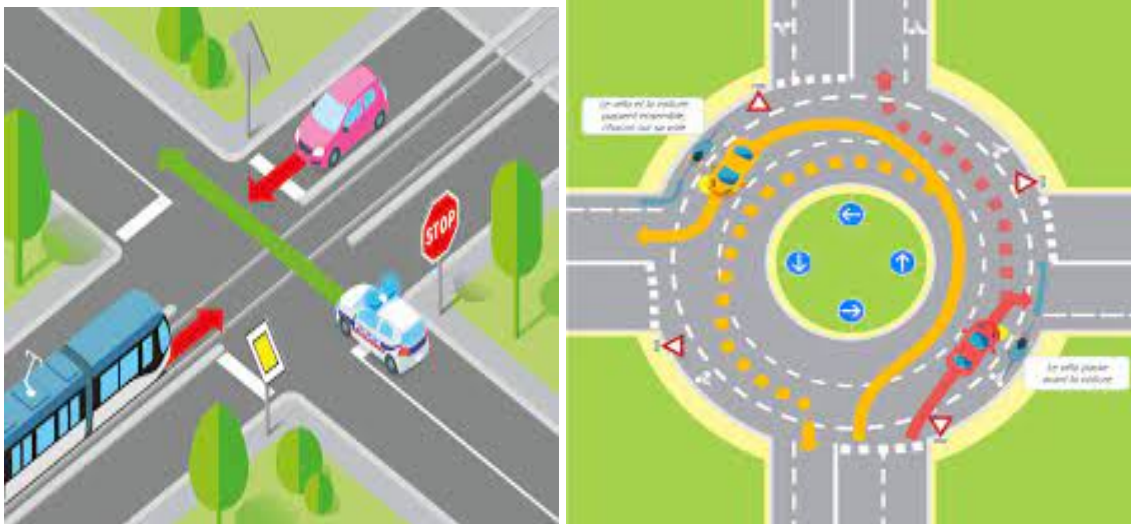


BUT Sciences des données 2023-2024

SAE3-03 Description et prévision de données temporelles



EVINA Alain, DIOP Coumba, NIKIEMA Anaïs, SECK Fatoumata

—

GROLLEMUND Paul Marie

Contexte et problématique

Dans l'optique d'améliorer la fluidité de la circulation et de répondre aux défis de mobilité urbaine, la ville desservie par la route nationale a introduit une nouvelle intersection, inaugurée le 1er janvier 2017. Cette étude, menée dans le cadre de la Science des données en 2ème année à l'IUT d'Aurillac, se concentre sur l'impact de cette modification sur le trafic aux abords de la ville. L'objectif est triple : évaluer l'efficacité de la nouvelle intersection I4 dans la réduction de la congestion aux intersections préexistantes, identifier les tendances anormales de circulation avant et après son introduction, et prédire le flux de trafic pour les périodes à venir.

L'analyse se fonde sur des données précises de flux véhiculaires, recueillies aux quatre intersections (I1, I2, I3, I4) et fournies dans le fichier `route_intersection.csv` sur la plateforme Moodle. Ce projet de SAE 3-03 en Description et prévision de données temporelles consiste à examiner minutieusement ces données à travers un prisme méthodologique robuste, impliquant des techniques de lissage statistique et des modèles prédictifs.

Notre démarche analytique se décompose en plusieurs étapes clés : la transformation des données brutes en informations exploitables pour chaque point d'entrée (E1, E2, E3, et E4), l'application de méthodes de lissage pour cerner la dynamique des données, la détection d'anomalies et la quantification de leur fréquence relative à l'introduction de l'intersection I4. Nous effectuerons également une répartition des flux de véhicules pour chaque intersection et chaque créneau horaire, suivie d'une agrégation temporelle des données pour en extraire des tendances diurnes et mensuelles. Enfin, ces informations nous serviront à élaborer des prévisions à court et moyen terme pour chaque point de passage.

Prétraitement

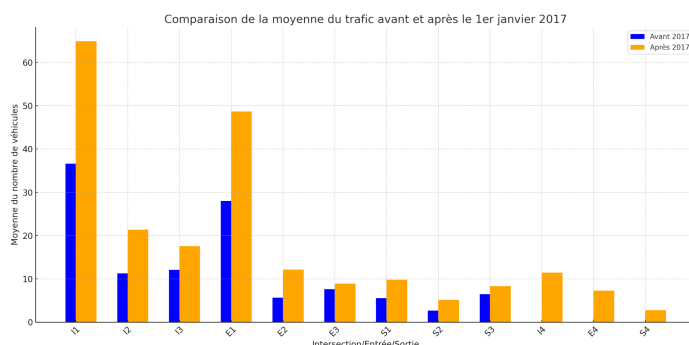
Dans le cadre du projet pour la SAE 3-03 sur la "Circulation et Nouvelle Intersection de Routes", une étape cruciale a été le prétraitement des données extraites du fichier `route_intersection.csv`. Ce processus a joué un rôle fondamental dans l'assurance de la qualité et de la précision des analyses subséquentes. Initialement, les données ont été agrégées par date et par type d'intersection, une démarche essentielle pour appréhender la dynamique de circulation à chaque point de contrôle. Cette agrégation a permis de structurer les données de manière à refléter avec exactitude les flux de trafic à travers les intersections I1, I2, I3, et la nouvelle intersection I4. Par la suite, une transformation approfondie du Data Frame a été entreprise pour distinguer les valeurs spécifiques des entrées (E1, E2, E3, E4) et des sorties (S1, S2, S3, S4) de la ville. Ce raffinement des données a été fondamental pour isoler l'effet de chaque intersection sur le flux global de trafic. Dans ce contexte, il était impératif de tenir compte des spécificités de la circulation :

chaque nombre de véhicules enregistré à un créneau horaire donné dans une intersection I_j correspondait uniquement aux véhicules provenant de l'intersection précédente (I_{j-1}) et de la ville (S_j) durant le même créneau. De même, les véhicules sortant de l'intersection I_j étaient considérés comme se dirigeant soit vers la ville (E_j) soit vers l'intersection suivante (I_{j+1}) dans le même créneau horaire. Ensuite, nous avons effectué une troisième transformation pour dériver les valeurs spécifiques des entrées ($E1, E2, E3, E4$) à partir des données des intersections et des sorties. Cette étape impliquait l'application de formules détaillées :

- $E1$ a été calculé comme $I1 - (I2 - S2)$
- $E2$ comme $I2 - (I3 - S3)$
- $E3$, différemment selon la période, soit comme $I3 - N$ (avant le 1er janvier 2017) ou $I3 - (I4 - S4)$ (après le 1er janvier 2017)
- et $E4$, introduit après le 1er janvier 2017, comme $I4 - N$.

Analyse descriptive des données

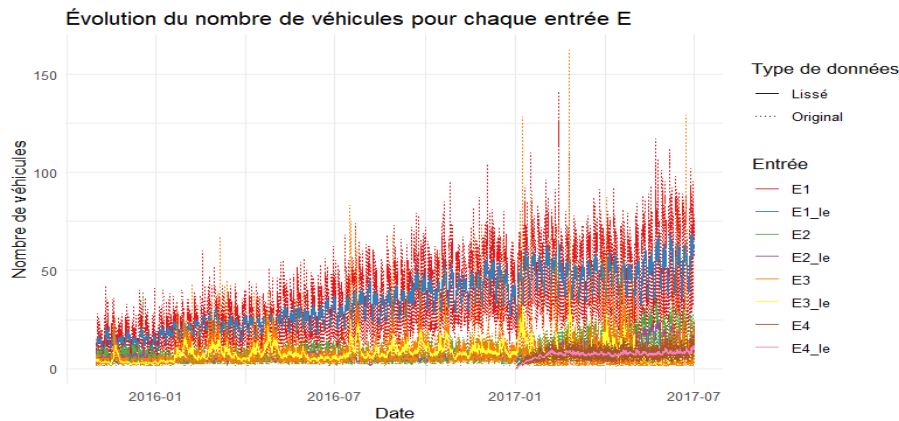
Avant le 1er janvier 2017, les intersections $I1, I2$ et $I3$ affichaient des moyennes de trafic de 36.62, 11.25 et 12.05 véhicules respectivement, avec des pics de congestion notables atteignant jusqu'à 120 véhicules pour $I1$, 40 pour $I2$ et 134 pour $I3$. Après cette date, une augmentation significative a été observée : $I1$ est passé à 64.94 véhicules, $I2$ à 21.34 et $I3$ à 17.56. L'introduction de l'intersection $I4$ a introduit une moyenne de 11.44 véhicules, avec des pics de trafic culminant à 156 véhicules pour $I1$, 48 pour $I2$, 180 pour $I3$ et environ 37 pour $I4$.



En observant ces tendances, il est clair que toutes les intersections et entrées ont enregistré une augmentation significative du trafic après l'introduction de l'intersection $I4$. Cette variation importante avant et après 2017 suggère une grande instabilité du trafic, avec des pics élevés persistants malgré la redistribution induite par l'intersection $I4$. Toutes les intersections et entrées ont montré une augmentation notable de la moyenne du trafic après le 1er janvier 2017, avec des doublement de trafic pour certaines, comme $I1$ et $E1$. L'introduction de l'intersection $I4$ a significativement affecté le flux de trafic, absorbant une part importante. De même, les sorties $S2$ et $S3$ ont connu une hausse de trafic, indiquant une augmentation du trafic quittant la ville après 2017. Seule l'entrée $E3$ a montré une légère diminution, suggérant un rééquilibrage suite à l'introduction de l'intersection $I4$. Ces tendances

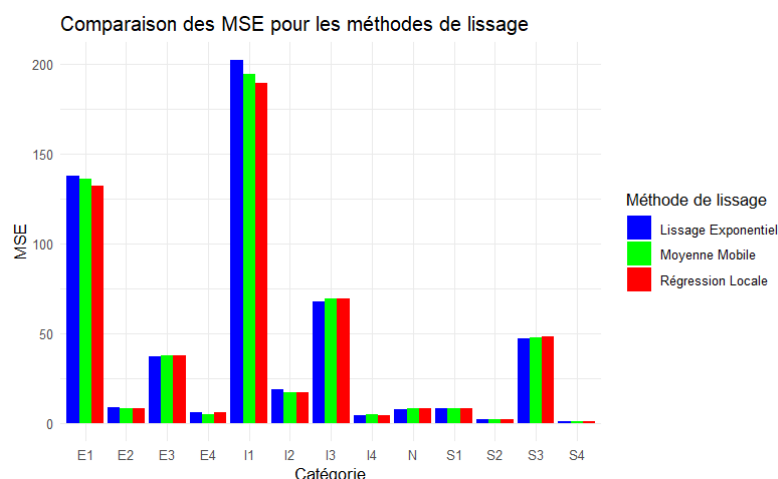
suggèrent un accroissement général du trafic, peut-être dû à une croissance de la population ou à une utilisation croissante des véhicules, nécessitant une attention particulière en matière de gestion de la circulation et de planification urbaine.

Modélisation

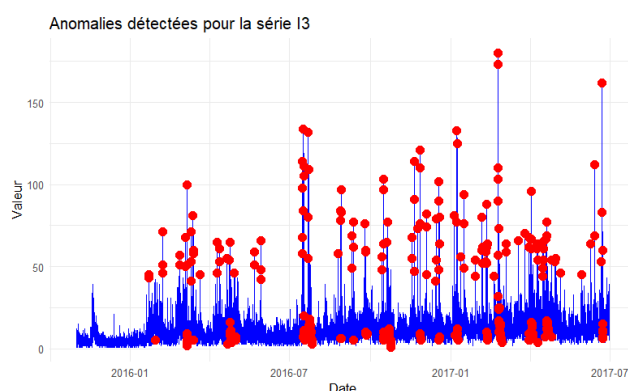


Il existe une variabilité considérable dans le nombre de véhicules passant par chaque entrée (E1, E2, E3, E4), ce qui peut refléter les heures de pointe et les périodes creuses habituelles. En outre, une tendance générale à l'augmentation du trafic est observable, en particulier à l'approche de l'introduction de la nouvelle intersection E4 en janvier 2017. Cette augmentation pourrait indiquer une croissance du trafic ou une réponse aux changements d'infrastructure. L'introduction de l'intersection E4 a entraîné un pic notable dans le volume du trafic, suggérant que l'E4 a absorbé une partie significative du trafic ou a encouragé davantage de véhicules à utiliser cet itinéraire. En comparant les données originales et lissées, on constate que les données originales présentent de nombreuses variations à court terme, tandis que les données lissées révèlent une tendance plus claire sur une période plus longue. Enfin, avant l'introduction de E4, E1, E2 et E3 présentaient des distributions assez similaires, E1 ayant tendance à avoir un volume légèrement supérieur. Après l'introduction de E4, le trafic semble être devenu plus équilibré entre les entrées.

Évaluation des méthodes de lissage basée sur le MSE



On remarque une grande variabilité des MSE entre les différentes intersections et méthodes de lissage, signalant que certaines méthodes conviennent mieux à certaines séries de données. On constate des valeurs élevées de MSE pour E4 et I4, suggérant des difficultés des modèles de lissage à saisir la dynamique de ces séries, peut-être liées à l'introduction récente de l'intersection E4 et à des schémas de trafic instables. L'analyse montre que le lissage exponentiel a le MSE le plus bas pour les séries E1, I1 et I3, suggérant une meilleure performance pour ces séries. Pour S1 et S3, on observe que la moyenne mobile et la régression locale respectivement semblent plus efficaces. En général, on constate que la régression locale affiche des MSE relativement bas pour la plupart des séries, suggérant sa cohérence et sa fiabilité comme méthode de lissage dans ce contexte.

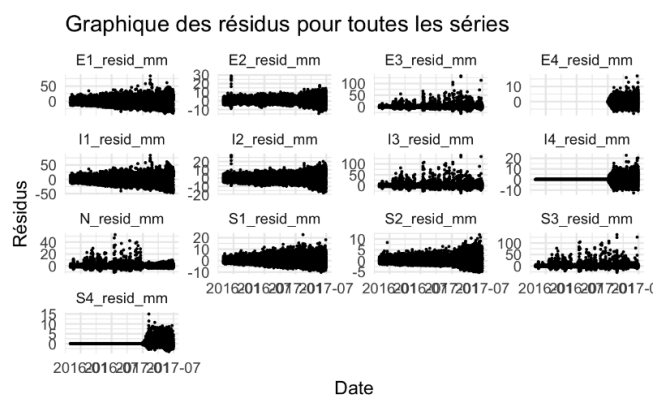


Le premier graphique illustre les anomalies détectées pour l'intersection I3, marquées par des points rouges. Ces anomalies, identifiées entre 1% et 99% des quantiles, suggèrent des valeurs atypiques qui pourraient correspondre à des événements exceptionnels ou à des erreurs de mesure. La concentration de ces points autour de certaines périodes peut indiquer des incidents routiers, des changements de signalisation ou des augmentations temporaires du trafic.

Les tables fournissent un récapitulatif des anomalies avant et après l'introduction de l'intersection E4. Une augmentation nette des anomalies est observée après cette

date pour la plupart des séries, en particulier pour I1, I2, I3 et S1, ce qui pourrait refléter l'impact des changements de circulation ou l'effet d'une augmentation générale du trafic. Notamment, E4 montre une forte augmentation des anomalies après son introduction, passant de 0 à 88, ce qui indique un ajustement significatif dans le trafic à cette intersection. Pour N, il y a une diminution drastique des anomalies après 2017, passant de 290 à 2, suggérant que la situation de circulation s'est stabilisée ou que les mesures de trafic ont été normalisées.

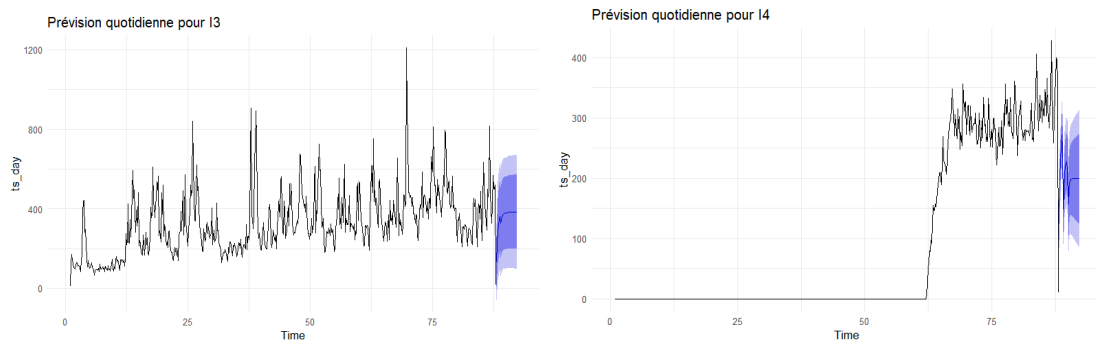
Cette analyse des anomalies révèle l'importance d'une surveillance continue du trafic et de l'ajustement des infrastructures et des politiques de circulation pour accommoder les évolutions dans le comportement des conducteurs et les conditions de route.



On a utilisé le graphique des résidus pour évaluer la performance du modèle de lissage. Cela nous a permis de voir comment les résidus sont répartis autour de zéro pour différentes séries de données. On a observé une dispersion uniforme des résidus, cela indique que la moyenne mobile s'ajuste bien aux données sans biais systématique.

Cependant, on remarque que certaines séries, comme I3 et S3, présentent une dispersion plus large, signalant une variabilité non capturée par le modèle. On peut identifier des points éloignés de zéro dans des séries comme E3, I3 et S3, qui pourraient indiquer des événements exceptionnels. D'autres séries, telles que E1, E2, I1, I2 et S1, montrent une répartition stable des résidus, suggérant un trafic régulier et prévisible bien capturé par la moyenne mobile. On a observé des résidus significatifs dans certaines séries, cela suggère qu'on devrait réviser le modèle de lissage ou ajouter des variables pour améliorer la prédiction du trafic. La série S4 montre très peu de données, ce qui suggère que cette intersection a été introduite récemment ou qu'il y a une disponibilité limitée de données historiques. En résumé, l'analyse des résidus nous aide à identifier où le modèle ne correspond pas aux données, permettant ainsi des ajustements pour améliorer nos modèles de prévision du trafic.

Prévisions et visualisations



Ces deux graphiques illustrent les prévisions quotidiennes pour les intersections I3 et I4, respectivement, utilisant un modèle de série temporelle tel que ARIMA.

Prévision Quotidienne pour I3:

Le trafic historique montre des fluctuations notables avec des pics élevés, indiquant des jours de trafic particulièrement intense. Le modèle prévoit une continuité de la variabilité du trafic avec une légère tendance à l'augmentation. L'intervalle de confiance s'élargit sur la période de prévision, signifiant une incertitude croissante dans les prédictions futures.

Prévision Quotidienne pour I4:

Les données historiques commencent plus tard dans la série temporelle, ce qui correspond à l'ouverture de l'intersection I4. La prévision montre une stabilisation du trafic après une période initiale d'augmentation. Comme pour I3, l'incertitude augmente avec le temps, comme le montre l'éventail plus large de l'intervalle de confiance.

Pour les deux intersections, les prévisions doivent être interprétées avec prudence en raison de l'incertitude inhérente aux modèles prédictifs, surtout lorsqu'elles sont appliquées à des données de trafic qui peuvent être influencées par de nombreux facteurs imprévisibles.

Conclusions et Interprétations

Après une analyse approfondie des données de trafic avant et après l'introduction de l'intersection I4, il est clair que cette modification a considérablement impacté la dynamique de circulation. Les résultats mettent en évidence une augmentation significative du trafic à toutes les intersections, avec des pics plus élevés et une variabilité accrue après 2017. Bien que l'introduction de l'intersection I4 ait contribué à redistribuer le trafic, elle n'a pas résolu les problèmes de congestion, comme en témoignent les pics de trafic plus importants et la variabilité persistante. Ces tendances soulignent la nécessité d'une gestion de la circulation plus proactive et d'une planification urbaine adaptative pour faire face à l'évolution du comportement des conducteurs et des conditions de la route.