur informant son utilisateur de la moyenne et de la variabilité d'un tronçon donné. Cet indicateur possède la capacité d'être simplifié, et donc se destine autant au gestionnaire qu'au simple usager de la route.

Enfin, nous avons étudié par application des méthodes, et observations des résultats issus des méthodes, l'année 2004 pour le circuit 1 face à l'ensemble des données. Cette partie correspond au patron du type d'analyse que l'on peut effectuer grâce à nos données.

CHAPITRE 9 : CONCLUSIONS ET PERSPECTIVES

La revue de littérature a présenté une liste exhaustive de technologies et d'indicateurs en soulignant une définition de la congestion. La fiabilité est donc la problématique actuelle de tous les gestionnaires de la circulation. C'est pourquoi ce mémoire a donné une évaluation de la fiabilité du réseau autoroutier montréalais. Commençons donc par donner les conclusions de notre étude et ses limites, puis évaluons les perspectives de travaux futurs sur ce projet.

9.1 Conclusions

9.1.1 Synthèse

Tout au long de ce mémoire, nous avons essayé de clarifier et de donner une vision globale et un cheminement de la démarche statistique employée. Tout d'abord, une analyse exploratoire a été effectuée pour mettre en avant la signification de différents facteurs (mois de l'année, période, conditions climatiques,...).

Toutefois, nous avons pu constater que leur signification dépendait très largement du circuit considéré, et qu'il était donc très difficile de généraliser les conclusions obtenues suite à l'analyse statistique descriptive d'un circuit à tous les circuits.

Plus particulièrement, pour un circuit, nous avons pu effectuer un *clustering* concernant la distribution fréquentielle des relevés de temps de parcours et donc proposer un échantillonnage pratique qu'il a fallu modifier afin de repartir équitablement sur le réseau montréalais.

L'analyse descriptive ainsi que l'analyse par *clustering* ont mis en évidence l'existence d'une circulation normale pour le circuit 1. Cette circulation normale correspondant ni plus ni moins à des conditions de circulation fluide a été appréhendée et modélisé au moyen de lois log-normales. Puis, ce fut au tour de la partie droite de la distribution fréquentielle des relevés de temps de parcours, de subir une modélisation.

La modélisation achevée, nous avons pu alors construire des arbres de décision, et simuler les moyennes et les variabilités pour chaque tronçon du circuit 1. Munis de ces valeurs, de nouveaux indicateurs ont été mis en place : probabilités d'incidents et Indicateur Moyenne-Variabilité. Le premier indicateur se relève être une mesure de la congestion non-récurrente et le deuxième présente deux niveaux d'utilisation puisqu'il peut être utilisé par le gestionnaire dans sa forme mathématique ou par l'usager dans sa forme catégorielle.

Enfin, afin de surveiller l'évolution des temps de parcours au fil du temps, grâce aux modèles que nous avions préalablement établis nous nous sommes attardés sur l'année 2004 et l'état de la circulation du circuit 1 pour cette année-ci. Les résultats semblent laisser croire que cette année n'a pas été significativement plus contraignante que les années précédentes.

9.1.2 Limitations

Plusieurs limitations ont été soulignées au cours de ce mémoire. Premièrement il y a le problème de la qualité de l'information dont nous disposons. Il peut s'agir d'une erreur expérimentale initiale, mais cette erreur peut aussi venir de notre segmentation en tronçon d'1 km. Il existe en effet toujours le problème du repérage exact du début et de la fin du circuit.

Ce problème sera, notons-le, complètement résolu par l'utilisation de la technologie Bluetooth.

Nous avons de plus observé les limites de nos travaux lorsque nous avons voulu modéliser la distribution fréquentielle des temps de parcours pour d'autres circuits, et généraliser nos résultats. Des circuits empruntent des boulevards, voire mixtes (autoroutes et boulevard) et l'analyse mathématique en est alors mauvaise. Il faudra alors effectuer une autre analyse ou un autre découpage des circuits (cf. 9.2.1).

9.2 Perspectives

9.2.1 Vers un nouveau groupement, et une vision à long terme du trafic

Afin de se rapprocher de la réalité et donc de mieux appréhender la fiabilité du réseau, nous avons choisi de considérer les relevés odométriques d'un circuit. Une généralisation de ces

travaux pour les autres circuits montre les limites de notre modèle. En effet le type de route prouve son importance dans le cas d'un circuit qui serait mi autoroute, mi boulevard.

Alors pourquoi ne pas considérer les relevés des tronçons pour tout le réseau montréalais et effectuer de nouveau une modélisation, puis une simulation. Les travaux concernant ce sujet sont actuellement en cours. Ceux-ci passent par 4 étapes :

1. Préparation de la base de données : nous rajoutons donc pour chaque tronçon des facteurs topographiques, comme le sens de déplacement, le type de route, les intersections qu'il porte...
2. *Clustering* vis-à-vis de la distribution fréquentielle des relevés de temps de parcours pour un tronçon. Nous choisissons alors un grand nombre de groupes afin de distinguer les 3 vitesses limites du réseau actuel 50, 70 et 100km/h.
3. A partir de ce moment, nous modélisons et définissons les bornes et catégories de relevés de temps de parcours avec une justification mathématique cette fois, puis pouvons simuler la moyenne et la variabilité des temps de parcours.
4. L'indicateur d'incidents voit donc sa nature changer, et devient plus pertinent.

L'espoir réside dans le fait que les mathématiques pourront donc différencier ou expliquer l'appartenance de tronçons à un groupe particulier sans avoir à se soucier de leurs caractéristiques topographiques.

A partir de ce moment, il deviendra certain que nous pourrons effectuer un nouvel échantillonnage, beaucoup plus succinct, car connaissant seulement les caractéristiques d'un tronçon quelconque nous pourrons déjà savoir quelle sera sa distribution fréquentielle de relevés de temps de parcours.

Dans le cadre du futur échantillonnage, ces travaux serviront à la comparaison entre l'état de la circulation actuelle et l'état de la circulation entre les années 1998 et 2004. Plusieurs outils devront être alors développés : carte de contrôle, modification de l'échantillonnage dans le cas de tronçons devenus problématiques. La difficulté latente sera donc de tenir à jour ces

données et d'actualiser constamment les résultats obtenus. Bien qu'une base d'échantillonnage initiale puisse être discutée avant le lancement de l'opération, la collecte des données devra être modifiée selon l'actualisation des résultats.

9.2.2 Vers une vision à court terme du trafic autoroutier montréalais

Grâce à l'utilisation de la technologie *Bluetooth*, le gestionnaire disposera d'une base de données plus effective et donc plus conséquente. Comme souligné dans la revue de littérature, les relevés seront de plus, beaucoup plus fiables.

Par conséquent, alors qu'actuellement les travaux débutent au moyen de nos bases de données, une modélisation à court-terme de la circulation autoroutière (minute par minute) sera envisageable. Cette nouvelle approche reposera non plus sur la modélisation de la distribution fréquentielle des relevés de temps de parcours d'un tronçon ou d'un groupe de tronçon, mais sur une analyse fréquentielle de Fourier des relevés de temps de parcours moyen sur un tronçon.

En effet, en raison des voyages origine-destination, une analyse de Fourier sur quelques tronçons du circuit 1, semble indiquer que le signal « temps de parcours » présente une périodicité de 20 minutes : on trouve toutes les 20 minutes une congestion dont l'amplitude varie selon l'heure de relevé. L'amplitude de ce signal sera alors la variable aléatoire du problème. On pourra donc y voir une analogie électrique, où le passage d'un tronçon à un autre s'effectue par une modification du gain (amplitude), et/ou un déphasage du signal (déplacement temporel des flux).

Cette approche est donc un prolongement de l'idée de matrices de transition dont nous avions discuté lors de la modélisation de la distribution fréquentielle des temps de parcours. Actuellement, à cause de nos bases de données, nous souffrons d'un manque d'information qui ne permet pas d'obtenir des résultats corrects. Espérons que la technologie *Bluetooth* sera la clé de cette analyse future.

En conclusion de ces deux très importantes perspectives, pourquoi alors ne pas concevoir une analogie économique ? Cette analogie proposera donc des solutions de choix de chemins selon un problème d'optimisation stochastique avec contraintes tel que le nombre de changement de voies. On en aura donc une vision à court terme et une vision à moyen et long terme.

RÉFÉRENCES

Alger, M. (2004). Real time traffic monitoring using mobile phone data, 49th European Study Group with Industry, University of Oxford (http://www2.maths.ox.ac.uk/ociam/Study-Groups/ESGI49/problems/vodafone/vodafone.pdf, page consultée en mars 2008)

Babin, A., Fournier, P., Gourvil, L. (2006). Base complète des relevés-tronçons des voitures flottantes du MTQ, Élaboration d'une base de données relative à la consommation de carburant et à l'émission de polluants à partir des informations des voitures flottantes, Ministère des transports du Québec (Service de la modélisation des systèmes de transport), Décembre 2006.

Babin, A., Gourvil, L. (2007). La base de données sur les relevés-tronçons des voitures flottantes du MTQ – Document technique, Ministère des transports du Québec (Service de la modélisation des systèmes de transport), Janvier 2007.

Babin, A., Gourvil, P., Fournier, P. (2007). Relevés des voitures flottantes du MTQ de 1998 à 2004. Exploitation par le Service de la modélisation des systèmes de transport de la Direction de la Planification du MTQ, Ministère des transports du Québec (Service de la modélisation des systèmes de transport), Octobre 2007.

Barceló, J. and e.a. J. Casas, (1999). Modeling advanced transport telematic applications with microscopic simulators: the case of AIMSUN2. Traffic and Mobility, Simulation, Economics, Environment. W.Brilon, F.Huber, M.Schreckenberg and H.Wallentowitz, Springer.

Bernard, M., J. Hackney and K. W. Axhausen (2006) Correlation of link travel speeds, 6th Swiss Transport Research Conference, Ascona.

Brookes, Mike (2005). Let's fight the traffic jams, Electronics World, June 2005, Vol. 111, pp.47.

Chen, Chao (2002). Freeway Performance Measurement System (PeMS), Ph.D. Thesis, Engineering – Electrical Engineering and Computer Sciences, University of California, Berkeley, 218 pages.

Cheng, Yudong, Zhang, Yi, Hu, Jianming, Li, Li (2007) Mining for Similarities in Urban Traffic Flow Using Wavelets, Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, Seattle, WA, USA, Sept. 30 - Oct. 3, 2007.

Chonchoi Lam, Terence (2000). The effect of Variability of Travel Time on Route and Time-of-Day Choice, Ph.D. Thesis, Economics, University of California, Irvine, 194 pages.

Cohen S., Bomier J, Schwab N. (2002) Estimation et prédiction des temps de parcours du réseau autoroutier, In : Modélisation du Trafic, Actes INRETS n° 83, pp 25-36, 2002 (www.inrets.fr/ur/gretia/publications/Pdfs/SC/Modelisation-2002-Texte.pdf, page consultée en décembre 2007).

Du, J., Aultman-Hall, L. (2007). Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues, Transportation Research Part A: Policy and Practice, Volume 41, Issue 3, Pages 220-232.

Emam, E. B. and Al-Deek, H. (2006). Using Rela-Life Dual-Loop Detector Data to Develop New Methodology fir Estimationg Freeway Travel Time Reliability, Transportation Research Record N° 1959, TRB, Washington, pp.140-150

FHWA (2004). Lessons Learned: Monitoring Highway Congestion and Reliability Using Archived Traffic Detector Data, Report n° FHWA-HOP-05-003, October 2004 http://www.ops.fhwa.dot.gov/publications/lessons_learned/index.htm.

FHWA (2005). Traffic Congestion and Reliability. Trends and Advanced Strategies for Congestion Mitigation, Final report prepared for FHWA by Cambridge Systematics, Inc. And Texas Transportation Institute, September 2005, 140 pages

(http://www.ops.fhwa.dot.gov/congestion_report/congestion_report_05.pdf, page consultée en mai 2008)

Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L. (2008). Understanding individual human mobility patterns, Nature, Vol. 453, June 2008, pp.779-782.

Gourvil, L., Joubert, F. (2004). Évaluation de la congestion routière dans la région de Montréal, Ministère des transports du Québec, Collection Études et Recherches en Transports – N° RTQ-04-01.

Guo, Huairui, Jin, Jionghua (Judy) (2006). Travel Time Estimation with Correlation Analysis of Single Loop Detector Data, Transportation Research Record: Journal of the Transportation Research Board, No. 1968, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 10–19.

Johnston, Scott (2005). Integrated congestion evaluatio and quantification : ICEQ, Thesis for Master of Applied Science, Graduate Department of Civil Engineering, University of Toronto, 179 pages.

Joubert, Gilles. (2000) « Les coûts de la congestion routière dans la région de Montréal », 35e Congrès annuel de l'AQTR - À l'heure des choix : du constat à l'action, Montréal, 23 p.

Ladier, C. (2005). Circulation routière : usages et qualités des indicateurs. Mémoire de master recherche Génie Civil (MEGA). Vaulx-en-Velin : École Nationale des Travaux Publics de l'État, 2005. Sous la direction de C. Buisson. (http://www.entpe.fr/fr/content/download/903/6637/file/Ladier-Master-2006.pdf, page consultée en décembre 2007)

Les Conseillers ADEC (1997). Évaluation des coûts de la congestion routière dans la région de Montréal : Rapport final. Pour le compte du ministère des Transports du Québec.

Litman, Todd (2008). Well Measured. Developing Indicators for Comprehensive and Sustainable Transport Planning, Report, Victoria Transport Policy Institute, 47 pages (http://www.vtpi.org/wellmeas.pdf, page consultée en mai 2008).

Lomax, Tim (2003). Measuring the distribution and costs of congestion, Texas Transportation Institute, Presentation in Washington, D.C., Traffic Congestion: Issues and Options Conference (http://tti.tamu.edu/helpful_downloads.htm, page consultée en mai 2008).

Lomax, Tim, Schrank, David, Turner, Shawn, Margiotta, Richard (2003). Selecting Travel Reliability Measures, Report of the Texas Transportation Institute, (http://tti.tamu.edu/documents/474360-1.pdf, page consultée en avril 2008)

Mazloumi, E., Currie, G., Sarvi, M. (2007) Assessing Measures of Transit Travel Time Variability and Reliability Using AVL Data, Paper presented at the 87th Annual Meeting of the Transportation Research Board, Washington D.C.

Mistel, Bryan (2007). Here and Now – How the North American Traffic Data Market Has Evolved, Traffic Technology International, June/July 2007, pp 46-50.

Philip J. Tarnoff, Jason S. Wasson, Stanley E, Young, Nicholas Ganig, Darcy M. Bullock, James R. Sturdevant, P.E., (2008). The Continuing Evolution of Travel Time Data Information Collection and Processing, Paper presented at the 89th Annual Meeting of the Transportation Research Board, Washington D.C., 2008, 14 pages

Moran Toledo, Carlos A. (2008). Framework for Estimating Congestion performance measures : from data collection to Reliability analysis: case study of Stockholm, Thesis, Division of trafic and Logistics, Department of Transport and Economics, School of Architecture and the Built Environment, Royal Institute of Technology, 223 pages. (http://www.infra.kth.se/~carlos/TTP6460/ALL_v2.pdf, page consultée en février 2008)

NCHRP (2003). Performance Measures of Operational Effectiveness for highway Segments and Systems, Synthesis 311, Transportation Research Board, Washington D.C., 2003. (http://trb.org/publications/nchrp/nchrp_syn_311.pdf page web consultée en mai 2008)

Nesamania, K.S., Chub, L., McNallyc, M.G., Jayakrishnan, R. (2007). Estimation of vehicular emissions by capturing traffic variations, Atmospheric Environment, IN PRESS.

OCDE (2001). Glossaire de l'OCDE. http://stats.oecd.org/glossary/detail.asp?ID=830, page consultée en mai 2008

Oh, Cheol (2003). Anonymous Vehicle Tracking for Real-Time Traffic Performance Measures, Ph.D. Thesis, Civil Engineering, University of California, Irvine, 209 pages.

Parimi, Venu (2004). Determination of reliability indices for congestion on urban arterial streets, Master of Science Degree in Engineering, Rajeev Gandhi Technical University, 105 pages.

Qu Li, Hu Jianming, ZHANG Yi (2007). A Flow Volumes Data Compression Approach for Traffic Network Based on Principal Component Analysis, Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, Seattle, WA, USA, Sept. 30 - Oct. 3, 2007.

Quiroga, C.A. (2000). Performance measures and data requirements for congestion management systems, Transportation Research Part C: Emerging Technologies, Volume 8, Issues 1-6, Pages 287-306.

Robinson, S., and J. W. Polak. Modelling Urban Link Travel-Time Using Data from Inductive Loop Detectors. Presented at 10th World Conference on Transport Research (WCTR 2004) (CD-ROM of unrevised papers), Istanbul, Turkey, July 4–8, 2004.

Robinson, Steve, Polak, John W. (2005). Modeling Urban Link Travel Time with Inductive Loop Detector Data by Using the k-NN MethodTransportation Research Record: Journal of the Transportation Research Board, No. 1935, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 47–56.

Robitaille, M., Nguyen, T. (2003). Évaluation de la congestion « De la théorie à la pratique » : Réseau routier de l'agglomération de Montréal, communication présentée au Congrès 2003 de l'Association des Transports du Canada, St John's – Terre-Neuve, http://www.tac-atc.ca/English/pdf/conf2003/robita_f.pdf.

Rumbaugh, J. (1995), OMT: Modélisation et conception orientées-objet, Masson.

Schmöcker, J.-D., Fonzone, A, Quddus, M., Bell, M.G.H. (2006). Changes in the frequency of shopping trips in response to a congestion charge, Transport Policy, Volume 13, pp. 217–228

Schneider W., Asamer J., Mrakotsky E., Toplak W. (2007). Influence of Environment Conditions on Traffic Flow, Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, Seattle, WA, USA, Sept. 30 - Oct. 3, 2007.

Schrank, D., Lomax, T. (2003). The 2003 Urban Mobility Report, Texas Transportation Institute, http://mobility.tamu.edu, May 2003.

Schrank, D., Lomax, T. (2005). The 2005 Urban Mobility Report, Texas Transportation Institute, http://mobility.tamu.edu, May 2005.

Schrank, D., Lomax, T. (2007). The 2007 Urban Mobility Report, Texas Transportation Institute, http://tti.tamu.edu/documents/mobility_report_2007_wappx.pdf (page consultée en mai 2008), September 2007.

Scott, Craig A. (1994). Improved GPS Positioning for Motor Vehicles Through Map Matching, University of Technology, Sydney, presented at ION-94, Salt Palace Convention Center, Salt Lake City, Utah, Sept. 20 - 23, 1994.

Shaw, T. et D. McLeod (1998). Mobility Performance Measures Handbook. Florida Department of Transportation, Systems Planning Office, Tallahassee, Florida, 1998.

Szekeres, D., Heckman, M. (2005). Statewide Congestion Measurement Study – Final Report, Pennsylvania Department of Transportation, Contract N° 030419, Juillet 2005.

Taylor, M.A.P., Woolley, J.E. Zito, R. (2000). Integration of the global positioning system and geographical information systems for traffic congestion studies, Transportation Research Part C: Emerging Technologies, Volume 8, Issues 1-6, Pages 257-285

Tremblay, G. et M. Robitaille (2005). Analyse des temps de parcours sur le réseau routier de la grande région de Montréal. Rapport final du Groupe Conseil GENIVAR au ministère des Transports du Québec, Laval-Mille-Îles, région de Montréal, 124 p. + 2 annexes.

Tremblay, P., Fournier, P. (2001). Développement d'un indicateur synthétique de congestion pour la région de Montréal – Document de travail, Ministère des transports du Québec (Service de la modélisation des systèmes de transport), Février 2001.

Trépanier, M., Chapleau, R. (2001). Analyse orientée-objet et totalement désagrégée des données d'enquêtes ménages origine-destination, Revue canadienne de génie civil, Ottawa, vol. 28, no 1, pp. 48-58.

Tu, Huizhao, Van Lint, J.W C., Van Zuylen, H.J. (2007). The Impact of Traffic Flow on Travel Time Variability of Freeway Corridors. Transportation and Planning Departement, Faculty of civil engineering and geosciences, Delft University of Technology, The Netherlands, presented at the 86th Annual Meeting of the Transportation Research Board.

Turner, Shawn, Richard Margiotta, and Timothy Lomax. (2004) Texas Transportation Institute and Cambridge Systematics, Inc. "Monitoring Urban Roadways in 2003: Current Conditions and Trends from Archived Operations Data." December 2004. (http://mobility.tamu.edu/mmp)

Van Lint, J. W. C., Van Zuylen, H. J. (2005). Monitoring and Predicting Freeway Travel Time Reliability, Using Width and Skew of Day-to-Day Travel Time Distribution, Transportation Research Record: Journal of the Transportation Research Board, No. 1917, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 54–62.

Washburn, Scott Stuart (1999). New Technologies for Data Collection and their Application for Empirical Investigation of Travel Time Measurement Issues, Ph.D. Thesis, Department of Civil and Environmental Engineering, University of Washington, 241 pages.

Whitehead, D., Fournier, P. (2000). Développement d'indicateurs descriptifs de la congestion routière – Document de travail, ministère des Transports du Québec (Service de la modélisation des systèmes de transport), Juillet 2000.

Wunnava, S., Yen, K., Babij, T., Zavalta, R., Romero, R., Archilla, C. (2007). Travel Time Estimation Using Cell Phones (TTECP) for Highways and Roadways, Final Report, project by the Florida Department of Transportation at Florida International University.

Xiang, Zhou, Jian, Rong, Jiancheng, Weng, Changqiao, Shao (2007). Network-wide Performance Assessment of Urban Traffic Based on Probe Vehicle Data, Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, Seattle, WA, USA, Sept. 30 - Oct. 3, 2007.

Zhu, Weihua, Boriboonsomsin, Kanok, Barth, Matthew (2007). Microscopic Traffic Flow Quality of Service from the Drivers' Point of View, Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, Seattle, WA, USA, Sept. 30 - Oct. 3, 2007.

ANNEXE A : ANALYSE STATISTIQUE DU CIRCUIT 1 AU MOYEN DE STATISTICA

A.1 Résultats descriptifs de la distribution des temps de parcours

Cette première partie a pour but de donner quelques résultats généraux quant à la distribution des temps de parcours observés sur le circuit 1.

A.1.1 Statistiques descriptives

Tableau A-1 Statistiques descriptives

Statistiques descriptives							
	Nombre de relevés	Moyenne	Minimum	Maximum	Percentile 10	Percentile 90	Ecart-type
temps de parcours	28296	61.28	4.72	3854.2	30.72	118.68	78.93

On observe donc un très grand écart entre minimum et maximum de cette distribution puisque le maximum vaut 3854.2s, ce qui correspond à presque 1h30 de temps passé sur un seul tronçon de 1 km, soit une vitesse de parcours de 0.93 km/h (cette observation pourrait être acceptée dans le cas où le tronçon sur lequel découle cette expérience est un tronçon sans sortie et ayant subi un incident très important, malheureusement d'autres observations sur les tronçons précédents ou suivants celui-ci à la même date semblent prouver qu'il s'agit d'une erreur expérimentale, à moins en dernier lieu que ce soit justement le véhicule flottant qui soit tombé en panne), alors que le minimum correspond à une vitesse aberrante de 762 km/h.

Néanmoins, la moyenne ainsi que les percentiles semblent confirmer l'idée générale d'une plage de temps de parcours raisonnable entre 30 et 120 s, avec une moyenne de 61,27s soit une vitesse de parcours égale à 59km/h. Par conséquent observons l'histogramme des temps de parcours.

A.1.2 Histogramme des temps de parcours

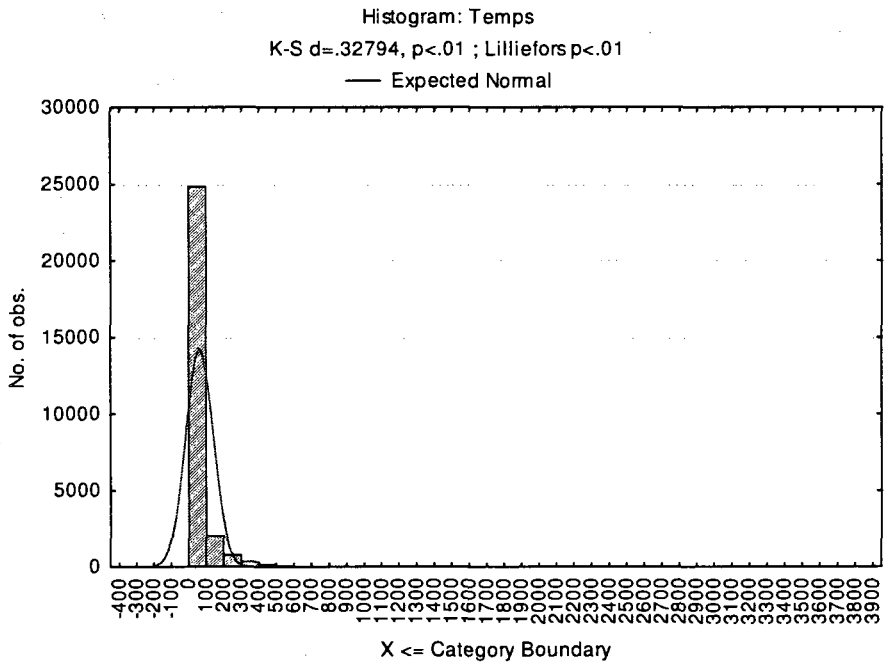


Figure A-1 Histogramme des temps de parcours

L'histogramme porte plusieurs conclusions pertinentes : premièrement, les observations semblent toutes être comprises dans un intervalle réduit. Deuxièmement, le test de Kolmogorov-Smirnoff donnant une distance de 0.32794 ainsi que la p-value très faible impliquent que l'hypothèse de normalité est à rejeter.

A.1.3 Tables de fréquence

Tableau A-2 Fréquence des observations par année

Tableau de fréquence : Année K-S d=.29592, p<.01				
	Nombre	**Cumulative**	**Pourcentage**	**Cumlative %**
1998	3950	3950	13.96	13.96
1999	1817	5767	6.42	20.38
2000	4596	10363	16.24	36.62
2001	2222	12585	7.85	44.48
2002	2595	15180	9.17	53.65
2003	0	15180	0	53.65

2004	13116	28296	46.35	100
Manquant	0	28296	0	

Tableau A-3 Fréquence des observations par mois

Tableau de fréquence : Mois K-S d=.13950, p<.01				
	Nombre	**Cumulative**	**Pourcentage**	**Cumulative %**
janvier	1248	1248	4.41	4.41
février	3622	4870	12.8	17.21
mars	3359	8229	11.87	29.1
avril	2267	10496	8.01	37.09
mai	1914	12410	6.76	43.86
juin	1868	14278	6.6	50.46
juillet	2285	16563	8.07	58.53
aout	1826	18389	6.45	64.99
septembre	3561	21950	12.58	77.57
octobre	3011	24961	10.64	88.21
novembre	2213	27174	7.82	96.03
décembre	1122	28296	3.96	100
Manquant	0	28296	0	

Ces tableaux dont l'utilité est certes réduite nous informent simplement du fait que sur ce circuit, le nombre d'observations est très largement suffisant pour effectuer les tests statistiques désirés.

A.1.4 Moyennes par tronçon

Observons désormais les différentes moyennes de tes temps de parcours selon le tronçon (sont rajoutés les intervalles de confiance à 95% de ces moyennes).

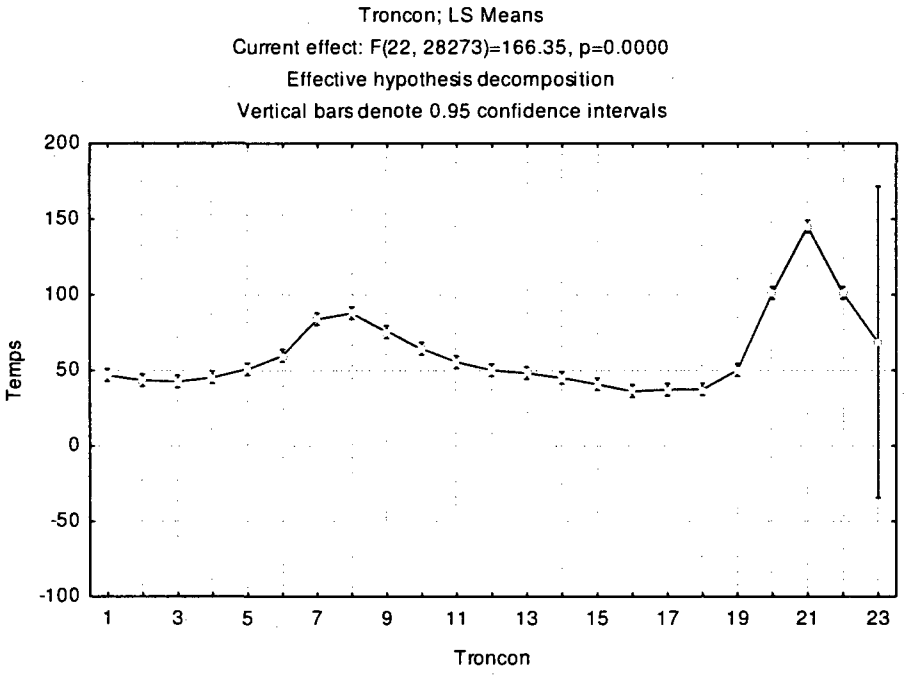


Figure A-2 Moyennes des temps de parcours selon le tronçon

On constate donc une nette augmentation du temps de parcours au niveau du tronçon 7, avec une dispersion au niveau du 10, alors que la fin du tronçon est problématique. On peut d'ores et déjà en donner la raison : Le tronçon 7 implique un échangeur et la fin du circuit correspond à l'arrivée au niveau de Montréal. En ce qui concerne le tronçon 21, nous pouvons émettre l'hypothèse raisonnable que ce tronçon est proche de la sortie de l'autoroute. Par conséquent, les usagers ralentissent, et le flux de voitures se densifie à l'approche de l'échangeur.

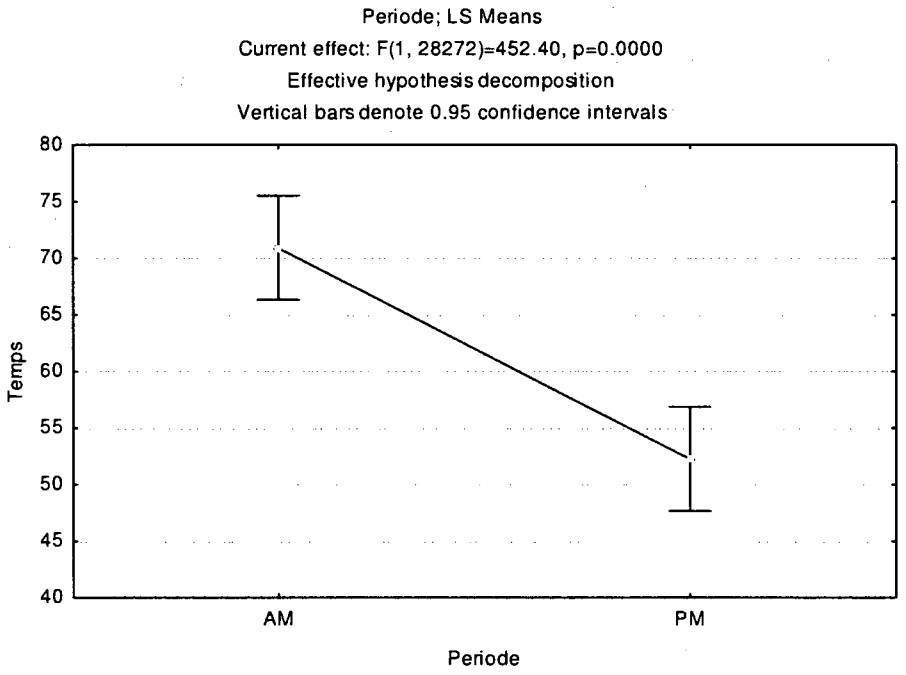


Figure A-3 Moyenne des temps de parcours selon la période

De plus, il y a une différence significative entre les relevés matinaux et de l'après-midi. En effet, pour la période AM, la moyenne des temps de parcours est de 71s alors qu'elle n'est que de 53s pour l'après-midi.

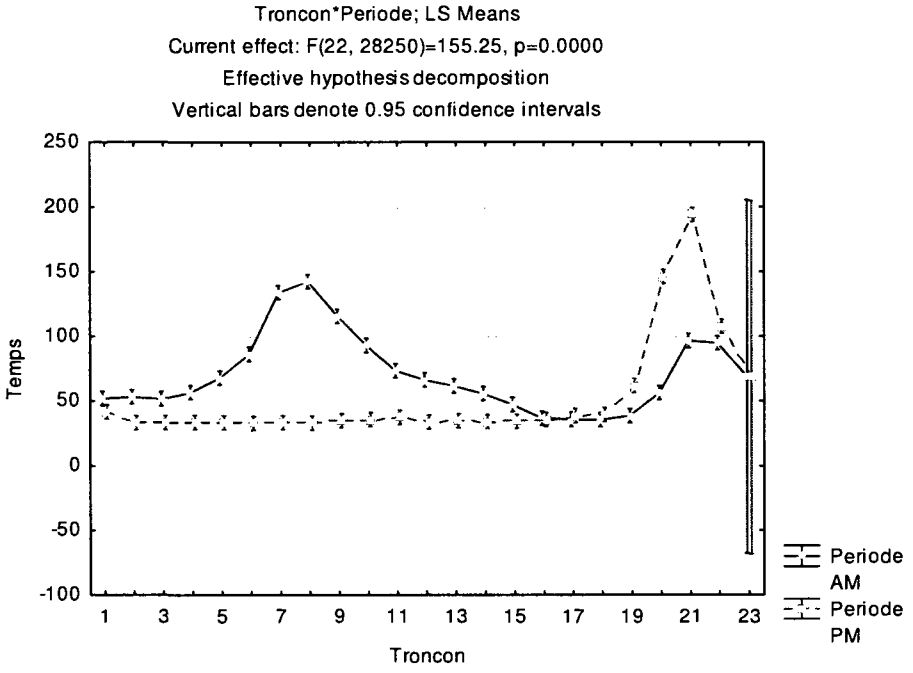


Figure A-4 Moyennes des temps de parcours selon le tronçon et la période

L'observation des temps de parcours selon le tronçon et la période, confirme cette constatation. En effet, on peut émettre l'hypothèse que cette différence – en particulier au niveau du tronçon 7 – correspond au départ des personnes vivant en zone urbaine périphérique et travaillant à Montréal.

Tableau A-4 ANOVA pour le tronçon et la période plus effet d'interaction

Tableau ANOVA					
	Deg de liberté	**Temps SS**	**Temps MS**	**Temps F-ratio**	**Temps p-value**
Constante	1	1.06E+08	1.06E+08	21902	0
Troncon	21	2.03E+07	9.66E+05	199	0
Periode	1	2.44E+06	2.44E+06	502	0
Troncon*Periode	21	1.66E+06	7.89E+05	163	0
Erreur	28250	1.37E+08	4851		
Total	28293	1.76E+08			

L'ANOVA confirme la signification des deux facteurs catégoriques tronçon et période ainsi que leur interaction.

A.2 Analyse de la variance selon l'année

Dans la perspective de dégager les évolutions des temps de parcours, nous choisissons de nous attarder en particulier sur les moyennes des temps de parcours selon l'année, facteur catégorique à 6 niveaux (1998-1999-2000-2001-2002-2004).

Tableau A-5 ANOVA selon l'année

Tableau ANOVA					
	Deg de liberté	**Temps SS**	**Temps MS**	**Temps F-ratio**	**Temps p-value**
Constante	1	7.02E+07	7.02E+07	11342	0
Année	5	1.30E+06	2.60E+05	42	0
Erreur	28290	1.73E+08	6185		
Total	28295	1.76E+08			

Le tableau d'analyse de la variance implique donc que le facteur année est très largement significatif.

Tableau A-6 Coefficients de la régression selon l'année

Estimation des coefficients						
	Niveau	**Colonne**	**Temps Coeff**	**Temps Std. Er.**	**Temps t**	**Temps p**
Constante		1	61.37	0.58	106.5	0
Année	1998	2	-13.83	1.17	-11.79	0
Année	1999	3	14.79	1.613	9.17	0
Année	2000	4	3.11	1.11	2.8	0.005
Année	2001	5	-1.75	1.48	-1.18	0.24
Année	2002	6	-4.28	1.39	-3.09	0.002

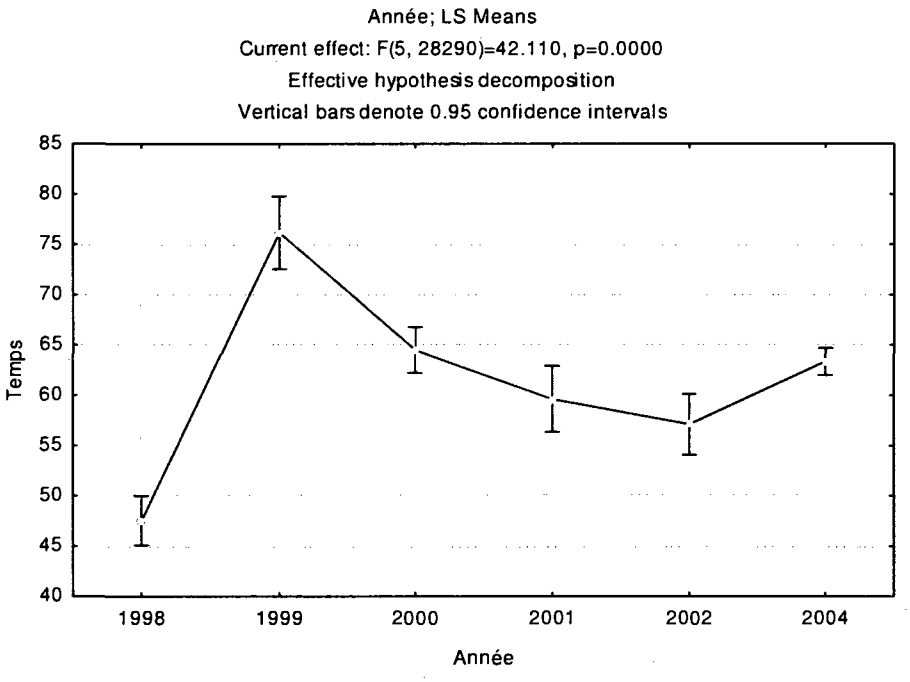


Figure A-5 Moyennes des temps de parcours selon l'année

On constate donc une amélioration de la moyenne des temps de parcours selon l'année. On ne doit toutefois guère s'avancer sur cette conclusion. En effet, tous les mois ne sont pas échantillonnés selon les années. De plus, l'année 2001 ne semble pas significative.

Tableau A-7 Moyennes et intervalles de confiances selon l'année

Statistiques Descriptives							
	Niveau	**N**	**Moyenne des relevés**	**Ecart-type**	**Erreur Std**	**Temps +95%**	**Temps - 95%**
Total		28296	61.28	78.93	0.47	60.36	62.2
Année	1998	3950	47.54	40.13	0.64	46.29	48.79
Année	1999	1817	76.16	85.53	2.01	72.23	80.1
Année	2000	4596	64.48	78.35	1.16	62.22	66.75
Année	2001	2222	59.62	65.44	1.39	56.9	62.35
Année	2002	2595	57.09	63.38	1.24	54.66	59.53
Année	2004	13116	63.33	90.41	0.79	61.79	64.88

Tableau A-8 Calcul des coefficients de signification

Coefficient de signification			
	R	**R^2**	**$R^{2\ ajusté}$**
Temps	0.09	0.01	0.01

On trouve de plus un R ajusté très faible.

Tableau A-9 Test de LSD selon l'année

Test LSD selon l'année (différence des moyennes entre années)							
	Année	**{1}**	**{2}**	**{3}**	**{4}**	**{5}**	**{6}**
1	1998		0.00	0.00	0.00	0.00	0.00
2	1999	0.00		0.00	0.00	0.00	0.00
3	2000	0.00	0.00		0.02	0.00	0.39
4	2001	0.00	0.00	0.02		0.27	0.04
5	2002	0.00	0.00	0.00	0.27		0.00
6	2004	0.00	0.00	0.39	0.04	0.00	

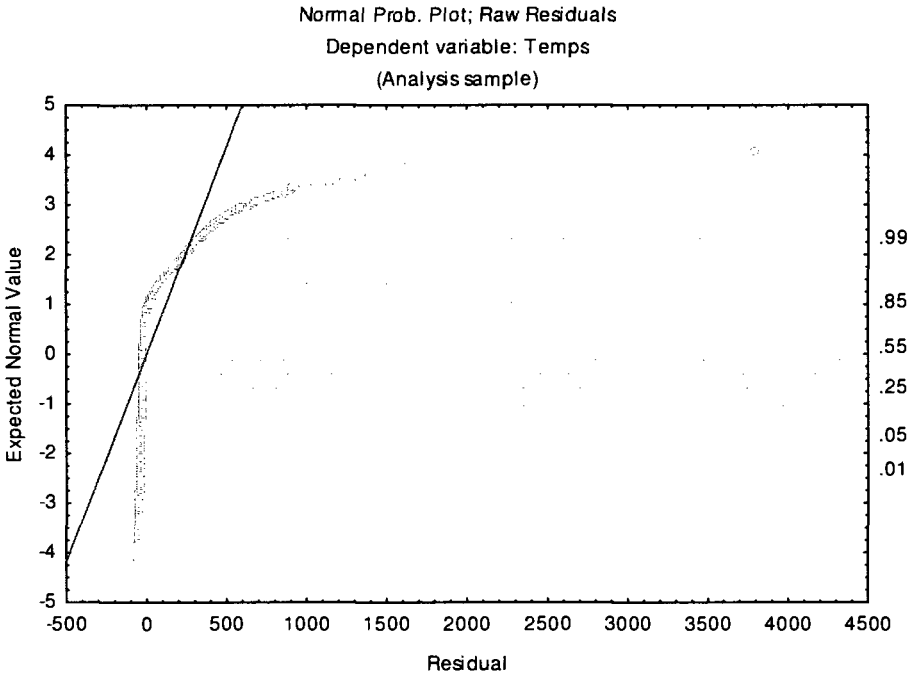


Figure A-6 Résidus sur échelle normale

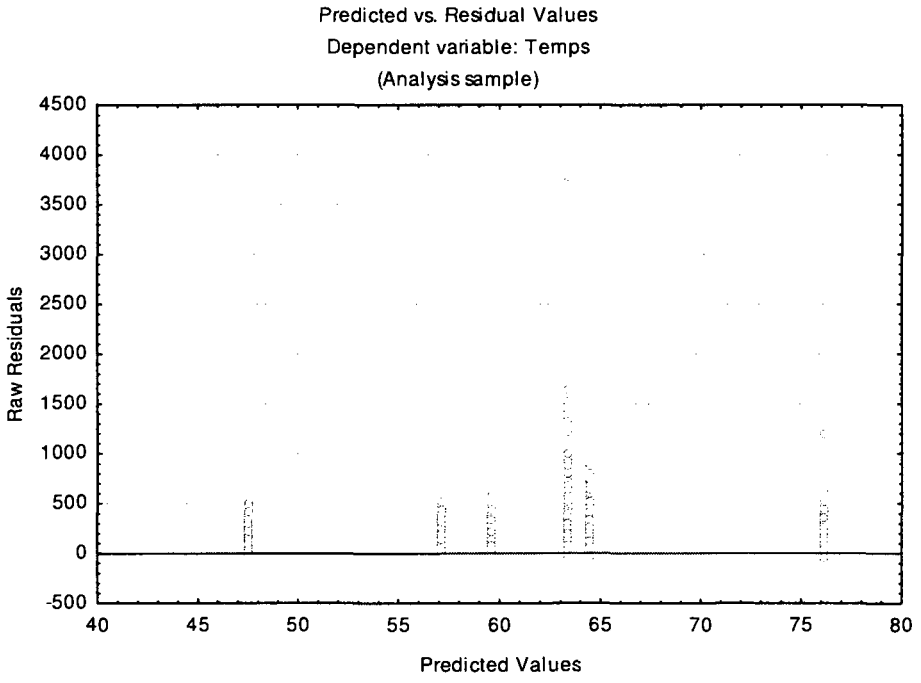


Figure A-7 Valeurs prédites vs résidus

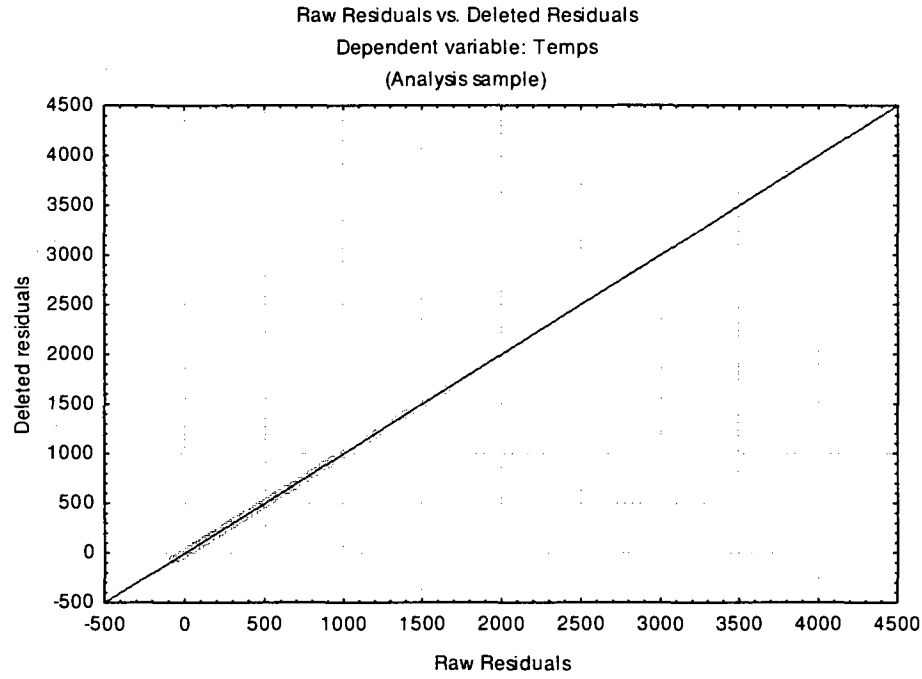


Figure A-8 Résidus vs Résidus exclus

Tableau A-10 Coefficients de corrélation

Tableau de Corrélations des relevés par année										
	Niveau	**Colonne**	**Effect**	**Col. 1**	**Col. 2**	**Col. 3**	**Col. 4**	**Col. 5**	**Col. 6**	**Col. 7**
Constantes		1	Fixe							
Année	1998	2	Fixe		1.00	0.78	0.71	0.76	0.75	-0.05
Année	1999	3	Fixe		0.78	1.00	0.77	0.81	0.80	0.00
Année	2000	4	Fixe		0.71	0.77	1.00	0.76	0.74	-0.01
Année	2001	5	Fixe		0.76	0.81	0.76	1.00	0.79	-0.02
Année	2002	6	Fixe		0.75	0.80	0.74	0.79	1.00	-0.03
Total		7			-0.05	0.00	-0.01	-0.02	-0.03	1.00

Le tableau des corrélations montrent que les observations annuelles sont relativement bien corrélées ce qui pourrait mettre en doute le choix de ce modèle. Par contre, l'observation des résidus semblent indiquer que donner un modèle linéaire du problème ne semble pas convenable.

Dans un deuxième temps nous effectuons une ANOVA selon l'année et la période.

Tableau A-11 ANOVA selon la période et l'année avec effet d'interaction

Tableau ANOVA					
	SS	**Degré de liberté**	**MS**	**F**	**p**
Constante	70085431	1	70085431	11533.57	0
Periode	1325251	1	1325251	218.09	0
Année	1307821	5	261564	43.04	0
Periode*Année	676240	5	135248	22.26	0
Erreur	171859737	28282	6077		

Les résultats sont donc que ces deux facteurs sont significatifs, de plus la courbe représentant les moyennes par année selon la période semble confirmer l'hypothèse d'une amélioration de la qualité de la circulation au fil des années.

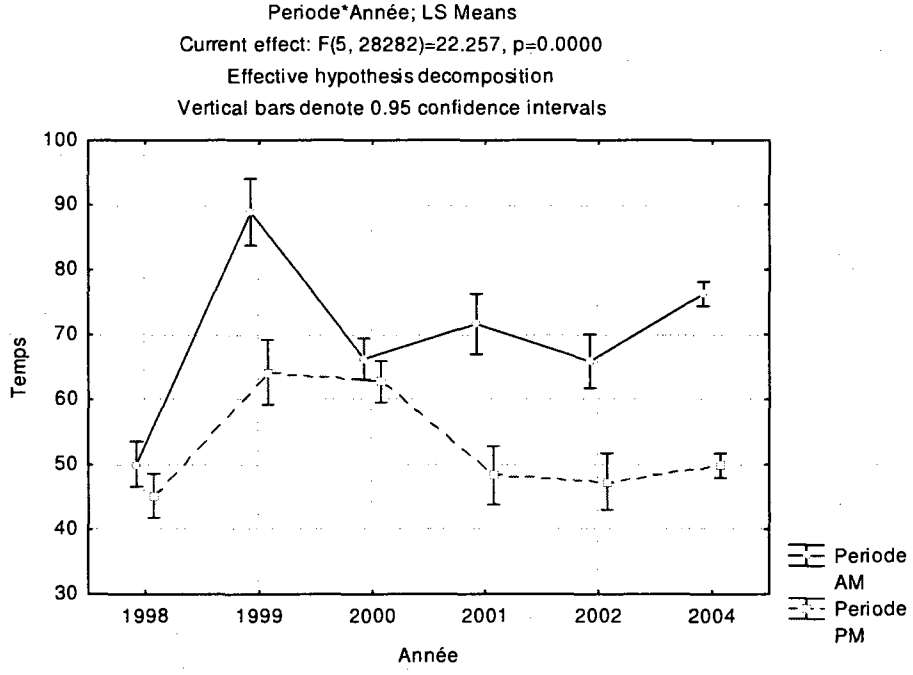


Figure A-9 Moyennes selon année et période

A.3 Analyse de la variance selon le mois

Désormais dans le but d'observer les différences significatives entre les mois, commentons les résultats obtenus après une analyse de la variance des temps de parcours selon le mois d'observations

Tableau A-12 ANOVA selon le mois

Tableau ANOVA					
	SS	**Deg de liberté**	**MS**	**F**	**p**
Constante	9.38E+07	1	9.38E+07	15390.02	0
Mois	3.97E+06	11	3.61E+05	59.17	0
Erreur	1.72E+08	28284	6.09E+03		

L'ANOVA donne directement le mois comme un facteur significatif.

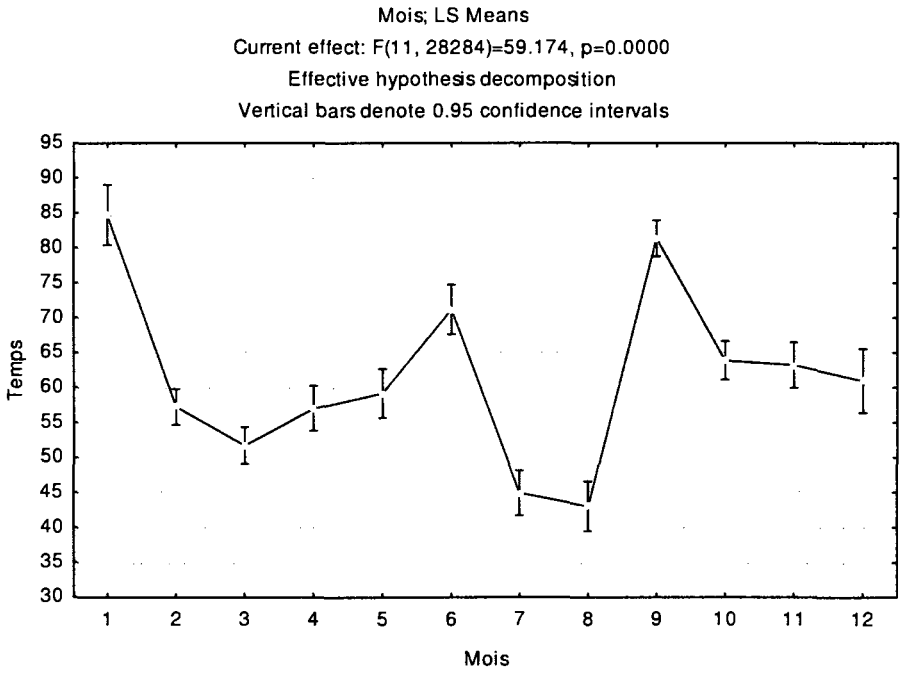


Figure A-10 Moyennes des temps de parcours selon le mois

Tableau A-13 Calcul du coefficient de signification

Coefficient de signification			
	R	**R^2**	**$R^{2\ ajusté}$**
Temps	0.15	0.02	0.02

Le R ajusté demeure faible, par contre nous pouvons voir une courbe de tendance intéressante. Les mois de juillet et aout possèdent de très largement inférieures moyennes de temps de parcours en raison surement des vacances estivales, alors que justement le mois de juin et de septembre, correspondraient aux départs et à la rentrée.

Tableau A-14 Coefficients et intervalle de confiance par mois

Estimation des coefficients								
	Niveau	**Colonne**	**Temps Coeff**	**Temps Err. Std.**	**Temps t**	**Temps p**	**- 95.00%**	**95.00 %**
Constante		1	61.54	0.50	124.06	0.00	60.56	62.51
Mois	1	2	23.19	2.08	11.17	0.00	19.12	27.26
Mois	2	3	-4.26	1.28	-3.32	0.00	-6.78	-1.75
Mois	3	4	-9.82	1.33	-7.40	0.00	-12.41	-7.22
Mois	4	5	-4.51	1.58	-2.86	0.00	-7.61	-1.42
Mois	5	6	-2.36	1.70	-1.38	0.17	-5.69	0.98
Mois	6	7	9.62	1.72	5.59	0.00	6.25	13.00
Mois	7	8	-16.60	1.57	-10.57	0.00	-19.68	-13.52
Mois	8	9	-18.52	1.74	-10.65	0.00	-21.93	-15.11
Mois	9	10	19.81	1.29	15.32	0.00	17.28	22.35
Mois	10	11	2.34	1.39	1.69	0.09	-0.38	5.07
Mois	11	12	1.66	1.59	1.04	0.30	-1.46	4.79

Tableau A-15 Test LSD sur les moyennes par mois

Test LSD sur les moyennes par mois													
	Mois	**{1}**	**{2}**	**{3}**	**{4}**	**{5}**	**{6}**	**{7}**	**{8}**	**{9}**	**{10}**	**{11}**	**{12}**
1	1		0	0	0	0	0.000002	0	0	0.188013	0	0	0
2	2	0.00		0.00	0.90	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.16
3	3	0.00	0.00		0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	4	0.00	0.90	0.01		0.37	0.00	0.00	0.00	0.00	0.00	0.01	0.17
5	5	0.00	0.39	0.00	0.37		0.00	0.00	0.00	0.00	0.04	0.10	0.54
6	6	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00
7	7	0.00	0.00	0.00	0.00	0.00	0.00		0.43	0.00	0.00	0.00	0.00
8	8	0.00	0.00	0.00	0.00	0.00	0.00	0.43		0.00	0.00	0.00	0.00
9	9	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00
10	10	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00		0.76	0.29
11	11	0.00	0.00	0.00	0.01	0.10	0.00	0.00	0.00	0.00	0.76		0.44
12	12	0.00	0.16	0.00	0.17	0.54	0.00	0.00	0.00	0.00	0.29	0.44	

Les valeurs du test de différences entre les moyennes des temps de parcours par mois sont relativement faibles, hormis quelques cas. On peut citer par exemple le mois d'avril et le mois de février (90%), mai et février (38%), avril et mai (37%), décembre et mai (54%), aout et juillet (43%). Peut-on alors en dégager une forme de conclusion ? Il semble que l'on confirme l'idée faite sur les mois de juillet et aout. Par contre, il semblerait inopportun de dégager une conclusion sur les autres valeurs car ceci dépend peut-être des conditions météorologiques rencontrées durant ces périodes.

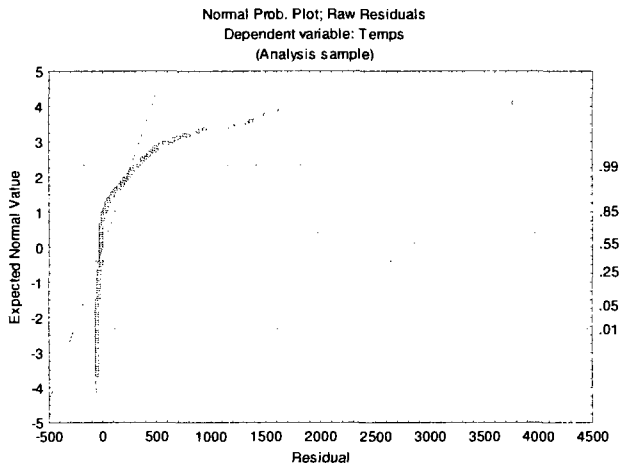


Figure A-11 Résidus sur échelle normale

De nouveau, de par l'observation de la courbe de résidus, un modèle linéaire ne semble pas être pertinent.

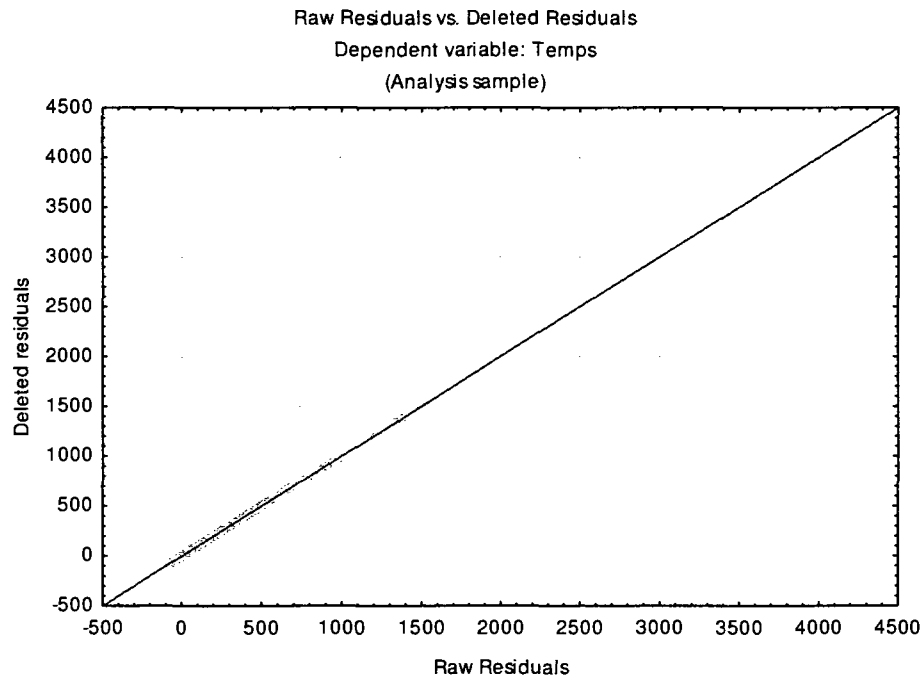


Figure A-12 Résidus vs résidus exclus

Tableau A-16 Coefficients de corrélation

Tableau de correlation des relevés par mois																
	Niveau	Colonne	Effet	Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	Col. 6	Col. 7	Col. 8	Col. 9	Col. 10	Col. 11	Col. 12	Col. 13
Constante		1	Fixe													
Mois	1	2	Fixe		1.00	0.34	0.35	0.40	0.42	0.42	0.40	0.42	0.34	0.36	0.40	0.05
Mois	2	3	Fixe		0.34	1.00	0.21	0.26	0.28	0.29	0.26	0.29	0.20	0.22	0.27	-0.02
Mois	3	4	Fixe		0.35	0.21	1.00	0.27	0.29	0.30	0.27	0.30	0.21	0.23	0.28	-0.04
Mois	4	5	Fixe		0.40	0.26	0.27	1.00	0.34	0.35	0.32	0.35	0.26	0.29	0.32	-0.01
Mois	5	6	Fixe		0.42	0.28	0.29	0.34	1.00	0.37	0.34	0.37	0.29	0.31	0.35	-0.01
Mois	6	7	Fixe		0.42	0.29	0.30	0.35	0.37	1.00	0.35	0.37	0.29	0.31	0.35	0.03
Mois	7	8	Fixe		0.40	0.26	0.27	0.32	0.34	0.35	1.00	0.35	0.26	0.28	0.32	-0.05
Mois	8	9	Fixe		0.42	0.29	0.30	0.35	0.37	0.37	0.35	1.00	0.29	0.31	0.35	-0.05
Mois	9	10	Fixe		0.34	0.20	0.21	0.26	0.29	0.29	0.26	0.29	1.00	0.23	0.27	0.08
Mois	10	11	Fixe		0.36	0.22	0.23	0.29	0.31	0.31	0.28	0.31	0.23	1.00	0.29	0.01
Mois	11	12	Fixe		0.40	0.27	0.28	0.32	0.35	0.35	0.32	0.35	0.27	0.29	1.00	0.01
Total		13			0.05	-0.02	-0.04	-0.01	-0.01	0.03	-0.05	-0.05	0.08	0.01	0.01	1.00

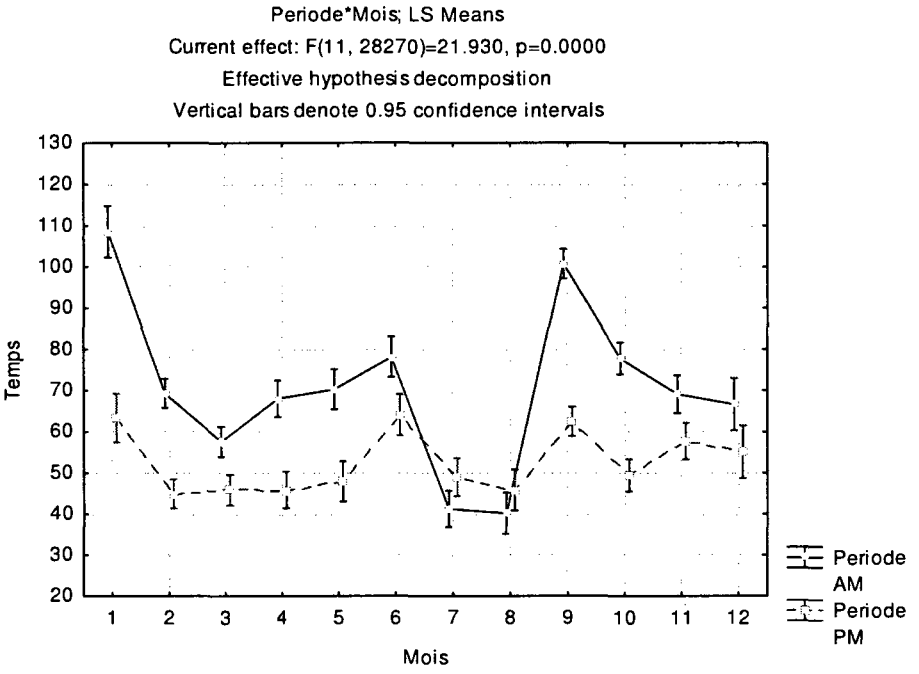


Figure A-13 Moyennes des temps de parcours selon le mois et la période

L'observation des temps de parcours selon le mois et la période expose les mêmes conclusions que celles citées précédemment avec exactement les tendances identiques pour les deux périodes que la tendance générale.

A.4 Analyse de la variance selon les conditions météorologiques

Nous effectuons désormais une analyse de la variance des temps de parcours, avec comme facteur catégorique les conditions météorologiques rencontrées sur le circuit.

Tableau A-17 ANOVA selon les conditions météo

Tableau ANOVA					
	SS	**Degré de liberté**	**MS**	**F**	**p**
Constante	5.74E+06	1	5.74E+06	930.4558	0
Cond	1.79E+06	5	3.58E+05	57.9769	0
Erreur	1.74E+08	28290	6168		

De nouveau, l'analyse donne ce facteur significatif. Le graphe suivant confirme l'idée entendue, d'une augmentation des temps de parcours avec la dégradation des conditions de circulation.

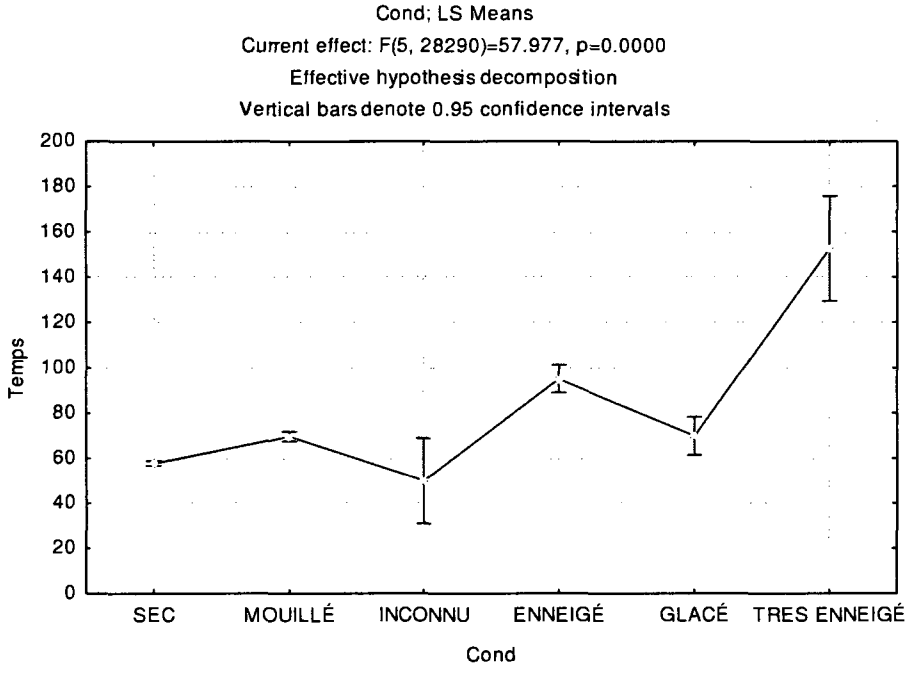


Figure A-14 Moyennes des temps de parcours selon les conditions météo

Tableau A-18 Coefficients de la régression et intervalles de confiance

Estimation des coefficients								
	Niveau	**Colonne**	**Temps Coeff**	**Temps Err. Std.**	**Temps t**	**Temps p**	**-95.00 %**	**95.00 %**
Constante		1	82.52	2.71	30.50	0.00	77.22	87.82
Cond	SEC	2	-24.73	2.74	-9.02	0.00	-30.10	-19.36
Cond	MOUILLÉ	3	-12.80	2.84	-4.51	0.00	-18.36	-7.24
Cond	INCONNU	4	-32.56	8.34	-3.90	0.00	-48.91	-16.20
Cond	ENNEIGÉ	5	12.60	3.72	3.39	0.00	5.31	19.88
Cond	GLACÉ	6	-12.65	4.45	-2.84	0.00	-21.37	-3.93

Tableau A-19 Calcul des coefficients de signification

Coefficient de signification			
	R	**R^2**	**$R^{2\ ajusté}$**
Temps	0.10	0.01	0.01

Tableau A-20 Test LSD

Test LSD sur les moyennes par conditions météo							
	Cond	**{1}**	**{2}**	**{3}**	**{4}**	**{5}**	**{6}**
1	SEC		0.00	0.42	0.00	0.01	0.00
2	MOUILLÉ	0.00		0.04	0.00	0.97	0.00
3	INCONNU	0.42	0.04		0.00	0.06	0.00
4	ENNEIGÉ	0.00	0.00	0.00		0.00	0.00
5	GLACÉ	0.01	0.97	0.06	0.00		0.00
6	TRES ENNEIGÉ	0.00	0.00	0.00	0.00	0.00	

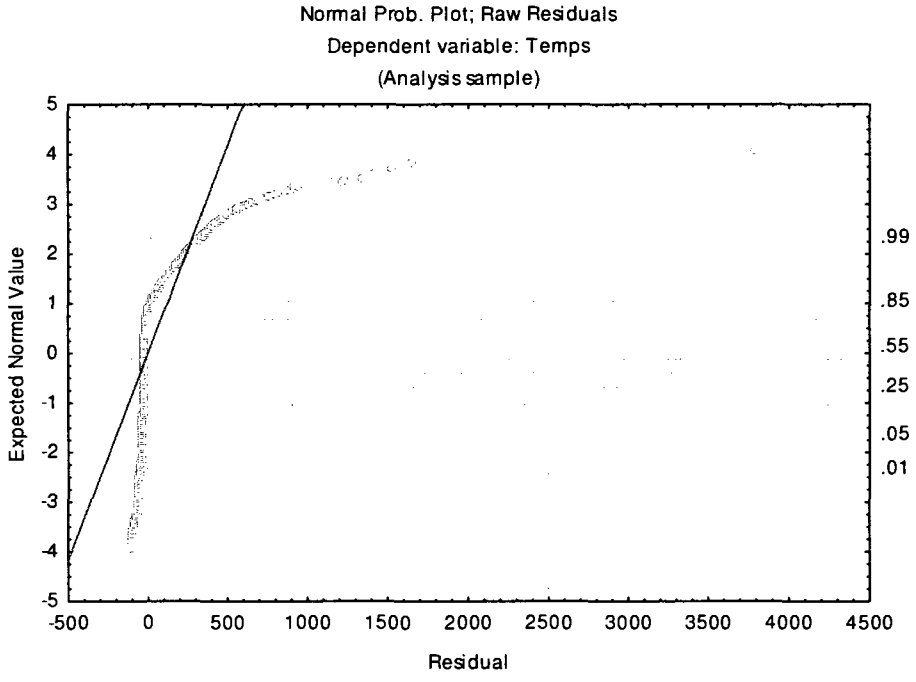


Figure A-15 Résidus sur échelle normale

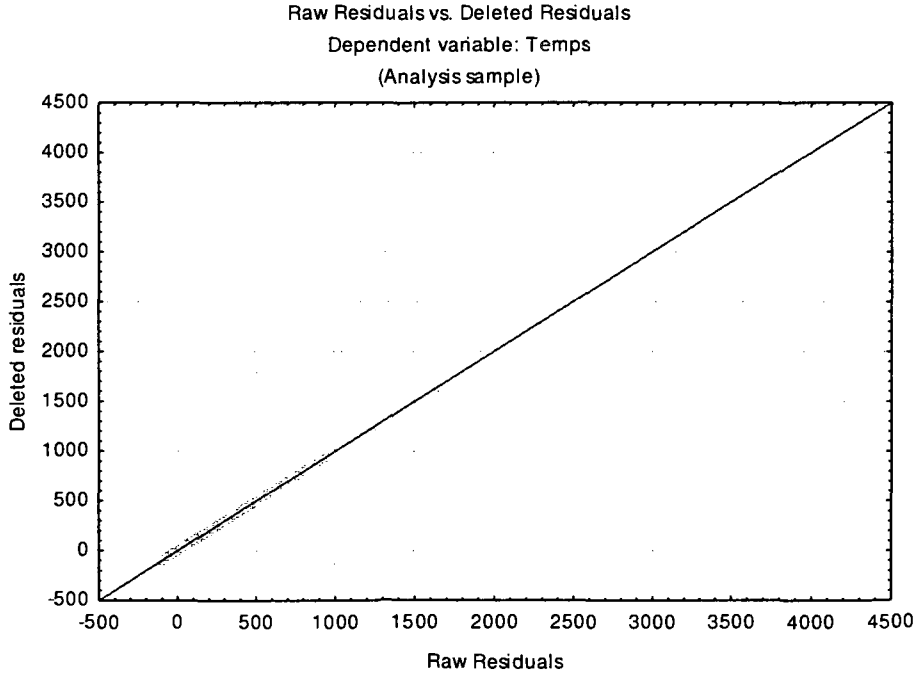


Figure A-16 Résidus vs résidus exclus

Tableau A-21 Coefficients de corrélation

Tableau de corrélations des relevés par conditions météo										
	Level	**Column**	**Effect**	**Col. 1**	**Col. 2**	**Col. 3**	**Col. 4**	**Col. 5**	**Col. 6**	**Col. 7**
Constante		1	Fixed							
Cond	SEC	2	Fixed		1.00	-0.86	0.04	-0.22	-0.12	-0.08
Cond	MOUILLÉ	3	Fixed		-0.86	1.00	0.06	-0.04	-0.01	0.05
Cond	INCONNU	4	Fixed		0.04	0.06	1.00	0.16	0.22	-0.03
Cond	ENNEIGÉ	5	Fixed		-0.22	-0.04	0.16	1.00	0.08	0.05
Cond	GLACÉ	6	Fixed		-0.12	-0.01	0.22	0.08	1.00	0.00
Total		7			-0.08	0.05	-0.03	0.05	0.00	1.00

Le tableau des corrélations donnent une très forte corrélation entre les observations des temps de parcours pour temps sec et pluvieux.

A.5 Temps de parcours avec mois et année

Cette étude est effectuée dans le but de remarquer les tendances et de pouvoir évaluer la tendance observée lors de l'ANOVA des temps de parcours selon l'année.

Tableau A-22 Moyennes selon Mois par Année

Moyennes selon mois de l'année							
	Année	**Mois**	**Temps Mean**	**Temps Std. Err.**	**Temps -95%**	**Temps +95%**	**N**
1	1998	2	49.31	4.20	41.09	57.54	342
2	1998	3	40.50	3.90	32.85	48.14	396
3	1998	4	58.31	4.99	48.53	68.08	242
4	1998	5	40.33	3.90	32.68	47.97	396
5	1998	6	54.14	4.01	46.27	62.00	374
6	1998	7	41.40	3.90	33.76	49.04	396
7	1998	8	39.41	3.90	31.76	47.05	396
8	1998	9	62.28	4.27	53.90	70.65	330
9	1998	11	47.20	2.93	41.46	52.93	704
10	1998	12	50.19	4.01	42.32	58.06	374
11	1999	1	103.18	4.59	94.18	112.17	286
12	1999	9	78.18	4.01	70.32	86.05	374
13	1999	10	66.32	3.90	58.68	73.97	396
14	1999	11	70.13	2.81	64.62	75.65	761
15	2000	1	65.96	4.01	58.10	73.83	374
16	2000	2	61.54	3.90	53.89	69.18	396
17	2000	3	50.03	3.90	42.39	57.67	396
18	2000	4	54.79	3.90	47.15	62.44	396
19	2000	5	79.07	3.90	71.43	86.72	396
20	2000	6	73.40	4.02	65.51	81.28	372
21	2000	7	54.69	3.90	47.05	62.34	396
22	2000	8	50.39	4.14	42.28	58.50	352
23	2000	9	74.14	3.90	66.50	81.79	396
24	2000	10	49.23	3.90	41.58	56.87	396
25	2000	11	86.32	4.01	78.46	94.19	374
26	2000	12	75.76	4.14	67.65	83.86	352
27	2001	7	48.34	3.90	40.69	55.98	396
28	2001	8	40.72	4.01	32.86	48.59	374
29	2001	9	87.79	3.90	80.14	95.43	396
30	2001	10	67.78	4.59	58.79	76.78	286
31	2001	11	56.10	4.01	48.23	63.96	374

Tableau A-23 Moyennes selon Mois par Année (suite et fin)

32	2001	12	58.02	3.90	50.38	65.67	396
33	2002	1	67.36	4.01	59.49	75.22	374
34	2002	2	59.96	4.01	52.10	67.83	374
35	2002	3	45.47	3.90	37.83	53.12	396
36	2002	4	59.50	3.02	53.58	65.42	660
37	2002	6	71.76	3.90	64.12	79.41	396
38	2002	7	37.43	3.91	29.76	45.11	393
39	2004	1	123.22	5.31	112.83	133.62	214
40	2004	2	57.28	1.55	54.24	60.31	2509
41	2004	3	55.21	1.67	51.95	58.48	2171
42	2004	4	55.91	2.49	51.02	60.80	968
43	2004	5	58.81	2.32	54.27	63.35	1122
44	2004	6	78.45	2.88	72.80	84.09	726
45	2004	7	43.70	2.93	37.97	49.44	704
46	2004	8	42.57	2.93	36.83	48.30	704
47	2004	9	85.11	1.71	81.77	88.46	2065
48	2004	10	65.80	1.77	62.34	69.26	1933

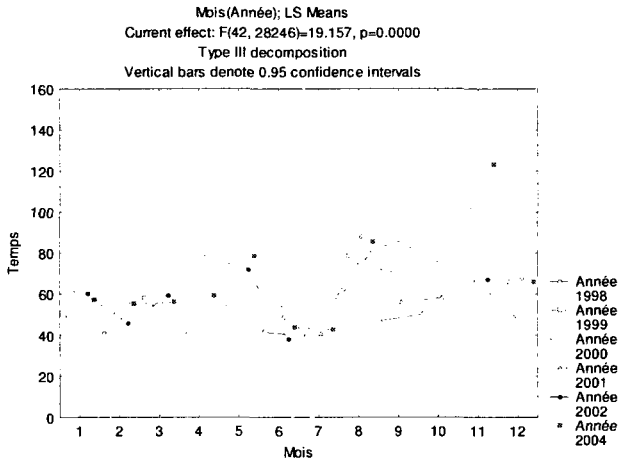


Figure A-17 Moyennes selon mois par année

On constate donc que les tendances sont relativement identiques selon les années. Les différences que nous avons pu alors constater auparavant entre les années semblent

implicitement liées aux mois d'observations selon l'année. En effet, en 1999, seulement les mois d'août, septembre, octobre, novembre et décembre ont été étudiés.

Tableau A-24 ANOVA Nested design

Tableau ANOVA					
	Degré de libertés	**SS**	**MS**	**F-Ratio**	**P-value**
Constante	1	6.96E+07	6.96E+07	11548.77	0
Annee	5	1.46E+06	2.91E+05	48.35	0
Mois(Annee)	42	4.85E+06	1.15E+05	19.15	0
Erreur	28244	1.70E+08	6.02E+03		
Total	28291	1.76E+08			

On constate donc que les deux facteurs sont significatifs. Cherchons alors à différencier les mois selon la période d'observation.

Tableau A-25 Moyenne et intervalle de confiance mois(année*periode)

Moyennes selon mois de l'année et période								
	Periode	**Annee**	**Mois**	**Temps**	**Temps**	**Temps**	**Temps**	**N**
1	PM	1998	2	43.06	6.15	31.01	55.11	154
2	PM	1998	3	44.05	5.42	33.42	54.68	198
3	PM	1998	4	56.31	5.75	45.04	67.58	176
4	PM	1998	5	37.71	5.42	27.09	48.34	198
5	PM	1998	6	58.87	5.75	47.60	70.15	176
6	PM	1998	7	38.45	5.42	27.82	49.07	198
7	PM	1998	8	41.16	5.42	30.53	51.78	198
8	PM	1998	9	55.94	5.42	45.32	66.57	198
9	PM	1998	11	40.03	4.07	32.06	48.00	352
10	PM	1998	12	42.64	5.42	32.01	53.27	198
11	PM	1999	1	67.18	5.75	55.90	78.45	176
12	PM	1999	9	66.85	5.44	56.20	77.51	197
13	PM	1999	10	50.57	5.42	39.95	61.20	198
14	PM	1999	11	68.96	3.99	61.14	76.77	366
15	PM	2000	1	52.90	5.42	42.27	63.53	198
16	PM	2000	2	42.75	5.42	32.12	53.38	198
17	PM	2000	3	58.35	5.42	47.72	68.97	198

Tableau A-26 Moyenne et intervalle de confiance mois(année*periode) (suite)

18	PM	2000	4	40.32	5.42	29.70	50.95	198
19	PM	2000	5	53.63	5.42	43.00	64.25	198
20	PM	2000	6	94.12	5.78	82.78	105.46	174
21	PM	2000	7	67.08	5.42	56.45	77.71	198
22	PM	2000	8	65.14	5.75	53.86	76.41	176
23	PM	2000	9	69.41	5.42	58.78	80.04	198
24	PM	2000	10	50.37	5.44	39.71	61.02	197
25	PM	2000	11	78.01	5.14	67.92	88.09	220
26	PM	2000	12	89.13	6.15	77.08	101.18	154
27	PM	2001	7	59.06	5.42	48.44	69.69	198
28	PM	2001	8	39.55	5.75	28.27	50.82	176
29	PM	2001	9	59.45	5.42	48.82	70.07	198
30	PM	2001	10	44.00	5.75	32.73	55.28	176
31	PM	2001	11	45.28	5.42	34.66	55.91	198
32	PM	2001	12	41.09	5.42	30.47	51.72	198
33	PM	2002	1	63.11	5.75	51.84	74.38	176
34	PM	2002	2	52.29	5.75	41.01	63.56	176
35	PM	2002	3	40.85	5.42	30.22	51.48	198
36	PM	2002	4	41.38	4.51	32.53	50.22	286
37	PM	2002	6	54.95	5.42	44.32	65.58	198
38	PM	2002	7	36.28	5.46	25.57	46.99	195
39	PM	2004	1	76.62	7.38	62.16	91.07	107
40	PM	2004	2	44.38	2.15	40.16	48.60	1255
41	PM	2004	3	44.78	2.32	40.23	49.34	1078
42	PM	2004	4	46.94	3.55	39.99	53.90	462
43	PM	2004	5	49.41	3.25	43.03	55.79	550
44	PM	2004	6	57.30	3.95	49.57	65.03	374
45	PM	2004	7	45.53	4.35	37.01	54.05	308
46	PM	2004	8	42.05	3.95	34.31	49.78	374
47	PM	2004	9	62.09	2.40	57.38	66.79	1010
48	PM	2004	10	49.96	2.54	44.97	54.95	899
49	AM	1998	2	54.43	5.56	43.53	65.34	188
50	AM	1998	3	36.94	5.42	26.31	47.57	198
51	AM	1998	4	63.63	9.39	45.22	82.04	66
52	AM	1998	5	42.94	5.42	32.31	53.57	198
53	AM	1998	6	49.93	5.42	39.30	60.56	198

Tableau A-27 Moyenne et intervalle de confiance mois(année*periode) (suite)

54	AM	1998	7	44.35	5.42	33.73	54.98	198
55	AM	1998	8	37.66	5.42	27.03	48.29	198
56	AM	1998	9	71.77	6.64	58.76	84.79	132
57	AM	1998	11	54.36	4.07	46.39	62.33	352
58	AM	1998	12	58.68	5.75	47.41	69.96	176
59	AM	1999	1	160.78	7.27	146.52	175.04	110
60	AM	1999	9	91.26	5.75	79.99	102.54	176
61	AM	1999	10	82.08	5.42	71.45	92.70	198
62	AM	1999	11	71.22	3.84	63.70	78.75	395
63	AM	2000	1	80.66	5.75	69.39	91.93	176
64	AM	2000	2	80.32	5.42	69.69	90.95	198
65	AM	2000	3	41.71	5.42	31.09	52.34	198
66	AM	2000	4	69.26	5.42	58.64	79.89	198
67	AM	2000	5	104.52	5.42	93.89	115.15	198
68	AM	2000	6	55.18	5.42	44.55	65.81	198
69	AM	2000	7	42.31	5.42	31.68	52.94	198
70	AM	2000	8	35.64	5.75	24.37	46.92	176
71	AM	2000	9	78.88	5.42	68.25	89.51	198
72	AM	2000	10	48.32	5.42	37.69	58.95	198
73	AM	2000	11	98.21	6.15	86.16	110.26	154
74	AM	2000	12	65.35	5.42	54.73	75.98	198
75	AM	2001	7	37.61	5.42	26.99	48.24	198
76	AM	2001	8	41.77	5.42	31.15	52.40	198
77	AM	2001	9	116.13	5.42	105.50	126.76	198
78	AM	2001	10	105.83	7.27	91.57	120.09	110
79	AM	2001	11	68.26	5.75	56.99	79.53	176
80	AM	2001	12	74.96	5.42	64.33	85.58	198
81	AM	2002	1	71.13	5.42	60.50	81.76	198
82	AM	2002	2	66.79	5.42	56.16	77.41	198
83	AM	2002	3	50.10	5.42	39.47	60.72	198
84	AM	2002	4	73.36	3.95	65.63	81.09	374
85	AM	2002	6	88.57	5.42	77.95	99.20	198
86	AM	2002	7	38.57	5.42	27.94	49.19	198
87	AM	2004	1	169.83	7.38	155.37	184.29	107
88	AM	2004	2	70.18	2.15	65.96	74.41	1254
89	AM	2004	3	65.50	2.31	60.98	70.03	1093

Tableau A-28 Moyenne et intervalle de confiance mois(année*periode) (suite et fin)

90	AM	2004	4	64.09	3.39	57.44	70.74	506
91	AM	2004	5	67.85	3.19	61.60	74.11	572
92	AM	2004	6	100.92	4.07	92.95	108.89	352
93	AM	2004	7	42.28	3.83	34.76	49.79	396
94	AM	2004	8	43.15	4.20	34.92	51.39	330
95	AM	2004	9	107.16	2.35	102.56	111.76	1055
96	AM	2004	10	79.58	2.37	74.93	84.23	1034

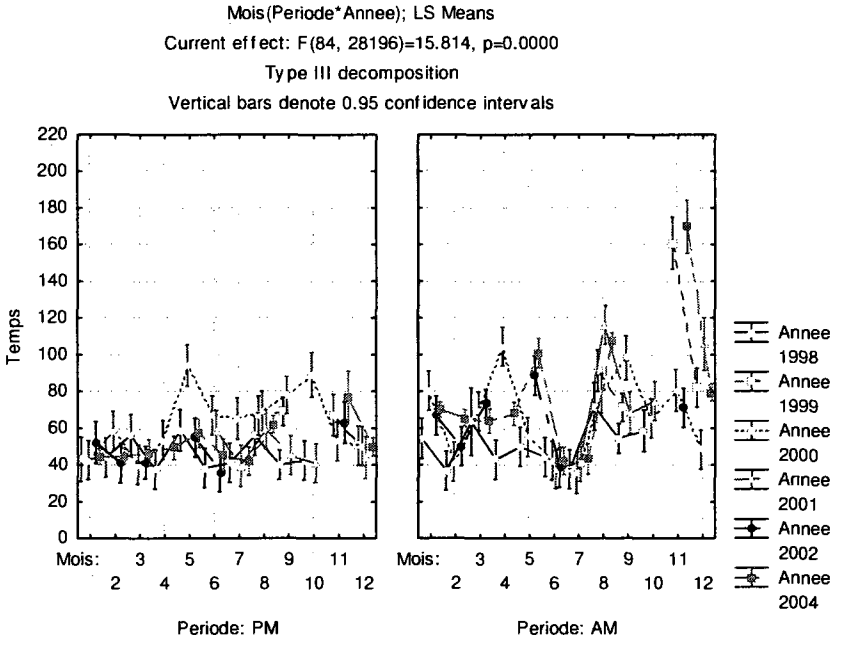


Figure A-18 Moyennes et intervalles de confiance Mois(année*période)

Tableau A-29 ANOVA Mois(Année*Période)

Tableau ANOVA Nested Design					
	Degrés de libertés	**SS**	**MS**	**Rapport F**	**p-value**
Constante	1	7.00E+07	7.00E+07	12019.74	0
Periode	1	1.64E+06	1.64E+06	282.21	0
Annee(Periode)	10	2.26E+06	2.26E+05	38.86	0
Mois(Periode*Annee)	84	7.73E+06	9.21E+04	15.81	0
Erreur	28196	1.64E+08	5821		
Total	28291	1.76E+08			

Tableau A-30 Calcul du coefficient de signification

Coefficient de signification			
	R	**R^2**	**$R^{2\ ajusté}$**
Temps	0.26	0.07	0.07

En clair, nous pouvons souligner un meilleur R, mais aussi des tendances qui ne varient pas trop entre les années, en particulier pour la période PM, ou pour les mois de juillet aout AM. Par contre, il semble que le mois de novembre soit atypique avec une moyenne des temps de parcours plus élevée.

ANNEXE B : TABLEAUX

Tableau B-1 ANOVA des temps de parcours selon Période et tronçons

Tableau ANOVA					
	Degrés de libertés	**SS**	**MS**	**Rapport F**	**p-value**
Constante	1	1.06E+08	1.06E+08	2.19E+04	0.00
Troncon	21	2.03E+07	9.66E+05	1.99E+02	0.00
Periode	1	2.44E+06	2.44E+06	5.02E+02	0.00
Troncon*Periode	21	1.66E+07	7.89E+05	1.63E+02	0.00
Erreur	28250	1.37E+08	4.85E+03		
Total	28293	1.76E+08			

Tableau B-2 ANOVA Nested Design

Tableau ANOVA Nested Design					
	Degrés de libertés	**SS**	**MS**	**Rapport F**	**p-value**
Constante	1	6.96E+07	6.96E+07	1.15E+04	0.00
Annee	5	1.46E+06	2.91E+05	4.84E+01	0.00
Mois(Annee)	42	4.85E+06	1.15E+05	1.92E+01	0.00
Erreur	28244	1.70E+08	6.02E+03		
Total	28291	1.76E+08			

Tableau B-3 ANOVA selon conditions de la chaussée

Tabelau ANOVA					
	SS	**Degrés de liberté**	**MS**	**Rapport F**	**p-value**
Constante	5.74E+06	1	5.74E+06	9.30E+02	0.00
Cond	1.79E+06	5	3.58E+05	5.80E+01	0.00
Error	1.74E+08	2.83E+04	6.17E+03		

Tableau B-4 Groupement des tronçons du circuit 1

Circuit	Circuit1							
	essaim1	éléments1	essaim2	éléments2	essaim3	éléments3	essaim4	éléments4
20-25	0.000	1	0.000	2	0.000	8	0.000	16
25-30	0.000	15	0.050	3	0.002	9	0.020	17
30-35	0.067		0.311	4	0.031	10	0.411	18
35-40	0.190		0.178	5	0.054	11	0.357	19
40-45	0.173		0.063	6	0.078	12	0.094	20
45-50	0.219		0.029	7	0.050	13	0.040	
50-55	0.145		0.014	21	0.059	14	0.017	
55-60	0.067		0.011		0.064	22	0.011	
60-65	0.046		0.012		0.049		0.005	
65-70	0.017		0.014		0.043		0.001	
70-75	0.014		0.008		0.051		0.004	
75-80	0.000		0.010		0.044		0.002	
80-85	0.000		0.010		0.033		0.001	
85-90	0.005		0.006		0.040		0.001	
90-95	0.002		0.001		0.025		0.002	
95-100	0.005		0.006		0.019		0.000	
100-105	0.005		0.008		0.022		0.002	
105-110	0.000		0.008		0.016		0.002	
110-115	0.005		0.003		0.020		0.003	
115-120	0.005		0.008		0.017		0.001	
120-125	0.000		0.011		0.014		0.001	
125-130	0.000		0.007		0.014		0.002	
130-135	0.005		0.010		0.019		0.000	
135-140	0.002		0.005		0.014		0.000	
140-145	0.002		0.008		0.017		0.001	
145-150	0.000		0.007		0.010		0.002	
150-155	0.002		0.002		0.010		0.001	
155-160	0.002		0.008		0.007		0.001	
160-165	0.000		0.003		0.006		0.001	
165-170	0.007		0.008		0.008		0.001	
170-175	0.002		0.005		0.008		0.000	
175-180	0.000		0.003		0.011		0.001	
180-185	0.002		0.006		0.005		0.000	
185-190	0.000		0.004		0.010		0.002	
190-195	0.000		0.005		0.004		0.000	
195-200	0.000		0.001		0.007		0.000	
200-205	0.000		0.003		0.008		0.001	
205-210	0.000		0.004		0.004		0.000	
210-215	0.000		0.005		0.005		0.000	
215-220	0.000		0.003		0.002		0.000	
220-225	0.000		0.004		0.004		0.000	

Tableau B-5 Groupement des tronçons du circuit 1 (suite et fin)

225-230	0.000	0.004	0.006	0.000
230-235	0.000	0.006	0.005	0.000
235-240	0.000	0.002	0.006	0.000
240-245	0.000	0.002	0.002	0.001
245-250	0.000	0.001	0.004	0.001
250-255	0.000	0.005	0.004	0.001
255-260	0.000	0.004	0.005	0.001
260-265	0.000	0.008	0.006	0.000
265-270	0.000	0.003	0.003	0.000
270-275	0.000	0.001	0.005	0.000
275-280	0.000	0.004	0.005	0.000
280-285	0.000	0.007	0.002	0.001
285-290	0.002	0.003	0.002	0.001
290-295	0.000	0.001	0.001	0.000
295-300	0.000	0.001	0.002	0.000
300-305	0.000	0.005	0.004	0.000
305-310	0.002	0.003	0.001	0.000
310-315	0.000	0.003	0.001	0.000
315-320	0.000	0.003	0.001	0.000
320-325	0.002	0.001	0.002	0.001
325-330	0.000	0.004	0.001	0.000
330-335	0.000	0.001	0.003	0.000
335-340	0.000	0.001	0.001	0.000
340-345	0.000	0.003	0.002	0.001
345-350	0.000	0.001	0.001	0.001
350-355	0.000	0.005	0.001	0.001
355-360	0.000	0.003	0.003	0.000
360 -	0.000	0.000	0.000	0.000

Tableau B-6 Variabilité des tronçons du circuit 1

Circuit1	
Periode	AM
Mois	(Plusieurs éléments)

	Valeurs		
Tronçons	Moyenne de SG_TEMPS_A	Écartypep de SG_TEMPS_A	Variabilité
1	56.91	41.12	0.72
2	69.68	99.55	1.43
3	61.75	66.73	1.08
4	67.14	71.86	1.07
5	79.50	76.95	0.97
6	112.38	116.25	1.03
7	188.86	174.65	0.92
8	194.65	131.16	0.67
9	148.87	88.28	0.59
10	112.07	55.51	0.50
11	80.90	33.52	0.41
12	74.90	33.24	0.44
13	68.40	31.63	0.46
14	61.60	24.58	0.40
15	48.58	11.13	0.23
16	37.76	5.71	0.15
17	37.26	8.15	0.22
18	36.94	8.75	0.24
19	43.43	34.83	0.80
20	70.43	98.51	1.40
21	121.59	150.58	1.24
22	113.28	84.27	0.74
Total général	85.8025592	92.3545262	

Tableau B-7 Évolution du nombre d'échantillons nécessaires pour le circuit 1 période AM selon le nombre de relevés par tronçon initial

N	10	50	100	200	500	1000	2000	3000	4000
Tronçon									
1	4	18	28	40	52	58	62	63	64
2	1	6	12	21	38	53	65	71	74
3	1	6	12	21	38	53	65	71	74
4	1	6	12	21	38	53	65	71	74
5	1	6	12	21	38	53	65	71	74
6	1	6	12	21	38	53	65	71	74
7	1	6	12	21	38	53	65	71	74
8	1	5	8	13	20	23	26	27	27
9	1	5	8	13	20	23	26	27	27
10	1	5	8	13	20	23	26	27	27
11	1	5	8	13	20	23	26	27	27
12	1	5	8	13	20	23	26	27	27
13	1	5	8	13	20	23	26	27	27
14	1	5	8	13	20	23	26	27	27
15	4	18	28	40	52	58	62	63	64
16	1	9	16	28	48	63	75	80	83
17	1	9	16	28	48	63	75	80	83
18	1	9	16	28	48	63	75	80	83
19	1	9	16	28	48	63	75	80	83
20	1	9	16	28	48	63	75	80	83
21	1	6	12	21	38	53	65	71	74
22	1	5	8	13	20	23	26	27	27

Tableau B-8 Proportions de relevés du tronçon 2 du circuit 1 période AM selon l'heure de départ

Periode	AM
SG_SEQ	2

	Code (0 = circulation normale, 1 = ventre mou, 2 = valeurs hautes de relevés)			
Créneau horaire	**0**	**1**	**2**	**Total général**
05:00:00	100.00%	0.00%	0.00%	100.00%
05:50:00	100.00%	0.00%	0.00%	100.00%
06:00:00	92.77%	4.82%	2.41%	100.00%
06:10:00	66.67%	33.33%	0.00%	100.00%
06:20:00	74.07%	9.88%	16.05%	100.00%
06:30:00	83.33%	0.00%	16.67%	100.00%
06:40:00	72.50%	12.50%	15.00%	100.00%
06:50:00	88.89%	0.00%	11.11%	100.00%
07:00:00	85.00%	6.25%	8.75%	100.00%
07:10:00	83.33%	0.00%	16.67%	100.00%
07:20:00	80.00%	3.75%	16.25%	100.00%
07:30:00	100.00%	0.00%	0.00%	100.00%
07:40:00	75.00%	2.50%	22.50%	100.00%
07:50:00	87.50%	0.00%	12.50%	100.00%
08:00:00	89.39%	1.52%	9.09%	100.00%
08:10:00	81.25%	6.25%	12.50%	100.00%
08:20:00	90.91%	3.03%	6.06%	100.00%
08:30:00	83.33%	0.00%	16.67%	100.00%
08:40:00	86.49%	5.41%	8.11%	100.00%
08:50:00	100.00%	0.00%	0.00%	100.00%
09:00:00	100.00%	0.00%	0.00%	100.00%
Total général	**83.77%**	**5.26%**	**10.97%**	**100.00%**