

Légende du voyageur transfrontalier. J'y vais, ou pas?

Tanja Armenski, Zdenek Patak et Steve Matthews¹

Résumé

Cet article intègre des données sur la météo et la circulation pour permettre de mieux comprendre l'incidence de la météo sur la circulation transfrontalière et d'élaborer des techniques de modélisation pouvant être adaptées à la production de prévisions précises. Des données sur la circulation sont recueillies par l'Agence des services frontaliers du Canada (ASFC) et sont produites par Statistique Canada dans le cadre du Programme de dénombrement à la frontière (DF), qui regroupe des données administratives sur toutes les arrivées internationales au Canada. Les données météorologiques comme la température, les précipitations et les bordées de neige ont été obtenues auprès d'Environnement et Changement climatique Canada. Des modèles ARIMA intégrant des régresseurs liés à la météo et aux jours fériés confirment que les phénomènes météorologiques importants comme les bordées de neige et la pluie verglaçante entraînent une diminution du nombre de voyageurs qui traversent la frontière. Les jours fériés représentent également des écarts importants par rapport à la norme et doivent être pris en compte au moment de modéliser et de prévoir les débits de circulation transfrontalière.

Mots clés : Intégration des données, ARIMA, météo, circulation transfrontalière.

1. Introduction

La circulation transfrontalière est un indicateur économique clé du commerce international et du tourisme. Les résidents des États-Unis qui arrivent au Canada en voiture forment la plus grande part de toutes les arrivées internationales. La météo – et en particulier les conditions météorologiques extrêmes – est souvent considérée comme un facteur qui influence le débit de circulation. De longues périodes de températures inhabituellement basses ou élevées ou des bordées de neige importantes peuvent avoir une incidence sur la décision d'une personne de traverser la frontière. Les données météorologiques peuvent expliquer au moins en partie la variation du débit automobile dans les séries chronologiques sur la circulation.

La présente étude vise deux objectifs. En premier lieu, elle intègre des données de deux sources distinctes, i) les données sur la circulation de l'Agence des services frontaliers du Canada (ASFC), compilées par Statistique Canada dans le cadre du Programme de dénombrement à la frontière (DF), et ii) les données météorologiques d'Environnement et Changement climatique Canada. En second lieu, elle vise à fournir une approche plus systématique intégrant la météo à l'estimation des débits de circulation transfrontalière entre le Canada et les États-Unis. Les principaux résultats révèlent que les bordées de neige importantes et la pluie verglaçante ont une incidence significative sur la circulation transfrontalière. Chaque phénomène météorologique entraîne une diminution du débit de circulation automobile transfrontalier. La circulation transfrontalière diminue aussi de manière significative lors de certains jours fériés au Canada tandis qu'elle augmente à l'occasion d'autres fêtes.

1.1 Examen de la documentation

La relation entre la météo et la circulation est largement reconnue. Deux constatations sont invariablement faites. En premier lieu, il existe un lien entre les débits de circulation et les conditions météorologiques extrêmes. Datla et coll.

¹ Tanja Armenski, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa, Canada, K1A 0T6; Zdenek Patak, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa, Canada, K1A 0T6; Steve Matthews, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa, Canada, K1A 0T6

(2013) font la preuve que le lien entre les débits de circulation en Alberta et le temps froid et enneigé varie selon l'intensité des conditions météorologiques. Les températures chaudes en été entraînent une augmentation des débits de circulation tandis qu'en hiver, les températures froides accompagnées de bordées de neige entraînent une diminution de la circulation. Roh et coll. (2016) révèlent qu'il existe une plus grande diminution des débits des voitures que des débits des camions par temps froid en Alberta. La diminution des débits de circulation des voitures et des camions s'accroît avec de grandes quantités de neige compte tenu de l'interaction du froid et de la neige.

En second lieu, l'incidence des mauvaises conditions météorologiques sur la circulation varie selon le jour de la semaine. Datla et coll. (2013) révèlent que les schémas de débits sont davantage touchés par le temps froid les fins de semaine comparativement aux jours de semaine. On s'intéresse aujourd'hui de plus en plus à la modélisation de l'incidence des conditions météorologiques sur les systèmes de gestion de la circulation en temps réel, soit la prévision immédiate. Cette approche de modélisation a au départ été appliquée aux conditions météorologiques afin d'accroître la précision spatiale des conditions météorologiques et de leur incidence, mais a ensuite été adoptée pour les systèmes de gestion de la circulation en vue d'accroître l'exactitude de la prévision de la circulation en temps réel (Lin et coll., 2015). Étant donné la disponibilité de données de haute fréquence (quotidiennes) diffusées seulement dix jours après la fin du mois de référence ou six semaines avant le Programme de dénombrement à la frontière, le Programme moderne de la statistique du tourisme franchit aussi une autre étape vers la collecte de données sur la circulation en temps réel. Le présent document est organisé comme suit : la section 2 décrit les données sur la circulation et les données météorologiques; la section 3 explique différentes méthodes d'estimation et les résultats obtenus, et est suivie d'une conclusion présentée à la section 4.

2. Sources des données et variables

2.1 Données sur la circulation

L'Agence des services frontaliers du Canada (ASFC) recueille des renseignements sur les voyageurs qui entrent au pays. Le Système intégré de gestion de la ligne d'inspection primaire (SIGLIP) est un système automatisé qui enregistre le nombre quotidien de véhicules qui entrent au Canada par les points d'entrée terrestres. Tous les grands points d'entrée de l'ASFC ont intégré le SIGLIP pour remplacer le formulaire papier d'entrée à la frontière dans le but de moderniser leurs systèmes statistiques.

Le SIGLIP génère automatiquement des renseignements en format numérique, ce qui réduit les besoins en matière de traitement des données et accroît l'actualité des données sur la circulation. Le SIGLIP est une composante importante des données administratives sur le dénombrement à la frontière (DF), car il couvre tous les grands points d'entrée terrestres de la circulation automobile internationale au Canada et il est utilisé dans la création d'un indicateur avancé du volume de voyageurs transfrontaliers, enregistrant les voyageurs qui entrent et qui reviennent au Canada en voiture. La circulation transfrontalière des voitures canadiennes qui reviennent des États-Unis est analysée dans le présent document. Les partenaires du tourisme et les usagers provinciaux peuvent maintenant observer les variations quotidiennes dans la circulation transfrontalière des véhicules. L'intégration de données de remplacement à l'appui des besoins statistiques est une des principales initiatives de modernisation du Programme moderne de la statistique du tourisme.

2.2 Point d'entrée terrestre d'intérêt

Le point d'entrée terrestre choisi pour cette étude est Cornwall², un poste frontalier terrestre situé dans le sud-est de l'Ontario qui relie l'autoroute 138 en Ontario à l'autoroute 37 dans l'État de New York. Le passage frontalier est associé au pont international de la Voie maritime, qui traverse les deux voies navigables de chaque côté de l'île de Cornwall située dans le fleuve Saint-Laurent. Des données sur la circulation ont été recueillies quotidiennement pour la période allant du 1^{er} mars 2016 au 30 avril 2018. Des données météorologiques quotidiennes sont disponibles auprès

² Tous les points d'entrée terrestres de l'Ontario ont été étudiés. Cornwall a été choisie pour illustrer les types d'analyses effectuées sur les données.

d'Environnement Canada pour la période correspondante. Les statistiques météorologiques moyennes ainsi que les chiffres de la circulation transfrontalière correspondants pour la période à l'étude sont présentés au tableau 2.2-1.

Tableau 2.2-1

Données sur la circulation et le climat pour la période à l'étude (1^{er} mars 2016 au 30 avril 2018)

Temp max moyenne °C	Précipitations annuelles moyennes mm	Chutes de neige annuelles moyennes cm	Nombre moyen de voitures par jour	Nombre moyen de voitures jour de semaine	Nombre moyen de voitures fin de semaine
12,3	798,1	234,3	1 407	1 400	1 424

Lorsque le nombre moyen de voitures ayant une plaque d'immatriculation canadienne qui reviennent des États-Unis est réparti selon le jour de la semaine, le vendredi présente la plus forte circulation et le mardi, la plus faible. La différence entre les deux jours est d'environ 150 voitures.

2.3 Données météorologiques

Les données météorologiques ont été obtenues auprès d'Environnement et Changement climatique Canada. Les observations météorologiques quotidiennes comme les températures maximale et minimale, les précipitations et les bordées de neige sont obtenues à partir de chacune des stations météorologiques dans tout le Canada. Les coordonnées géographiques ont été utilisées pour choisir les stations météorologiques situées le plus à proximité des points d'entrée terrestres à la frontière entre les États-Unis et le Canada en Ontario. Les variables météorologiques utilisées dans cette analyse comprennent a) la température maximale (degrés Celsius); b) des indicateurs de bordées de neige importantes (égaux à un lorsque la bordée de neige dépassait 5 cm et à zéro autrement) les jours de semaine et les fins de semaine; c) un indicateur de pluie verglaçante (égal à un lorsque les précipitations étaient supérieures à 0 et la température de l'air était inférieure à 2 ° C³ et à zéro autrement); et d) des indicateurs de l'accumulation de neige la veille les jours de semaine et les fins de semaine.

2.4 Effets des jours fériés

Des études initiales ont été menées pour étudier l'incidence des jours fériés sur la circulation transfrontalière malgré les conditions météorologiques. Une circulation considérablement plus dense a été observée pendant les jours fériés comme la fête de la Reine, le Congé civique, la fête du Canada, la fête du Travail, le lundi de Pâques et l'Action de grâce, et une circulation plus faible le jour de Noël et la veille du jour de l'An. Deux variables nominales ont été utilisées pour représenter l'effet des jours fériés sur les débits de circulation. La première variable nominale prend la valeur de un lorsque le jour férié a une incidence positive sur la circulation et la valeur de zéro autrement. La seconde variable nominale prend la valeur de un lorsque le jour férié a une incidence négative sur la circulation et la valeur de zéro autrement. Le regroupement des jours fériés n'est peut-être pas idéal, mais la longueur limitée de la série ne permettait pas de construire des régresseurs pour chacun des jours fériés afin de tirer avantage de leur pleine capacité de prévision. À mesure que davantage de données deviendront disponibles, les régresseurs de jour férié seront affinés.

3. Méthodes et résultats

3.1 Déterminer les décalages importants (ou fréquences)

Pour analyser des données tirées de séries chronologiques, la première étape consiste généralement à déterminer l'ordre des composantes autorégressives et de moyennes mobiles non saisonnières et saisonnières des séries en analysant les fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF) à divers décalages (p. ex., une

³Pour de plus amples renseignements concernant les normes de formation, la répartition et la fréquence de la pluie verglaçante en Ontario, veuillez consulter les travaux de Cortinas (2000).

tendance hebdomadaire représente un décalage de sept jours). Pour y arriver de façon fiable, une série doit être faiblement stationnaire, c.-à-d. que la série chronologique doit présenter une moyenne constante et une autocovariance par rapport au temps (voir Box et Jenkins, 1976). Si une tendance est présente dans les données, le critère de la stationnarité est violé. Il s'agit d'une étape importante, car l'absence de stationnarité peut mener à des conclusions erronées. La stationnarité moyenne peut être vérifiée au moyen du test de Dickey-Fuller augmenté (DFA).

Les graphiques des fonctions ACF et PACF de la série chronologique de la circulation transfrontalière à Cornwall indiquaient une saisonnalité aux décalages 7 (hebdomadaire) et 365 (annuelle). Cette constatation appuie la notion selon laquelle les jours de la semaine et les mois de l'année ne sont pas tous égaux. On s'attendrait à plus de déplacements les fins de semaine et en été. D'autres paramètres autorégressifs ou de moyennes mobiles peuvent être obtenus en analysant les graphiques des fonctions ACF et PACF des valeurs résiduelles du modèle. Un modèle bien ajusté doit présenter des erreurs indépendantes et à peu près normalement distribuées, et les graphiques des fonctions ACF et PACF ne doivent pas présenter de décalages importants. La section suivante décrit les différents modèles utilisés dans l'analyse et les résultats correspondants en fonction d'une séquence hiérarchique conçue pour montrer comment l'erreur quadratique moyenne résiduelle est réduite tandis que le modèle de base est modifié pour inclure d'autres variables explicatives.

3.2 Modèle ARIMA

Le modèle ARIMA bien établi peut être utilisé pour expliquer les variations de la circulation transfrontalière. Le modèle de base débute par une formule ARIMA purement saisonnière sans régresseurs qui peut être exprimée comme suit :

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D y_t = \theta(B)\Theta(B^s)\varepsilon_t, d, D \geq 0 \quad (1a)$$

où $\phi(\cdot)$ et $\Phi(\cdot)$ spécifient les composantes autorégressives non saisonnières et saisonnières du modèle, et $\theta(\cdot)$ et $\Theta(\cdot)$ indiquent les composantes de moyennes mobiles non saisonnières et saisonnières du modèle. L'opérateur de décalage est désigné par B . Les différences non saisonnières et saisonnières sont désignées par d et D . La variable y_t est la valeur de la série chronologique au temps t , et ε_t représente des termes d'erreur satisfaisant $\varepsilon_t \sim N(0, \sigma^2)$, $Cov(\varepsilon_t, \varepsilon_s) = \begin{cases} 0, & \text{pour } t \neq s \\ \sigma^2, & \text{pour } t = s \end{cases}$.

Le modèle de base final a été établi comme étant $(1,1,1)(1,0,1)_7(0,0,1)_{365}$. Cela signifie que, suivant l'application de la méthode des différences premières à la série chronologique, les composantes AR(1), AR(7), MA(1), MA(7) et MA(365) ont été incluses. Le modèle correspondant peut être exprimé comme suit :

$$Y_t = (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + \Phi_7 Y_{t-7} - \Phi_7 Y_{t-8} + \theta_1 \varepsilon_{t-1} + \Theta_7 \varepsilon_{t-7} + \Theta_{365} \varepsilon_{t-365} + \varepsilon_t \quad (2a)$$

Les valeurs résiduelles du modèle qui en ont résulté étaient indépendantes et à peu près normalement distribuées. Cependant, les tests diagnostiques sur les valeurs résiduelles indiquaient une distribution à queue lourde et la présence de valeurs aberrantes qui pourraient signaler une inadéquation du modèle. Cette possibilité est abordée aux étapes suivantes tandis que le modèle de base évolue.

On suppose que la météo, en particulier les périodes de froid extrême et les grandes accumulations de neige et de pluie verglaçante, influence la décision des gens de se déplacer. Pour vérifier cette hypothèse, un certain nombre de variables liées à la météo ont été ajoutées au modèle ARIMA de base. Ces variables comprenaient la température quotidienne maximale, un indicateur de pluie verglaçante, des indicateurs de bordée de neige importante les jours de semaine et les fins de semaine, et des indicateurs de bordée de neige importante la veille les jours de semaine et les fins de semaine.

Le modèle augmenté peut être exprimé comme suit :

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D (y_t - \sum_i \beta_{1i} x_{1it}) = \theta(B)\Theta(B^s)\varepsilon_t, d, D \geq 0 \quad (1b)$$

où x_{1it} est la valeur de la variable explicative $1i$ au temps t , β_{1i} est le coefficient de régression associé à la variable $1i$. Les autres paramètres sont identiques à ceux de l'équation 1a.

Le modèle augmenté final intégrant les régresseurs météo était $(1,1,1)(1,0,1)_7(0,0,1)_{365}$, ou

$$Y_t = (1 + \phi_1)(Y_{t-1} - \sum_i \beta_{1i} x_{1i,t-1}) - \phi_1(Y_{t-2} - \sum_i \beta_{1i} x_{1i,t-2}) + \Phi_7(Y_{t-7} - \sum_i \beta_{1i} x_{1i,t-7}) - \Phi_7(Y_{t-8} - \sum_i \beta_{1i} x_{1i,t-8}) + \theta_1 \varepsilon_{t-1} + \theta_7 \varepsilon_{t-7} + \theta_{365} \varepsilon_{t-365} + \varepsilon_t \quad (2b)$$

3.3 Résultats

La distribution des valeurs résiduelles était toujours à queue lourde, car les valeurs aberrantes n'avaient pas encore été abordées à ce moment dans l'évolution du modèle. Toutes les variables météorologiques étaient significatives. Les estimations du modèle révèlent que les bordées de neige présentent un obstacle majeur pour les déplacements transfrontaliers. L'incidence de la neige était plus importante les fins de semaine, alors que les gens se déplacent davantage pour le plaisir que pour les affaires et peuvent être plus susceptibles de pouvoir changer leurs plans de déplacements en conséquence. La température maximale, bien que significative sur le plan statistique, n'est pas pertinente d'un point de vue empirique, étant donné que le nombre de voitures traversant la frontière change de façon marginale (de l'ordre de trois voitures par degré Celsius).

L'ajout de variables météorologiques au modèle de base réduisait la lourdeur des queues de la distribution des valeurs résiduelles, mais un certain nombre de valeurs aberrantes demeuraient. Certains écarts importants par rapport à la tendance hebdomadaire pouvaient être attribués à différents jours fériés au Canada. Le modèle intégrant un régresseur de jour férié était identique à l'équation 2b, $\beta_{1i} x_{1it}$ étant remplacé par $\beta_{2i} x_{2it}$. En incluant deux variables indicatrices de jour férié (une pour l'incidence positive des jours fériés et l'autre pour l'incidence négative), les régresseurs réduisaient davantage la lourdeur des queues de la distribution des valeurs résiduelles. Les estimations pour les variables météorologiques sont semblables à celles obtenues avec le modèle précédent, mais les deux régresseurs de jour férié présentent maintenant une grande incidence sur les déplacements. Les jours fériés « positifs » (comme la fête de la Reine, la fête du Travail, etc.) augmentent le débit de circulation de 574 voitures en moyenne, tandis que les jours fériés « négatifs » (comme le jour de Noël, le lendemain de Noël, etc.) présentent une diminution moyenne de 421 véhicules par rapport à la tendance hebdomadaire bien établie. Il s'agit d'écarts importants par rapport à la norme qui doivent être pris en compte au moment de modéliser et de prévoir les débits de circulation transfrontalière.

3.4 Critère de la qualité de l'ajustement

La qualité de l'ajustement d'un modèle statistique décrit à quel point ce dernier s'ajuste à un ensemble d'observations. Les mesures de la qualité de l'ajustement résument généralement les divergences entre les valeurs observées et les valeurs attendues dans le modèle ajusté. Pour évaluer le rendement des trois modèles décrits ci-dessus, la variance des valeurs résiduelles du modèle et le critère d'information d'Akaike ($AIC = 2k - 2\ln\hat{L}$, où k est le nombre de paramètres estimés dans le modèle et \hat{L} est la valeur maximale de la fonction de vraisemblance pour le modèle) sont calculés. Le modèle parmi les trois candidats qui minimise la variance résiduelle et la perte d'information est choisi.

Tableau 3.4-1

Statistiques des mesures de la qualité de l'ajustement

	Modèle ARIMA de base	+ Régresseurs météo	+ Régresseurs de jour férié
Variance résiduelle	18 087	15 401	11 339
AIC	9 999	9 881	9 644

Le tableau révèle que l'ajout de régresseurs météo au modèle ARIMA de base réduit la variance résiduelle de 15 %. Une autre amélioration de 26 % peut être réalisée en intégrant deux régresseurs de jour férié. Les chiffres indiquent clairement que l'ajout de régresseurs pertinents entraîne d'importants gains d'efficacité. À mesure que le modèle

évoluera et suivra de plus près les données observées, il pourrait être utilisé pour produire des prévisions fiables et efficaces, ce qui est le but ultime de cet exercice de modélisation.

4. Conclusion

De nombreux facteurs ont une incidence sur la circulation transfrontalière des véhicules. Cette étude révèle deux grands facteurs, la météo et les jours fériés. Les bordées de neige importantes et la pluie verglaçante sont des phénomènes météorologiques qui entraînent une diminution importante du nombre de voyageurs qui traversent la frontière. Cette constatation n'est pas étonnante, étant donné que les conditions routières correspondantes deviennent plus dangereuses et que les autorités encouragent les voyageurs à réduire au minimum leurs déplacements. L'incidence des mauvaises conditions météorologiques sur la circulation variait selon le jour de la semaine, les schémas de débits étant davantage touchés par les bordées de neige les fins de semaine comparativement aux jours de semaine. Les personnes qui se déplacent la fin de semaine peuvent avoir plus de souplesse dans la décision de se déplacer selon les conditions météorologiques à un moment donné, tandis que ceux qui se déplacent régulièrement les jours de semaine, vraisemblablement pour le travail ou les affaires, portent moins attention aux variations météorologiques. L'effet décalé d'une bordée de neige sur la diminution de la circulation était évident, ce qui indique que, lorsque survient une importante bordée de neige tard en journée, les conditions routières le lendemain pourraient être encore suffisamment dangereuses pour avoir une incidence sur les plans de déplacements.

La diminution de la circulation automobile s'accroît avec l'interaction du temps froid et de la pluie. L'étude révèle une incidence négative de la pluie verglaçante sur les débits de circulation aux postes frontaliers. La chaussée glacée rend la conduite automobile potentiellement dangereuse et incite les gens à ne pas prendre la route. L'effet de la température a révélé que les débits de circulation quotidiens moyens augmentent de manière significative par temps chaud. La variable de température maximale, bien que significative, n'est pas vraiment pertinente étant donné que la circulation transfrontalière augmente de seulement quelques voitures tandis que les températures atteignent le maximum quotidien. Différents jours fériés des deux côtés de la frontière peuvent également entraîner une augmentation ou une diminution de la circulation transfrontalière. L'augmentation des débits de circulation la plus importante a été enregistrée le jour de la fête de la Reine. Pour de nombreux voyageurs, cette fête signifie l'arrivée du redoux et possiblement le début de la saison touristique estivale. En revanche, le jour de Noël ressort comme période de déplacements réduits.

L'analyse et les résultats correspondants présentés dans le présent document représentent une première étape en vue de comprendre les données et de déterminer les modèles qui fournissent le meilleur ajustement et qui pourraient être utilisés pour prévoir le dénombrement à la frontière en fonction de la météo, des jours fériés et d'autres régresseurs pertinents. C'est là le but ultime du projet, la prévision en général, et possiblement la prévision immédiate en particulier, car des prévisions fiables pourraient être utilisées dans la gestion des ressources aux postes frontaliers terrestres canadiens.

Bibliographie

Box, G.E.P., et G.M. Jenkins (1976), *Time Series Analysis, Forecasting and Control (Revised Edition)*, San Francisco: Holden-Day.

Cortinas, J. J. (2000), « A Climatology of Freezing Rain in the Great Lakes Region of North America », *Monthly Weather Review*, 128, p. 3574- 3588.

Datla, S., S. Prasanta, H. Roh, et S. Sharma (2013), « A Comprehensive Analysis of the Association of Highway Traffic with Winter Weather Conditions », *Procedia - Social and Behavioral Sciences*, p. 497 – 506.

De Livera, M. A., J. R. Hyndman, et D. R. Snyder (2011), « Forecasting time series with complex seasonal patterns using exponential smoothing », *Journal of the American Statistical Association*, 106:496, p. 1513-1527.

Gelper, S., R. Fried, et C. Croux (2010), « Robust forecasting with exponential and Holt–Winters smoothing », *Journal of forecasting*, 29(3), 285–300.

Lin, L., M. Ni, Q. He, J. Gao, et W. A. Sadek (2015), “Transportation Research Record: Journal of the Transportation Research Board”, 2482, Transportation Research Board, Washington, D.C., pp. 82–89.

Roh, H., P. K. Sahu, S. Sharma, et S. Datla (2016), “Statistical Investigations of Snowfall and Temperature Interaction with Passenger Car and Truck Traffic on Primary Highways in Canada”, *Journal of Cold Regions Engineering*, 30(2).