

Analyse de données spatio-temporelles du service d'autobus de la ville de Montréal : Structuration, visualisation et traitements

PFE présenté par : **Damien Grosjean**
Société d'accueil : *Ecole Polytechnique de Montréal*
Directeur(trice) de PFE : *Robert Chapleau*
Correcteurs(trices) : *Jacques Ledig*
Gilbert Ferhat



Contexte : Transport Collectif urbain et planification

Depuis les années 70, la planification des transports en commun est basée sur des échantillons de personnes représentatives des usagers du réseau, obtenu grâce à des enquêtes téléphoniques réalisées tous les quatre ans environ. Ces enquêtes, appelées « Origine-Destination », permettent par pondération de planifier les horaires de passage des autobus pour un jour de semaine typique. Couramment, le mercredi d'un jour d'automne est utilisé, lorsque les vacances d'été sont terminées et l'utilisation des transports en commun à son maximum sur l'année.

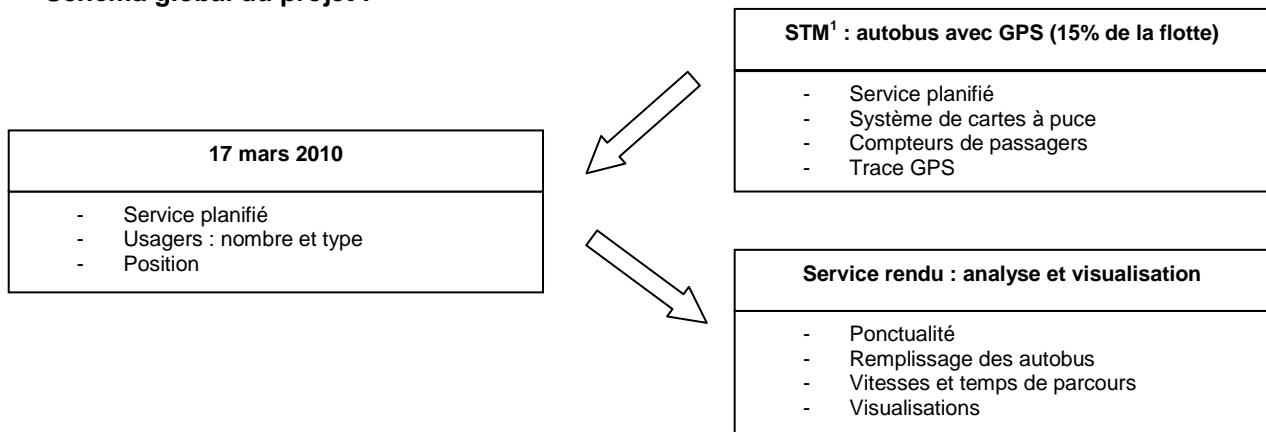
L'amélioration des systèmes informatiques, en capacité de stockage et de traitement, a rendu possible le traitement de données nouvelles et massives. Il est désormais possible de suivre un autobus par GPS, d'obtenir des informations sur les usagers, leur statut démographique, leur comportement, directement par analyse de leur carte à puce et la chronologie des transactions associées.

Le présent projet vise alors à réaliser une fusion complète des données disponibles et utilisables pour peaufiner la planification actuelle qui semble avoir atteint ses limites en termes de fiabilité.

Le but principal ici est de démontrer la faisabilité ainsi que l'intérêt de l'intégration de données nouvelles et très fines dans la planification. Elle est particulièrement cohérente avec un projet de suivi en temps réel de la ponctualité des autobus, annoncé pour 2014 par la Société de Transport de Montréal (iBus).

Pour ce faire, l'étude sera limitée à un premier sous-ensemble, à savoir la journée du 17 mars 2010, date à laquelle aucun événement majeur n'est à déplorer (pas de manifestation, de tempête de neige ou autre). Selon les traitements et expérimentations, d'autres sous ensembles moins importants seront utilisés.

Schéma global du projet :



Données spatio-temporelles à disposition

Pour réaliser ce projet d'envergure, les données disponibles sont discernables en trois grandes catégories. La première concerne le service planifié, avec les horaires de passage des autobus par arrêt, selon un parcours prédéfini. Ensuite, des données relatives aux passagers et usagers du réseau

¹ Société de Transport de Montréal

existent, tant sur leur nombre que sur leur caractéristiques sociodémographiques. Enfin, une partie de la flotte d'autobus est équipée d'un système de positionnement par GPS qui enregistre des positions toutes les quatre secondes ainsi que les activités du véhicule : haltes, ouvertures de portes.

Les données de service planifié vont servir de base pour la comparaison et l'analyse du service. Dans un premier temps, leur compréhension en dit long sur l'organisation du service, les incohérences qui existent malgré le caractère théorique de ces données. On pense ici à deux arrêts distincts et successifs avec un même horaire de passage par exemple. Le format GTFS² utilisé provient des données transmises à Google pour son calculateur de trajet.

Les informations concernant les usagers sont de deux types : dans un premier temps, des compteurs automatiques de passagers au dessus des portes fournissent des informations précieuses sur les lieux de montée et de descente et sur la charge (nombre de personnes) des autobus. A cela viennent s'ajouter les transactions de cartes à puces, qui valident les données de montée ou font ressortir des incohérences. Le problème majeur de ces données est le manque de détails sur les lieux de descente, qui devront être estimés selon des algorithmes utilisant les correspondances par exemple.

Enfin, enjeu majeur du projet, l'analyse spatio-temporelle des traces GPS des autobus va permettre d'évaluer la ponctualité, les lieux de congestion routière, et d'autres concepts.

Difficultés dues aux données

En vue d'une analyse automatique entre les services planifié et rendu, ce projet élabore une méthodologie de fusion de données, à partir de données imparfaites qui devront être maniées avec précaution. Les erreurs sont multiples, et les difficultés principales qui découlent de l'étude sont les suivantes :

Les données planifiées ont le mérite d'avoir une mise en forme et une cohérence quasi-parfaite. Le format GTFS fait preuve d'une rigueur organisationnelle et les traitements y sont de ce fait grandement facilités. Nous verrons que la confrontation avec la réalité va à l'encontre du caractère parfait de la planification, souvent impossible à respecter précisément par les chauffeurs.

Les compteurs automatiques de montées/descentes ne retournent pas une équivalence et on compte le 17 mars 2010 plus d'un pourcent de montants en plus de descendants.

Les cartes à puces n'ayant qu'un intérêt financier à l'origine, la mise en forme de ces données nécessite un traitement rigoureux, afin de les simplifier. Notons d'ailleurs que leur utilisation n'a été rendue possible qu'après suppression de toutes les informations relatives au détenteur de la carte, évitant ainsi tout problème d'anonymat, mais limitant du même fait l'intérêt des informations sur le type de passager.

Enfin, les traces GPS sont de bonne qualité sur la majeure partie du territoire, due à l'architecture basse et étalée de la ville, ou les immeubles dépassent rarement trois étages. Cependant, le centre-ville de Montréal est principalement constitué de gratte-ciels créant d'importants canyons urbains et source d'un problème majeur pour l'analyse de la trace à cet endroit

Signature GPS sur l'île de Montréal

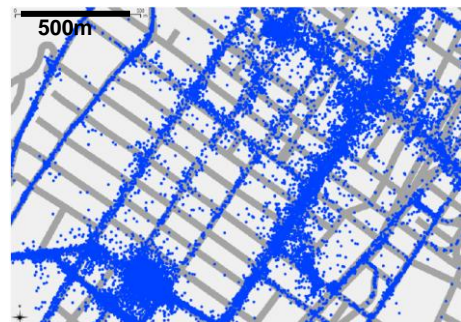
La réception GPS est un point d'analyse important sur l'île de Montréal. Deux zones aux caractéristiques opposées sont à signaler. Aucun problème majeur de réception n'existe sur l'île, excepté au centre-ville, comme énoncé précédemment.

La simple visualisation des enregistrements au centre-ville fait état d'une situation anormale, typique des canyons urbains.

La figure ci-contre présente cette dispersion des enregistrements, proche de l'université Guy-Concordia et du boulevard Maisonneuve.

Cette zone géographique est principalement constituée de grands immeubles de type gratte-ciel, et met en évidence la

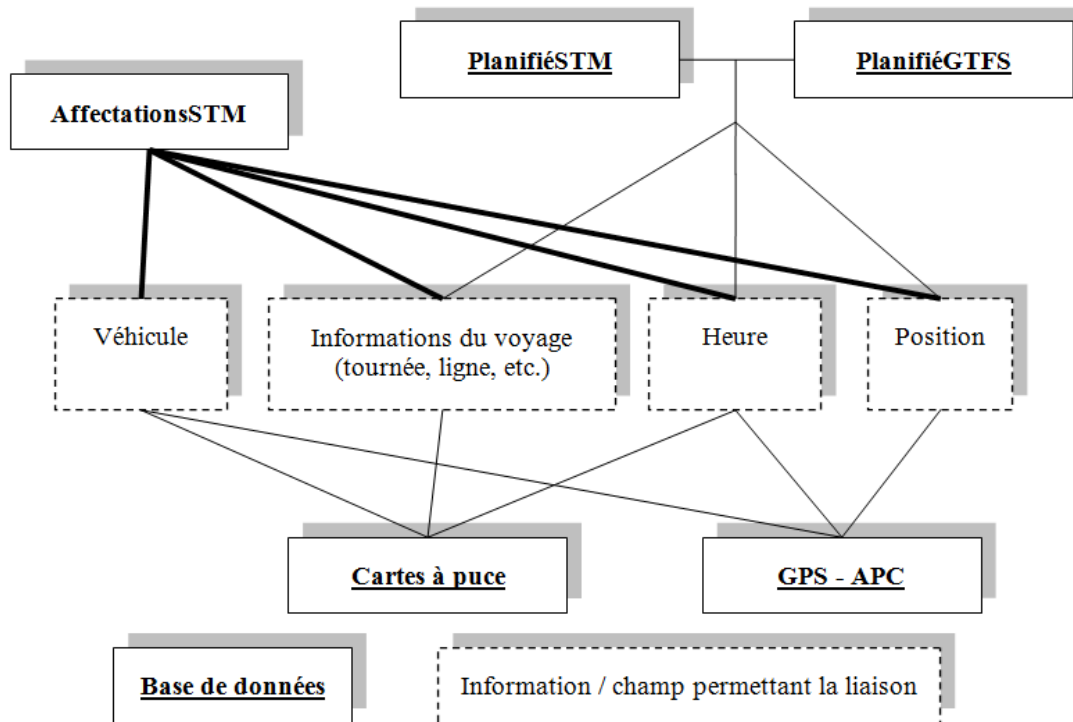
nécessité de traitements ou de systèmes de positionnement annexe dans certaines parties de la ville. Une ébauche de traitement est évoquée dans le rapport, ainsi que différentes cartes thématiques de visualisation des zones peu propices à l'usage du GPS par leur mauvaise réception du signal.



² « Google Transit Feed Specification »

Fusion des données et analyse du service rendu

L'ensemble de ces données est susceptible d'être fusionné spatialement et/ou temporellement. Ceci induit une synchronisation entre les horloges des différents systèmes (boîtes de perception de carte à puce et GPS). De plus, les données prises séparément font ressortir des erreurs mais la fusion entre elles met en évidence d'autres problèmes majeurs. Le schéma suivant résume les bases de données, ainsi que les liens entre elles, et sert de base à tous les traitements futurs :

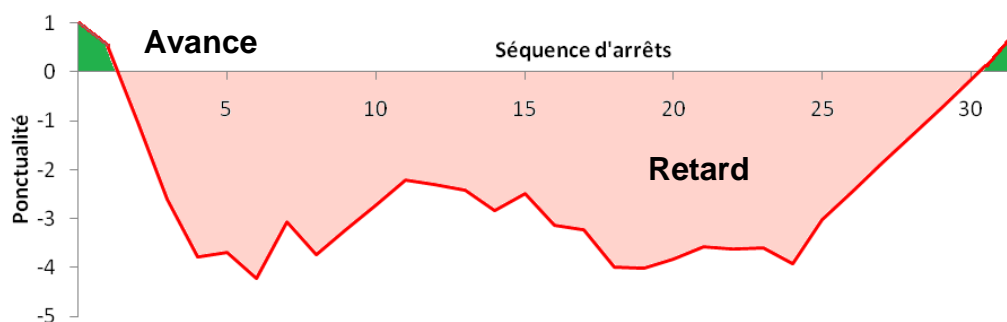


La méthodologie retenue suit ces grandes étapes:

- Détection et affectation des terminus, de la ligne empruntée, de l'heure de départ à une trace GPS
- Analyse de la tournée et de la ligne dominante
- Détection des arrêts au cours du voyage
- Comparaison des heures de passage et des heures planifiées
- Rattachement des données de carte à puce
- Analyses de service rendu

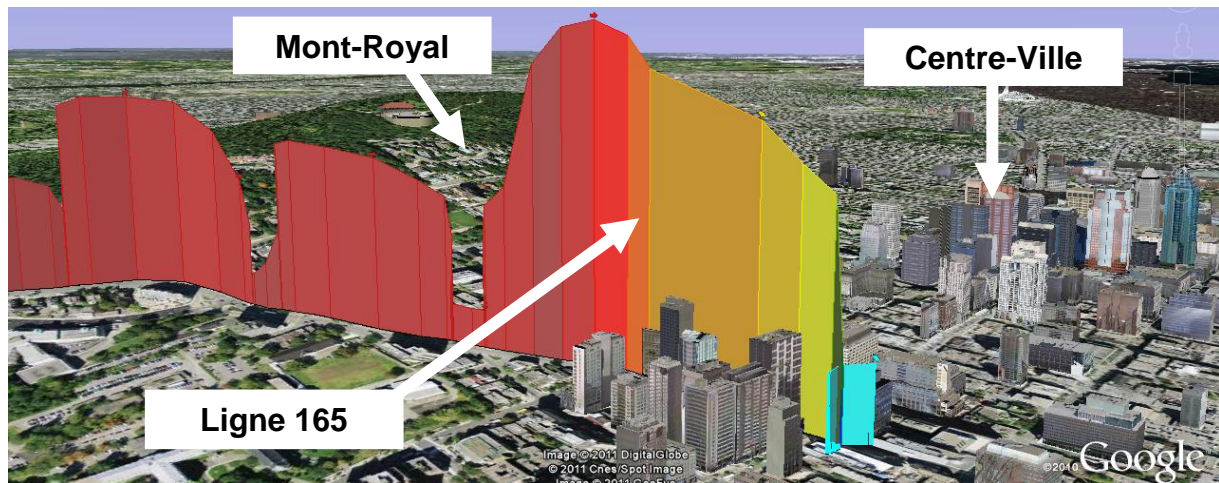
Outre les traitements d'extraction réalisés sur chaque base, la jointure spatiale entre planifié et GPS est rendue possible par l'usage d'un système d'information géographique, choisi libre de droit.

Cette fusion donne accès à une panoplie d'analyse et études de service rendu. La figure suivante illustre la ponctualité d'un autobus sur un sous-ensemble qui est détaillé dans le rapport.



Visualisation des données et analyses

Egalement point clef du projet, les données et analyses doivent être visualisées, permettant ainsi la compréhension intuitive d'enregistrements abondants et peu parlant d'eux-mêmes. L'usage de logiciels libres de droit limite parfois les possibilités, et le recours à des moyens détournés permet à chacun de créer des visualisations intéressantes. QuantumGIS permet la création aisée de visualisations et traitements en deux dimensions, le recours à Google Earth ouvre la voie de la troisième dimension. La figure suivante présente la représentation d'une analyse de service, pour une analyse visuelle aisée de la charge et de la vitesse en fonction des rues empruntées par l'autobus en question.



(Élévation : Vitesse du véhicule – Coloration : charge du véhicule)

Problématiques en suspend et avenir du projet

Ce projet mène à l'étude d'une masse très importante de données, afin de valider l'usage ou la discrétisation de certaines bases de données dans un premier temps, d'établir un maillage permettant de lier et former un bloc de données hétérogènes ensuite, puis de dresser un bilan des limites d'utilisation de chacune des bases.

Outre l'aspect structuration, il s'agit de l'étudier la méthodologie de fusion, et d'analyse par l'usage continu de visualisations afin d'en faciliter la compréhension.

En revanche, cette étude préliminaire pose les bases d'un projet global, mais des problématiques d'envergure restent en suspend. Le projet permet d'en formuler quelques unes mais aucun doute ne plane sur l'émergence de nouvelles au fur et à mesure de l'avancement.

Parmi ces problèmes majeurs, le traitement des cartes à puces reste un projet de taille, qui a été évoqué plus tôt dans ce résumé.

Le traitement des traces GPS fait également l'objet d'une réflexion à mener : les véhicules enregistrent une trace sans aucune information sur la ligne parcourue, tournée, etc. Cela révèle d'ailleurs un problème d'homogénéité entre les systèmes installés qui proviennent de fournisseurs différents, puisque les boîtes de perception installées dans les mêmes bus contiennent pour leur part cette information.

Enfin, l'automatisation et la programmation pourront venir compléter et finaliser ce projet, afin de rendre rapide et intuitive la visualisation du service sur un sous-ensemble précis choisi. En plus du suivi en temps réel des autobus pour l'information voyageurs envisagé par la Société de Transport de Montréal à l'horizon 2014, ces études permettront une compréhension minutieuse et des adaptations du service planifié pour le mener de plus en plus vers une adéquation avec les réalités du réseau.