



Université de Toulouse

MASTER 2 GEOMATIQUE

« **ScIences Géomatiques en environneMent et AménageMent** » (**SIGMA**)

<http://sigma.univ-toulouse.fr>

RAPPORT DE STAGE

Automatiser, analyser et visualiser les données de Floating Car de la région grenobloise

Vallet Baptiste

Agence d'Urbanisme de la Région Grenobloise



Maître de stage : Elsa Liégeois
Tuteur-enseignant : Laurent Jégou

Septembre 2019

Résumé

L'acquisition de floating car data représente un enjeu majeur pour les études de mobilités. En effet elles permettent d'acquérir des informations sur les temps de parcours moyens ou théoriques sur des itinéraires définis. Cela permet dès lors de proposer un suivi de la congestion sur certains axes clés, ou encore d'estimer l'impact d'un projet sur ces mêmes axes. Les données collectées peuvent servir à objectiver les débats sur la congestion.

La finalité du stage était de permettre l'utilisation de cette donnée neuve dans les projets de l'agence. L'objectif est pleinement rempli par la création de plusieurs outils. Ils offrent la possibilité de réaliser des suivis automatiques et pérennes, des études ponctuelles et de visualiser les indicateurs issus de l'acquisition de ces données. Même s'ils requièrent une certaine rigueur (notamment pour la constitution de la matrice), les programmes restent assez souples pour être adaptés à n'importe quel type d'étude.

Ces outils sont le fruit de nombreuses réunions que ce soit avec ma maître de stage, l'équipe déplacements, les administrateurs de bases de données, le responsable système et réseaux mais aussi les partenaires membres de l'observatoire des déplacements. Cela garantit que les outils répondent bien aux besoins des différents collaborateurs travaillant sur la mobilité, mais aussi que les normes de l'agence quant au stockage des données et à l'utilisation des serveurs soient bien respectées.

Abstract

The acquisition of floating car data represents a major challenge for mobility studies. Indeed, they make it possible to acquire information on average or theoretical travel times on defined routes. This makes it possible to propose congestion monitoring on certain key axes, or to estimate the impact of a project on these same axes. The data collected can be used to objectify debates about congestion.

The purpose of the internship was to allow the use of this new data in the projects of the agency. The goal is fully fulfilled by creating several tools. They offer the possibility of carrying out automatic and permanent follow-ups, ad hoc studies and visualizing the indicators resulting from the acquisition of these data. Even if they require a certain rigor (especially for the constitution of the matrix), the programs remain flexible enough to be adapted to any type of study.

These tools are the result of numerous meetings with my internship supervisor, the travel team, the database administrators, the system and network manager, as well as the members of the travel observatory. This ensures that the tools meet the needs of the various employees working on mobility, and that the agency's standards for data storage and server usage are well respected

Remerciements :

Je tiens à remercier,

ma maitre de stage Elsa Liégeois pour l'opportunité qu'elle m'a offerte de réaliser ce stage. Mais surtout pour son investissement, son aide et sa confiance. Elle m'a aussi fait progresser tant professionnellement qu'humainement.

Mon tuteur Laurent Jegou pour sa disponibilité, son support, ses conseils et nombreux retours pendant ma période de stage.

L'ensemble des collaborateurs de l'Agence d'Urbanisme de la Région Grenobloise pour leur accueil et l'ambiance au bureau. Je remercie aussi plus précisément Fatima Deffous qui a facilité mon intégration à l'agence.

L'ensemble du corps enseignant du Master SIGMA sans qui je n'aurai pu mener à bien cette mission.

SOMMAIRE

L'AGENCE D'URBANISME DE LA REGION GRENOBLOISE	7
<i>Une structure ancienne qui a connu de nombreuses évolutions</i>	7
<i>Une agence organisée en différents secteurs</i> :	9
LA MISSION	11

FLOATING CAR DATA : DEFINITION ET UTILISATIONS 12

I- FLOATING CAR DATA : DEFINITION ET UTILISATIONS	13
A) <i>Les données de Floating Car</i>	13
B) <i>Metrosat l'outil de l'agence d'urbanisme de Nantes</i>	16

LES SOURCES DE DONNEES 21

II- UNE DONNEE DONT LE MARCHÉ EST DÉJÀ BIEN DÉVELOPPÉ	22
A) <i>Des acteurs bien connus</i>	22
B) <i>L'acquisition des données</i>	22
C) <i>Comparaisons des données renvoyées</i>	26

LA CREATION DE L'OUTILS 28

III- LA CREATION DE L'OUTILS	29
A) <i>La conceptualisation de l'outil</i>	29
B) <i>La réalisation de l'outil</i>	30

L'INTEGRATION AUX PROJETS DE L'AGENCE 42

IV- L'INTEGRATION DE L'OUTIL AUX PROJETS DE L'AGENCE	42
A) <i>L'implantation d'un Bus à Haut Niveau de Service (BHNS) dans la commune de Meylan</i>	43
B) <i>Schéma de développement des espaces économiques (SDEE)</i>	43
C) <i>Etude de localisation d'aire de covoiturage</i>	44
D) <i>Schéma de mobilité Nord-Ouest (SMNO)</i>	44
E) <i>Les missions annexes</i>	45

CONCLUSION 47

<i>La finalité de l'outil</i>	48
<i>Un outil encore limité</i>	48
<i>...Mais qui n'a pas encore démontré tout son potentiel</i>	48

RETOUR D'EXPERIENCE 49

SOURCES : 50

<i>Annexe 1 : Note technique (non finalisée) de l'outil développée à l'AURG</i>	52
---	----

<i>Annexes 2 : Exploitation des données de temps de parcours obtenues par le biais de l'outil de requête automatisées développé à l'AURG pour le SDEE.....</i>	<i>63</i>
<i>Annexes 3 : Exploitation des données de temps de parcours du SMTC dans le cadre du BHNS.....</i>	<i>64</i>
<i>Annexes 4 : Exemple de tableau récapitulatif du nombre d'observations dans la base de données de l'AURAN pour une période donnée.</i>	<i>65</i>

Figure 1 : Cartographie des partenaires de l'agence. Réalisation AURG. 2019.....	8
Figure 2: Organigramme de l'Agence d'Urbanisme de la Région Grenobloise. Réalisation : AURG 2019.....	9
Figure 3 :Schéma des erreurs possibles de l'application du script Python aux FCD	15
Figure 4 : Graphiques de la synthèse réalisée par L'AURAN sur les temps de parcours moyen par jour des communes de l'aire urbaine de Nantes vers et depuis Nantes. Réalisation Auran 2018.....	17
Figure 5 : Courbes de la synthèse réalisée par L'AURAN sur l'évolution des temps de parcours moyen des mardis et jeudis des communes de l'aire urbaine de Nantes vers et depuis Nantes. Réalisation Auran 2018.	18
Figure 6 : Tableaux synthétique des indicateurs réalisés par L'AURAN par le biais de leur outil Métrosat. Réalisation AURAN. 2018.....	18
Figure 7 : Cartographies réalisées par l'AURAN pour représenter les indicateurs calculable par le biais de leur outil Métrosat. Réalisation Auran. 2018.	19
Figure 8 : Tableau récapitulatif des arguments de la requête à formuler pour l'API Calculate Route de TomTom. Réalisation Baptiste Vallet 2019.	23
Figure 9: Résultats renvoyés après requête à l' API "Calculate Route" de TomTom	24
Figure 10 : Tableau récapitulatif des arguments de la requête à formuler pour l'API Matrix Distance de Google. Réalisation Baptiste Vallet 2019.	25
Figure 11: Résultats renvoyés après requête à l' API "Distance Matrix" de Google	25
Figure 12 : Tableau comparatifs des données renvoyées par les différentes API. Réalisation Baptiste Vallet 2019.	26
Figure 13: Tableau de comparaison du coût des requêtes pour chaque espace potentiel d'étude. Vallet Baptiste 2019.....	27
Figure 14: Schéma de l'organisation de l'outil développé à l'Agence d'urbanisme de la Région Grenobloise pour exploiter les données de floating car Data	31
Figure 15 : Modèle de Matrice à renseigner pour l'obtention de données de temps de parcours par le biais de l'outil développé à l'agence d'urbanisme de Nantes. Réalisation Baptiste Vallet 2019.	32
Figure 16 : Organigramme des traitements permettant l'acquisition de données FCD à l'agence d'urbanisme de la région grenobloise.....	34
Figure 17 : Capture d'écran du premier panel de SateliteGre présentant le temps moyen de parcours pour se rendre d'une commune de l'aire urbaine de Grenoble à Grenoble pour les mardi et jeudi du mois de mars 2018 hors période de vacances.....	37
Figure 18 : Représentation graphique à travers l'outil SateliteGre des temps de parcours moyens des communes de l'aire urbaine de Grenoble vers Grenoble pour les mardis et jeudis du mois de Janvier 2018 (hormis jours en période de vacances scolaire)	38
Figure 19: Schéma de la synergie d'un onglet développé sous Rshiny: Réalisation Vallet Baptiste 2019.....	39
Figure 20: Organigramme des onglets de cartographies de l'outil SateliteGre. Réalisation Baptiste Vallet 2019. 40	40
Figure 21: Organigramme des onglets de représentation graphique de l'outil SateliteGre. Réalisation Baptiste Vallet 2019	41
Figure 22 : Exploitation des données de obtenues par le biais de l'outil de requête automatisé développé à l'agence pour le SDEE. Réalisation Vincent Escartin 2019.	63
Figure 23: Carte d'exploitation des données du SMTC dans le cadre de l'étude d'un BHNS dans la commune de Meylan. Baptiste Vallet 2019	64
Figure 24 : Extraction d'un tableau récapitulatif du nombre d'observations dans la base de données de l'AURAN pour une période donnée. Réalisation Vallet Baptiste. 2019	65

L'agence d'urbanisme de la région grenobloise.

Une structure ancienne qui a connu de nombreuses évolutions.

L'agence d'urbanisme de la Région Grenobloise (AURG), fondée en 1967, est l'une des premières agences d'urbanisme à avoir vu le jour en France. Elle fait suite à l'adoption de la loi d'orientation foncière, promulgué pour affronter les enjeux d'organisation de l'espace urbain dans un contexte de forte croissance démographique. Au fil du temps, cette structure a évolué et a dû s'adapter pour répondre aux différents rôles qui lui furent enjoins par les différentes politiques d'aménagement du territoire. La fonction première des Agences d'urbanismes aujourd'hui est de permettre aux intercommunalités de profiter d'un organisme d'ingénierie. Cela facilite pour les petites communes la gestion et la réalisation de projets d'envergures.

Le personnel de l'ancien nommée Agence d'Urbanisme de l'Agglomération Grenobloise a ainsi pu travailler sur différents documents d'urbanismes comme :

- Le Schéma de Cohérence Territoriale (SCoT) ;
- Le Plan Local d'Urbanisme (PLU), devenu plus tard le Plan Local d'Urbanisme Intercommunal (PLUI) ;
- Le Schéma directeur de la région grenobloise ;
- Le Contrat de ville d'agglomération ;
- Le Plan de déplacements urbains (PDU) ;
- Le Projet d'agglomération.

Le périmètre d'intervention de l'agence d'urbanisme s'est étendu à la suite de la loi ALUR. Aujourd'hui l'Agence a en charge des études sur 8 Etablissement Public de Communauté Intercommunal (EPCI) et de deux Parcs Naturels Régionaux (PNR).

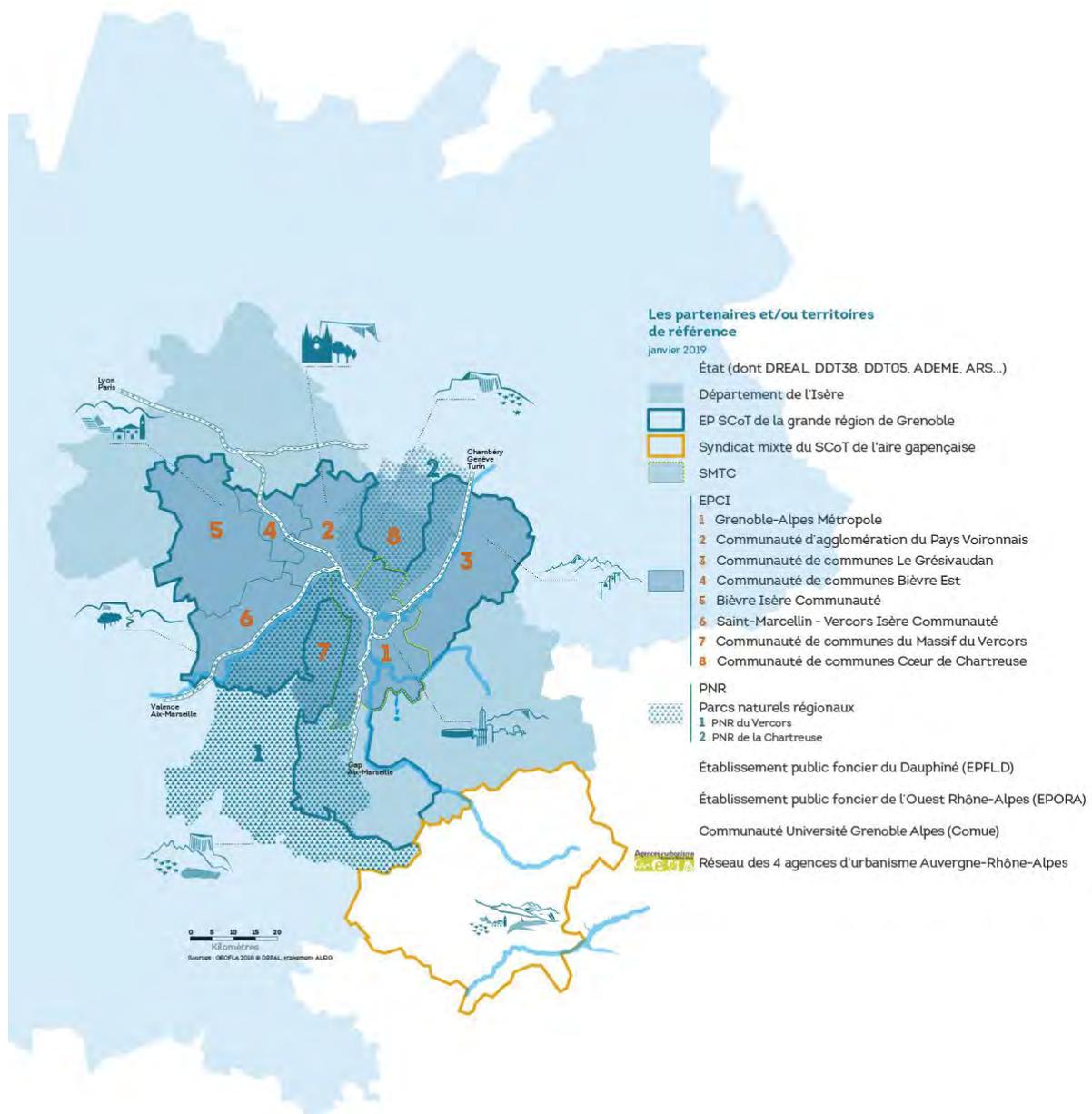


Figure 1 : Cartographie des partenaires de l'agence. Réalisation AURG. 2019

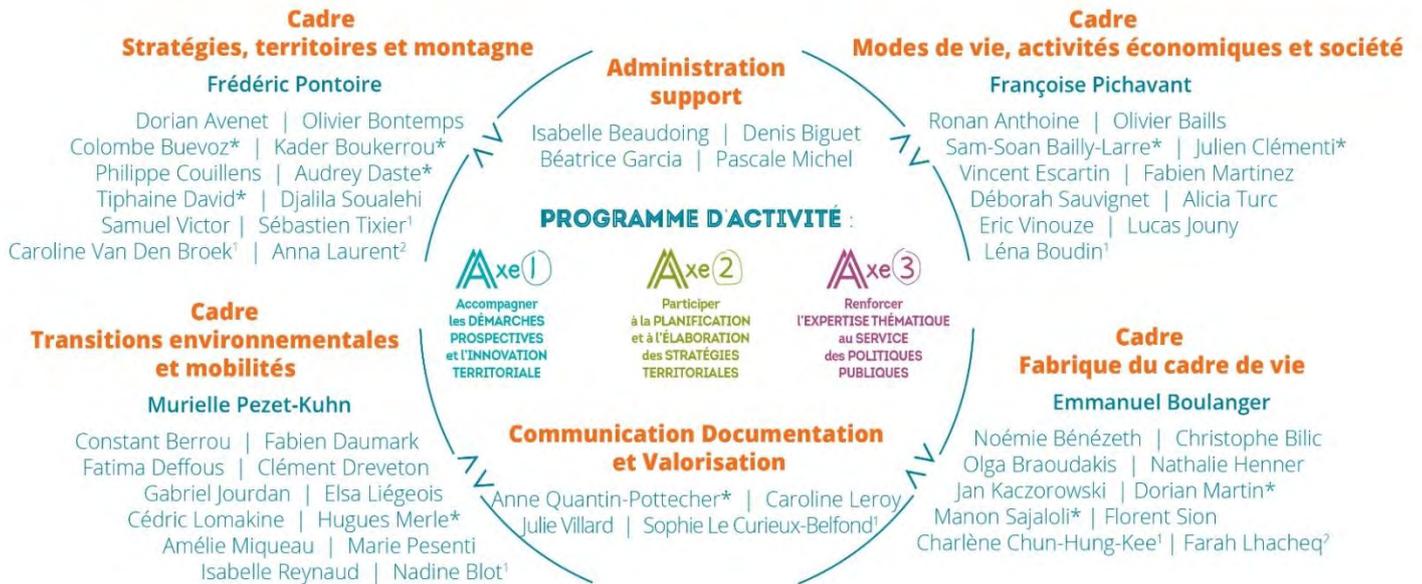
Aujourd'hui, les principales missions qui cristallisent l'organisation de l'agence sont la réalisation des 5 Plui et des 2 SCoT sur les territoires qu'ils couvrent. A cela s'ajoute de nombreux schémas de développement territoriaux, des expertises sur le foncier, le logement, l'énergie, l'environnement ou encore de la prospective à l'aide de modélisation.

L'AURG a aussi tissée de forts liens partenariaux tout au long de son histoire. Elle travaille principalement avec des structures comme la Métropole de Grenoble, l'ADEME, la DDT, la CCI, les communalités ou encore des bureaux d'études. Le réseau urba4 permet aussi la réalisation de projets entre les agences d'urbanismes de Lyon, Clermont Ferrand, St-Etienne et Grenoble. A l'échelon national la Fédération Nationale des Agences d'Urbanisme permet le partage de connaissances par des colloques, offre des aides financières et façonne ainsi par les outils numériques une forte cohésion entre les différentes Agences de France.

DIRECTION GÉNÉRALE

Benoît Parent
Renée Wessels

Florence Duband



* Membres du comité de programme

¹ CDD

² Apprentissage

L'AGENCE
D'URBANISME DE LA RÉGION GRENOBLOISE

Juillet 2019

Figure 2: Organigramme de l'Agence d'Urbanisme de la Région Grenobloise. Réalisation : AURG 2019

Une agence organisée en différents secteurs : Cadres et référents.

L'agence se compose de 60 employés (hors stagiaires et apprentis) aux profils variés. Ma structure d'accueil s'est ainsi dotée d'urbanistes, géomaticiens, démographes, urbanistes, architectes urbanistes administrateurs de base de données et d'un service de communications.

Chaque chargée d'étude de l'AURG se voit affecter un service nommé « cadre » dirigé chacun par un collaborateur nommé référent. Ils sont aux nombres de 4 :

- Stratégies, territoires et montagne
- Modes de vies, activités économiques et sociétés
- Transitions environnementales et mobilités
- Fabrique du cadre de vie

Chaque territoire possède aussi son référent en plus de trois autres nommés pour le SCOT, l'administration de base de données et la gestion réseaux. Leur principal rôle est de servir de personne ressource aux autres agents qui auraient besoin d'informations sur un territoire ou une thématique spécifique.

L'équipe mobilité et l'observatoire du déplacement

Dans le cadre de ce stage j'ai intégré l'équipe mobilité au sein du cadre Transitions environnementales et mobilités. J'étais alors sous la tutelle de la chargée d'étude en mobilité, observation et modélisation : Elsa Liégeois.

L'équipe mobilités est en charge d'études de différents types:

- Définition des politiques de mobilité et schémas d'organisation des offres de mobilité (maîtrise d'œuvre de l'élaboration du PDU)
- Aide à la coopération entre autorités organisatrices de mobilité
- Etudes sur tous les sujets de la mobilité (covoiturage, autopartage, transports collectifs structurants, modes actifs, intermodalité, logistique, espace public, urbanisme-transport, enquêtes mobilité...)

Lancé en 2008, l'Observatoire des déplacements, dont est responsable ma tutrice de stage, est financé par le SMTC (Syndicat mixte des transports en commun de l'agglomération grenobloise), la Direction départementale des territoires, le Département de l'Isère, le Grésivaudan et le Pays Voironnais. La Région Auvergne - Rhône-Alpes et Grenoble-Alpes Métropole apportent des contributions techniques. À travers cet outil commun, les partenaires souhaitent mutualiser et structurer leurs données, en garantir la fiabilité, favoriser des analyses communes et les échanges de bonnes pratiques. Il diffuse des analyses sur les mobilités et leurs évolutions à l'échelle de l'aire d'influence de Grenoble, à la fois en termes d'offre de déplacements, de demande de déplacement et de leurs déterminants.

L'observatoire des déplacements alimente plusieurs outils d'évaluation et de prospective en lien avec la mobilité et s'en enrichit. Ces outils sont principalement :

- Un modèle multimodal de déplacements,
- Un modèle de bruit,
- Des modèles d'émission et de dispersion des polluants atmosphériques.

La mission

Les enjeux de ce stage sont de créer un outil qui permettrait d'analyser et de visualiser les phénomènes de congestions dans l'agglomération grenobloise.

Si le sujet n'est pas novateur la particularité réside dans le type de données utilisées. En effet, les lectrices de plaques automatiques (LAPI) et les boucles magnétiques ayant largement montrées leurs limites, des chercheurs ont essayé d'exploiter d'autres sources de données.

La mission s'appuiera donc sur l'exploitation de Floating Car Data (FCD) issues de GPS. Ces nouvelles sources permettent de récupérer des données sur la vitesse des véhicules ainsi que leur localisation à un intervalle de temps régulier selon le fournisseur.

Le premier axe de l'étude aura pour objectif de réaliser un travail documentaire sur les FCD. Il sera ensuite nécessaire de comparer les différents fournisseurs afin de sélectionner le plus pertinent pour répondre aux besoins de l'observatoire.

Dans un second temps il sera nécessaire de constituer une base de données de temps de parcours et de l'intégrer aux données de l'agence. Il faudra ensuite définir des indicateurs permettant l'étude et l'observation des phénomènes de congestions sur le territoire d'action de l'AURG. De manière technique il est demandé par la mission d'automatiser à moindre frais la récupération de la donnée.

Enfin les études et observations de l'agence sont amenées à être présentées aux partenaires ou inclut dans des rapports. La question de la visualisation et de la représentation des indicateurs composera ainsi le dernier acte de la mission.

| - FLOATING CAR DATA : DEFINITION ET UTILISATIONS

I- Floating Car Data : Définition et utilisations

Avant toute étude il est nécessaire de rendre compte de l'existant sur le sujet. Il sera bon de s'appuyer sur les travaux déjà réalisés pour enrichir et améliorer la compréhension des enjeux de notre sujet d'étude. Il sera également possible de s'appuyer et de confronter les recherches déjà réalisées, à nos futurs résultats pour vérifier, invalider ou encore s'opposer à l'existant.

A) Les données de Floating Car

Les Floating Car Data (FCD) sont des données permettant d'acquérir de l'information sur :

- La vitesse des voitures.
- Le temps de parcours sur un trajet.
- Leur localisation

Si plusieurs méthodes d'obtention existent, deux grands courants se dégagent :

- Le courant Eulérien : les données seront obtenues par le biais de comptages réalisés sur des itinéraires précis. Pour cela sont utilisés des bornes captant les ondes Bluetooth des véhicules étant équipés de tels appareils ou encore les Lecteurs Automatiques de Plaques d'Immatriculation (LAPI).
- Le courant Lagrangien : ici l'obtention de données se fait par l'utilisation de GPS. Ces terminaux peuvent être des solutions professionnelles (GPS de camion fixant les itinéraires de livraisons), des applications (Google Maps) ou des boîtiers (TomTom, Coyote).

La méthode Eulérienne permet d'obtenir de l'information sur une grande partie des véhicules empruntant la voie. La principale faiblesse de ce courant, est son application à un espace réduit, prédéfini par la pose des détecteurs. Pour étudier la totalité d'un réseau, il faudrait donc disséminer sur tout le parcours, des outils de mesure qu'il faudra par la suite entretenir.

A l'inverse la méthode Lagrangienne ; permet d'acquérir de la donnée sur un plus grand nombre de tronçons. Cependant, les informations obtenues seront basées sur les trajets réalisés par les consommateurs des solutions GPS, et uniquement si ces terminaux sont actifs. Ainsi il est possible qu'un tronçon soit référencé, mais avec un nombre d'utilisateurs insuffisant l'ayant pratiqué. Lors d'une conférence au Cotita, des chercheurs ont estimé à 6% le taux de véhicule fournissant des FCD par le biais de terminaux Coyote pour une étude sur le Grand Lyon. Cette solution n'est donc pas préconisée pour les études de charge de flux.

Les deux méthodes ne sont pas antinomiques et au contraire peuvent être jointes pour combler les faiblesses de l'autre.

Guider les FCD jusqu'à la base de données.

Méthodologiquement, les serveurs ne reçoivent en réalité que des semis de points. Un point est créé à intervalles réguliers. Dans la majeure partie du temps le point contient des informations, sur :

- La vitesse de l'automobiliste,
- La position géographique exprimée en coordonnées.

Ces points sont ensuite utilisés pour retracer l'entièreté de la trace GPS du conducteur (méthode dites de « collier de perles »).

Force et faiblesse d'une donnée nouvelle

Les apports de cette donnée

Plusieurs études se sont déjà attelées à la comparaison des FCD issues de GPS à celles obtenues par des boucles magnétiques (Cerema, Cotita). Dans différentes analyses, il apparaît que les valeurs répertoriées dans les bases de données des fournisseurs, sont similaires à celles calculées par les bornes de détections. La méthode semble aussi moins onéreuse. Dans l'étude présentée au Cotita sur le grand Lyon, les requêtes TomTom ont été évalué à 20€ plus chère que la pose de borne Bluetooth sur le même espace. Cependant ils ne prennent pas en compte l'entretien, ce qui pourrait entraîner un coût final supérieur des détecteurs. L'un des avantages notables des données GPS est la possibilité d'acquérir de la donnée auprès de différents fournisseurs. Cette pluralité d'offres permet d'obtenir des informations servant à divers objectifs tels que la représentation graphique sur plateforme web, le développement d'application ou encore la création de Base de Données.

Il est possible de récupérer des données de vitesses sur des plages horaires définies, sur plusieurs jours ou dans le passé, ce qui permet des études diachroniques. Les indicateurs fournis divergent entre chacun des fournisseurs.

Ces jeux de données intéressent de plus en plus les métropoles. Comme la méthode d'acquisition est similaire sur tous les territoires pour un même fournisseur, cela facilite les comparaisons entre espaces.

Les principaux apports de cette donnée sont donc :

- Le coût,
- Une méthodologie homogène sur tous les espaces,
- Une grande diversité d'indicateurs et de données possibles.

Cependant comme toute méthode d'analyse certaines réserves peuvent être émises.

Une donnée présentant encore des biais

Les études réalisées sur l'utilisation de cette nouvelle source de données ont déjà pointé les différents biais qui risqueraient de fausser l'interprétation des résultats.

La méthode d'acquisition des positions GPS est la source du premier biais. Dans le cas où, un véhicule se situerait dans une zone faiblement couverte par le réseau, des erreurs de géolocalisation peuvent apparaître. Pour cela un script python est utilisé pour relocaliser sur la route le véhicule. La localisation de l'automobile est alors affectée à la voie dont la distance perpendiculaire et la plus courte. Cette réaffectation a un impact direct sur la reconstitution des itinéraires, notamment lorsque les voies se situent à proximité. En effet le point peut être ainsi géocodé sur la mauvaise voie. Des écarts peuvent alors apparaître entre le trajet réellement réalisé et le trajet renvoyé par le script sur le serveur.

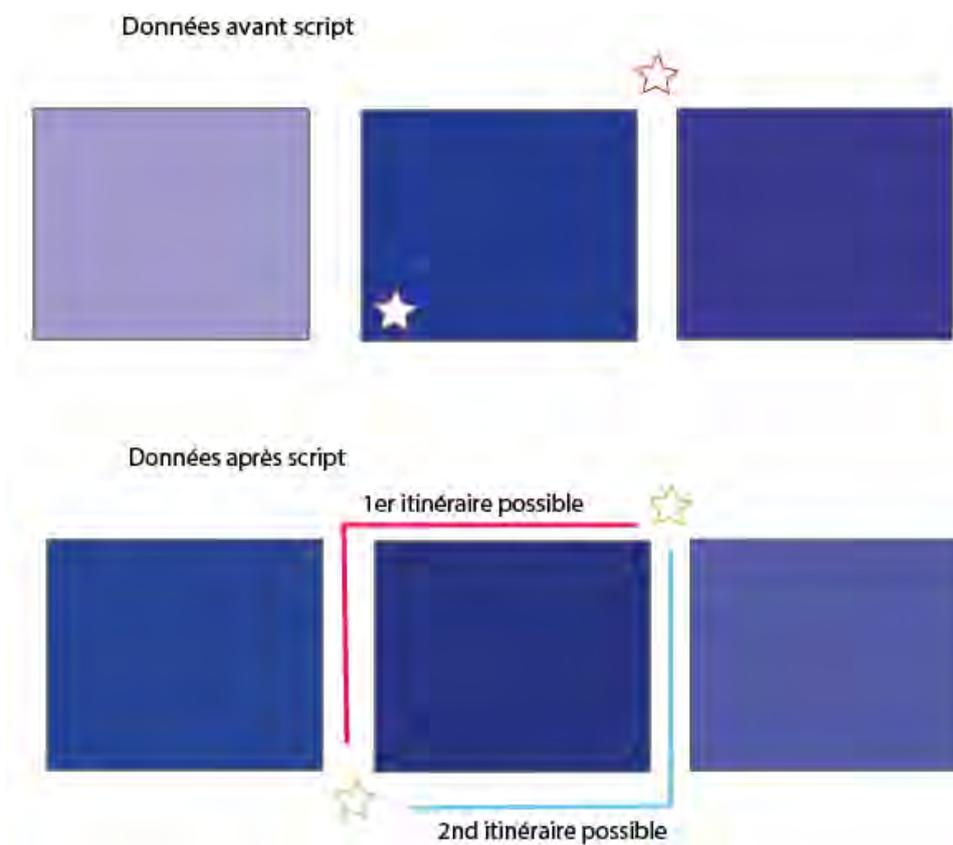


Figure 3 : Schéma des erreurs possibles de l'application du script Python aux FCD

Les différentes recherches en lien avec les FCD ont mis en exergue les faiblesses d'une analyse ne s'appuyant que sur ce type de donnée. En effet, étudier la congestion routière au seul prisme de la vitesse des véhicules peut engendrer de mauvaises analyses. Un conducteur peut, de lui-même décider de rouler lentement ou y être contraint par la météo. Certains axes peuvent souffrir de congestion, uniquement lors de périodes ponctuelles, comme les jours de matchs de foot ou de spectacles. Dans le même ordre d'idée, les manifestations (grèves SNCF, gilets jaunes) peuvent aussi altérer la vitesse ou le nombre de conducteurs.

De plus, il ne faut pas négliger le fait que les floating car data traitent d'informations personnelles. Même si les données des fournisseurs répondent aux normes RGPD. Dans le cas où une succession de tronçons possèderaient de faibles effectifs ; il deviendrait possible d'identifier les personnes en questions dès lors que l'on a accès à des informations complémentaires telles que : leur domicile, leur lieu de travail ou de loisir. Il devient alors possible d'identifier l'itinéraire d'un individu en particulier. Si certains fournisseurs comme TomTom fournissent les effectifs par tronçons, d'autres à l'exemple de Google Maps ont décidé de ne pas les délivrer dans leurs fichiers.

Une donnée en cours évolution

Ces faiblesses ont amené à repenser l'utilisation des Floating Car Data et à trouver des méthodes pour améliorer leurs qualités. Une nouvelle donnée, naquit de ces réflexions : les Extended Floating Car Data (XFCD).

Ces données utilisent les capteurs des véhicules ou des sources de données externes (bases de données météorologiques, etc...) pour apporter des informations supplémentaires comme :

- La Météorologie (pluie, brouillard),

- Des indications sur la Vitesse :
 - Accident
 - Arrêt du conducteur
 - Paramètre des régulateurs, limiteurs
- Volume de trafic,
- Contexte social.

Les Floating Car Data permettent donc d'étudier :

- Les temps de parcours de plusieurs itinéraires,
- Des études diachroniques (avant et après la réalisation de projet de mobilité),
- Les zones de congestions et leurs fréquences,
- Les déviations d'itinéraire.

Cependant seules, elles ne permettent pas :

- D'évaluer les charges de trafic,
- De pouvoir modéliser des flux,
- D'expliquer tous les phénomènes de ralentissement.

Cette donnée continue d'évoluer encore aujourd'hui. Ce qui n'est pas étudiable avec les FCD seules, peut le devenir par l'enrichissement à l'aide d'autres bases de données ou encore par l'augmentation du parc utilisateur.

B) Metrosat l'outil de l'agence d'urbanisme de Nantes.

La synthèse réalisée par L'agence d'Urbanisme de la Région Nantaise (Auran) développe la méthodologie mise en place pour la constitution de leur outil Métrosat. L'étude de ce document permet de faire ressortir les différentes analyses qui sont réalisables à partir d'une base de données constituée par l'aide de FCD.

L'objectif de leur outil est réalisé des études diachroniques sur les temps de parcours pour l'ensemble de l'aire urbaine.

Pour la mise en œuvre de Métrosat, l'agence utilise les données de Floating Car fournies par Google Maps à travers des requêtes API. Les requêtes ont été réalisées quotidiennement à différentes heures. L'agence récupère ainsi des données à 3h et toutes les demi-heures entre 7h et 10h et 16h30 et 19h30. Les itinéraires étudiés prennent leurs points d'origines au centre des 113 communes de l'aire urbaine nantaise et se dirigent en direction du centre de Nantes et inversement. Les requêtes sont réalisées depuis novembre 2016.

A partir des fichiers fournis par Google Maps, 4 indicateurs peuvent être réalisés :

- Le temps de référence,
- Le temps réel,
- L'écart relatif entre le temps de référence et le réel,
- L'écart absolu entre le temps de référence et réel.

De leurs analyses, ressort une congestion plus élevée le mardi à 8h et 8h30 en direction de Nantes et à 17h30 et 18h le vendredi dans le sens inverse.

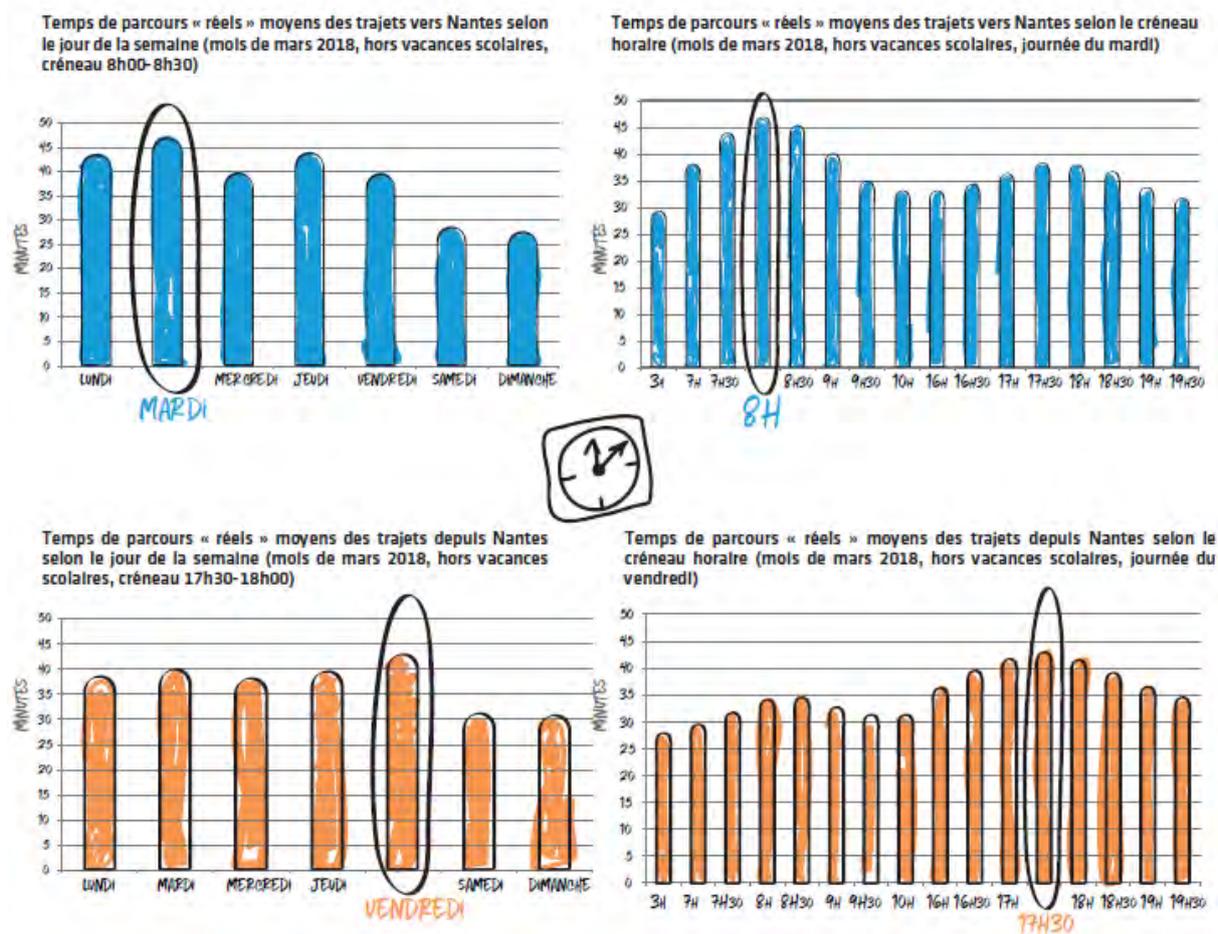
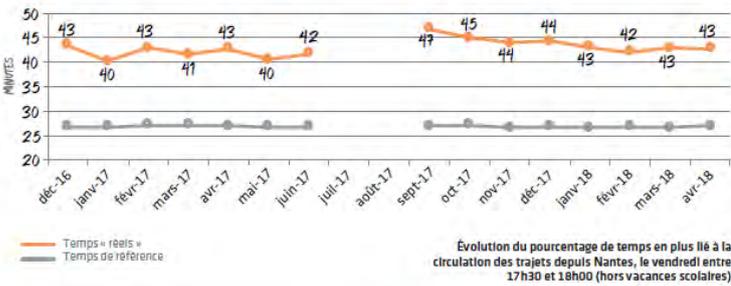


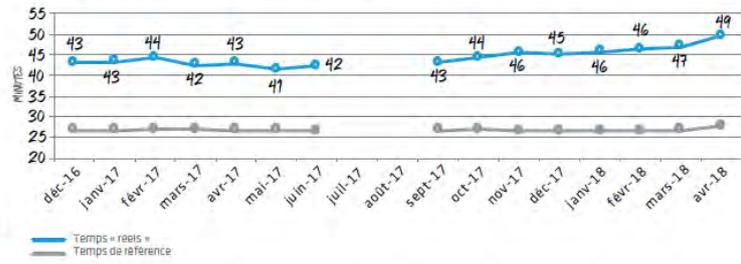
Figure 4 : Graphiques de la synthèse réalisée par L'AURAN sur les temps de parcours moyen par jour des communes de l'aire urbaine de Nantes vers et depuis Nantes. Réalisation Auran 2018

D'un point de vue diachronique les augmentations de temps de parcours sont en corrélation avec une augmentation générale de la charge de trafic. Si le mardi concentre des mobilités principalement pendulaires, le vendredi connaît en plus de ces mobilités des départs en weekend et des sorties loisirs qui sont difficiles à anticiper. Le second jour de la semaine voit une augmentation de ses temps de parcours. En revanche, celui précédent le week-end voit une légère baisse de ses temps de parcours dans le temps.

Évolution des temps de parcours « réels » moyens, et des temps de référence des trajets depuis Nantes, le vendredi entre 17h30 et 18h00 (hors vacances scolaires)



Évolution des temps de parcours « réels » moyens, et des temps de référence des trajets vers Nantes, le mardi entre 8h00 et 8h30 (hors vacances scolaires)



Évolution du pourcentage de temps en plus lié à la circulation des trajets depuis Nantes, le vendredi entre 17h30 et 18h00 (hors vacances scolaires)



Évolution du pourcentage de temps en plus lié à la circulation des trajets vers Nantes, le mardi entre 8h00 et 8h30 (hors vacances scolaires)

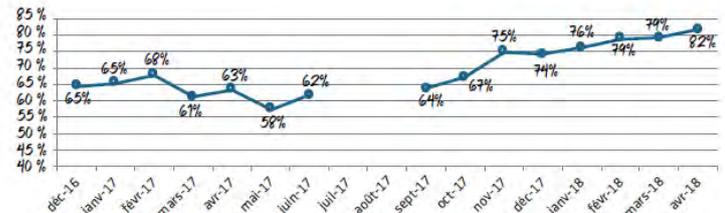


Figure 5 : Courbes de la synthèse réalisée par L'AURAN sur l'évolution des temps de parcours moyen des mardis et jeudis des communes de l'aire urbaine de Nantes vers et depuis Nantes. Réalisation Auran 2018.

En chiffre, cela correspond pour les mardis du mois de mars, à 47 minutes en moyenne pour rejoindre le centre de Nantes en période de congestion, soit 20 minutes de plus que le temps référence de Google. Les habitants des différentes communes de l'aire urbaine, mettent en moyenne 79% de temps en plus pour réaliser leurs trajets. Le vendredi, le temps moyen de parcours est de 43 minutes, l'écart avec le temps de référence est donc de 16 minutes. Du centre des communes de l'aire urbaine jusqu'au centre de la métropole il faut en moyenne 64% de temps en plus pour réaliser les trajets. Les périodes de vacances scolaires voient des chutes drastiques des temps de parcours.

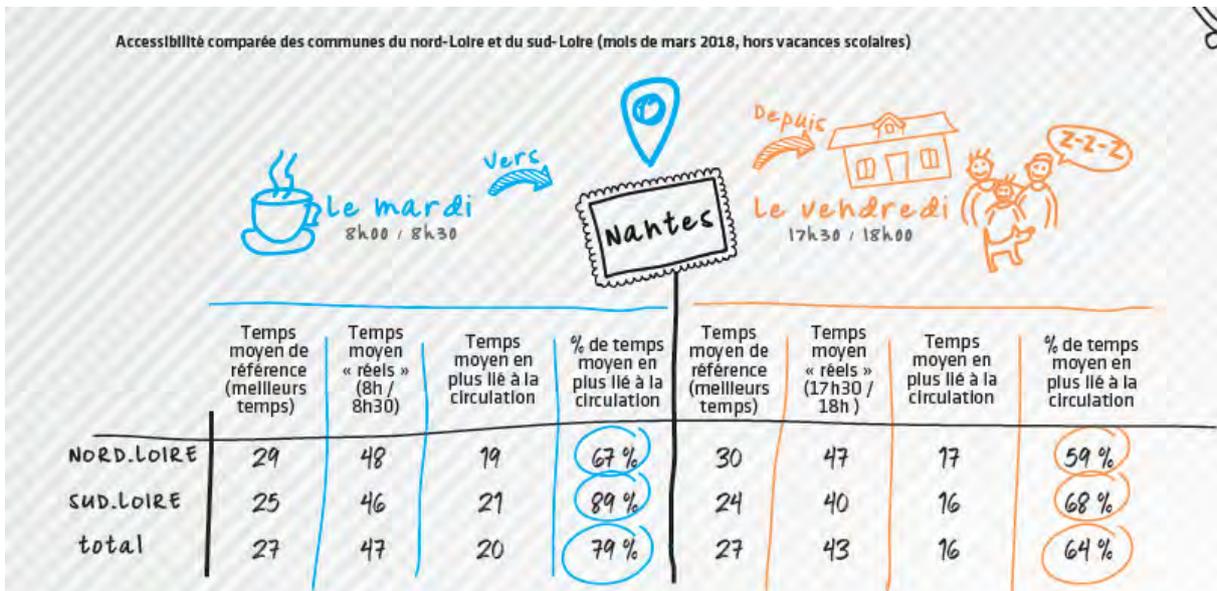


Figure 6 : Tableaux synthétique des indicateurs réalisés par L'AURAN par le biais de leur outil Métrosat. Réalisation AURAN. 2018

Spatialement ce sont les communes du sud-Loire qui subissent les plus gros écarts entre les temps de parcours réel et les temps de références. Enfin les communes desservies par les axes structurants tels que

les RN137, 165, la RD 723 ainsi que l'A11 voient aussi leurs temps de parcours progressés fortement lors des périodes congestion.

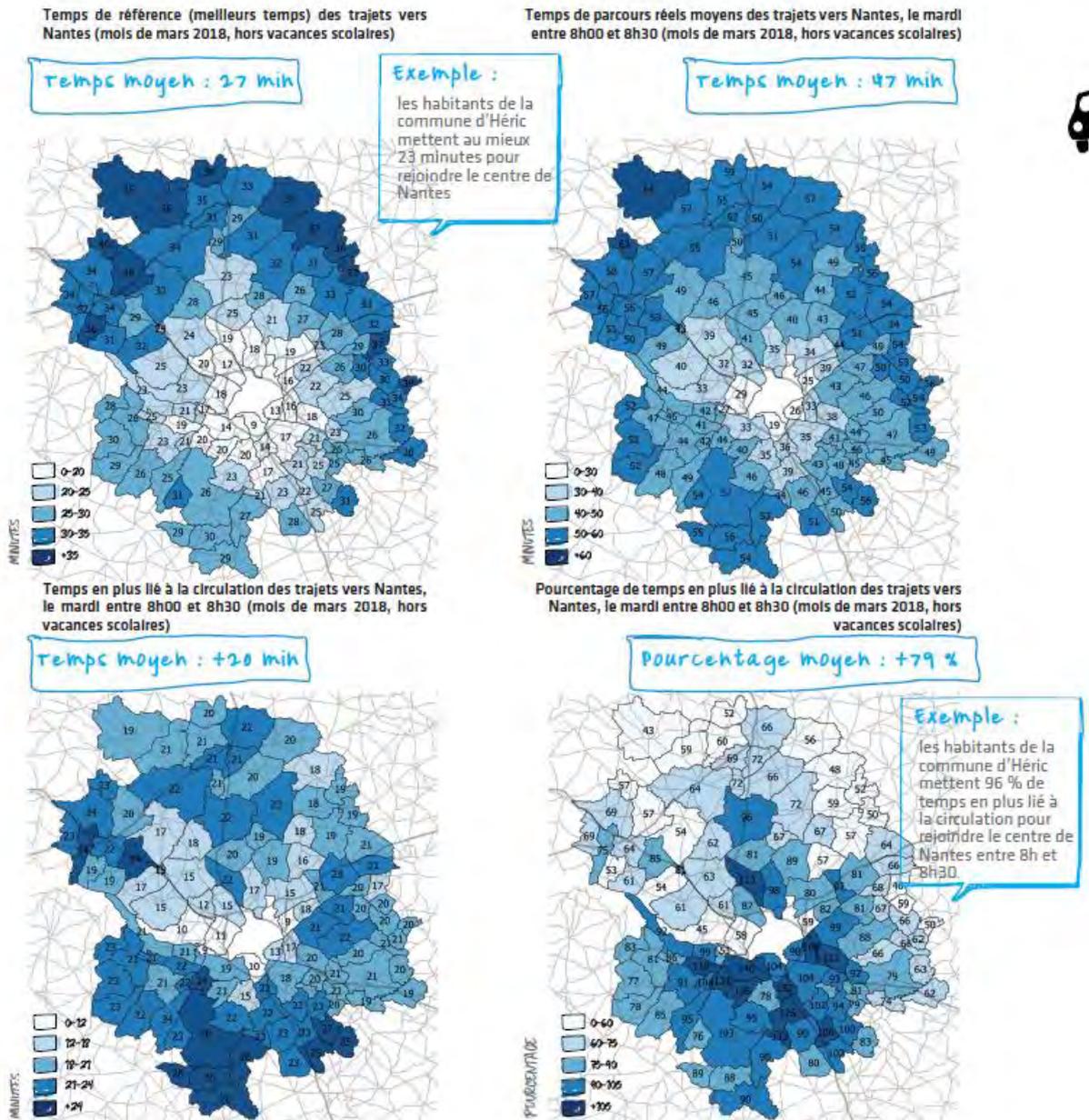


Figure 7 : Cartographies réalisées par l'AURAN pour représenter les indicateurs calculable par le biais de leur outil Métrosat. Réalisation Auran. 2018.

Une deuxième analyse est réalisée pour laquelle la métropole nantaise est divisée en 16 pôles structurants de nature diverses (Emplois, commerces, universitaires, etc...)

Si les temps de références pour rallier les divers points ne présentent que peu d'écart (de l'ordre de 5min au maximum), les résultats en temps réel se font plus disparates. Les lieux centraux sont ceux qui nécessitent le plus de temps à rejoindre. A l'inverse les pôles sud sont les espaces les plus facilement accessibles en termes de temps et de parcours. Le nord-ouest est plus impacté le matin par les conditions de circulations. A l'inverse, le soir se sont les espaces centraux qui subissent le plus, les effets de la congestion.

Les résultats retournés s'expliquent principalement par de fortes congestions sur les axes de franchissement de la Loire. Les mouvements de grève de la SNCF, lors de la période d'étude, ont aussi eu un impact non négligeable sur les mobilités. Les trains n'étant pas en service un report modal s'est effectué sur la voiture alourdissant ainsi la charge de trafic des axes routiers.

||- LES SOURCES DE DONNEES

II- Une donnée dont le marché est déjà bien développé

A) Des acteurs bien connus.

Ayant mieux cerné ce qu'était les données de flottes de véhicules il m'était ensuite nécessaire de rendre compte du marché de celles-ci.

Toutes entreprises proposant des services de navigation ou des solutions GPS est potentiellement un fournisseur de FCD. Ici nous ne nous appuyons que sur les deux principaux acteurs de ce marché en France :

- TomTom
- Google Maps (et par rachat Waze)

Plusieurs points ont motivé ce choix. Tout d'abord les deux fournisseurs sont reconnus dans ce domaine. Comme présenté en début de rapport, des jeux de données issus de ces plateformes ont déjà été utilisés en France. Enfin leur base utilisateur en France est assez développé pour garantir une certaine fiabilité de la donnée.

A savoir que l'étude s'appuyait à l'origine sur quatre délivreurs. La société Here, possède depuis un partenariat en 2017 des données FCD des boitiers Coyote. Le groupe INRIX qui présente chaque année un classement mondial des villes les plus congestionnées. Néanmoins ces fournisseurs ont été rapidement mit de côtés. Pour le premier les conditions générales d'utilisation ne permettaient pas de réaliser une base de données à but analytique. Pour le second même si la base utilisateur est très développée aux Etats-Unis il est difficile d'évaluer la part d'utilisateurs français.

B) L'acquisition des données.

La méthode d'acquisition des données est similaire chez les deux fournisseurs. Chacun possède une ou plusieurs API qui offrent la possibilité de formuler des requêtes. Pour toute acquisition de données il faudra préciser : l'espace d'étude (une aire ou un itinéraire) et la temporalité du jeu de données. La donnée renvoyée prend la forme de temps et de distance de parcours sur l'itinéraire défini.

Si les méthodes sont similaires en théorie, il y a en réalité de grandes divergences lors de la pratique.

TomTom :

La plateforme TomTom regorge d'outils pour étudier le trafic routier. Ne serons présentés ici que les 3 principaux.

Le premier nommé Custom Times Travel (CTT) renvoie des informations sur un itinéraire spécifique. L'itinéraire ne peut dépasser les 70km. Les points d'origines et de destinations sont décrits par des coordonnées, l'adresse postale ou un curseur placé sur leur carte. Il est possible de définir jusqu'à 3 points intermédiaires permettant l'étude de détours.

Le second Custom Area Analysis (CAA) est similaire, si ce n'est sous sa forme. Ici la requête renverra des tronçons d'une longueur variant entre 1 à 400m sur une aire définie lors de la commande par : les coordonnées du coin inférieur gauche et supérieur droit. Il est ensuite possible d'indiquer la hiérarchie maximale des routes retenues (ceci peut avoir un impact non négligeable sur le prix final).

Pour les deux outils suscités, il sera ensuite nécessaire de définir la temporalité de l'étude. Un maximum de 7 plages horaires peut être défini. Le pas de temps s'échelonne du quart d'heure à l'heure. Les données renvoyées seront alors des moyennes sur les intervalles de temps définis.

Avant validation de la commande, il est fourni un tableau des effectifs qui ont permis de calculer les indicateurs. Si le nombre de véhicule est jugé insuffisant, il est alors possible d'annuler gratuitement la commande.

Dans le cas où la requête formulée est validée, des agents des services de la firme hollandaise vont se charger d'acquérir les données (cela pouvant prendre plusieurs mois)

La donnée livrée prend la forme suivante :

- Un fichier Shape contenant les géométries (selon l'année demandée dans la requête est utilisée une base datant soit de 2016, soit de 2018) ;
- Un fichier Excel pour chaque plage horaire définies lors de la requête (pouvant être joint au Shape par l'identifiant de la colonne ID, identifiant variant à chaque commande selon le nombre de tronçons sélectionnés) ;
- Un dossier par jour.

Il est à savoir que le fichier de géométrie, dans le cas d'une commande CAA, superpose deux tronçons (un par sens de circulation) qu'il est ensuite possible de décaler avec l'outil de rendu de style « décaler » de Qgis. Il n'est pas nécessaire de l'appliquer pour traiter les voies rapides. Dans ces cas-là les deux géométries par sens de circulation ne sont pas superposées.

Le dernier : TomTom Calculate Route prend une tout autre forme. Il suffit d'entrer dans une requête API (url web) les coordonnées GPS du point de départ et de destination (possibilité d'intégrer des points intermédiaires) ainsi que l'heure à laquelle doit être faites la requête. Ici le fichier est renvoyé automatiquement et immédiatement par le serveur sous la forme d'un fichier json ou XML.

<i>Paramètre</i>	<i>Méthode de renseignement</i>
<i>L'URI</i>	https://api.tomtom.com/routing/1/calculateRoute/
<i>Le type de fichier en sortie</i>	Json? pour un fichier Json Xml? pour du xml
<i>Point d'origine et de départ séparé par :</i>	Origine :destination
<i>La clé API</i>	key=
<i>Une date de départ</i>	departureat=

Il est nécessaire d'utiliser l'opérateur « & » pour séparer chacun des arguments.

Figure 8 : Tableau récapitulatif des arguments de la requête à formuler pour l'API Calculate Route de TomTom. Réalisation Baptiste Vallet 2019.

Exemple de réponse à la requête :

```

formatVersion: "0.0.12"
routes:
  0:
    summary:
      lengthInMeters: 10664
      travelTimeInSeconds: 2009
      trafficDelayInSeconds: 0
      departureTime: "2019-08-29T09:27:56+02:00"
      arrivalTime: "2019-08-29T10:01:24+02:00"
      noTrafficTravelTimeInSeconds: 1509
      historicTrafficTravelTimeInSeconds: 2009
      liveTrafficIncidentsTravelTimeInSeconds: 2009
    legs:
      0:
        summary:
          lengthInMeters: 10664
          travelTimeInSeconds: 2009
          trafficDelayInSeconds: 0
          departureTime: "2019-08-29T09:27:56+02:00"
          arrivalTime: "2019-08-29T10:01:24+02:00"
          noTrafficTravelTimeInSeconds: 1509
          historicTrafficTravelTimeInSeconds: 2009
          liveTrafficIncidentsTravelTimeInSeconds: 2009
    points:
      0:
        latitude: 45.21241
        longitude: 5.77551
      1:
        latitude: 45.21244
        longitude: 5.77556
      2:
        latitude: 45.21273
        longitude: 5.77594
      3:
        latitude: 45.21283
        longitude: 5.77601

```

Figure 9: Résultats renvoyés après requête à l'API "Calculate Route" de TomTom

Google

La firme Google possède une plateforme nommée Google Cloud Platform. Elle contient tous les outils disponibles pour le développement web et l'intégration des applications Google sur les sites. Dans ces outils deux nous intéresseront ici : « Matrix Distance » et « Direction ».

Dans les deux cas il est nécessaire d'avoir un compte sur Google Cloud Platform, et réalisée une requête API. Cependant les variables demandées sont différentes selon l'outil. La principale différence entre les deux et le fait que le premier outil ne renvoie pas de données de géométries relative aux parcours réalisés.

Pour l'outils Matrix :

Paramètre	Méthode de renseignement
L'URI	https://maps.googleapis.com/maps/api/distancematrix/
Le type de fichier en sortie	Json? pour un fichier Json Xml? pour du xml
L'unité de mesure	units=Imperial ; pour s'appuyer sur le système d'exploitation units=Meters ; pour des distances en mètre
Point d'origine renseigné par des coordonnées GPS	origins=
Point de destination renseigné par des coordonnées GPS	destinations=
La clé API	key=
Une date de départ exprimée en seconde depuis le 1 ^{er} Janvier 1970 à minuit	departure_time=

Il est nécessaire d'utiliser l'opérateur « & » pour séparer chacun des arguments.

Figure 10 : Tableau récapitulatif des arguments de la requête à formuler pour l'API Matrix Distance de Google. Réalisation Baptiste Vallet 2019.

Pour ce qui est de l'outil direction, les mêmes paramètres sont nécessaires. A la différence que les paramètres « origins » et « destinations » ne présentent plus de « s » pour leur appel dans la requête. Cet outil permet aussi d'intégrer des points d'étapes renseignés comme suit :

- waypoints=optimize:true|45.181514%2C5.725737
 - o optimize correspond à un booléen qui permet ou non selon le choix d'optimiser le tracé des parcours.

Ces paramètres restent optionnels et ne peuvent être qu'au nombre de 3 maximum. Il est aussi possible de renseigner le trajet à l'aide d'une polyligne encodée selon la méthode Google.

Exemple de réponse à la requête :

```

▼ destination_addresses:
  0: "10 Rue de l'Industrie, 38640 Claix, France"
▼ origin_addresses:
  0: "E713, 38340 Voreppe, France"
▼ rows:
  ▼ 0:
    ▼ elements:
      ▼ 0:
        ▼ distance:
          text: "13,7 miles"
          value: 22100
        ▼ duration:
          text: "16 minutes"
          value: 953
        ▼ duration_in_traffic:
          text: "29 minutes"
          value: 1730
        status: "OK"
    status: "OK"

```

Figure 11: Résultats renvoyés après requête à l'API "Distance Matrix" de Google

C) Comparaisons des données renvoyées

Pour chaque fournisseur l'offre de données diffère. Il était donc nécessaire de dresser le bilan de l'offre et de sélectionner celui qui offrirait le jeu de données le plus à même de répondre aux besoins de l'agence.

Fournisseurs	TomTom (Ana aire) (SMTC)	TomTom Calculate Route (AURG)	GM Matrix distance (AURAN)
Objets observés	Tronçons fins par sens sur une aire	Géométrie parcours	Pas de géométrie
Finesse Horaires	Maximum de 7 plages horaires découpées en quarts d'heures	A l'heure définie dans la requête	A l'heure définie dans la requête
Temporalité de la requête	Choix d'une liste de jours et des plages horaires ; Passé (situation réelle)	Instantané; Futur (estimation)	Instantané; Futur (estimation)
Attributs renvoyés	Longueur tronçon	Distance du parcours (m)	Distance du parcours (m)
	TP moyen/médian	Temps de trajet historique	Temps de trajet référence
	Ratio tp moyen/tp réf	Temps de trajet dans le trafic	Temps de trajet dans le trafic
	Vitesse limite	Temps de trajet en écoulement libre	
	Vitesse moyenne	Heure de départ	
	Vitesse du 15e percentile, 15% des vitesses sont au-dessus de cette valeur sur la période étudiée	Heure d'arrivée	
	Vitesse du 85e percentile, 85% des vitesses sont au-dessus de cette valeur sur la période étudiée		
	Nombre de voitures référencées		

Figure 12 : Tableau comparatifs des données renvoyées par les différentes API. Réalisation Baptiste Vallet 2019.

Le tableau permet ainsi de mettre en valeur les grandes différences entre l'outil CAA et les deux API web et que la plateforme Google est celle qui fournit le moins de données. De plus les temps de trajets référence et historique sont des temps dont la méthodologie même sur les documentations des API reste obscurs. Dans les deux cas que ce sont des moyennes des temps de parcours pour l'itinéraire. Cependant la période sur laquelle est calculée cette moyenne n'est pas indiquée. Il devient ainsi difficile d'exploiter un temps qu'on ne peut réellement expliquer.

Étude comparative des prix

Le facteur coût est toujours un paramètre important dans le choix d'une nouvelle source de données. Il est donc bon de comparer le coût monétaire des requêtes de chacun des outils.

Les méthodes CAA et CTT ne seront pas traitées. Ces outils n'ont pas vocation à fournir des données avec un fort taux de fréquence. De plus le coût dépend expressément de la requête formulée. Il est donc sujet à de nombreuses variations.

Les calculs sont basés sur les informations de paiements présentes dans la documentation de l'API TomTom Calculate Route et la facture envoyée par Google tous les mois aux personnes inscrite sur sa plateforme.

	TomTom Calculate Route	Google Matrix Distance
Sud Isère :	387,5€	8162€
Aire urbaine :	62€	1735€
Métropole :	0€	217€

Figure 13: Tableau de comparaison du coût des requêtes pour chaque espace potentiel d'étude. Vallet Baptiste 2019.

Le choix du fournisseur

Après comparaison des prix et du contexte il a été décidé de former la base de données autour d'un tri type.

L'agence continuera à faire des demandes de données à l'agence d'urbanisme de Nantes pour les requêtes réalisées sur l'aire urbaine de Grenoble.

Des requêtes quotidiennes seront réalisées par l'agence d'urbanisme de Grenoble pour couvrir des tronçons plus fins par le biais de l'API TomTom Calculate Route. L'objectif sera ici d'analyser la congestion sur des tronçons à fort enjeu.

Enfin, ponctuellement seront formulées des requêtes à TomTom sur l'API Custom Analysis Area. Ces données seront utilisées pour répondre aux besoins des projets.

III - LA CREATION DE L'OUTILS

III- La création de l'outils

A) La conceptualisation de l'outil.

Un outil pour quoi ?

L'objectif de la mission est de pouvoir créer une base de données de temps de parcours afin de pouvoir étudier la congestion routière sur le territoire grenoblois. Ces temps de parcours seront utilisés dans le cadre de l'observatoire des déplacements, de la modélisation et des diverses études de l'agence :

- Les niveaux de congestion sur les axes structurants
- L'évolution de la congestion
- La fiabilité des temps de parcours
- Amélioration de la finesse du modèle de déplacement grâce à la comparaison des données recueillies avec les résultats du modèle lors de son recalage.

Ces analyses répondent à un enjeu d'objectivation des débats sur la congestion routière.

Au début de mon stage nous nous sommes rendus avec ma tutrice à un colloque organisé par la FNAU. Lors de ce colloque les travaux autour du projet Métrosat de Nantes ont été présentés. Il a ainsi été possible de rencontrer les développeurs de l'application. De cette rencontre a été signée une convention entre les deux agences. Cet accord définit les règles de partage et d'utilisation de la base de données que constituait l'AURAN sur les données de temps de parcours allant des communes de l'aire urbaine de Grenoble vers Grenoble et inversement. Cette base de données nous permet de réaliser des analyses d'accessibilité territoriale.

En plus de la base de données, les agents de Nantes ont fourni une notice décrivant le programme permettant de réaliser les requêtes.

Un outil pour qui ?

L'outil développé aurait pu être mis à disposition de trois cibles :

- Les chargés d'études du pôle mobilité de l'agence ;
- Les partenaires de l'agence ;
- Le grand public.

Après discussion avec le directeur de l'Agence d'urbanisme de la région Grenobloise. Il a été décidé qu'en premier lieu il ne serait utilisable que par le premier groupe, mais qu'il était nécessaire d'en parler aux partenaires. Cependant la mise à disposition des données ne se fera qu'après une période d'appropriation. Cette phase visera à mieux maîtriser les données et en évaluer la pertinence (aléas liés à la collecte, compréhension des résultats). La communication et la publication des analyses de la congestion devra faire l'objet de toutes les précautions nécessaires pour un sujet éminemment sensible politiquement. Cependant ces chiffres

Du fait de ces choix les chargés d'études et les partenaires ont pu apporter leur avis lors de différentes phases de la réalisation de l'outil (choix des indicateurs, itinéraires) réfléchis avec ma tutrice. En effet les agents de l'agence ont été sollicités à travers quatre réunions brainstorming et des points individuels réalisés avec chacun. Les partenaires de l'agence ont pu se prononcer à travers deux cotech modélisation auxquels les avancées du stage ont pu être présentées.

Les prémices de la conception

L'outil constitué à l'agence d'urbanisme de la Région Grenobloise s'est donc grandement inspiré de celui développé à l'agence d'urbanisme de Nantes. Plusieurs points fondamentaux divergent cependant :

- Le fournisseur de donnée permettant de constituer la base de données n'est plus Google (API Matrix OD) mais TomTom (API TomTom Calculate Route). Comme dit précédemment la donnée de TomTom permet d'obtenir de l'information sur le tracé des itinéraires recommandés par le GPS. Dans certaines conditions il devient possible d'analyser des temps de parcours.
- L'AURAN utilise pour ses calculs d'indicateurs un temps référence correspondant au temps en écoulement libre (requête réalisée à 3h du matin). Cependant après plusieurs points avec les agents de l'observatoire du déplacement, il apparaît que le temps à vide représente un temps fictif très rarement réalisable en journée. Ce choix a donc tendance à surestimer les écarts de temps en période de congestion, ce qui n'est pas souhaitable dans une démarche scientifique d'objectivation des temps de parcours. De ce fait l'outil développé lors de ce stage ne possède pas de temps référence mais un temps à vide et un temps en heure creuse (moyenne des requêtes réalisées à 15h pour un itinéraire).
- La troisième différence se situe dans le choix des itinéraires. Si dans le cas de l'agence de Nantes les requêtes sont réalisées de communes à communes. Le principal objectif de la base de données est d'étudier l'impact de la congestion sur les principaux axes routiers de la métropole. De ce fait il paraît plus pertinent de requêter des tronçons plus restreints mais avec une forte propension à être congestionnés.
- Enfin l'application de l'AURAN a été développée en C++ sous linux alors que celle de l'AURG a été développée sous R avec un système d'exploitation Windows. Le script ne peut donc être récupéré tel quel mais doit être adapté aux besoins et au système d'exploitation

B) La réalisation de l'outil.

Les outils de requêtes automatisées et manuelles

D'un point de vue technique les chargés d'étude de l'observatoire des déplacements étant plus des thématiciens que des géomaticiens il était nécessaire de réaliser un outil qui soit facile d'utilisation. L'outil a donc été segmenté en plusieurs parties permettant à chaque type de profil de pouvoir l'utiliser en fonction de ses compétences.

Il prend la forme suivante :

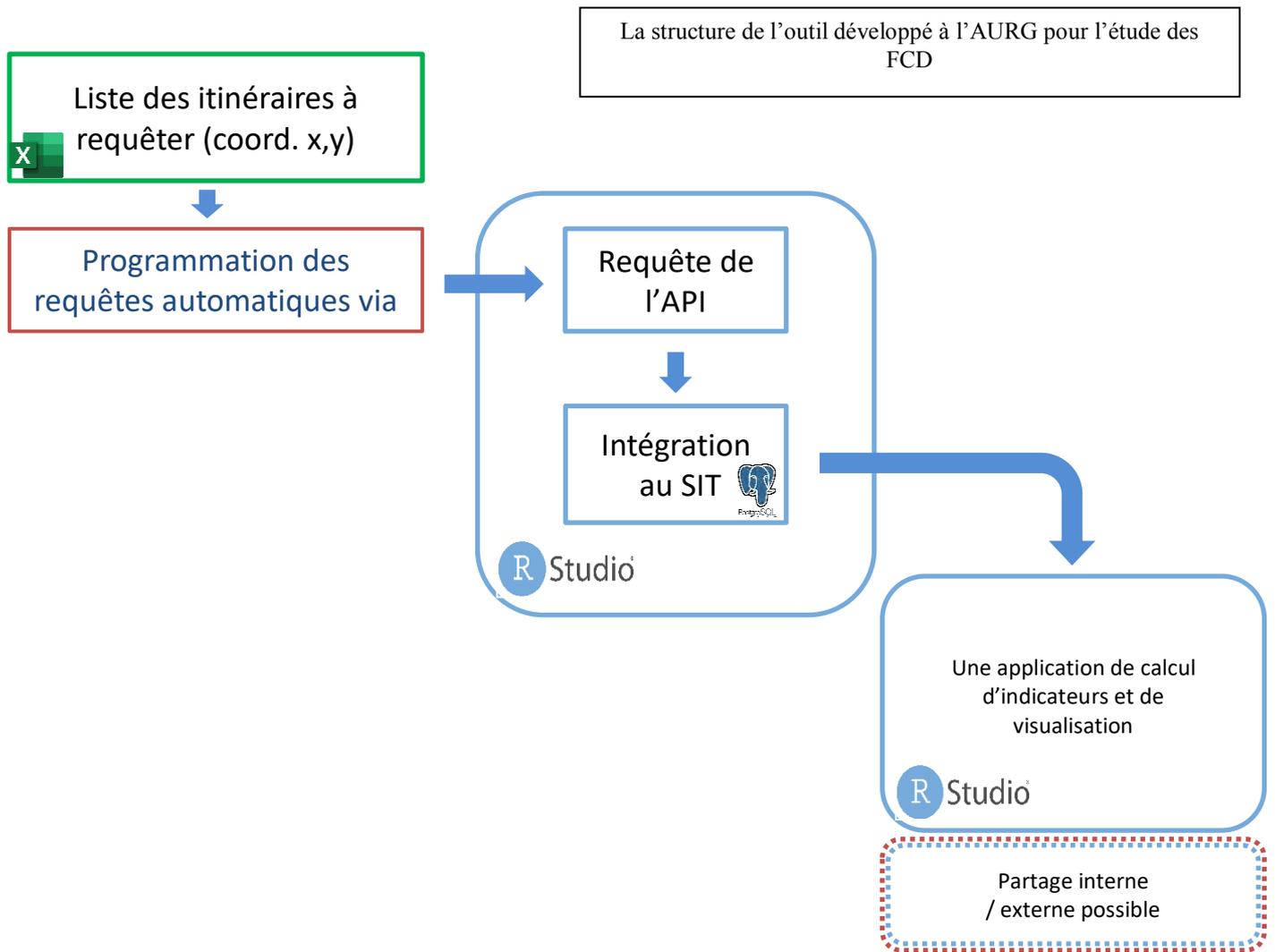


Figure 14: Schéma de l'organisation de l'outil développé à l'Agence d'urbanisme de la Région Grenobloise pour exploiter les données de floating car Data

La matrice

La matrice d'origine destination est décrite dans un fichier csv. Ce fichier permet de faire évoluer la matrice sans avoir à aller modifier le code du programme. Le csv doit cependant répondre à certains critères pour ne pas générer d'erreur lors de son utilisation dans le script (voir notice en annexe). La matrice prend la forme suivante :

coord_origine	coord_interm	coord_destination	lib_origine	code_communes_origine	lib_destination	code_communes_destination	ID_OD
45.202,5.763	45.188,5.7411: 45.1897,5.742	45.185,5.731	Carronnerie	38185	Chavant	38185	14

Figure 15 : Modèle de Matrice à renseigner pour l'obtention de données de temps de parcours par le biais de l'outil développé à l'agence d'urbanisme de Nantes. Réalisation Baptiste Vallet 2019.

Cette matrice est ensuite utilisée comme entrée dans un script R.

Le script des requêtes

La sélection de ce langage de programmation est motivée par le fait que plusieurs agents travaillent déjà avec ce langage. De plus Le logiciel Rstudio est déjà installé sur certaines machines, notamment celles empruntées pour les présentations.

L'outil permet de réaliser les requêtes API pour l'ensemble des itinéraires du csv et d'intégrer les résultats obtenus dans une table du Système d'information Territorial (SIT, base de données de l'AURG). Il a été défini avec l'administratrice de la base de données qu'un schéma exclusif serait créé pour pouvoir stocker et les données de requête automatisées et les données acquises ponctuellement dans le cadre d'une étude spécifique. Chaque étude posséderait dès lors une table spécifique, permettant aux chargés d'études de pouvoir réaliser des analyses sur les temps de parcours étudiés. Principalement lorsque les requêtes ont vocation à être réalisées plusieurs fois sans avoir une récurrence assez forte pour que cela devienne nécessaire de l'automatiser sur un serveur.

Le script R fait appel à deux fonctions. La première fonction attend en argument :

- Le csv contenant la matrice et
- La clé API liée au compte TomTom permettant de s'identifier sur leur serveur.

Pour chaque ligne de la matrice une extraction des coordonnées d'origines, de destinations et, s'il y a, des points intermédiaires sera faites. Ces points sont ensuite reformatés pour être intégrés en tant qu'argument de la requête API, adressée à la plateforme TomTom. Pour chaque requête la fonction utilisée renvoie le code du statut http informant sur la bonne réalisation ou non de la requête. Cette information est stockée dans un objet est écrite dans un log en fin de boucle pour créer un suivi des erreurs. Une fois les résultats de la requête obtenus et stocké dans un objet une série de traitement sont réalisés. Sont extraites de l'objet les valeurs chiffrées pertinentes comme :

- Le temps de parcours en situation réelle ;
- Le temps de parcours historique (chiffre peu documenté mais qui s'apparenterai au temps moyen pour réaliser le parcours) ;
- Le temps en écoulement libre ;
- La distance ;
- L'heure de la requête.

Dans les données retournées par la requête API, le parcours n'est pas renvoyé sous la forme d'une ligne mais sous un tableau contenant l'entièreté des points du tracé. Il est donc nécessaire de reconstituer une polyligne à partir de ces points pour ajouter la géométrie des itinéraires.

Une fois résultats d'une requête retraités, ils sont alors enregistrés dans un dataframe contenant une seule ligne. Cette ligne est ajoutée à un second dataframe qui contiendra l'ensemble des données de chaque requête.

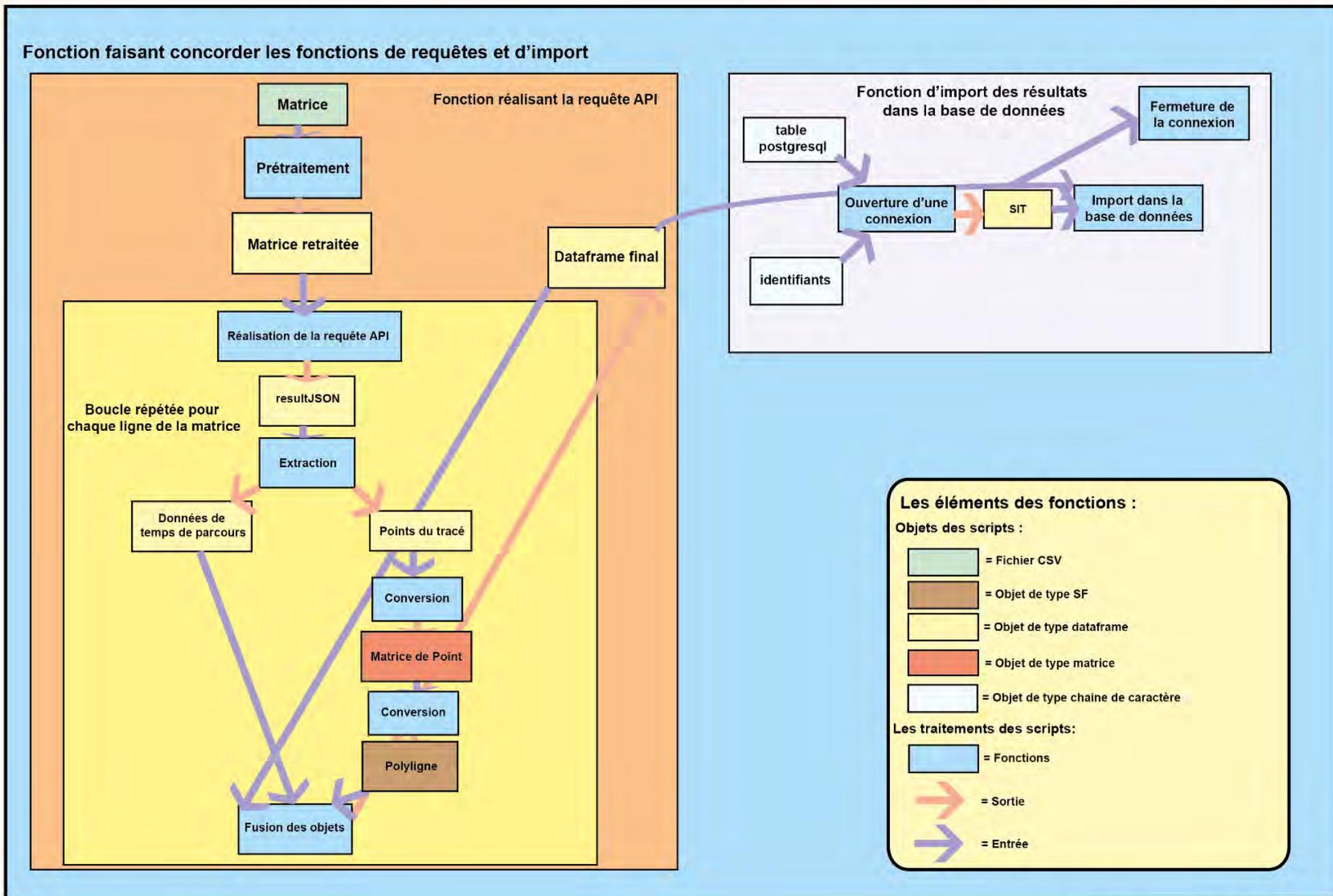
Ce tableau de résultat est l'objet renvoyé par la fonction.

Le script de l'outil permet l'exploitation de ce tableau dans la seconde fonction. Cette dernière va intégrer les données sur la base de données PostGresSQL ou renvoyer le message d'erreur dans un log dans le cas de l'échec de l'import. Elle prend en argument :

- Un tableau de donnée contenant les informations des requêtes API
- Le nom d'un schema
- Le nom d'une table
- Un nom d'utilisateur de la base PostGreSQL
- Le mot de passe attribué à cet utilisateur

Elle renvoie le statut de la connexion au SIT (fermée ou ouverte).

Représentation schématique de la fonction permettant l'acquisition des données de Floating Car Data



Réalisation : Vallet Baptiste.2019

Figure 16 : Organigramme des traitements permettant l'acquisition de données FCD à l'agence d'urbanisme de la région grenobloise

Plusieurs réunions avec le chargé réseau de l'agence ont été nécessaires pour automatiser le lancement du script au même horaire tous les jours. Il a été nécessaire de réaliser un crontab qui lance automatiquement le script sur le serveur linux de l'agence aux plages horaires suivantes :

- Tous les quarts d'heure de 6H45 à 9H ;
- À 15H
- Toutes les demies heures de 16h à 20h.

Cela a aussi permis de dresser une liste des erreurs à éviter et de localiser les principaux points d'adaptation des scripts et fichiers pour passer d'un système d'exploitation windows à linux comme l'encodage des fichiers (voir annexe).

Ce script possède une variante qui peut être lancée manuellement. Ce programme secondaire permet de répondre aux demandes ponctuelles des agents qui ont besoin de données temps de parcours théoriques ou sur une matrice conséquente.

L'application de visualisation

Etude comparative de plusieurs solutions

Certains agents n'ont pas la nécessité d'exploiter entièrement la base de données mais seulement les indicateurs qui peuvent être calculés. Il était donc nécessaire pour ce type de profil, pour qu'ils puissent profiter de la base de données, de créer une application visuelle générant automatiquement la visualisation de ces indicateurs de manière dynamique et interactive.

Une étude sur les différents logiciels permettant de réaliser ces traitements a donc été faite. Trois logiciels ont été confrontés :

- Glik Sense utilisé par l'agence de Nantes ;
- PowerBI : logiciel qui avait déjà été envisagé pour d'autres rendus à l'agence ;
- Rshiny : package de R qui offrirait la possibilité de n'avoir à recourir qu'à un seul logiciel pour l'intégralité de l'outil.

Une analyse comparative des trois plateformes a permis de sélectionner la plateforme qui serait la plus à même de répondre aux besoins de l'agence. Les logiciels Glik Sense et PowerBI présentent une plus-value certaine quant à la facilité d'utilisation de leurs outils. Néanmoins les indicateurs réalisables ne correspondaient pas à ceux définis suite aux différentes discussions avec les collaborateurs. Il aurait été possible de les développer par l'utilisation des langages de développement BNF (Backus-Naur Formalism) ou DAX (Data Analysis Expression). Mais ce sont des langages propriétaires et peu utilisés. De plus les cartographies réalisées sur les deux plateformes présentaient plusieurs complications lors de la discrétisation. Les outils ne proposaient que très peu de méthodes de discrétisations qui n'étaient pas forcément les plus pertinentes au vu des séries statistiques à analyser. De plus des manquements à la sémiologie graphique par l'utilisation de couleurs chaudes pour des valeurs négatives et inversement était courant. Enfin la charte agence, comprenant les différentes normes quant aux rendus des chargés d'études (couleurs, polices, logo etc...), était difficilement incorporable aux éléments visuels renvoyés.

Le choix s'est ainsi arrêté sur le package Rshiny. Ce dernier est certes plus difficile à développer mais il permet de réaliser les mêmes traitements que les autres logiciels tout en étant plus flexible. De plus le langage employé reste plus répandu que ses deux homologues. En outre ce package est une alternative gratuite permettant de n'avoir qu'un seul logiciel pour l'entièreté de l'outil. Enfin à l'aide d'un fichier style en cascade (css) il a été possible d'intégrer la charte agence à l'application. Du fait de l'utilisation du package leaflet il est cependant obligatoire d'utiliser une version 3.6 de R.

Un large temps du stage a donc été nécessaire pour le développement de cette application. La mouture finale a longuement été discutée avec ma tutrice pour le choix des indicateurs et la manière de les représenter. Les agents du pôle communication de l'agence ont pu apporter leur aide notamment pour l'intégration de la charte agence.

L'armature de l'application

Une application développée sous shiny est similaire à un site internet.

Pour cette application on utilisera la fonction nommée `NavbarPage`. Cette fonction permet de faire apparaître dans le corps de la page des éléments en fonction de l'onglet (Panel) sélectionné dans une barre de navigation. On retrouve ensuite toutes les différentes possibilités d'organisations d'une page web (grilles, colonnes, lignes, div).

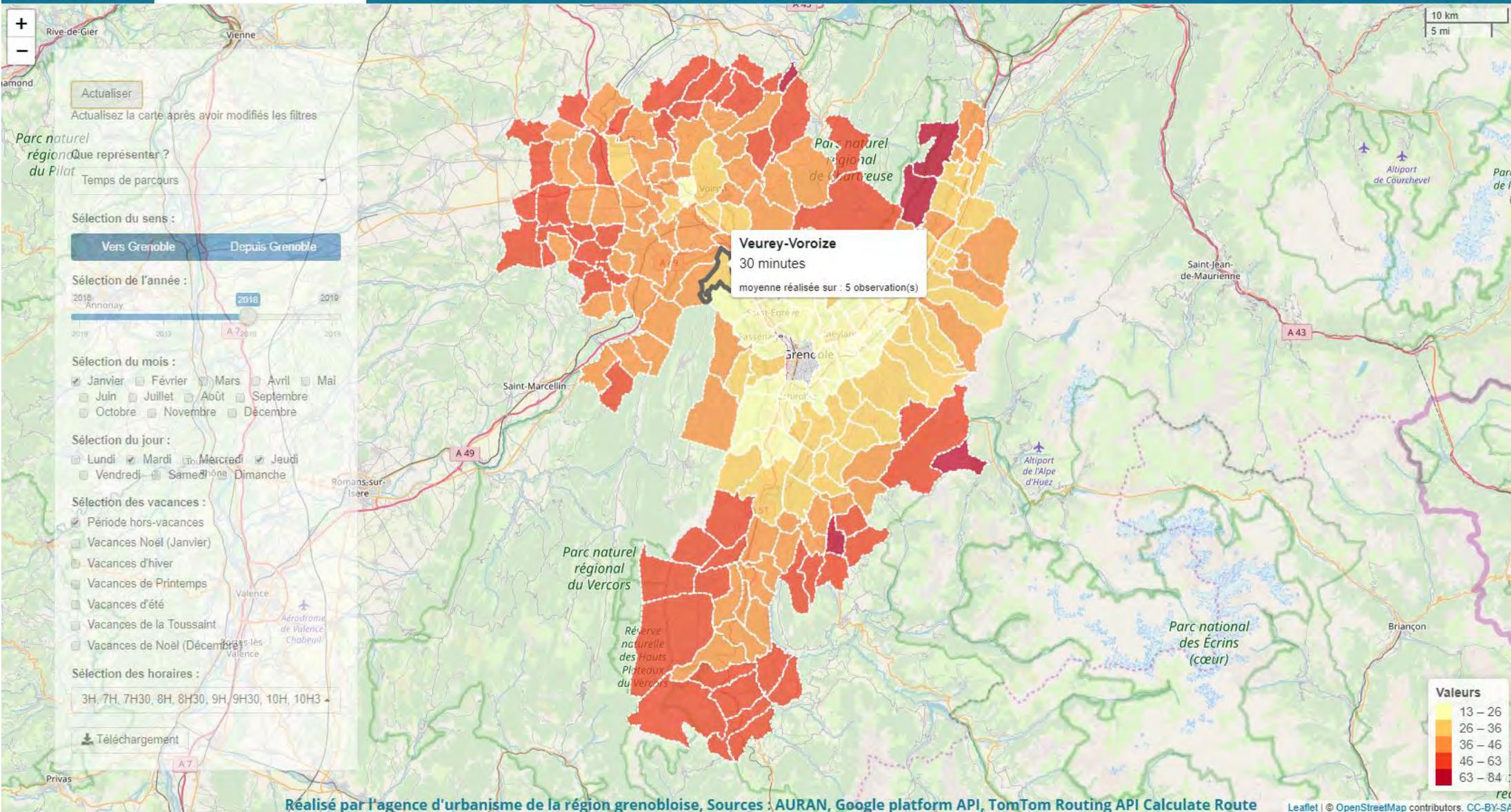
Globalement chaque onglet (Panel) est constitué de filtres (input) présent dans une barre latérale ou une div et d'éléments visuels (output) pouvant prendre la forme de cartes ou graphiques dans le corps de la page. Les éléments visuels sont ainsi influencés par les filtres.

Ici deux onglets étudient les données fournies par l'agence d'urbanisme de Nantes (AURAN). L'objectif est de représenter par des cartes différents indicateurs :

- Temps de parcours
- Écart de temps de parcours en fonction de celui en écoulement libre (valeur absolue) ;
- Écart de temps de parcours en fonction de celui en heure creuse (valeur absolue) ;
- Écart de temps de parcours en fonction de celui en écoulement libre (valeur relative) ;
- Écart de temps de parcours en fonction de celui en heure creuse (valeur relative) ;
- Evolution des temps de parcours SF en valeurs absolue
- Evolution des temps de parcours SF en valeurs relative

Ces cartographies sont dynamiques et réactives selon plusieurs filtres :

- L'heure
- Le jour
- L'année
- Le mois
- Les vacances



Réalisé par l'agence d'urbanisme de la région grenobloise, Sources : AURAN, Google platform API, TomTom Routing API Calculate Route

Leaflet | © OpenStreetMap contributors, CC-BY-SA

Figure 17 : Capture d'écran du premier panel de SateliteGre présentant le temps moyen de parcours pour se rendre d'une commune de l'aire urbaine de Grenoble à Grenoble pour les mardi et jeudi du mois de mars 2018 hors période de vacances

Le troisième onglet a été conçu pour analyser les données de l'AURAN sous forme, cette fois-ci, de graphique et de courbes. Les données représentées sont ici similaires à la note présentée en première partie.

Le 4^{ème} panel permet quant à lui d'étudier les données des requêtes réalisées à l'issue de mon stage.

Il propose 4 graphiques en bar représentant :

- L'écart de temps parcouru moyen par heure ;
- Le temps de parcours moyen par heure ;
- L'écart de temps de parcours moyen par jour ;
- Le temps moyen de parcours par jour.

Mais aussi deux courbes, prenant une forme similaire à celles développées pour le panel des graphiques des données de l'AURAN. Les écarts sont calculés entre la moyenne des temps contenus dans la base de données pour une OD (par heure/jour) et le temps à vide (donnée présente dans les requêtes TomTom qui correspond à la moyenne des temps pour l'itinéraire choisie à 3h du matin) ou le temps en heure creuse, défini à 15h.



Figure 18 : Représentation graphique à travers l'outil SatelliteGre des temps de parcours moyens des communes de l'aire urbaine de Grenoble vers Grenoble pour les mardis et jeudis du mois de Janvier 2018 (hormis jours en période de vacances scolaire)

Le fonctionnement de l'application :

La fonction Rshiny prend en argument :

- Un objet ui contenant l'armature de l'application ainsi que les balises input et output
- Une fonction server qui va exploiter les valeurs des inputs dans des fonctions réactives pour renvoyer des informations à visualiser dans les outputs.

La fonction Rshiny permet ainsi de faire communiquer les deux objets en argument pour renvoyer une application prenant la forme d'une page web.

Les deux onglets des cartes ont une armature similaire composée de deux éléments :

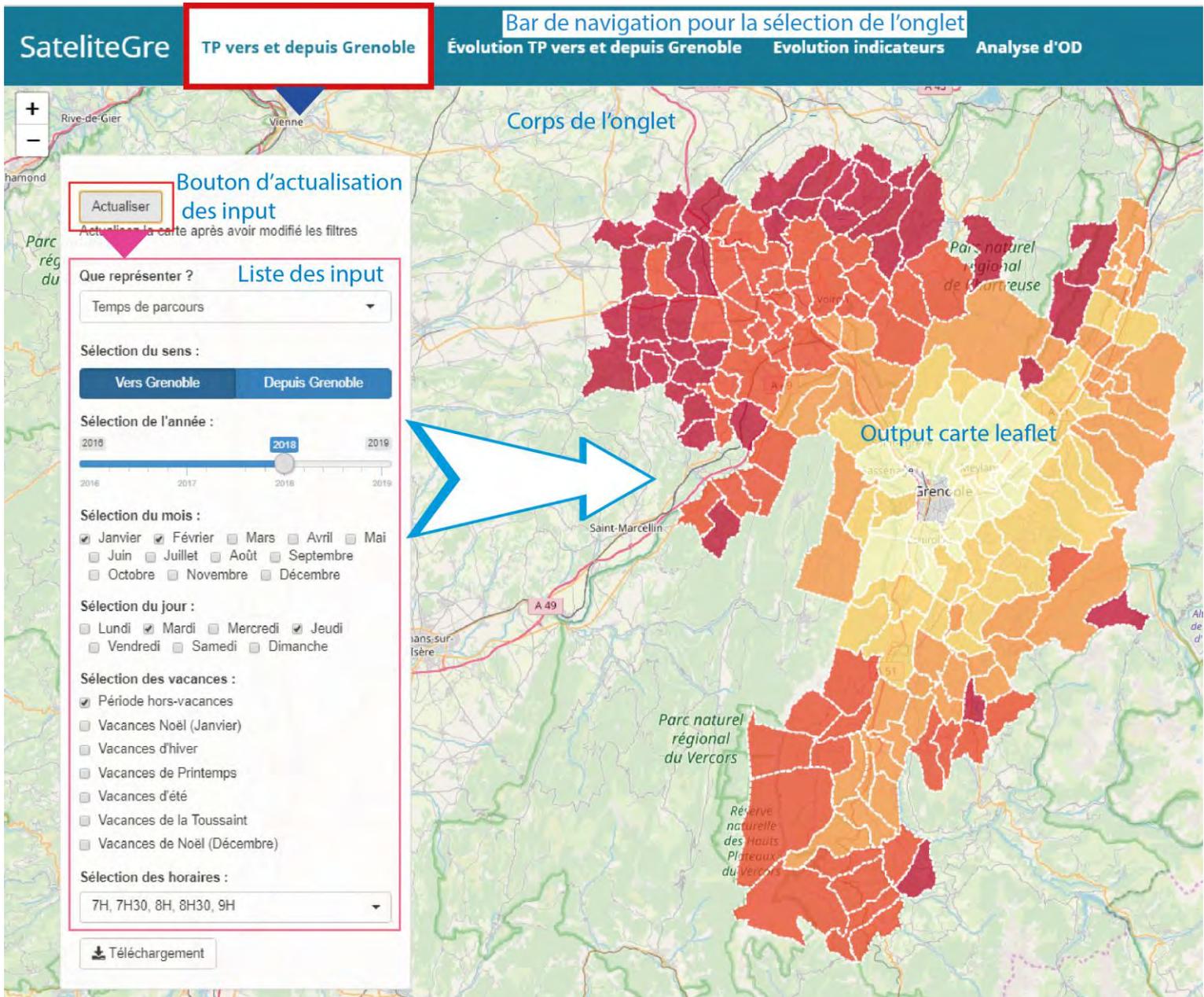


Figure 19: Schéma de la synergie d'un onglet développé sous Rshiny: Réalisation Vallet Baptiste 2019.

- Une div dont la largeur et la longueur sont égale à 100% de l'espace alloué à l'écran_ contenant l'output de la carte. Cet output prend la forme d'une fonction nommée leafletoutput() elle attend en argument l'id qui servira à l'appeler dans les fonctions réactives et des informations sur sa taille et son placement.
- Un absolutePanel qui accueille les inputs qui vont servir de filtres
- Un bouton filtre qui permettra de mettre à jour dans la fonction server la valeur des inputs.

La fonction server pour ces deux panels réalisent des traitements similaires pour la création des outputs.

Tout d'abord il stocke la valeur des inputs dans des objets réactifs qui ne se mettent à jour que lorsque l'on clique sur le bouton. Les différents inputs servent ensuite à filtrer le tableau de données qui est différent selon le sens choisi. Le tableau avec les données filtrés est alors utilisé dans différentes fonctions pour calculer les indicateurs. Indicateurs qui dans le cas des données de Nantes doivent être joint à une couche de géométries des contours des limites communales à l'aide des codes communes. Cet objet géographique est ensuite discrétisé par la méthode

jenks pour le premier onglet et la méthode pretty (jolies ruptures) pour le second. Chaque entrée du tableau se voient ensuite attribuée une couleur selon la palette définie et la classe de discrétisation dans laquelle elle se situe. L'ensemble des objets créés à la suite des traitements réalisés sur l'objet géographiques servent d'argument à la fonction générant la carte leaflet et aux observer, servant à optimiser les temps de traitements, dans l'output défini.

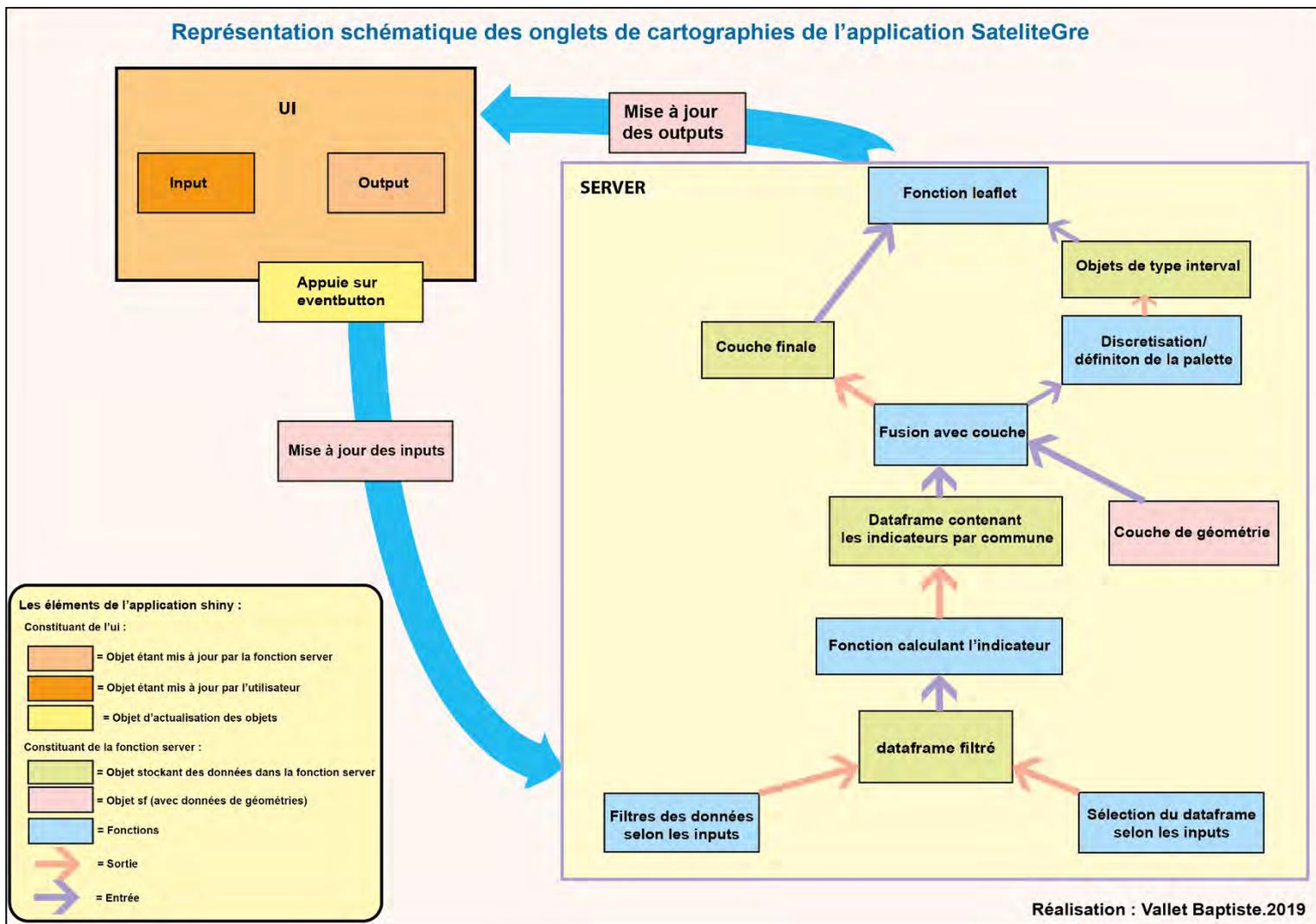


Figure 20: Organigramme des onglets de cartographies de l'outil SatelliteGre. Réalisation Baptiste Vallet 2019.

Les onglets de représentations graphiques ont aussi une armature similaire qui s'appuie sur les fonctions développées pour les applications shiny.

Ils sont générés par la fonction Fluidrow(). Cette dernière reçoit en argument trois fonctions column() :

- La première de taille 2 est l'équivalent des absolute Panel des onglets de cartographie, les mêmes inputs sont enregistrés à l'intérieur à l'exception de celui permettant la sélection des indicateurs à représenter (remplacer pour le panel AURG par un input similaire mais contenant les libellés des OD).
- Les deux autres de tailles 5 contiennent les outputs des graphiques qui seront affichés.
- Le panel Aurg contient en plus un output Leaflet pour faire apparaître le tracé de l'itinéraire sélectionné.

Les traitements de la fonction server sont similaires à ceux pour les onglets de cartographie à la différence qu'il n'est pas nécessaire de stocker les valeurs des inputs dans des objets réactifs (les inputs sont appelés directement dans les fonctions) car il n'y a pas de bouton réactifs. De plus les données des requêtes réalisées à l'agence de Grenoble contenant la géométrie des tracés et l'onglet de représentation graphique ne renvoyant pas de carte, il n'y a pas de nécessité à fusionner avec un objet contenant des données géométriques. La principale différence réside dans la fonction renseignant les informations à afficher dans l'output renvoyant ici des graphiques. C'est

le package ggplot2 du tidyverse qui sera mis à contribution. Ce package prend en entrée un tableau de données et une série d'arguments permettant de définir l'esthétique de l'objet. Par la création d'un thème les graphiques sont conformes à la charte agence. Ils sont donc instantanément intégrables à un rapport de l'agence une fois tracé.

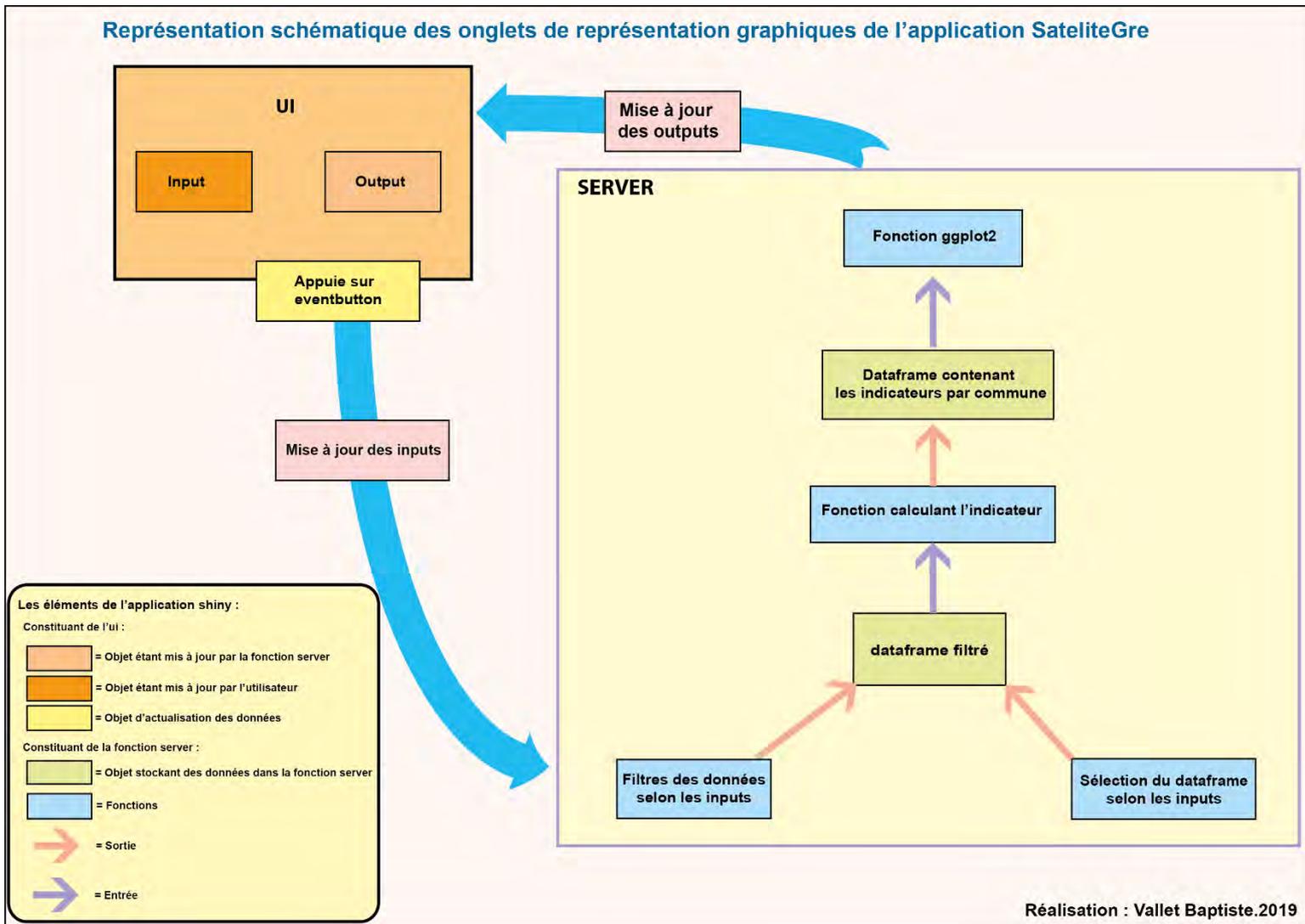


Figure 21: Organigramme des onglets de représentation graphique de l'outil SatelliteGre. Réalisation Baptiste Vallet 2019

IV- L'INTEGRATION AUX PROJETS DE L'AGENCE

A) L'implantation d'un Bus à Haut Niveau de Service (BHNS) dans la commune de Meylan

Au début de ce stage, l'expertise de l'observatoire du développement a été souhaitée par la commune de Meylan. La commune étant l'une des portes d'entrées de la métropole grenobloise à l'est, souffre de fortes congestions sur ses axes routiers. Pour essayer d'endiguer le phénomène un projet de développement de BHNS est en cours de réalisation. Les questions qui se posent sont sur le tracé que ce bus devra emprunter ainsi que les horaires et la fréquence de passage.

Dans ce cadre un des collaborateurs de l'agence m'a demandé s'il était possible de réaliser une carte de congestions de la commune. Pour traiter la demande il a été nécessaire d'utiliser le jeu de données TomTom présent sur le SIT. Il se base sur une commande faite en 2018 par le SMTC à l'outil Custom Area Analysis de TomTom. Il contient ainsi des données de temps de parcours pour des tronçons fins sur l'ensemble de la métropole.

Après discussion avec l'agent pour définir ses besoins il a été décidé d'analyser les temps de parcours obtenus les Mardis et Jeudis de 7H à 9H et 16H et 19H. Seront ainsi représentés les ratios entre les moyennes de temps aux horaires cités précédemment et ceux recensés à 7H.

Je lui ai présenté un premier résultat intermédiaire, ce qui m'a permis de le mettre en garde face aux quelques biais que peut présenter ce type de carte :

- Tout d'abord le fait que les ratios se basent sur des moyennes de temps récoltés les mardis et jeudis. Le jeudi ayant selon les statistiques des temps de congestion plus faible que le mardi cela entraîne des temps de parcours légèrement diminués.
- Ensuite les ratios sont faits sur de petits tronçons. De ce fait, les chiffres de temps de parcours sont extrêmement faibles. Il y a donc un décalage qui peut se créer entre la perception d'un ratio élevé sur la carte mais qui correspondrait au final à une perte de temps pas si significative.

Si la carte a suscité l'intérêt du chargé d'études, une refonte du planning de ce projet a été faite, repoussant les discussions sur ce projet au mois d'octobre, lorsque la mission sera finie.

(Carte finale en annexes)

B) Schéma de développement des espaces économiques (SDEE)

L'agence promouvant l'interdisciplinarité, les agents de l'équipe mobilité sont parfois amenés à travailler avec ceux de l'équipe en charge de l'économie. L'un des projets portés par le pôle économique est le schéma de développement des espaces économiques. Il consiste à produire un diagnostic territorial sur les intercommunalités du Grésivaudan et du SMVIC. L'objectif serait de faire un classement des zones d'activités à l'aide d'un scoring prenant en compte une quinzaine de paramètres. Dans ses paramètres les deux agents voudraient inclure un facteur accessibilité. Pour ce faire ils avaient besoin de données de temps de parcours des itinéraires partant des zones d'activités pour se rendre vers des centres urbains, les gares et les échangeurs autoroutiers. La valeur du scoring était donc défini selon le temps le plus court pour se rendre d'une zone d'activité à l'un des points de destination.

Le challenge reposait principalement sur un échéancier court. Les deux agents avaient un point la semaine suivante avec des élus pour présenter l'avancée de leurs travaux. Il fallait ainsi dans ce laps de temps obtenir les chiffres et rédiger une notice qui leur permettrait d'expliquer les données lors de cette réunion.

Pour réaliser la commande il a tout d'abord été nécessaire de construire les trois fichiers de matrice. Si des couches contenant les coordonnées des gares et polarités existaient déjà sur le serveur de l'agence, ce n'était pas le cas pour les échangeurs. Les données d'OpenStreetMap ont donc été sollicitées pour obtenir ces informations. Le tag « Motorway junction », regroupe toutes les sorties et entrées d'autoroutes. Un filtre a été appliqué pour ventiler

toutes les sorties et entrées vers des aires de repos. Il a ensuite été nécessaire de rédiger un script R pour la réalisation des matrices. Ces fichiers sont ensuite tour à tour utilisés comme argument de la fonction de l'outil de requête ponctuelle. Ces temps sont ensuite stockés dans une table du schéma alloué sur le serveur. On extrait ensuite pour chaque zone d'activité l'entrée contenant le temps le plus court pour se rendre à un des types de destination.

C) Etude de localisation d'aire de covoiturage

Du fait des retours positifs des élus et des agents à propos de la commande pour le SDEE un des agents en charge des études sur le covoiturage a formulé une demande similaire. L'objectif de l'étude était de carroyer la métropole de Grenoble et d'affecter à chaque carreaux un temps de parcours théorique pour se rendre à plusieurs zones d'activités. La finalité serait ensuite de cartographier le résultat, sans représenter les carreaux (carte de chaleur, interpolation ...).

Le principal obstacle à cet exercice est le quota de requêtes gratuites imposé par TomTom (2500 par jour). Cette barre a tendance à être très vite atteinte avec ce type de travail. Une concertation a donc été nécessaire pour pouvoir sélectionner la maille la plus adéquate. Le maillage devrait permettre une analyse scientifique pertinente sans avoir recours à un nombre trop conséquent de requêtes. A l'origine le travail se basait sur une maille de 100m sur 100m. Cependant, si fine les carreaux ne présenteraient pas d'écarts significatifs. Ce qui aurait pour conséquence d'alourdir de façon peu pertinente le nombre de requêtes. Je lui ai donc proposé de ne pas descendre en dessous de 200m mais de ne pas aller au-dessus de 500.

La métropole de Grenoble étant un territoire assez vallonné, beaucoup d'espaces sont scientifiquement peu intéressants car non habités ou ne possédant tout simplement pas de routes. Pour réduire le nombre de requêtes, les carreaux dont l'occupation du sol n'est pas constituée d'au moins 20% de « bâti -résidentiel » ont été retiré. La couche « bâti » permettant l'application du filtre s'est appuyée sur un jeu de données du spot themia, conseillé par un géomaticien de l'agence. De cette couche ont été extraite les zones résidentielles.

Une fonction a été développée permettant de réaliser l'entièreté des traitements. Elle permet d'appliquer le filtre aux carreaux, de récupérer les coordonnées de leurs centroïdes, de constituer la matrice, de requêter l'API TomTom et d'intégrer les résultats sur la base de données. Les données stockées dans la base comprennent ainsi les temps de parcours et les tracés des itinéraires. Avoir une information sur les géométries des parcours permet de pouvoir vérifier la fiabilité de la donnée. Les carreaux étant placés arbitrairement sans tenir compte de la typologie du réseau, les centroïdes de deux carreaux adjacents peuvent se situer sur des sens différents. Dans le cas des autoroutes, pour lesquelles un changement de sens induit une longue distance, cela peut avoir un impact non négligeable sur les temps de parcours. De forts écarts pourront donc être constatés sur des carreaux pourtant à proximité.

Pour une exploitation cartographique, il a été décidé avec l'administratrice de la base de donnée de stocker les géométries des carreaux et les données dans deux tables différentes. Il faut donc procéder à un filtre en sélectionnant que les entrées de la zone d'activité qui doit être représentée. Une requête SQL a donc été rédigée pour importer les données sur QGIS.

D) Schéma de mobilité Nord-Ouest (SMNO)

Dans le cadre de la restructuration du schéma de mobilité nord-ouest (SMNO). Une des collaboratrices de l'observatoire du déplacement désirait des temps de parcours théoriques sur 5 itinéraires pour un mardi de novembre en heure de pointe. L'outil a permis de pouvoir répondre à la commande. La chargée d'étude fut ainsi intéressée par le fait de pouvoir réaliser un suivi sur ces trajets. Les requêtes ont donc à la suite de ce travail, été intégrées à celles automatisées.

E) Les missions annexes

Toutes les missions de ce stage n'étaient cependant pas tournées vers les outils qui ont été développés. Certaines commandes étaient m'ont été faites pour pouvoir exploiter les compétences acquises lors de mon cursus universitaire.

L'étude des vitesses

Il m'a été demandé d'analyser l'un des principaux problèmes des études de mobilités routières : la fiabilité des vitesses légales des voies dans les bases de données. L'objectif était de pouvoir intégrer cette donnée dans les indicateurs de l'outil SatelliteGRE.

J'ai ainsi réalisé une comparaison entre les vitesses renseignées par TomTom (données fournies dans le fichier de réponse de la méthode CAA) et celles d'une couche OSM déjà présente dans le Système d'Informations Territorial. Par pur hasard, et sans qu'il ait été possible de l'expliquer, la nomenclature des voies était identique entre les deux couches. Il a donc été possible de réaliser une jointure attributive sur ce champ. Cela a permis de quantifier le nombre de tronçons du fichier de géométrie TomTom dont la vitesse limite légale ne correspondait pas à celle issue de OSM. Il s'avère que les tronçons présentant des erreurs étaient les tronçons limités à 30kmh. Or la métropole de Grenoble est un des fers de lance du projet Métropole apaisée. Ce projet a pour objectif de limiter la grande majorité des axes routiers urbains à 30kmh. Aujourd'hui 47 des 49 communes de la métropole grenobloise participent à ce projet. Le nombre de tronçon étant limité à 30km dans la couche TomTom a donc été compté. Celui-ci était inférieur à 20, ce qui peut s'expliquer principalement par les paramètres de la requête faites à TomTom. Il est en effet possible de filtrer les voies des classes les plus « faibles ».

Au vu du nombre de voies qui devraient être limitées à 30 et du fait que les voies rapides et axes intermédiaires (supérieur à 80km/h) ne présentent que peu de contradictions, il devient plus pertinent d'inverser la réflexion. Plutôt que rechercher une couche avec des zones 30 bien référencées, il devient plus facile de rechercher les voies à 50 et d'attribuer à tout axe en zone urbaine une valeur 30. Cependant l'étude n'est pas allée plus loin. En effet même s'il est possible de pointer les différences entre les sources de données, il est beaucoup plus complexe de définir qu'elle est la valeur juste. Seule une étude de terrain permettrait de départager les deux. Mais cela serait long et coûteux et ne pouvait être réalisé dans le cadre de ce stage.

L'études des API

Plusieurs demandes m'ont ensuite été faites quant à l'acquisition de nouvelles données auprès d'API.

Le Grand A numérique

Le « Grand A », est une publication de l'agence vers ses partenaires, dont le sujet cette année est : « le numérique ». L'objectif était de voir s'il était possible de faire une publication similaire à ce qui a été fait par l'agence d'urbanisme de Lille. Elle s'appuyait sur des API telles que TripAdvisor, AirBNB et booking, pour réaliser plusieurs cartographies en lien avec le tourisme.

L'aide à l'estimation de la taxe d'habitation

L'agence mène une étude sur l'hébergement touristique entre particuliers, basée en partie sur les données du site airBNB. Leur demande consistait à savoir s'il était possible de requêter l'API airBNB (ou à défaut AirDNA contenant des informations de la plateforme airBNB). Deux principaux enjeux se dégagent de l'obtention de ces données. Pour la métropole de Grenoble cela offrait un outil permettant de mieux estimer la taxe de séjour à percevoir et s'assurer qu'elle le soit totalement. Pour l'agence cela permet de faire un diagnostic du logement touristique et donnerai la possibilité de localiser les lieux à fort potentiel ou qui ont besoin de projets de rénovations. En outre ce projet constituerait le temps 0 de l'observation de la présence de AirBNB sur le territoire.

Dans les deux cas j'ai rapidement effleuré les bases du webscraping, mais je me suis rapidement rendu compte que je n'aurai pas assez de temps pour mener ces missions à leur terme. J'ai donc simplement rendu une note explicative dans laquelle je mettais en garde les agents sur l'évolution des API. Par exemple, dans le cas des

plateformes AirBNB et TripAdvisor ces dernières interdisent aujourd'hui (les données de la publication de Lille datent soit de 2016, soit de 2017) l'exploitation de leurs données API pour des analyses mais aussi l'utilisation de webscrapping pour créer une base de données depuis leurs sites.

Il semblerait aussi, d'après les différentes sources bibliographiques, que l'entrée en vigueur des normes du RGPD et le scandale Cambridge Analytica ont eu un fort impact sur les politiques des fournisseurs. Il semble dès lors qu'il était possible de récupérer aisément des données issues d'API gratuitement entre 2016 et 2018. Mais qu'aujourd'hui, devenues trop sensibles, ces données sont délivrées de manière payante, sous partenariat ou alors gratuitement mais avec bien moins d'informations, comme la géolocalisation ou le nombre de personnes/utilisateurs sur lequel se base les chiffres.

V- CONCLUSION

La finalité de l'outil

L'outil sert aujourd'hui plusieurs buts. Tout d'abord, il permet par le biais de requête d'itinéraires fixes et automatisés, de constituer une base de données de temps de parcours. Cela permettra alors au pôle mobilité de l'agence de pouvoir étudier sur un temps long la congestion des axes routiers de la métropole grenobloise. Il permettra aussi d'étudier l'impact des différents projets de réaménagements à posteriori de leur réalisation. Ce sera notamment le cas de l'autoroute A480.

Ensuite, à l'aide de requêtes ponctuelles, l'outil offre la possibilité aux chargés d'études de tout pôle de faciliter l'intégration de temps de parcours théoriques à leurs analyses. Ce qui est déjà le cas pour le pôle économique qui a intégré dans son scoring un facteur temps de parcours des zones d'activités vers plusieurs destinations. Ou encore pour l'analyse des lieux les plus pertinents pour l'implantation de parking de covoiturage.

L'application de visualisation développée permettra une exploitation simplifiée de la base de données. Cela même pour les agents n'ayant pas de compétences particulières en géomatique. De plus l'intégration de la charte agence dans les visuels, rend aisée l'insertion de ces derniers dans un rapport.

Une notice a été fournie à l'agence ciblant les principaux points sur lesquels il est nécessaire d'être attentif. Trois agents ont aussi été formés pendant une journée et demi à la maintenance et au développement de l'outil. Cela assure un outil stable et pérenne.

Un outil encore limité...

Comme dit précédemment, la mission consistait à la constitution d'une base de données. La principale complication pour l'exploitation d'une base de données est le temps nécessaire pour sa création. L'outil proposé n'aura de valeurs scientifiques que d'ici six mois voire un an. Il sera cependant de plus en plus intéressant au fur et à mesure que le temps passe. La seconde difficulté est celle de la fiabilité des données. La base issue des requêtes de Nantes a pu démontrer que certains manques apparaissaient dans les données (aucune donnée en novembre 2017, voir tableaux annexes 4). La constitution de la base de données devra donc faire l'objet d'un suivi sérieux pour pouvoir rendre compte des manques dans les données.

Une des limites les plus importantes de l'outil est la manière même d'acquérir la donnée. Le fait d'être soumis à un fournisseur de donnée peut avoir de graves conséquences. Comme les API Google la politique tarifaire peut évaluer, de façon négative ou positive pour le client. Les différents scandales sur les données privées comme Cambridge analytica ou encore la mise en place des normes RGPD peuvent aussi avoir un impact sur les données renvoyées par les fournisseurs. À tout moment la donnée peut évoluer et devenir inexploitable, même si ce mémoire à présenter plusieurs sources de données possibles il sera nécessaire de rester attentif sur ces différents points.

Les informations liées à la congestion routière présentent de forts enjeux politiques. Il est donc nécessaire de les traiter avec prudence. La communication des chiffres doit donc faire suite à un protocole de fiabilisation de la donnée.

...Mais qui n'a pas encore démontré tout son potentiel

Cet outil peut encore être amélioré par l'injection de données exogènes. Des informations sur les accidents ou la météo pourraient notamment permettre de mieux comprendre les valeurs calculées.

Certaines optimisations du code pourraient permettre à l'outil d'être plus performant. Cela pourrait passer par une meilleure gestion de l'import des bases de données ou encore des éléments visuels qui se chargeraient plus rapidement.

Il serait bon aussi de réaliser quelques prétraitements sur la base de données pour éviter les incompatibilités de formats. Cela pourrait être réalisé à la création même de la donnée dans les scripts de requête ou par des requêtes SQL pour les données de Nantes.

Retour d'expérience

Techniquement, le stage réalisé à l'agence d'urbanisme de la région grenobloise a dépassé de loin mes attentes. J'ai pu tout d'abord approfondir mes connaissances des langages R. et SQL. J'ai pu aussi découvrir le vaste monde des API, leur fonctionnement ainsi que le moyen d'exploiter les données. Je connais ainsi mieux les points d'alertes comme le besoin de toujours se référer aux Conditions générales d'utilisation. Cela m'a aussi permis de me rendre compte que le monde de la géomatique est un monde qui évolue sans cesse dans une aire où le numérique devient un réel enjeu pour les politiques, que ce soit économiques ou législatives. J'ai pu percevoir une pléthore de nouveaux logiciels notamment comme GlikSense ou PowerBI pour de la représentation visuelle.

Humainement, ce court moment passé dans une agence m'a appris à travailler avec des personnes aux profils variés. Il était ainsi nécessaire de présenter à travers des oraux de manière simple et concise les tenants et aboutissants des missions de mon stage. Le choix de présentations fait lors de ces points est très important devant un public hétéroclite qui n'a pas un niveau technique homogène et qui va souvent attendre plus les résultats que la méthode. Je me suis aussi familiarisé avec la vie en entreprise, les relations avec collègues et prendre l'initiative de trouver quelle personne est la plus susceptible de répondre à mes questions. Tout cela sous des contraintes temporelles diverses (fermeture de l'agence, congés, peu de temps libre).

La combinaison des deux m'a permis de mieux comprendre et d'appréhender le besoin de connaître un grand nombre d'outils pour pouvoir s'intégrer aux méthodes de travail déjà présente dans la structure d'accueil. Cela passe aussi par le besoin de se conformer aux chartes déjà disponibles (charte graphiques, interopérabilité des formats et bonne dénomination des tables ou fichiers).

Sources :

LA CIRCULATION SUR NANTES, ÇA COINCE OU PAS ? Vincent Caillard, AURAN

La circulation sur Nantes, ça coince ou pas ? [en ligne] [consulté le 30/08/2019]
https://www.auran.org/sites/default/files/publications/documents/synth_metrosat_vf.pdf

Site API TomTom [consulté le 30/08/2019]
<https://developer.tomtom.com/>

Site API Google [consulté le 30/08/2019]
<https://cloud.google.com/maps-platform/products/?hl=fr>

Site AURG [consulté le 30/08/2019]
<https://www.aurg.org/>

Site Rshiny [consulté le 30/08/2019]
<https://shiny.rstudio.com/>