

PRÉPARÉ PAR

JEHANNE BALEYDIER, MARIE CHALLET,
CÉDRIC JESTIN, JIMMY YOBO

SAE 3-03

DESCRIPTION ET PRÉVISION DE
DONNÉES TEMPORELLES

CIRCULATION ET NOUVELLE INTERSECTION DE ROUTES

RAPPORT
D'ÉTUDE

PROFESSEUR REFERENT
PAUL MARIE GROLLEMUND

**NOVEMBRE
2023**

Introduction

La mobilité urbaine est un enjeu majeur dans de nombreuses villes à travers le monde. La circulation routière, en particulier, est un facteur clé qui influence la qualité de la vie des citoyens, l'efficacité économique et l'environnement.

Dans le cadre de ce projet en groupe, nous avons entrepris une étude approfondie de la circulation entre une ville donnée et la route nationale qui la dessert, les données dont on dispose s'étendent du 01/11/2015 au 30/06/2017 et comprennent des relevés du nombre de véhicules passant à un point de mesure pendant un intervalle de temps d'une heure.

Notre étude se concentre sur la ville qui est reliée à la route nationale par trois intersections : (I1, I2 et I3) qui permettent d'entrer (E1, E2 et E3) ou de sortir (S1, S2 et S3) de la ville par diverses routes. Les entrées et les sorties, cependant, ne sont pas exemptes de défis. L'engorgement du trafic a poussé la municipalité à la mise en place de la quatrième intersection nommée I4 en janvier 2017.

Nous avons cherché à déterminer si la construction de cette intersection a permis de réduire la congestion des trois premières.

Pour cela nous avons traité puis analysé les données afin de détecter des anomalies et d'identifier s'il y a eu des fluctuations inhabituelles dans la circulation après la date du 1er janvier 2017, date à laquelle l'intersection I4 est entrée en service. Nous avons essayé de déterminer si ces variations semblent se prolonger dans le temps.

Pour aider la municipalité à anticiper la gestion du trafic, nous avons fourni des prévisions sur l'évolution de la circulation selon le meilleur modèle qui nous a permis de décrire les données. Ces prévisions seront accompagnées d'une analyse critique sur nos hypothèses et nos résultats obtenus.

I- Le prétraitement

L'ensemble des traitements a été réalisé en utilisant le langage R et des modules dont 'dplyr', 'lubridate' et 'tidyverse'. Les enregistrements temporels ont été standardisés au format date-heure pour permettre une manipulation temporelle précise. À l'aide de la bibliothèque 'tidyverse', nous avons restructuré nos données pour distinguer le nombre de véhicules par intersection et par heure pour chaque date, en prenant soin d'attribuer les valeurs correctes aux entrées (Ei) en fonction du nombre de véhicules passant par l'intersection et du nombre de véhicules sortant selon les créneaux horaires spécifiés. Nous avons également construit des data-frames en regroupant les données par jour et par mois, en prenant comme fonction d'agglomération (summarize()) la somme (sum()). Cela nous semblait le meilleur moyen de rassembler nos données mais nous pouvions choisir la moyenne ou encore le maximum dans ce contexte.

Avec ces traitements nous obtenons trois tables 'data', 'data_journaliere' et 'data_mois' contenant les données regroupées par heure, par jour et par mois avec les entrées (Ei), les intersections (Ii), les sorties (Si), ainsi que d'autres colonnes ajoutées qui nous semblaient pertinentes (ou "suggérées") avec notamment le rapport entrée sur intersection (Ei/Ii) et la somme des entrées 1 à 4, en ajoutant N (somme).

Nous utiliserons par la suite les ratios d'entrées par intersection pour une analyse plus approfondie des flux de circulation, ces ratios permettent de mesurer l'engorgement des entrées sur chaque intersection.

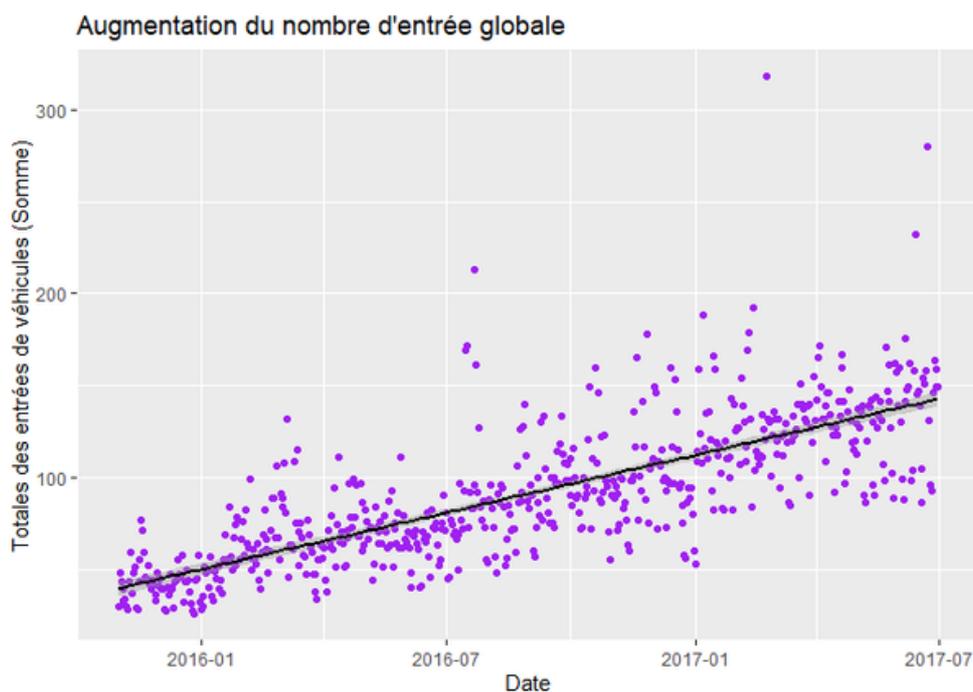
Avec les data frames résultant, nous avons créé des visualisations initiales pour évaluer les tendances générales, en testant des lissages, des modèles linéaires des périodicités.

En résumé, le prétraitement des données a permis d'obtenir des données cohérentes et exploitables pour l'analyse subséquente. Les transformations, agrégations réalisées ont permis de rendre les données plus faciles à traiter et également de nous interroger sur les modèles qui pourraient au mieux les résumer. Les colonnes ajoutées permettront de préciser les analyses futures et contribueront à une meilleure compréhension de l'efficacité des infrastructures de circulation de la ville.

2- Modélisation

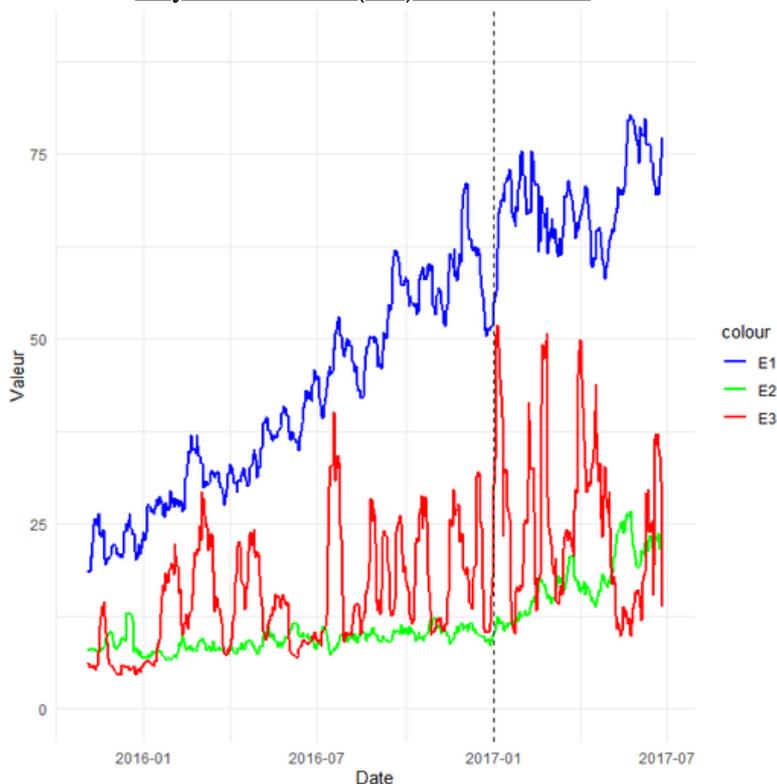
zone de mesure et ratio E_i/I_i	I1	I2	I3	I4	E1/I1	E2/I2	E3/I3	E4/I4
moyenne journalière	63	21	29	5	0,76	0,56	0,61	0,57

Après avoir regrouper les données par jour on peut ici observer les moyennes pour les éléments qui nous intéressent et voir que l'intersection 1 (I1) comprend la majorité du trafic avec en moyenne 63 véhicules par jours contre 55 pour la somme des 3 autres (pour I4 nous avons la moyenne sur la période ou l'intersection est utilisé). Puis nous avons également calculé le rapport $E1/I1$, qui est de $3/4$ environ ce qui nous a amené à penser que l'entrée E1 était donc le point d'engorgement qui posait problème, c'est pour cela que par la suite nous nous sommes focalisés sur cette entrée afin de répondre à l'une des problématiques initiales que notre client ou plutôt notre professeur nous a indiquée : les intersections ont-elles été désengorgées suite à la mise en place de la quatrième ?



Après s'être assuré de la cohérence de nos traitements de données, nous avons commencé à les modéliser. Nous avons rassemblé les données afin d'observer la tendance "globale", et nous pouvons voir que la circulation a globalement augmentée d'une manière régulière. Le graphique ci-dessus montre les entrées E1, E2, E3, E4, N, cette somme permet de mesurer le flux de voitures passant par les intersections de la ville. La tendance concernant les intersections est similaire si l'on rassemble (I1, I2, I3, I4) ou si l'on représente I1 ou E1 seul : on obtient un nuage de points et un modèle linéaire assez similaire. Ensuite nous avons commencé à tester des méthodes de lissage et de régression. Nous en avons sélectionné quelques une dans les pages suivantes.

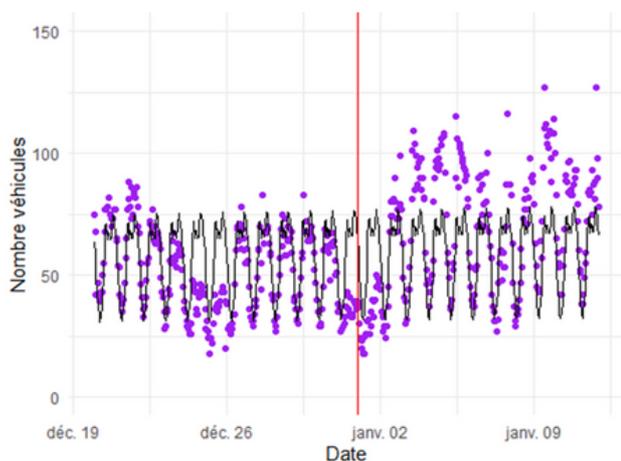
Moyennes Mobiles (k=3) sur les entrées



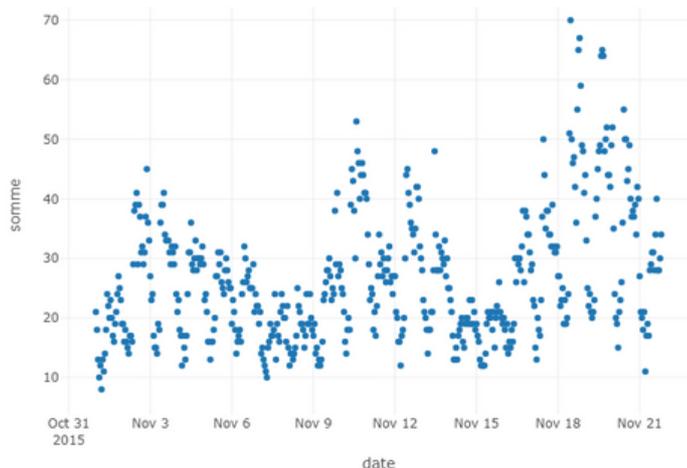
Nous avons testé l'ensemble des méthodes à notre disposition et choisis de mettre en avant cette méthode de lissage : la moyenne mobile. Nous avons choisis un degré k relativement faible de 3 car bien que les données soient chaotiques, nous avons trouvé que ce degré était plus intéressant en terme de représentation visuelle des données. Le lissage exponentielle était également plutôt correct pour visualiser les données. Nous avons observé une certaine forme de périodicité que nous avons explorée par la suite. Ici nous voyons encore que la majorité de la circulation est concentrée sur l'entrée E1.

On peut observer concernant E1 et E2 une très forte variation autour de la date du 1er janvier 2017, ce qui peut nous amener à penser que la mise en place de l'intersection I4 a fortement impacté le trafic pendant plusieurs jours suivant son introduction et peut être qu'elle a impacté et impactera probablement de manière durable la circulation de l'ensemble des intersections autour de cette ville.

Modèle multiplicatif sur une période de 24 heures



Visualisation de la 'somme' sur les trois premières semaines



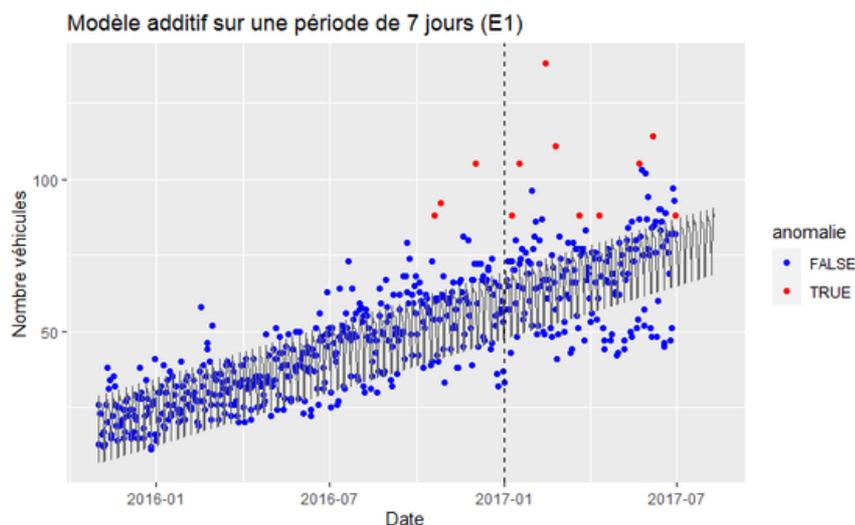
Nous avons réalisé un modèle multiplicatif sur une période de 24 heures sur les données de 'somme' (somme de E1, E2, E3, E4, N), les dates sont comprises entre ("2016-12-20") et ("2017-01-12")) pour rendre le graphique plus lisible car nous avons de très nombreux points et avec une échelle sur l'ensemble des données on a finalement un amas de point avec un modèle qui ressemble plutôt à une grande bande suivant l'amas central de données.

Ce modèle suit plutôt bien la tendance sur les données et nous semblait approprié car le volume de véhicules suit logiquement le rythme de vie des êtres humains et donc, naturellement, le trafic est plus dense autour de 9h et 17h. Aussi, il nous a semblé que la période avait tendance à devenir de plus en plus large sur les données dont nous disposons voila pourquoi nous avons choisis un modèle multiplicatif de saisonnalité de 24 heures.

Finalement parmi l'ensemble des modèles que nous avons pu tester, le meilleur semble être un additif avec une saisonnalité de 7 jours que nous avons exploré par la suite plus en détail en calculant les erreurs notamment.

Le graphique ci-dessus (à droite) montre le flux des trois premières semaines sur l'ensembles des données. on observe un modèle qui se répète hebdomadairement.

3- Anomalies, cassure et prédictions



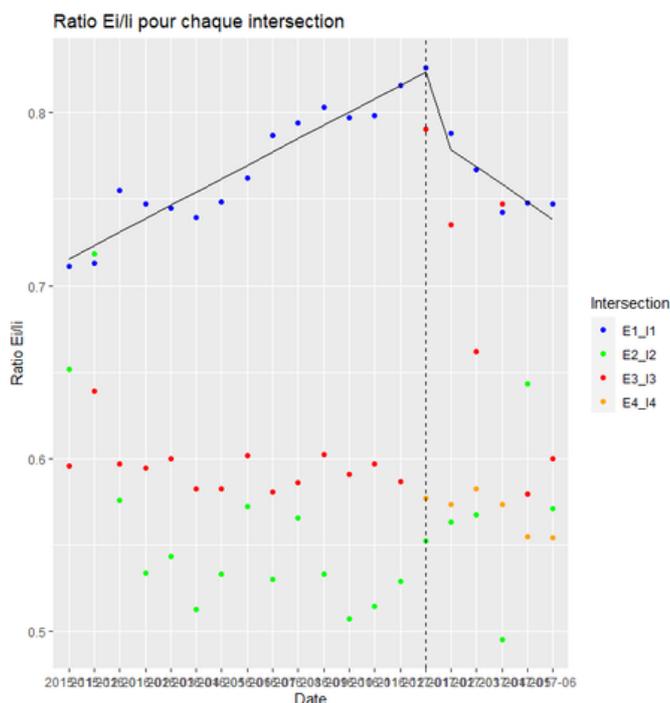
L'un des meilleurs modèles que nous avons pu réaliser est le suivant, c'est un modèle additif composé donc d'une tendance linéaire précédemment observé et d'une saisonnalité (d'une semaine ici). Le modèle (en gris) suit les données de manière assez fidèle, on a également des prédictions sur 40 jours en fin de graphique, et en jaune 2 modèles linéaires permettant de visualiser la cassure qui sera établi de manière précise par la suite.

Concernant les anomalies au vue du contexte, nous avons décidé de ne garder que les valeurs extrêmement hautes car les valeurs basses n'ont pas d'intérêt. On se place dans l'idée que les données ou plutôt ici les résidus du modèle additif suivent une loi normale. Les erreurs de notre modèle considérées comme anormales seront donc les valeurs étant supérieur à 99% des autres valeurs. Autrement dit les résidus ou erreurs (e_i) les plus hautes au seuil de 1%. On peut voir que sur les données de E1, il y a plus d'anomalies après le 2017-01-01 soit le jour d'introduction de la nouvelle intersection I4.

Nous avons également évalué si le moment indiqué par un trait pointillé noir correspond à une cassure à l'aide d'un test de Chow au niveau de risque de 5%. Pour cela nous avons créé deux modèles linéaires en prenant les données avant et après la date de cassure (t_0 : 2017-01-01). Ensuite en utilisant la fonction 'chow.test()' du module gap nous obtenons une F-value que nous avons comparé au quantile de la loi de Fisher à 0.95 avec des degrés de libertés (df_1 et df_2) correspondant à $df_1 = k$ ($k=2$) et $df_2 = n_1 + n_2 + 2k$ (ici n_1 et n_2 sont les tailles d'échantillons). Nous obtenons une F-value de -2.58 qui n'est dans l'intervalle $[3.01; +\infty]$.

Ce test a donc été concluant dans le sens qu'il nous a indiqué qu'aucune cassure existait au risque de 5%, on ne rejette pas l'hypothèse H_0 .

Les prédictions ont été réalisé sur l'ensemble des données et concernent les 40 jours suivant le 30/06/2017 et donc montrent une probable augmentation du nombre de véhicules.



Nous avons représenté le rapport E_i/I_i avec des données regroupées par mois, nous avons représenté l'ensemble des ratios pour chaque intersection. On peut voir autour de l'introduction de l'intersection I4 une forte variation concernant E_1/I_1 et E_2/I_2 .

Le ratio concernant l'intersection I1 a très fortement baissé suite au 1er janvier 2017 et semble continué à diminuer de manière durable sur les mois suivant. Concernant l'utilisation de l'intersection I4 et de son entrée elle est autour de 0.58 et semble plutôt stable.

Les rapports entrée/intersection pour l'intersection 2 et 3 n'ont pas ou peu changé depuis l'arrivé de l'intersection 4, on peut supposer que la diminution du nombre d'entrée dans la ville (donc du ratio E_1/I_1) a été causé par l'utilisation de l'entrée E4 et ainsi que cette nouvelle entrée et plus généralement cette intersection I4 a eu un effet positif sur le trafic. Elle a permis de désengorger l'entrée E1 qui était la plus prisée par les conducteurs souhaitant entrée dans la ville.

Conclusion

Notre projet d'étude a abordé la question de la mobilité urbaine, en se concentrant sur l'impact de l'introduction de la quatrième intersection (I4) sur la circulation d'une ville. Ce projet nous a permis de développer/consolider nos compétences en programmation sur RStudio avec l'utilisation de packages populaires tels que 'dplyr', 'tidyverse' ou encore 'ggplot2'. Il nous a dans le même temps permis d'utiliser notre esprit critique afin de choisir quelles méthodes de lissage ou de modélisation étaient les plus adaptées afin de représenter nos données, tout en gardant à l'esprit l'objectif de répondre à des problématiques précises, telles que la détection d'anomalies, l'évaluation du désengorgement des intersections, et aussi la réalisation de prédictions sur les jours suivant la fin de nos données. Suite aux tests de diverses approches, nous avons finalement sélectionné quelques graphiques issus de l'utilisation d'outils variés afin de répondre de la manière la plus pertinente possible aux demandes par le biais de ce rapport.

À travers l'analyse de données temporelles, nous avons observé une redistribution du trafic, avec une réduction notable de la congestion au niveau de l'intersection I1, grâce à l'utilisation de méthodes statistiques. Cependant, il ne faut pas négliger l'augmentation constante de la circulation qui montre que finalement l'intersection 1 n'a pas été totalement désengorgée, mais l'ouverture de l'intersection 4 a permis de réduire l'augmentation du flux sur l'entrée 1.

Le test de Chow, appliqué à nos séries temporelles, n'a pas montré de cassure significative, cependant il semble que l'augmentation de la circulation au niveau de l'entrée E1 est baissée quelque peu suite à la mise en place de l'intersection I4. Il est important de noter que ces résultats, qui semblent légèrement positifs, ne couvrent qu'une courte période et appellent à une analyse sur le long terme pour confirmer ces premières observations.

Nous avons pu réaliser un test de Chow sur les données d'entrées de E2, qui nous a permis de voir qu'il y a une cassure nous montrant que l'augmentation des entrées en E2 a été plus forte après l'ouverture de la 4e intersection.

Pour conclure l'ouverture de l'intersection 4 n'a pas diminué le flux de données passant par chacune des intersections mais a permis de mieux répartir les flux sur chaque intersection. Cependant les données nous offrent très peu de recul dans le temps suite à l'ouverture de l'intersection (seulement 6 mois).

En élargissant notre perspective, nous reconnaissons que les comportements de mobilité urbaine sont influencés par une multitude de facteurs, y compris les changements sociétaux et technologiques. Il serait donc intéressant d'étendre l'étude à l'impact des nouvelles formes de mobilité, telles que les véhicules autonomes et les services de covoiturage, sur les infrastructures urbaines existantes. Bien que notre étude soit un exercice fictif, elle jette les bases pour des recherches futures et souligne la pertinence d'une planification urbaine éclairée par les données pour répondre aux défis de la circulation dans les villes en croissance.