

TELEBILLETTIQUE : LA VALIDATION AU SERVICE DE LA DECISION



01/03/2010

Présentation du système d'information décisionnel du STIF s'appuyant sur les données de validation

Description du système d'information décisionnel s'appuyant sur les données de validation : données, exemples d'exploitation...

Télébillettique : la validation au service de la décision

PREAMBULE ET TABLE DES MATIERES

Ce document est destiné à l'ensemble des acteurs des transports ayant une démarche d'analyse du trafic sur la base des données de validation sur leur réseau. Il décrit la manière dont le STIF a mis en place son système d'information décisionnel et donne des exemples des premières exploitations réalisées.

Table des matières

I. INTRODUCTION	3
II. CONTEXTE DU PROJET.....	4
Les transports en Île de France	4
La région Île de France	4
Le réseau de transport	5
Le STIF : Syndicat des Transports d'Ile de France.....	6
III. LE SYSTEME TELEBILLETIQUE	7
Le système.....	7
La télébillettique.....	7
Objectifs du projet	7
Chronologie	8
Les flux des données.....	9
Quelles informations ?.....	9
Remontées des données	9
Anonymisation	10
Architecture logique	10
L'organisation de l'information	12
Un système décisionnel	12
Modèle	12
IV. LES DONNEES.....	14
Les données de mobilité.....	14
Enquêtes et études de mobilité	14
Thèmes du SIDV vs Enquêtes.....	15
Les données de validation	15
Organisation des données.....	16
Les données et leurs limites.....	16
V. UTILISATIONS DES DONNEES DE VALIDATION	19
Analyse des validations, estimer des fréquentations.....	19
Présentation	19
Mise en application générale	19
Questions traitées	21
Analyse des cartes, estimer le nombre de voyageurs	24
Présentation	24
Mise en application générale	24
Questions traitées	24
Analyse de la mobilité, quantifier et qualifier les déplacements.....	26
Présentation	26
Mise en application générale	27
Questions traitées	28
VI. CONCLUSION.....	33

I. INTRODUCTION

Le STIF - Syndicat des Transports de l'Ile de France - a depuis plusieurs années décidé de généraliser le passage des titres magnétiques sur support télébillettique, le passe Navigo. Il a dans ce cadre conçu un système d'information décisionnel lui permettant une meilleure connaissance de la mobilité des usagers.

Ce système permet la remontée des données depuis les opérateurs jusqu'au STIF, puis leur uniformisation et leur archivage. Dès lors, à chaque fois qu'un usager valide son titre de transport sur le réseau francilien, une information anonyme est remontée à un système central permettant la collecte et le calcul de statistiques sur la mobilité.

La mission d'autorité organisatrice des transports en Ile de France intègre notamment l'adaptation du système de transport aux besoins des usagers et l'optimisation des crédits publics consacrés à son fonctionnement. Grâce à ce système, le STIF a pour vocation de consolider sa connaissance de l'usage du réseau et de la mobilité des porteurs des différents titres de transports.

L'utilisation d'un tel outil décisionnel donnant une vision fine de la fréquentation des gares ou des lignes de bus et une reconstitution des déplacements des cartes peut dans ce cadre avoir des applications multiples.

Néanmoins ce système est encore jeune et les applications qui en sont faites montrent des limites, que ce soit par les défauts de qualité inhérents à la remontée de données, ou également par l'expertise qu'il est nécessaire de développer sur les différents redressements et les analyses à mener pour passer d'une donnée simple - la validation - à des modèles plus complexes de trafic.

Ce document présente l'ensemble du système et quelques-uns des axes d'analyse qui y sont mis en œuvre. Pour illustration, quelques exemples d'exploitation sont donnés, mettant l'accent sur des résultats concrets mais aussi sur certains biais du système et les méthodes mises en place pour les contourner.

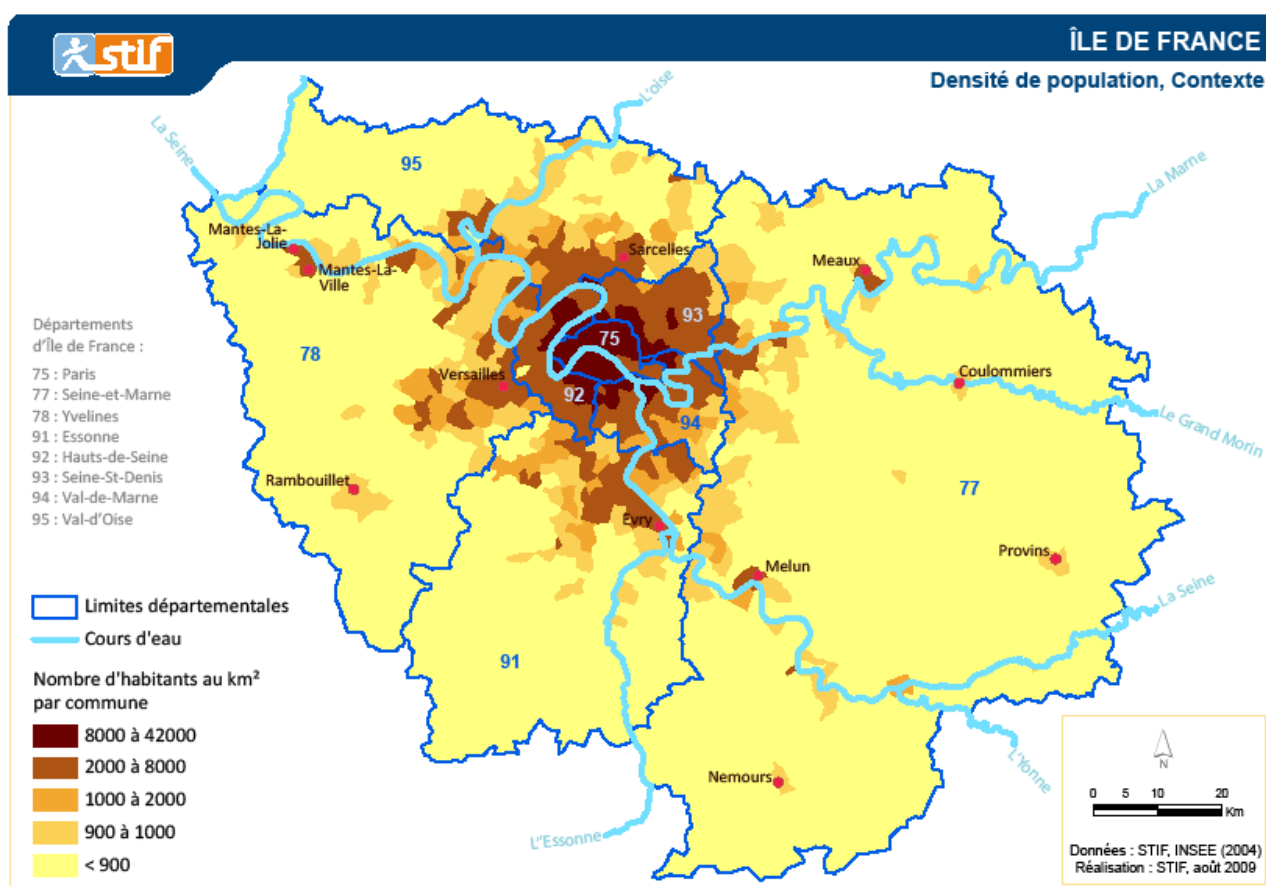
II. CONTEXTE DU PROJET

Les transports en Île de France

Le système d'information décisionnel du STIF et son utilisation sont intimement liés aux spécificités des transports en Ile de France. L'organisation multi-opérateurs et multi-modale ainsi que la richesse des déplacements rendent particulièrement difficile l'uniformisation des concepts et des données.

La région Île de France

Le Syndicat des Transports d'Île de France a pour mission d'organiser des services de transport intégrés et performants pour toute la région Île de France (IDF) et tous les franciliens. Il coordonne chaque jour le transport d'une population de **11,5 millions d'habitants**. Il doit répondre au besoin de mobilité de cette population sur un territoire qui s'étend sur **12 000 km²**, 8 départements et 1 280 communes.







La population mais aussi l'offre d'emploi se concentre essentiellement à Paris et en Petite Couronne (Hauts-de-Seine, Seine-St-Denis, Val de Marne). Au point de vue démographique, la région IDF offre donc un paysage très asymétrique entre Paris et la Petite Couronne d'un côté, et la Grande Couronne de l'autre. Cette asymétrie se retrouve naturellement dans la structure de son réseau de transport.

Le réseau de transport

MODES DE TRANSPORT

Pour répondre à la demande de mobilité très importante, le STIF supervise un réseau de transport en commun (TC) extrêmement dense, proposant un important choix de modalité, souterrain ou en surface.

Transilien + RER	Méto	Tram et Tram-train	Bus
			
1 401 km 444 gares	211 km 381 stations	39 km 67 arrêts	22 670 km (étendue) 28 950 arrêts

USAGE

Alors que les Français mettent en moyenne 32 minutes pour se rendre au travail le matin, les habitants d'IDF y consacrent plus de 40 min (source : INSEE 2006). D'après l'Enquête Globale Transport 2001-2002, sur les 37 millions de déplacements par jour (soit 65% des voyageurs français qui circulent) :

- **8 millions** se font en transport en commun
- 12,2 millions se font à pied
- 16 millions en automobile
- 800 000 en deux-roues

La consommation de transports en IDF (source 2004) est la suivante :

Transilien + RER	Méto	Tramway	Bus
614 millions de voyages/an	1350 millions de voyages/an	44 millions de voyages/an	RATP Paris : 350 millions de voyages / an RATP banlieue : 590 millions Optile : 250 millions

GESTION

Si le STIF assure la coordination, le fonctionnement opérationnel du réseau est assuré quant à lui par la **SNCF**, la **RATP** et les 90 entreprises privées de bus regroupées au sein de l'association **Optile**.

Les opérateurs mettent en œuvre une offre de transport et une qualité de service prescrites par le STIF, au travers de contrats d'une durée de quatre ans. En contrepartie, ils reçoivent une rémunération prédéfinie, avec des mécanismes d'intéressements financiers positifs ou négatifs. Les modalités des **remontées des données** de validation ainsi que la **politique d'incitation à valider** sont comprises dans ces contrats.

Le STIF : Syndicat des Transports d'Ile de France

Le Syndicat des Transports d'Ile-de-France (STIF) est l'**autorité organisatrice des transports publics** (AOT) en Île-de-France. C'est un établissement public associant la Région IDF et les huit départements franciliens, et dont la vocation est d'organiser et développer des transports publics adaptés aux besoins des franciliens (intégration, système multimodal, qualité, efficacité). Ainsi, le STIF :

- définit les conditions générales d'exploitation, crée les **titres** de transport et fixe les **tarifs**
- définit l'**offre** de transport et le niveau de **qualité des services** dans le cadre de contrats signés avec les entreprises de transports
- coordonne la **modernisation du système** de transports publics franciliens, en définissant les principes des grands investissements publics
- assure le contrôle de la **maîtrise d'ouvrage** des projets et cofinance la modernisation ou la création des **équipements** nécessaires à l'amélioration de la qualité de service

III. LE SYSTEME TELEBILLETTIQUE

Le **Système d'Information Décisionnel basé sur les Validations** (SIDV) est un outil mis en place par le STIF pour suivre l'utilisation du réseau de transports francilien et réaliser des études de toutes sortes (trafic, utilisation des titres, fréquentation des gares...).

Le système

La télébillettique

Une carte télébillettique est un **passé sans contact** équipé d'une carte à microprocesseur. Cette carte à puce contient les informations relatives à l'abonnement de l'usager et une antenne permettant la transmission radio des données aux **valideurs**. Les valideurs sont situés à l'entrée des bus, des stations de métro, de RER et à certaines sorties RER. Cette technologie permet une plus grande rapidité et fluidité des passages, une simple présentation de la carte sans contact devant la cible violine permettant une lecture rapide du titre.



En Île de France, le passé télébillettique se nomme « **Navigo** » ; il permet de charger un forfait tarifaire de type hebdomadaire, mensuel ou annuel, Solidarité Transport et Gratuité Transport (forfaits pour les allocataires du RSA) ou encore Imagine R (forfaits pour les scolaires et étudiants d'IDF).

Objectifs du projet

Avec la généralisation de la télébillettique, le STIF a mis en place un **système d'information décisionnel** (SID) s'appuyant sur les données issues de ces validations. Celui-ci permet d'observer la saisonnalité des comportements ainsi que la fluctuation pour différentes périodes des choix de mobilité des usagers. L'enjeu est une meilleure connaissance des systèmes de transports en commun franciliens, de leur utilisation, de l'utilisation des titres, etc.

En résumé, les données de validation offrent l'opportunité de :

- **Compléter ou se substituer aux outils de collecte existants** à un coût moindre aux enquêtes/comptages
- Apporter des **mesures précises et « indépendantes »** aux différents stades des prises de décisions de l'Autorité Organisatrice
- Disposer d'une **information homogène** :
 - Grâce à l'intégration tarifaire (possibilité d'utiliser sa carte Navigo sur l'ensemble du réseau)
 - Par l'acquisition de données issues de l'ensemble des opérateurs de transport

Si l'information est globalement homogène, elle est cependant **incomplète**. En effet, d'autres supports billettiques persistent toujours. Le titre le plus connu est le ticket ++ (magnétique), pratique pour les voyageurs occasionnels, puisque la carte à puce est destinée aux forfaits. Enfin, objectif affiché lors du lancement du projet, la **validation systématique en entrée** (VSE) n'est pas encore en place. Il suffit d'emprunter le mode bus pour constater que, malgré les rappels des transporteurs et l'obligation de valider, beaucoup d'usagers ne valident toujours pas, le valideur ne présentant pas un obstacle.

Chronologie

En IDF, l'intégration tarifaire très rapide du réseau avec la **carte Orange** (permettant les déplacements multimodaux, partout dans la région) fut un déclencheur du projet télibillettique Navigo.

C'est en 1999 que le STP (Syndicat des Transports Parisiens, actuel STIF) décida la généralisation de la télibillettique. La diffusion du passe Navigo, depuis **fin 2001**, a concerné dans un premier temps les abonnés à la carte Intégrale, rejoints en 2003 par les détenteurs de la carte imagine R et s'est étendu aux deux premières zones de la carte Orange en 2006. Parallèlement, ces décisions ont nécessité l'installation d'équipements dans les bus (plus de 4000 bus équipés de valideurs).



Le nombre d'abonnés avec Passe Navigo ayant atteint plus de 600 000 abonnés début 2005 (contre 100 000 abonnés en 1999), a commencé une réflexion autour des données : leur remontée, le système pour les stocker, leur gestion, leur usage, etc. Le **portail d'accès aux données** du STIF fut opérationnel fin 2008 et les données sont disponibles pour la totalité des opérateurs depuis janvier 2009.

Aujourd'hui, tous les opérateurs doivent remonter leurs données de validations au STIF. Toutefois, il subsiste des soucis dans la remontée des données de validations. Par exemple, certaines générations de valideurs déployés par la SNCF ne remontent aujourd'hui pas de données de validations, et en particulier pour de grandes gares parisiennes. Il existe néanmoins d'autres **irrégularités** ponctuelles, plus difficiles à détecter puisque le STIF ne peut expertiser finement le volume de validations attendu à chaque gare ou à chaque arrêt. Ainsi une forte variation dans les volumes peut être due à un **problème technique** mais peut aussi être liée à un **facteur contextuel** (jour férié, manifestation, grève,...).

Aujourd'hui le système télibillettique remonte quotidiennement plus de 8 millions de validations qui sont transformées, nettoyées, stockées et analysées.

Les flux des données

Quelles informations ?

Les informations du SIDV sont issues des données contenues dans le passe Navigo remontées via les valideurs des différents transporteurs. Les données enregistrées sont composées des **données contrats** (sur la carte) et des **données de localisation et d'horodatage** (sur le valideur) :

Données issues de la carte Navigo			Données issues du valideur	
Contrat	Type de contrat		Date et heure des validations	Date de validation
N° de carte anonymisé	N° de série de la carte anonymisé	Nature de la validation	Nature de la validation	
Profil	Profil du porteur si contrat social	Lieu de validation	Lieu de validation	
Validité du contrat	Date de début du contrat	Référence de la ligne	Identité de l'exploitant	
	Date de fin du contrat		Identificateur de l'équipement	
	Zones de validité		Référence de la ligne	
			Référence de la mission	
			Identificateur du véhicule	

Des différences existent sur les informations contenues dans une validation pour les modes ferrés et bus/tram.

Dans le cas du ferré, les validations sont rattachées à une gare et pas une ligne. En effet, la validation se fait à l'entrée de la station, indépendamment de la ligne précise empruntée. De plus, une fois entré dans le réseau métro ou RER, l'utilisateur n'a en général pas besoin de revalider lorsqu'il change de ligne (sauf en correspondance fer lourd – métro).

Pour les bus, le valideur est mobile puisqu'il se situe dans le bus lui-même. La validation peut donc être facilement associée à une ligne, mais la liaison à un arrêt est plus complexe. Les coordonnées de l'arrêt sont en fait repérées par GPS, méthode de radiolocalisation ou, lorsque le service est dégradé, manuellement par le conducteur. Il existe néanmoins des problèmes lorsque la validation est réalisée entre deux arrêts.

Remontées des données

Les **flux de données brutes** de chaque réseau sont remontés tous les jours par les différents opérateurs au STIF. Les données télébilletiques sont composées de deux fichiers pour chaque opérateur : un fichier de **validations** et un fichier **référentiel**. Les fichiers référentiels permettent l'harmonisation des données. Ils fournissent une correspondance entre leur référentiel et le référentiel STIF pour les entreprises, réseaux, lignes et sous-lignes, arrêts ou encore la codification des types de contrats.

Après **harmonisation**, les données télébilletiques sont stockées dans une **zone de stockage temporaire (zone tampon)** avant d'être poussées dans l'**entrepôt de données** (ou Data Warehouse) au bout de 30 jours. Ces données permettent de construire des **vues métier** que pourra ensuite interroger l'utilisateur.

Anonymisation

Si le STIF récupère les données des opérateurs, ceux-ci sont également d'exploiter celles-ci. Les bases de données des transporteurs servent ainsi à lutter contre les tentatives de contrefaçon.

L'ensemble du système télébillettique a été défini avec la **CNIL** (Commission nationale de l'informatique et des libertés) qui a rendu un avis positif sur le système (avis du 22 décembre 2004). Avec cette technologie, on touche en effet ici à un droit fondamental qui est celui de se déplacer librement ; le risque étant la capacité à « suivre » un individu dans les transports.

L'anonymat est total et garanti de deux façons :

Tout d'abord par l'organisation sécurisée mise en place. Les données de validation sont toujours anonymes, elles ne sont jamais associées au nom du voyageur. L'organisation est conçue sur la séparation du circuit commercial (fichier clients des transporteurs) de celui du traitement de données de validations (utiles aux transporteurs pour lutter contre les tentatives de contrefaçon et au STIF pour réaliser des études statistiques afin d'adapter l'offre de transport aux besoins des voyageurs). Les accès aux données sont strictement réglementés et compartimentés.

Ensuite par les choix technologiques retenus. Il est impossible d'associer une validation à un passe. Les numéros de passe sont chiffrés de façon irréversible une première fois par les transporteurs dans un délai maximum de 48 heures (autorisé par la CNIL pour lutter contre la contrefaçon) et une deuxième fois par le STIF.

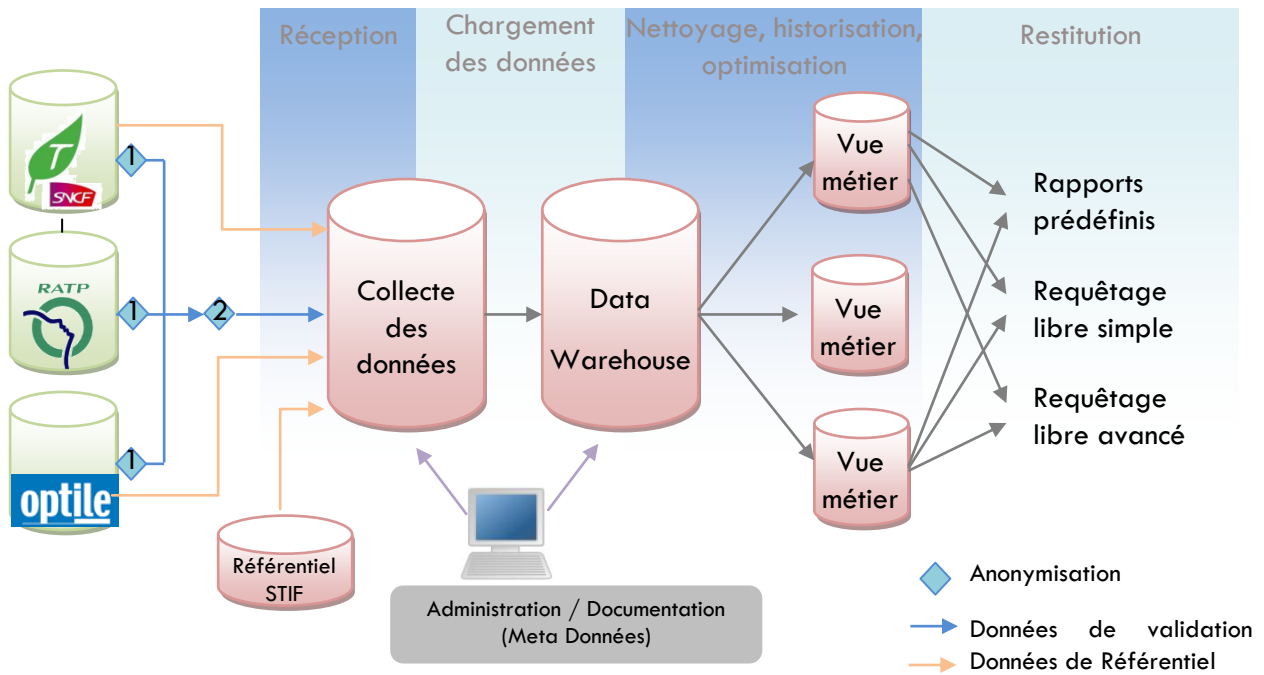
Ce système de double **anonymisation** du numéro de carte, avec un changement trimestriel des clés d'anonymisation interdit donc de suivre un usager particulier en faisant le lien entre la base clients et le numéro anonymisé utilisé pour les données de validation.

Architecture logique

L'architecture logique du système est présentée ci-après. Elle est classique pour un système d'information décisionnel.

Elle se décompose en :

- une zone de **réception** des différents flux opérateurs (flux quotidiens) ;
- une zone dite de **chargement** permettant l'homogénéisation des données, le contrôle de conformité de celles-ci avec les référentiels (gares, titres de transports, transporteurs...) et le stockage de manière standardisée dans un entrepôt de données ;
- une zone **métier** constituée de différentes vues de l'information dédiées aux analyses et structurées en fonction des besoins des utilisateurs ;
- un espace de **restitution** permettant le requêtage du système.



L'organisation de l'information

Un système décisionnel

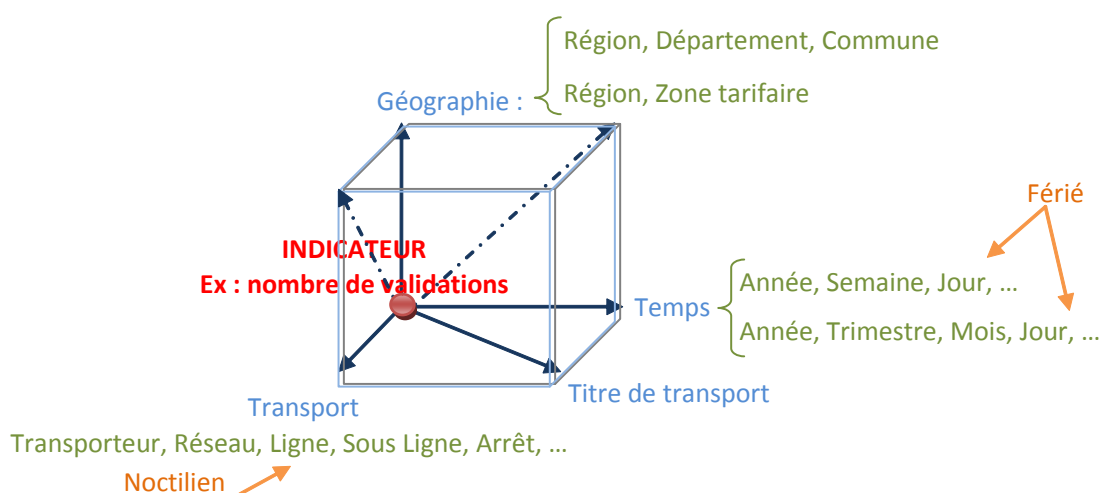
L'**informatique décisionnelle** (ou Business Intelligence) est à la base l'exploitation des données d'une entreprise dans le but de faciliter la prise de décision, c'est-à-dire la compréhension du fonctionnement actuel et l'anticipation de ses actions.

Un système d'information décisionnel est alimenté grâce à l'extraction de données diverses. Une première étape consiste à extraire les données dans différentes sources, les nettoyer et les charger dans un « **entrepôt de données** ». Une seconde étape dite d'analyse décisionnelle s'appuie sur la modélisation de représentations à base de requêtes afin de constituer des tableaux de bord ; on parle de « **reporting** ». Un tel système a pour objectif de proposer aux utilisateurs des **vues métiers** correspondant à des notions et termes connus par les utilisateurs et organisés de manière à optimiser les performances (temps de réponse des requêtes).

Modèle

Le modèle de données choisi pour le SIDV est un **modèle dit « en étoile »**. Celui-ci offre des performances adaptées aux fortes volumétries via des opérations de transformation et d'agrégation. Ce type de modèle repose sur les notions de dimension (ou axe d'analyse) et d'indicateur, soit la variable à analyser.

L'ensemble des **dimensions** représente l'offre d'analyse pour un besoin donné. Plus le nombre de dimensions est important, plus les possibilités d'étude sont nombreuses mais plus la volumétrie va, en général, s'accroître. Concrètement, les dimensions sont des ensembles logiques d'informations constitués d'une ou plusieurs **hiérarchie**. Par exemple, pour l'axe temps, une semaine ne pouvant être rattachée à un seul mois, certaines étant à cheval sur deux mois, on a deux hiérarchies indépendantes : année/semaine/jour ou année/trimestre/mois/jour. En plus des éléments définissant une ou plusieurs hiérarchies, des **attributs** supplémentaires, transverses à la hiérarchie, peuvent être définis pour un axe, relatifs à un ou plusieurs niveaux (ex : attribut férié pour l'élément jour).



Les indicateurs (ou faits ou mesures) sont généralement numériques et destinés à mesurer un élément représentatif du métier de l'entreprise. L'indicateur est en général additif, ou peut être **agrégé** selon une opération simple. Exemples : nombre de voyageurs ; nombre de déplacements ; nombre de validations télébillettiques... Un indicateur est défini par les propriétés suivantes : nom, définition fonctionnelle, axes d'analyse par lesquels il est étudié, règle de calcul, sources de données. Quand les axes d'études sont associés à un indicateur, il faut aussi définir le **grain** correspondant, i.e. le niveau de la hiérarchie le plus fin sur lequel peut être restitué l'indicateur. Par exemple, un nombre de courses sera défini sur une ligne de bus, et non par arrêt, donc au grain ligne sur l'axe transport.

IV. LES DONNEES

Une fois rapatriées, traitées, stockées et analysées, les données de validation peuvent être utilisées pour des analyses diverses.

On les compare traditionnellement à d'autres sources de données. Ainsi les données d'enquête ou de comptage pourront servir soit à compléter les analyses, soit à redresser les données de validations pour combler les phénomènes structurels liés à la validation télébillettique (sous validation dans les bus, absence de valideurs dans certaines gares...).

Il convient donc de replacer les données de validation dans un contexte plus global de mesures.

Les données de mobilité

Enquêtes et études de mobilité

Le STIF dispose de plusieurs enquêtes réalisées à des fréquences plus ou moins importantes. Nombre de prévisions s'appuient sur les données de l'EGT (**Enquête globale transport**), qui fournit des informations diverses sur les habitudes de transport (en commun et autres) des franciliens. Mais le STIF effectue aussi ses études à partir des données d'**enquêtes** ou de **comptage**, basées sur un nombre plus limité de voyageurs. Les résultats de ces enquêtes peuvent servir à définir la rémunération des opérateurs, ou pour par exemple à répartir les subventions (exemple : estimer la part du titre Imagine R financée par la région à partir d'une enquête réalisée tous les 4 ou 5 ans).

Trois types d'enquêtes sont à distinguer :

- Les **comptages**, axés autour du point d'arrêt :
 - Elles recensent pour un point d'arrêt, une ligne, un mode... le nombre de montants et de descendants au cours d'une période de référence. L'unité statistique est ici le voyage, et non les utilisateurs uniques
 - Elles sont réalisées par les transporteurs eux-mêmes ; et chaque mode ayant ses spécificités, les informations que l'on y retrouve ne sont pas homogènes d'une enquête à l'autre
- Les **enquêtes mobilité/tarifification**, axées sur l'utilisation d'un titre :
 - Elles recensent pour un individu donné, l'ensemble des déplacements réalisés avec un titre donné, sur une période de référence (semaine, mois...) variant selon le titre utilisé
 - Elles permettent d'estimer des indicateurs individuels de mobilité (fréquence et volume de déplacement des porteurs, répartition spatiale de leurs déplacements, etc.)
- Les **enquêtes OD** (Origine – Destination), axées autour de l'utilisation d'un mode ou d'une ligne spécifique, mais qui ont trait à l'ensemble du déplacement associé à l'utilisation de ce mode :
 - Comme les enquêtes de Mobilité/tarifification, elles rendent compte des phénomènes de multi modalité et de répartition spatiale des déplacements
 - Elles permettent notamment d'évaluer les taux de correspondance, de rabattement ou de diffusion pour une gare X
 - Elles sont réalisées par les transporteurs eux-mêmes : toutes les lignes ne sont pas forcément enquêtées, et pas forcément mises à disposition du STIF

Les transporteurs réalisent également eux-mêmes des comptages sur leur réseau, auquel peut avoir parfois accès le STIF, mais ces comptages ne sont pas toujours homogènes.

Thèmes du SIDV vs Enquêtes

Les grandes vues métier (axes d'analyse) du SIDV sont : la fréquentation (ou plutôt la validation), la vitesse commerciale bus, le rabattement en gare et la mobilité.

Pour le moment, l'analyse de la **fréquentation** se faisait jusqu'à présent principalement via les comptages OPTILE (tous les 2 ans, 600 lignes enquêtées sur 1 000) et les enquêtes par titre (réalisées tous les 4 ans pour un titre). Concernant la **vitesse commerciale** des bus, il n'y avait actuellement aucune restitution permettant de la suivre à part celles communiquées par les transporteurs. Seules quelques lignes de bus OPTILE sont suivies. Pour les pôles, le **rabattement** est actuellement analysé à l'aide des rapports des bureaux d'études dans le cadre des études de pôles PDU (Plan de déplacement urbain). Enfin les **enquêtes de mobilité** sont réalisées au travers des enquêtes sur les catégories de titres existants ou sur des catégories de personnes (pour la création de nouveaux titres).

Les enquêtes et le système télébillettique ont tous deux des avantages comme des inconvénients, comme indiqué ci-après :

Enquête de mobilité	Remontées des données TLB
<ul style="list-style-type: none"> + Connaissance fine des profils des personnes enquêtées (CSP, Age, lieu de résidence...) + Connaissance des chaînes modales empruntées, des titres utilisés... - Echantillonnage rendant difficile l'analyse des titres rares et des cas limites - Périodes d'enquêtes nécessairement limitées dans le temps 	<ul style="list-style-type: none"> - Peu d'information sur le profil du porteur - Modélisation des comportements de validation pour approcher les trajets et déplacements + Remontées de l'exhaustivité des données de validation permettant l'analyse de cas minoritaires + Mesure continue dans le temps, permettant de suivre les évolutions et d'établir des comparaisons temporelles

Les données de validation

Les données collectées depuis la construction du système décisionnel offrent de nouvelles opportunités d'analyse ou, le cas échéant, la possibilité de mettre à jour des données issues d'enquêtes anciennes en observant l'évolution de certains indicateurs. Différents thèmes métier ont été identifiés et des résultats sont ainsi produits régulièrement.

Néanmoins, les résultats produits peuvent montrer des différences plus ou moins importantes avec les informations issues d'autres sources. Il s'agit dans ces cas-là d'identifier la source la plus pertinente et de s'en servir pour recalibrer les données en identifiant l'origine de ces différences.

Comme on le verra ci-après, le système a des biais, qui peuvent pour certains d'entre eux être redressés assez aisément. Pour d'autres, les méthodes à mettre en œuvre sont plus complexes et ne seront totalement efficaces que dès lors que l'équipement en gares et le taux de validation auront progressé.

Organisation des données

Dans le SIDV, les différentes vues métier disponibles ont été construites après analyse des besoins potentiels d'étude. On peut regrouper les vues métiers en **quatre thèmes** : Validation, Nombre de cartes, Rabattement, Vitesse commerciale et Mobilité. Le thème Validation donne une image de l'action de validation lui-même, alors que les indicateurs de Rabattement et Mobilité sont reconstitués/modélisés à partir de ces validations pour obtenir une vue de :

- L'inter-modalité bus/ferré, soit comment les usagers utilisent la complémentarité de ces deux modes pour se déplacer. Il s'agit du phénomène de rabattement ou diffusion en gare.
- Des déplacements/trajets effectués sur le réseau et de leur particularité (origine, destination, chaîne modale, durée, etc.)

Les données et leurs limites

LES BIAIS

Il existe plusieurs biais dans le système ou ses principes qui viennent dégrader la qualité de l'information et la capacité à l'analyser de manière directe pour fournir une information exhaustive et fiable. Les biais du SIDV peuvent être classés en deux catégories :

- des biais liés aux remontées de données
- des biais liés aux règles métiers

LES BIAIS LIES AUX DONNEES

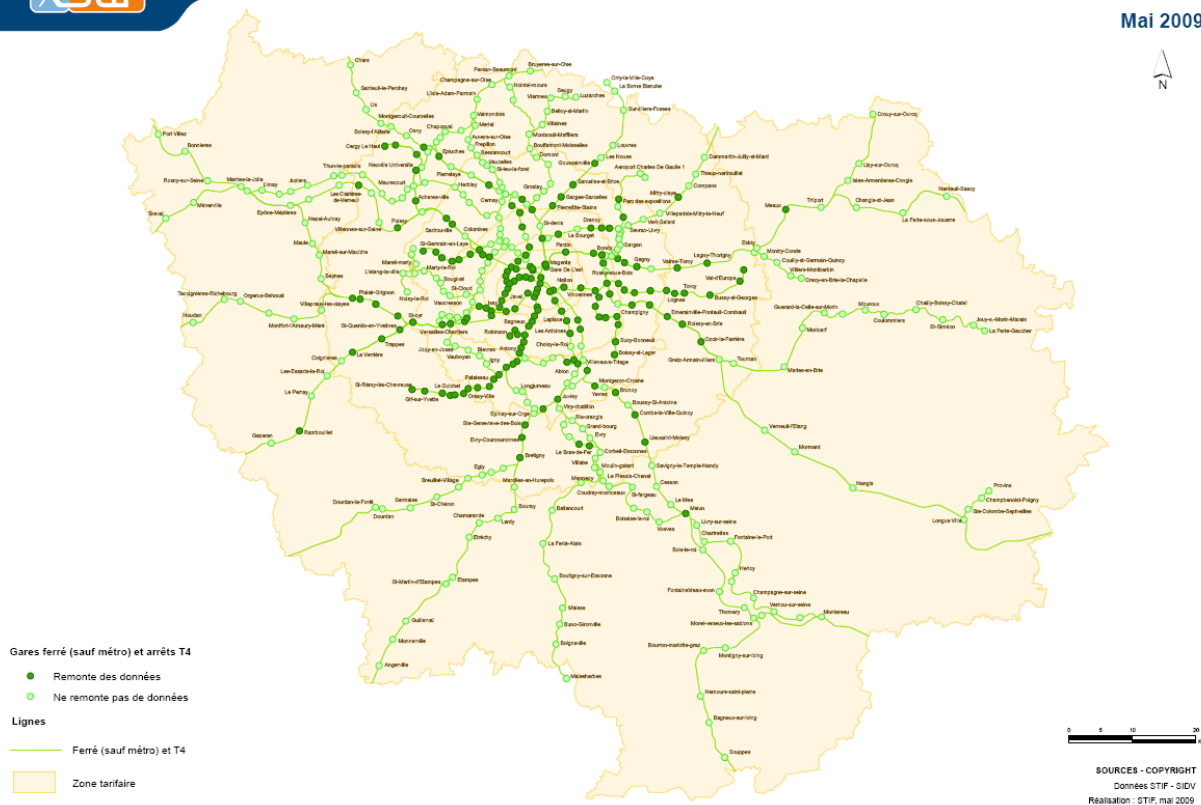
Biais de type gare ouvertes :

Sur 389 gares du réseau SNCF, 277 ne remontent aucune donnée. Parmi elles, nous retrouvons celles que l'on appelle « **Gares ouvertes** » et les gares équipées en CAB de technologie M1. Les gares ouvertes ne sont pas équipées pour remonter des données aux SIDV, on les retrouve sur le réseau SNCF notamment en grande couronne, mais également à Paris (ex : Gare Saint Lazare). Les gares équipées de valideurs de technologie récente (CAB M1) ou de bornes de validations (SNCF) ont commencé à remonter des données à partir de mars 2010 seulement (en raison d'un problème sur le système d'information de la SNCF).



Gares remontant les validations - Mode ferré (sauf métro) et T4

Mai 2009



Gares remontant (vert foncé) ou non (vert clair) des données de validation

Autre biais liés aux équipements :

Les **validations en correspondance** sont peu fiables en regard des équipements mis en place et/ou du fonctionnement dans les gares en heures pleines. De plus, en heure de pointe, il arrive souvent que les portes soient ouvertes en correspondance RER-métro dans Paris.

Validité et complétude des référentiels :

L'incohérence des référentiels entre transporteurs est à l'origine du manque de fiabilité sur les données. La non-remontée des **coordonnées (X, Y)** des arrêts par certaines entreprises OPTILE impacte quant à elle le calcul de cohérence des trajets, déplacements, distance, dézouage...

LES BIAIS LIES AUX REGLES METIERS OU AU COMPORTEMENT DES USAGERS

Biais ayant trait aux habitudes des usagers :

- Non-validations sur le réseau routier (Bus et Tramways) : l'entrée des bus et tramways étant **non bloquante**, la validation peut être optionnelle de la part des voyageurs. Ce phénomène de sous-validation induit une sous-représentation dans le SIDV des trajets bus.
- Bornes de validations : cet équipement ne constitue aucun obstacle à l'entrée dans une gare. La validation en ces points est optionnelle et donc assujettie à de nombreux facteurs externes (conditions climatiques, bon vouloir de l'utilisateur, si le train est sur le quai...). Les données sont remontées, mais la variabilité d'un jour à l'autre peut être importante.

Indisponibilité ponctuelle des remontées dues à la **maintenance ou aux travaux** dans certaines gares : joue directement sur la fréquentation de la gare/station et la ligne concernée ;

Jours exceptionnels par rapport au plein trafic : ils peuvent être de plusieurs types : mouvements sociaux, jours fériés, ponts, etc.

Titres en magnétiques : La coexistence de titres magnétiques et télébillettiques est un frein à l'étude du comportement d'un voyageur dans le sens où il est impossible de voir si l'utilisateur utilise ou non un titre complémentaire à son titre télébillettique. La généralisation à venir du support télébillettique pour l'ensemble des titres franciliens permettra de régler ce problème.

CONCEPTS IMPACTES

L'ensemble de ces biais impacte directement ou indirectement les notions suivantes :

- la fréquentation : impossibilité de connaître le nombre de validations réelles pour certaines gares ou dans les bus
- la définition des déplacements : rupture non pertinente entre les différents trajets
- les caractéristiques des déplacements (erronées ou manquantes) : origine/destination, valorisations associées, chaînes modales biaisées
- l'étude du rabattement, et de la multi modalité

A cause de ces biais, on n'obtient donc actuellement jamais l'exhaustivité des données. N'ayant pas de situation de référence, il est difficile de savoir parfois si les résultats obtenus sont normaux ou anormaux.

Les exemples d'utilisation décrits ci-après montrent, d'une part, l'utilisation des données de validation pour répondre à des questions concrètes, et d'autre part, la manière de redresser ces données pour obtenir un indicateur fiable.

V. UTILISATIONS DES DONNEES DE VALIDATION

Analyse des validations, estimer des fréquentations

Présentation

Le thème Validation offre une vue de l'acte validation lui-même au travers de l'indicateur simple : « **Nombre de validations** ».

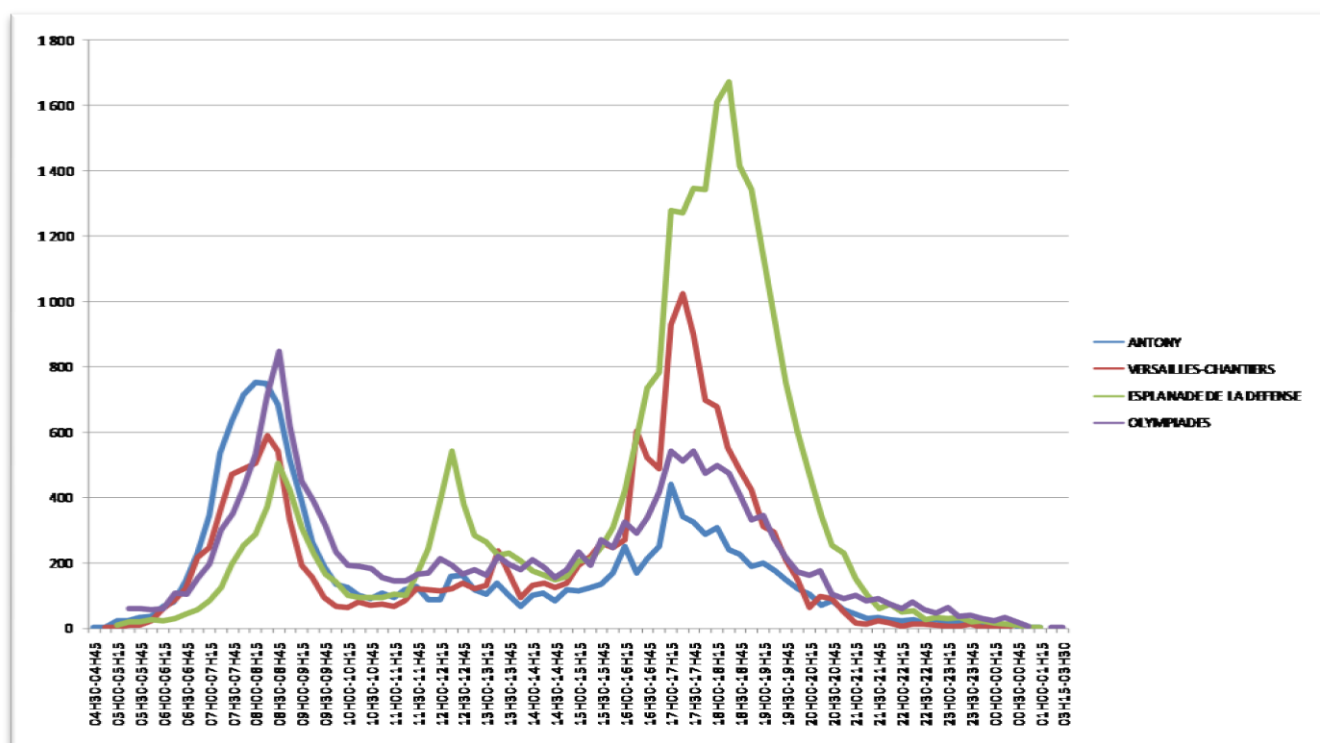
A terme, après redressement des données, l'objectif est de pouvoir établir une vue de la **fréquentation** générale pour notamment :

- établir le point de plus forte charge d'une ligne,
- apprécier la fréquentation d'une gare,
- évaluer le trafic sur les jours exceptionnels, etc.

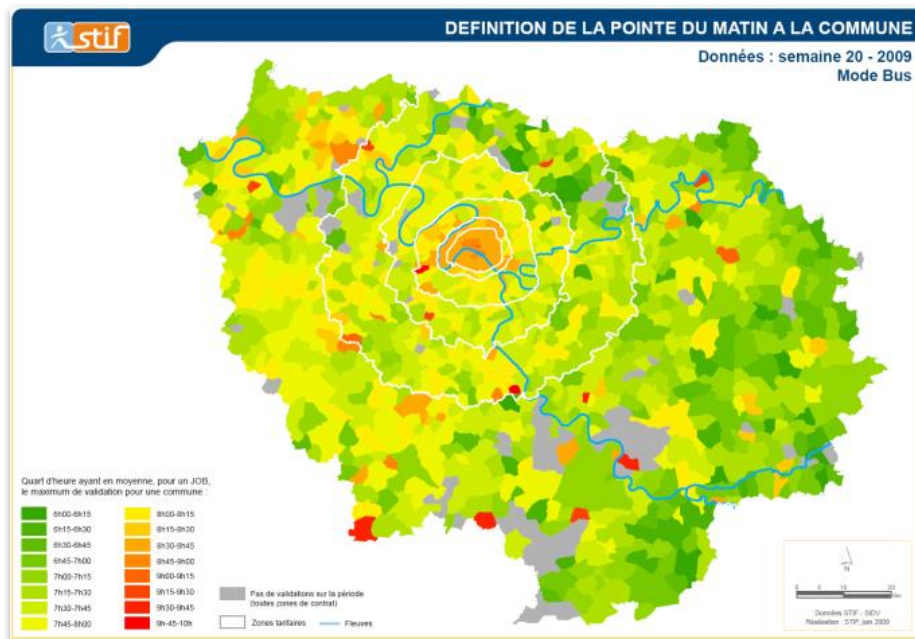
Mise en application générale

A ce stade, l'analyse des validations permet déjà de répondre à des problématiques particulières, en ce qui concerne la charge (lignes de bus par exemple) ou l'appréciation des points prioritaires d'un réseau (nombre de montants par arrêt, trafic entrants/sortants sur des gares).

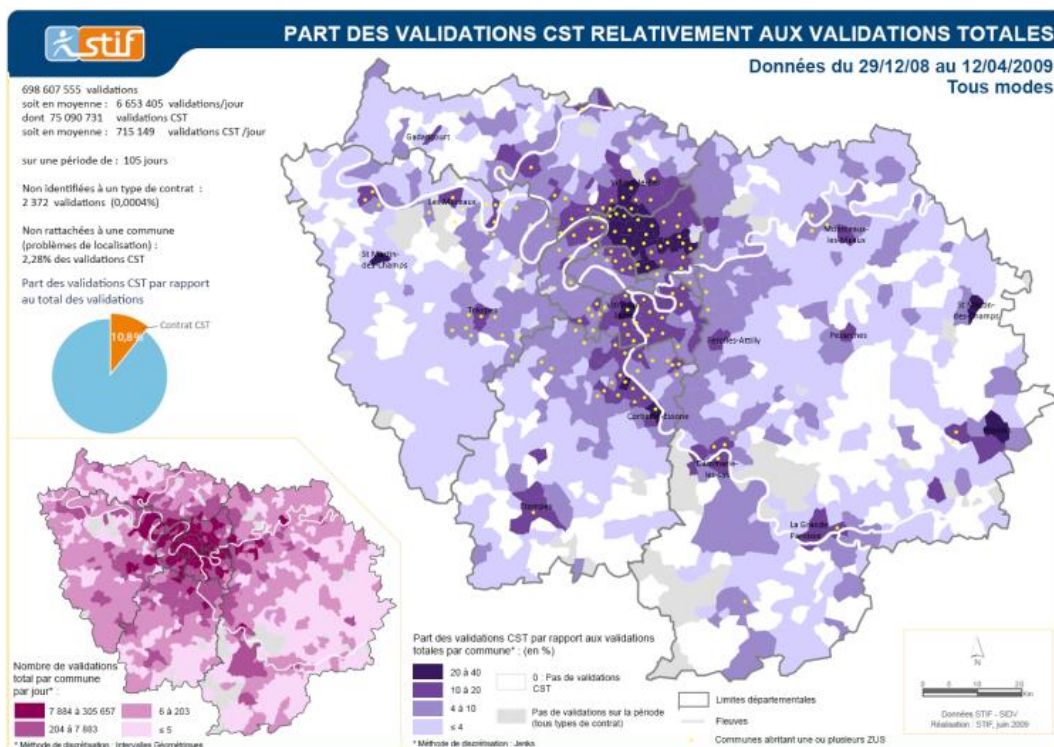
L'utilisation de l'axe temps permet de dégager des tendances sur des gares ou des zones géographiques. Ainsi il est possible d'avoir une idée de la fréquentation des gares aux différentes heures de la journée, ou plus globalement d'approcher l'heure de pointe en fonction de la commune de validation.



Dans l'exemple suivant, le SIDV permet ainsi de donner des chiffres précis et des tendances sur les horaires de début de déplacements bus. Ainsi, alors qu'on considère habituellement que l'heure de pointe est homogène pour l'ensemble de l'Île de France, on observe plus précisément que les habitants de grande couronne ont une heure de pointe plus matinale que les parisiens.



L'analyse des validations permet aussi, en utilisant les informations sur les contrats détenus par les porteurs des passes, d'approcher leurs lieux de validation et la typologie de leurs déplacements.



Questions traitées

ESTIMATION DU TRAFIC SUR UN RESEAU FERRE

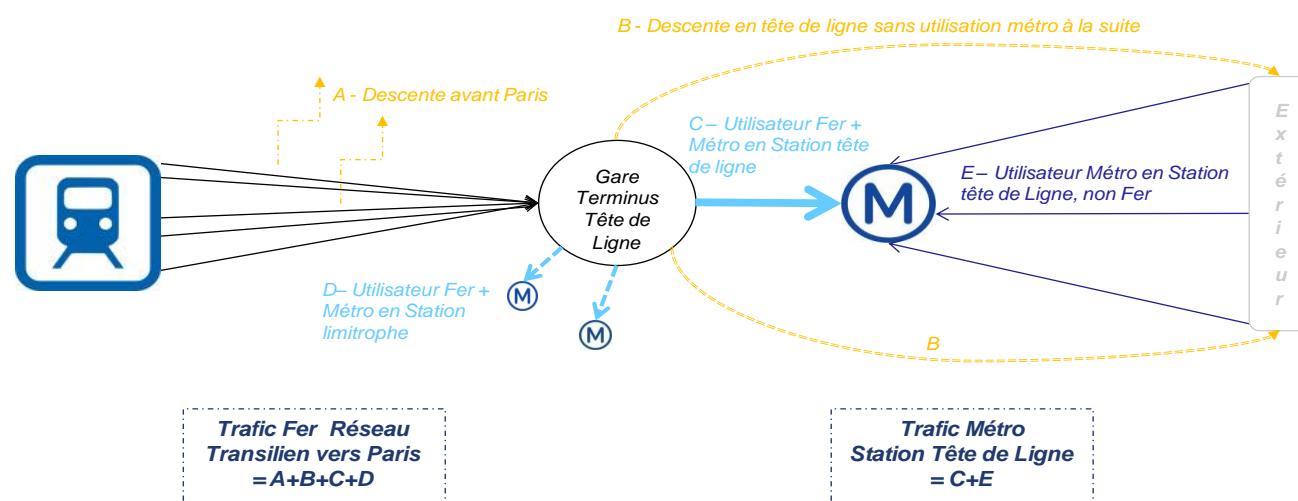
Enjeu

Si on prend l'exemple du réseau St-Lazare, la principale difficulté d'estimation du trafic sur un tel réseau vient du fait que ce réseau, le principal d'Ile de France est quasiment intégralement composé de gares ouvertes (à commencer par la gare St Lazare elle-même). De ce fait, la simple comptabilité des données de validation ne permet pas une reconstitution de la mobilité sur celui-ci.

Mise en œuvre

Une idée a donc été de s'appuyer sur le volume d'entrants sur la station de réseau de tête de ligne (métro St Lazare) pour reconstituer le trafic global sur le réseau à l'aide de données de comptages et d'enquêtes.

Sous réserve d'une certaine stabilité structurelle, l'observation des données de validation de cette station de métro (et des stations alentour) permet de déterminer de manière assez fiable les fluctuations de trafic sur le réseau.



Le principe de l'estimation consiste à extrapoler l'ensemble du trafic généré par le réseau, c'est-à-dire environ $2 \times (A+B+C+D)$, à partir des données de validations métro en « station pivot », c'est-à-dire $C+E$. Pour ce faire on s'appuie sur les informations issues d'enquêtes, notamment RATP, afin de déterminer les valeurs de chacun des coefficients.

Ce raisonnement est valable pour l'ensemble des réseaux disposant d'une station pivot, dotée d'un volume conséquent de données et surtout dont la part du trafic en provenance du réseau concerné est suffisamment représentative.

Dans d'autres cas, on dispose d'un nombre suffisant de gares équipées et remontant des données de validation sur le réseau pour en extrapoler le trafic global à partir de coefficients de redressement construits sur la base d'enquêtes et de comptages. C'est le cas par exemple du RER A pour lequel la principale difficulté résidera donc dans l'identification de la part de trafic montants – descendants des

stations à forte correspondance sans revalidation (ex : correspondance quai à quai RER A – RER B à Châtelet - les Halles).

Résultats

En utilisant l'une ou l'autre de ces méthodes, on peut reconstruire et observer les évolutions (variations hebdomadaires, annuelles...) du trafic sur les réseaux suivants :

- St Lazare (Métro St Lazare)
- Montparnasse (Métro Montparnasse + gares fermées)
- Paris Est (Métros Gare de l'Est + Château Landon)
- RER E (Métros St Lazare + Gare de L'Est)
- RER A (ensemble des stations sauf stations complexes)

Par contre, nous observons aujourd'hui des difficultés à quantifier correctement le trafic sur les réseaux suivants :

- RER B, C & D : lignes traversantes sans tête de ligne importante et manquant de gares fermées ou remontant des données
- Paris Nord : actuellement, aucune validation en correspondance vers le métro ne remonte
- Paris Sud Est : station tête de ligne Gare de Lyon drainant trop de trafic pour être utilisée comme gare pivot

Pour illustration, l'étude des validations sur les gares du RER A permet d'estimer finement le trafic lors d'un jour ouvré sur cette ligne. Ainsi on déduit des validations que le RER A représente en moyenne 855 000 utilisateurs par jour en télébillettique, dont 75 000 montants en zone SNCF et 780 000 montants en zone RATP, avec 511 000 montants en branche centrale (A0).

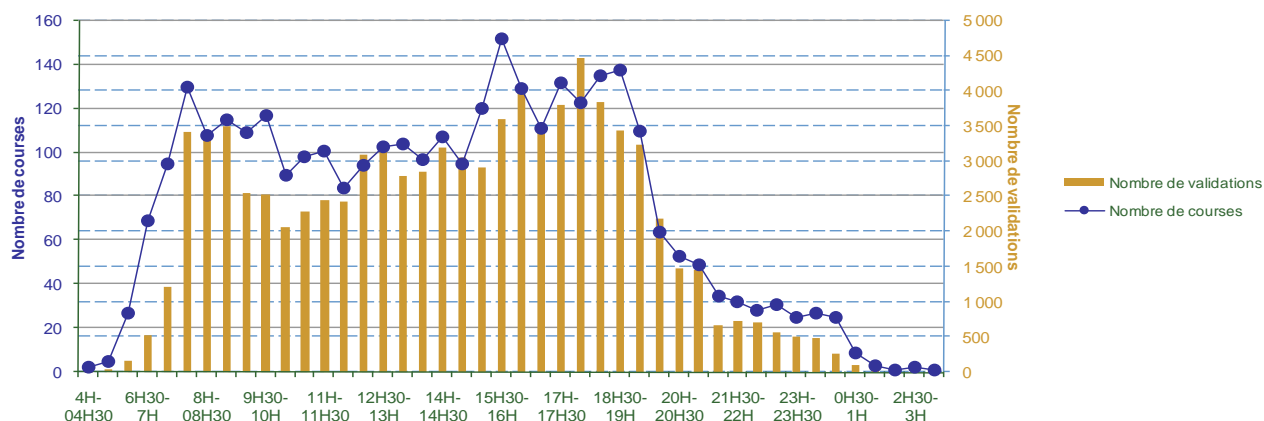
L'estimation des déplacements effectués avec des titres magnétiques sur le RER A sur un jour ouvré (environ 25% du total) permet de redresser ce chiffre pour obtenir un trafic global d'1,15 millions de déplacements chaque jour ouvré.

COMPARAISON OFFRE ET DEMANDE

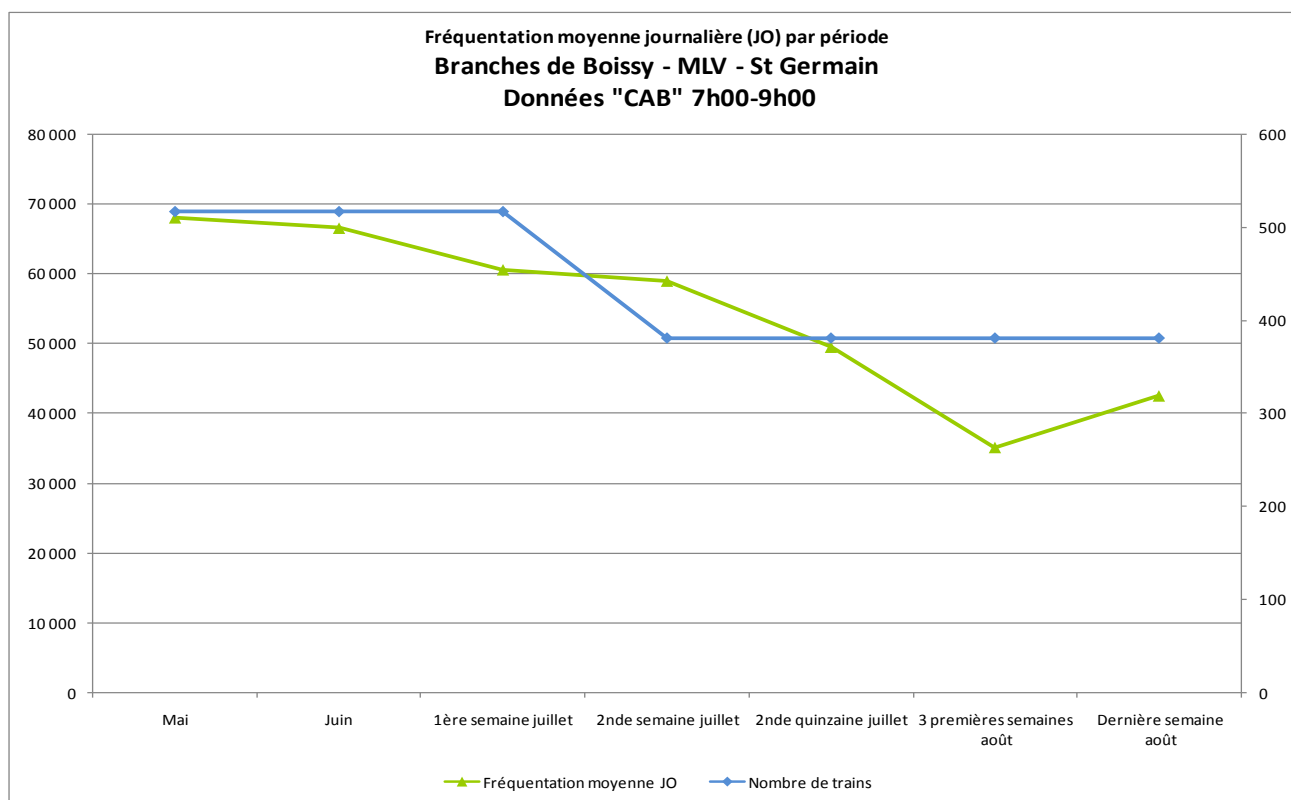
La problématique d'adéquation entre l'offre et la demande est au cœur des contrats avec les opérateurs. Il s'agit d'ajuster le plus finement possible ces deux paramètres de manière à offrir une qualité de service satisfaisante à l'usager tout en maîtrisant les coûts.

Ce point peut être utile pour un ajustement horaire en fonction du moment de la journée. Par exemple, il peut être traité de manière assez basique sur le bus en faisant une comparaison horaire du nombre de courses et du nombre de montants dans le bus sur une journée entière.

Par exemple, sur le graphique suivant on mesure l'adéquation entre le nombre de courses, d'une part, et le nombre d'usagers (a priori proportionnel au nombre de validations) pour une ligne de bus sur l'ensemble des jours ouvrés d'une semaine.



Mais, le système peut également permettre d'étudier certaines périodes de l'année, comme les périodes de vacances et de vérifier que la baisse d'offre pendant ces périodes est en phase avec la consommation réelle de transport sur une ligne.



Par exemple, le système a été utilisé pour mesurer l'adéquation de l'offre de trains sur une branche de RER pendant les vacances d'été avec le nombre d'usagers empruntant cette branche aux heures de pointe du matin.

Analyse des cartes, estimer le nombre de voyageurs

Présentation

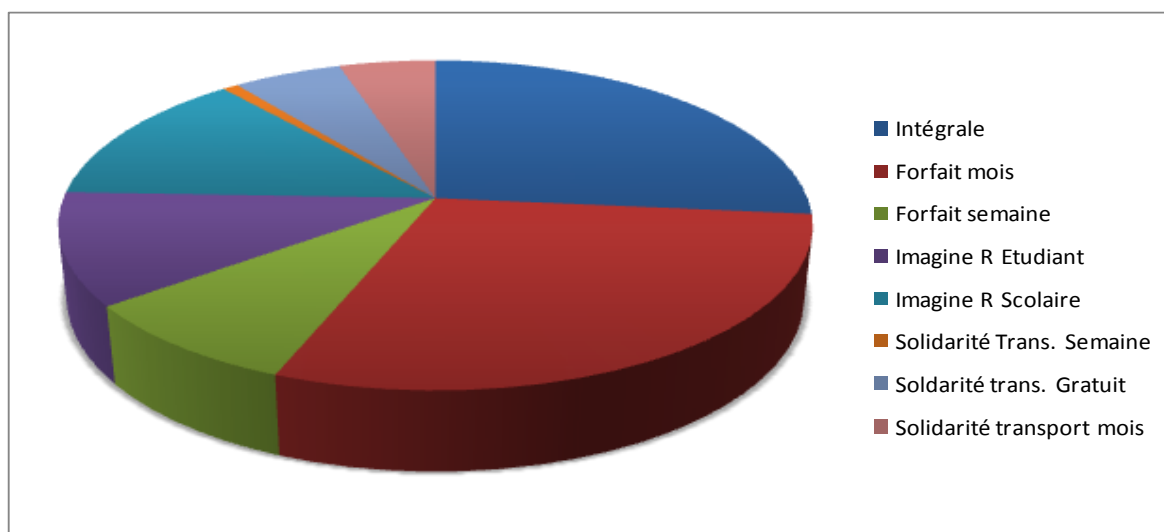
Le thème relatif aux contrats et aux cartes offre une possibilité, sur la base du numéro anonymisé, d'historiser sur une période de 3 mois maximum le contrat utilisé pour une validation et d'approcher de ce fait les contrats détenus sur une carte pendant cette période.

Ces chiffres, qu'on peut affiner suivant divers axes (par exemple le couple de zones du contrat) permettent d'observer l'utilisation des contrats et leur complémentarité pour un même individu et de valider l'utilisation des titres obtenue traditionnellement à partir des chiffres de ventes. Mieux, ils permettent, à partir de l'observation réelle sur le réseau du nombre de cartes porteuses de contrats, d'approcher très tôt une prévision des ventes concernant les titres télébillettiques.

Mise en application générale

A ce stade, l'analyse des contrats et des cartes permet déjà d'extrapoler pour certains titres les chiffres de ventes, communiqués traditionnellement par les transporteurs un mois après la fin du mois concerné.

Elle permet de donner en volumes, la répartition des contrats circulant sur le réseau (type de contrats, couple de zones...).



Questions traitées

On observe quelques disparités entre les contrats vues en circulation et les contrats réellement vendus. Une fois redressés, ces données permettent une **prévision** assez juste (avec 0,5% de précision) **des ventes de titres télébillettiques sur un mois donné** et ce, simplement à partir de l'analyse des validations sur les 15 premiers jours du mois.

En pratique, on observe plus de cartes circulant avec un forfait Navigo mois que de forfaits vendus. Le différentiel, assez faible, s'explique par les pertes et changements de cartes (une carte perdue étant reconstituée avec le même contrat mais un numéro différent, elle est comptée une fois dans les chiffres de vente et deux fois dans les validations) ainsi que les changements de zone en cours de mois. Globalement, l'intégralité de ces forfaits valide au moins une fois dans le mois concerné sur un appareil dont les validations remontent au système.

A l'inverse, certains titres comme Imagine R scolaire ne sont pas vus en totalité ; en effet certains n'empruntent que des bus Optile dans des zones où le taux de validation est faible. D'autres cartes, financées en partie ou en totalité pour certains élèves, sont achetées mais inutilisées certains mois. C'est également le cas pour certaines Cartes Intégrale (Navigo annuel) qui ne circulent pas du tout au mois d'août.

Globalement, cette méthode permet d'arriver à des résultats fiables comme le montre le tableau suivant qui pour le forfait Navigo mois et le forfait annuel indique chaque mois les ventes effectives, les circulants vus dans le système et l'extrapolation qui en est faite sur la base de l'observation des redressements des mois précédents.

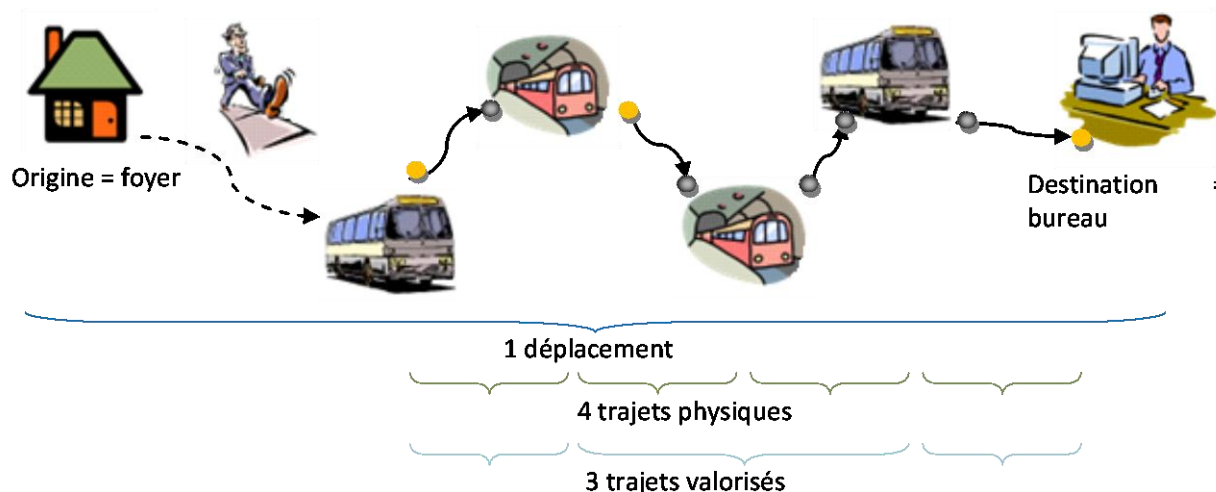
Mois	Navigo mois			Navigo annuel		
	Unités			Unités		
	Ventes effectives	Circulants SIDV	Estimation des ventes	Ventes effectives	Circulants SIDV	Estimation des ventes
janv.-09	1 007 306	1 028 060		869 802	871 636	
févr.-09	1 000 821	1 009 646	989 264	873 072	868 588	866 760
mars-09	1 039 140	1 051 056	1 035 854	874 495	873 097	874 432
avr.-09	1 000 386	1 012 844	999 250	873 774	870 644	871 996
mai-09	984 556	993 606	980 549	872 875	871 377	873 175
juin-09	1 018 735	1 021 472	1 008 872	871 278	877 840	879 591
juil.-09	833 588	854 829	845 661	874 664	850 883	851 237
août-09	565 729	570 227	564 111	870 418	798 939	799 272
sept.-09	952 839	963 728	953 392	876 481	874 904	875 268
oct.-09	1 034 598	1 046 478	1 035 168	886 155	883 646	884 189
nov.-09	1 025 735	1 032 787	1 021 555	890 299	886 435	887 226
déc.-09	980 551	991 599	981 261	890 744	887 928	889 062
janv.-10	1 003 344	1 009 788	999 188	894 193	890 691	891 998
févr.-10		1 005 992	995 808		895 105	896 619

Analyse de la mobilité, quantifier et qualifier les déplacements

Présentation

Le thème Mobilité repose sur les notions suivantes :

- La notion de **déplacement** : Fait d'aller d'un point à un autre pour un motif déterminé (par exemple : travail, course, sortie, ...), en utilisant un ou plusieurs modes de transport successifs. Un déplacement est défini notamment par un motif, une origine et une destination. Un déplacement se décompose en un ou plusieurs trajets.
- La notion de **trajet** : Fait d'utiliser un seul véhicule / mode / exploitant, sauf sur le réseau métro où une entrée correspond à un trajet quel que soit le nombre de correspondances. Donc tout changement de véhicule, de mode ou d'exploitant génère un nouveau trajet sauf pour le métro.



Les principales difficultés pour reconstituer ces trajets et déplacements à partir des données télébillettiques sont :

- la **détermination de la destination** : délicate par manque d'information sur les sorties du réseau transport (seules certaines sorties du réseau RER sont fournies)
- l'identification des validations correspondant à un déplacement : difficile à effectuer car le temps séparant 2 déplacements peut sensiblement varier (journée de travail, course de quelques minutes le midi, ...)
- par manque d'information dans les données télébillettiques (validations en sortie très rares, pas de connaissance des lignes ferrées empruntées, pendularité complexe à identifier au niveau des déplacements)

Afin de définir des **règles de modélisation** des déplacements en se basant sur les données télébillettiques, une étude préalable sur les déplacements contenus dans l'Enquête Globale Transport (EGT) a été réalisée. Des règles ont été alors proposées pour reconstituer les trajets et déplacements à partir des données télébillettiques.

La modélisation des trajets est ainsi définie entre autres par les règles suivantes :

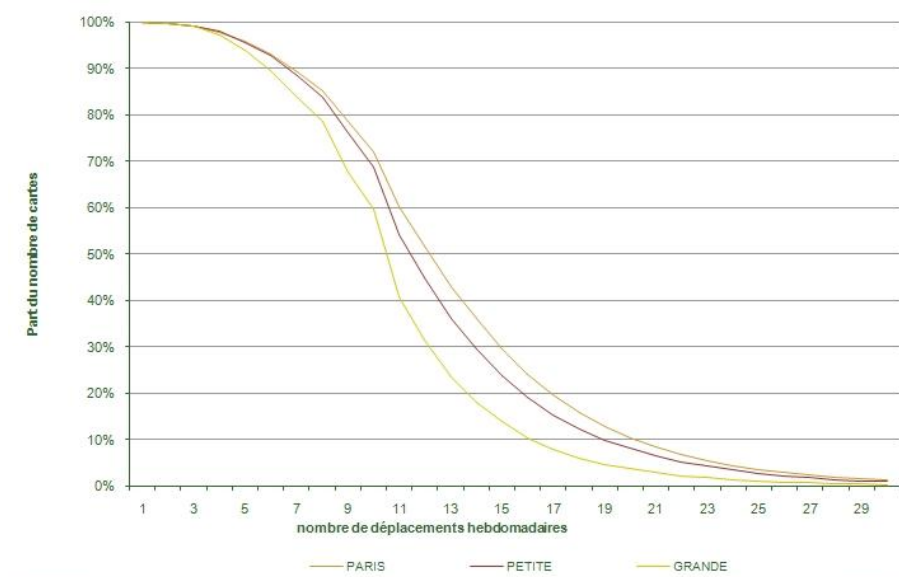
- un déplacement est défini par toutes les validations en entrée ou correspondance dans un délai de **60 minutes** (délai standard, dépend de la localisation et de l'heure de début de déplacement)
- la destination d'un trajet est le lieu de validation en sortie si elle est fournie, sinon le début du trajet suivant,
- deux validations en entrée métro, ou sur une même ligne de bus, ne peuvent appartenir à un même déplacement,
- utilisation d'une **matrice origine-destination** générée à partir du modèle de prévision de trafic du STIF pour déterminer la distance et les zones traversées

Mise en application générale

L'analyse de la mobilité a pour objectifs, entre autres, de :

- adapter et créer de nouveaux produits tarifaires répondant aux besoins des voyageurs
- aider à la définition de la rémunération des entreprises en déterminant les trajets par titre et sur chacun des réseaux
- connaître les déplacements des titres pris en charge par les collectivités pour identifier le prix de revient réel de ces titres

Par exemple, l'observation des trajets et déplacements des porteurs d'un forfait mensuel ou annuel permet de définir au mieux la pertinence économique du tarif de ces titres. Sur des mesures de septembre 2009, on observe que la mobilité hebdomadaire moyenne d'un porteur de forfait mensuel est de 11,7 déplacements, contre 10,5 pour un porteur de carte intégrale. Le système permet également de distinguer les parisiens ou habitants de petite couronne des habitants de grande couronne. En effet, il n'y a aucune information sur le lieu de résidence du porteur mais on approxime un « lieu usuel de première validation » qui donne une localisation approximative de résidence du porteur. On se rend de ce fait compte qu'un parisien ou habitant de proche banlieue va avoir une mobilité supérieure en TC, en termes de nombre de déplacements, un peu au-delà du simple aller-retour quotidien domicile-travail. A l'opposé, un habitant de grande couronne a une mobilité proche des 5 aller-retours hebdomadaires.



















Questions traitées

La mobilité est aujourd'hui la notion métier qui est la plus approximative dans le système de par les biais évoqués précédemment.

De nombreux déplacements sont jugés non cohérents par le système quand on les compare aux correspondances théoriquement effectuées pour un couple origine-destination, ou tout simplement car on est incapable de reconstituer leur origine ou leur destination. De plus, on constate une déformation de la décomposition en chaînes modales des déplacements, certains modes étant surreprésentés de par leur meilleure visibilité dans le système (métro, au détriment du bus et du fer).

Sur le tableau suivant on peut visualiser l'effet déformant apporté par l'analyse de la mobilité à partir des validations.








Si on lit par exemple la 4^{ème} ligne de celui-ci, on voit que les déplacements constitués d'un trajet bus et d'un trajet métro sont pour une bonne partie d'entre eux pertinents dans le SIDV mais que beaucoup, parmi eux se transforment en déplacements métro (de par le faible taux de validation dans le bus). Quelques-uns sont vus en bus seul, d'autres enfin ne sont pas du tout vus dans le système.

								
	●							●
		●						●
			●					●
	●	●		●				●
		●	●		●			●
	●		●			●		●
	●	●	●	●	●	●	●	●

Plus globalement, la part de trajets non vus dans le SIDV actuellement est de 22%. La part de déplacements non vus est de 10%.

De plus, si on compare avec une enquête récente (par exemple l'enquête Imagine R), on constate que les déplacements sont « simplifiés » par le SIDV puisque la part de déplacements monomodes y est de 84% contre 75% en réalité.

L'analyse de la déformation apportée par le système sur une population donnée (exemple ici des porteurs de cartes Imagine R 1-2 et 1-3) est la suivante :








	Structure théorique (Enquête IR)	Structure reflétée par SIDV
	53%	64%
	14%	9%
	8%	11%
	7%	6%
	4%	2%
	11%	7%
	3%	1%

Le travail de redressement effectué a donc consisté à modéliser une matrice théorique de passage Réel → SIDV et à l'appliquer à la répartition des nombres de déplacements par chaîne modale pour la population concernée.

Celle-ci a été construite à partir d'un postulat sur les taux de visibilité par mode de la population cible :

- 96% de visibilité dans le métro
- 60% de visibilité dans le bus
- 80% de visibilité sur le fer

Une fois appliquée aux données, cette matrice donne la structure de déplacements suivante :

	Structure théorique (Enquête IR)	Structure reflétée par SIDV	Structure corrigée post SIDV
	53%	64%	54%
	14%	9%	14%
	8%	11%	9%
	7%	6%	7%
	4%	2%	4%
	11%	7%	9%
	3%	1%	3%

Cette méthode de redressement offre un certain nombre d'avantages. En effet elle permet :

- Une rectification de l'assiette de déplacements par extrapolation ;
- Le redressement simultané des chaînes modales et du nombre de trajets par déplacement ;
- L'application d'un modèle théorique simple, qui peut toujours être calé par des observations terrain (enquêtes de « calage » permettant la comparaison entre validations et déplacements réels) ;
- Une adaptation potentielle à différentes cibles, zones géographiques, où les taux de visibilité par mode seraient différenciés.

Cependant elle ne permet que des redressements macroscopiques et, comme toute extrapolation, elle n'est valable que si l'univers de données d'origine est suffisamment large. De ce fait, il est impossible avec cette méthode de redresser la mobilité sur un niveau géographique trop restreint et sur lequel les données de validation seraient quasiment absentes (mauvais équipement en gare, taux de validation bus faible).

L'utilisation des données de reconstitution de trajets et déplacements permet néanmoins de concevoir des indicateurs sur des domaines variés affectant la valorisation financière d'un titre ou encore la qualité de service.

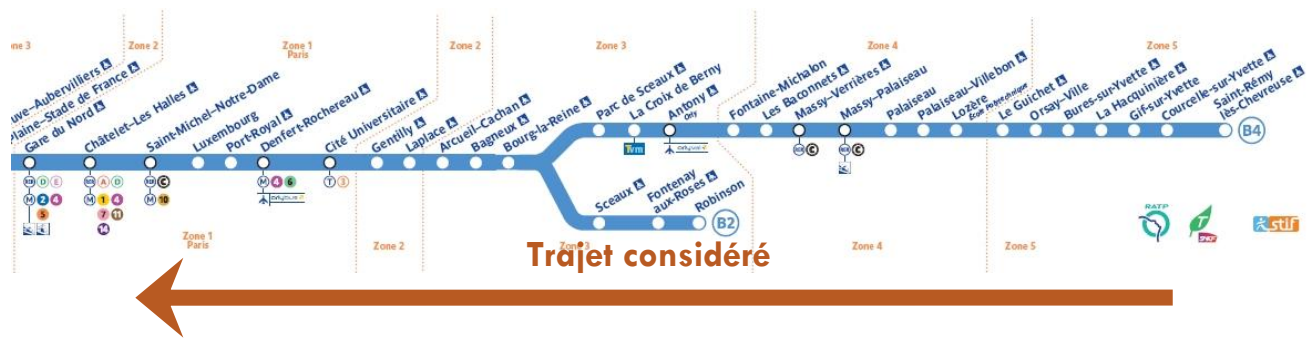
ESTIMATION DE LA REGULARITE DES TRAINS

Enjeu

Les données de validation permettent d'apprécier de manière précise la durée effective d'un trajet pour une carte donnée. Par exemple sur le réseau ferré, on peut mesurer le temps séparant la validation en entrée du réseau pour un voyageur et la validation de sortie / correspondance suivante.

L'exemple pris ici est celui de la branche Sud (St-Rémy lès Chevreuse) du RER B. Le tableau figurant ci après donne, jour après jour, le temps moyen d'accès à Paris pour une personne qui entre dans une gare en bout de ligne (en l'occurrence ici une gare entre St Rémy et Bures sur Yvette) les jours de semaine aux heures de pointe. L'accès à Paris s'interprète par l'arrivée dans toute gare entre Cité Universitaire et Châtelet.

Le temps de parcours s'obtient en prenant la durée des trajets RER, à savoir le temps entre la validation d'entrée dans une des gares de départ considérées et la validation de sortie/correspondance dans une des gares parisiennes.



Le tableau présente la moyenne, par tranche horaire des temps de déplacement des usagers concernés.

Le fait de s'appuyer sur plusieurs gares au départ comme à l'arrivée permet de lisser les effets « locaux » d'un problème relatif à un arrêt précis.

Tranche horaire / les parcours	18/11/2009	17/11/2009	16/11/2009	13/11/2009	12/11/2009	10/11/2009	09/11/2009	06/11/2009	05/11/2009	04/11/2009	03/11/2009	02/11/2009	30/10/2009	29/10/2009	28/10/2009	27/10/2009	26/10/2009	23/10/2009	22/10/2009	21/10/2009	20/10/2009	19/10/2009
06:30-07:00	0.54	0.42	0.48	1.17	1.13	1.00	1.09	0.53	0.47	0.48	0.48	0.50	0.48	0.44	0.49	0.41	0.47	0.48	0.48	0.51	0.40	0.46
07:00-07:30	1.04	0.53	0.53	1.00	1.14	0.59	1.04	0.51	0.47	0.52	0.50	0.47	0.45	0.52	0.45	0.57	0.47	0.51	0.49	1.14	0.46	0.46
07:30-08:00	1.07	0.47	0.46	1.13	1.07	1.00	1.00	0.50	0.48	0.51	0.47	0.48	0.48	0.57	0.44	0.43	0.44	0.48	0.47	1.03	0.49	0.47
08:00-08:30	1.09	0.46	0.44	1.08	0.59	1.02	0.54	0.48	0.51	0.47	0.42	0.49	0.47	0.49	0.41	0.43	0.45	0.47	0.48	0.53	0.59	0.47
08:30-09:00	1.32	0.49	0.47	1.19	0.57	1.01	0.51	0.48	0.51	0.45	0.48	0.51	0.47	0.53	0.43	0.42	0.45	0.45	0.45	0.57	1.11	0.53
09:00-09:30	1.40	0.43	0.43	1.13	1.00	0.56	0.50	0.50	0.50	0.45	0.58	0.56	0.50	0.49	0.40	0.42	0.43	0.42	0.47	0.58	1.02	0.49
09:30-10:00	1.48	0.53	0.47	1.11	1.03	0.54	0.58	0.49	0.58	0.47	0.57	0.48	0.45	0.47	0.41	0.43	0.45	0.44	0.48	0.54	0.58	0.49
10:00-10:30	1.47	0.57	0.46	1.10	0.57	0.51	0.57	0.51	0.53	0.50	0.50	0.51	0.48	0.48	0.45	0.44	0.44	0.44	0.45	1.00	0.58	0.52
10:30-11:00	1.47	1.23	0.46	1.12	1.13	0.56	0.52	0.50	0.56	0.48	0.53	0.54	0.45	0.44	0.48	0.47	0.47	0.47	0.48	0.59	0.59	0.58
11:00-11:30	1.38	1.08	0.43	1.10	1.18	0.58	1.00	0.55	0.53	0.51	0.59	0.48	0.51	0.46	0.44	0.49	0.47	0.47	0.44	0.57	1.02	0.59
11:30-12:00	1.32	0.57	0.46	1.04	1.18	1.07	0.58	0.48	0.51	0.48	0.55	0.50	0.50	0.47	0.43	0.48	0.45	0.50	0.47	0.55	1.03	0.50
12:00-12:30	1.22	0.58	0.58	1.12	1.13	1.05	0.47	0.51	0.56	0.46	1.00	0.50	0.44	0.48	0.44	0.43	0.44	0.50	0.48	0.57	1.04	0.43
12:30-13:00	1.18	0.54	0.53	1.13	1.07	0.58	0.49	0.50	0.58	0.49	0.56	0.52	0.48	0.48	0.44	0.43	0.45	0.52	0.47	0.52	1.01	0.47
13:00-13:30	1.10	0.54	0.53	1.11	0.88	1.03	0.55	0.48	0.48	0.49	1.03	0.46	0.48	0.48	0.49	0.42	0.43	0.58	0.54	0.48	0.56	0.52
13:30-14:00	1.01	0.53	0.53	1.10	1.02	0.57	0.53	0.49	0.48	0.49	0.58	0.46	0.47	0.49	0.45	0.42	0.50	0.49	0.50	0.54	0.55	0.48
14:00-14:30	0.57	0.50	0.49	1.10	1.28	1.02	1.01	0.48	0.49	0.48	1.08	0.49	0.45	0.48	0.45	0.43	0.52	1.05	0.48	0.48	0.59	0.45
14:30-15:00	0.55	0.47	0.58	1.04	1.24	0.58	1.00	0.47	0.45	0.45	0.58	0.46	0.46	0.58	0.43	0.47	0.48	0.50	0.43	0.47	0.48	0.48
15:00-15:30	0.48	0.54	0.46	0.44	1.24	0.57	0.57	0.40	0.42	0.47	0.53	0.45	0.46	0.58	0.43	0.41	0.41	0.43	0.41	1.01	0.47	0.47
Moyenne	1.22	0.57	0.48	1.12	1.07	0.58	0.55	0.50	0.53	0.48	0.54	0.50	0.46	0.43	0.43	0.45	0.45	0.48	0.47	0.57	0.59	0.50

Le tableau ci-dessus comporte en abscisse les jours consécutifs et en ordonnées la tranche horaire de part (heures de pointe du matin, tranches horaires de 10 minutes).

Si l'on prend en compte qu'aux heures étudiées, le temps de parcours théorique moyen est de 39 minutes. En incluant un temps moyen d'attente de 6 minutes (un train toutes les 12 minutes), on peut considérer que toute valeur entre 40 minutes et 50 minutes est acceptable (vert). En jaune, on a les valeurs entre 50 minutes et 1h, en rouge on est au-delà.

L'indicateur calculé présente deux avantages :

- La validation d'entrée symbolisant l'entrée sur le quai et celle de sortie/correspondance la sortie du quai, il s'agit donc d'un indicateur complet intégrant à la fois le temps d'attente du train ainsi que le temps passé effectivement dans le train.
- Comme il prend en compte la totalité des trajets concernés, les temps de parcours sont naturellement pondérés par le nombre de voyageurs montant/descendant aux gares considérées. Il est donc représentatif du ressenti « global » de la population étudiée.

Perspectives

Ce tableau de « ressenti usager » sur la régularité des réseaux ferrés peut être généralisé, modulo certaines contraintes :

1. Le modèle est perfectible, notamment pour prendre en compte la représentativité des informations (suppression des valeurs anormalement basses ou élevées, seuil sur le nombre de trajets par tranche pour garantir la représentativité de l'indicateur).
2. Ce travail de généralisation ne peut être fait que si l'on dispose sur la branche considérée d'un nombre suffisant de gares équipées en valideurs. Cela exclut notamment le réseau St Lazare. Par contre cela permet une étude fine, branche par branche pour les RER A et B ou des réseaux comme Montparnasse.
3. Etudier la régularité sur le « trajet retour » est plus délicat si on ne dispose pas de gares fermées en sortie.
4. La généralisation peut être effectuée, sur une période relativement longue pour les réseaux pour lesquels les remontées sont fiables. Ainsi on peut mettre en évidence les évolutions, dans le temps, des temps de parcours moyens et du nombre d'incidents significatifs affectant les usagers.

VI. CONCLUSION

Le STIF a développé un outil très utile pour l'analyse de la mobilité des porteurs de titres de transports franciliens. Cet outil encore jeune ne concrétise pas encore pleinement tout son potentiel d'exploitation. En particulier, on voit bien que les biais qui affectent le système nécessitent en permanence de resituer les données dans leur contexte, de les analyser, et que les exploitations « brutes » sont délicates.

En réalité, l'analyse nécessaire du contexte, le redressement des données manquantes, la comparaison avec des données tierces, incitent à penser que le système décisionnel ne permet pas aujourd'hui de passer totalement outre les traditionnelles enquêtes. Cependant il offre, dans de nombreux cas, la possibilité de se passer de la phase terrain, les données brutes étant déjà remontées et agrégées automatiquement.

Néanmoins, quelle que soit la problématique posée, il convient d'effectuer sur chaque étude un double travail d'analyse :

- En amont de l'extraction des données utiles, afin de constituer un échantillon représentatif de la problématique à étudier (sélection de la période pertinente, quota pour garantir la visibilité des populations de grande couronne dont les validations sont sous-représentées dans le système...).
- En aval de la sélection des validations ou des déplacements de la catégorie de porteurs considérée, afin de reconstituer les déplacements réels (les approximations effectuées par le système décisionnel posant des difficultés pour les déplacements complexes ou lorsqu'une validation est manquante), de redresser les modes sous-représentés, et de corriger les anomalies dues à la non validation sur certaines gares ou certains réseaux.

On le voit, les utilisations du système décisionnel sont potentiellement nombreuses. Néanmoins, ce système nécessite un degré d'expertise élevé, loin de la production « presse-bouton » de rapports imaginée au moment de sa conception.

L'arrivée progressive d'autres validations (gares équipées de CAB M1 SNCF, bornes de validation) va permettre de donner enfin de la visibilité à certaines zones blanches du réseau. L'ensemble des déplacements (origines-destinations) sera progressivement représenté dans le système. Il ne restera plus qu'à être capable d'en ajuster la pondération, notamment en fonction d'une étude sur les taux de validation.

De ce fait, les prochaines étapes vont permettre de sécuriser les données, leur qualité et leur interprétation. Ainsi le travail entrepris sur l'analyse fine des règles de mobilité permet déjà d'imaginer une reconstitution plus fiable des trajets et des déplacements grâce notamment à une interprétation des types de validation et une prise en compte plus adaptée des lieux et modes de validation.

L'ensemble de ces travaux devrait permettre de construire prochainement des indicateurs stratégiques pour le STIF. Ainsi les données de validation seront utilisées afin de déterminer à partir de 2011 la fréquentation des lignes de bus OPTILE dans le cadre des contrats de type 2 ou encore pour constituer un ou plusieurs indicateur(s) de mesure de trafic sur la base des données télibillettiques pour l'intéressement de la RATP et de la SNCF.

Ces prochains pas permettront progressivement de mobiliser l'ensemble des acteurs sur la qualité et la fiabilité de l'information et de faire monter en expertise l'ensemble du STIF sur son interprétation.