



HAL
open science

Fiabilité des données de comptage et tendance d'évolution du vélo à Nantes (2014-2019)

Florent Bédécarrats

► **To cite this version:**

Florent Bédécarrats. Fiabilité des données de comptage et tendance d'évolution du vélo à Nantes (2014-2019). 2021. hal-03105063

HAL Id: hal-03105063

<https://hal.science/hal-03105063>

Preprint submitted on 10 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fiabilité des données de comptage et tendance d'évolution du vélo à Nantes (2014-2019)

Florent Bédécarrats (Nantes Métropole)*

Version du 10/07/2020

Synthèse : La Métropole de Nantes effectue depuis 2006 des comptages de passages de vélos sur certaines voies cyclables. Ces données sont disponibles depuis 2014 et, comme l'ensemble des informations produites par la collectivité, elles ont vocation à être ouvertes sur le portail data.nantesmetropole.fr. Mais ces données comportent des erreurs : des dysfonctionnements divers (pannes ou dérèglement des capteurs, perturbations ou réorganisations provisoires des axes de circulation...) provoquent sporadiquement des relevés nuls ou massivement surestimés. Sur vingt compteurs, les relevés erronés sont minoritaires par rapport aux relevés valides, mais l'ampleur des erreurs suffit encore à fausser les tendances d'évolution observées si on se contente d'agrèger les données brutes en sommes ou en moyennes mensuelles ou annuelles. Nantes Métropole publie donc les données brutes de ces compteurs exploitables pour la période 2014-2019 en ajoutant deux variables : un indicateur qui signale pour chaque point de donnée si celui-ci est vraisemblablement erroné et une valeur de comptage ajustée. La détection d'anomalie et l'estimation de la valeur ajustée sont fondées sur un algorithme statistique que nous présentons dans cette note. La publication des données de comptage vélo s'accompagne d'une mise en garde et d'une notice explicative afin de mettre en garde les utilisateurs contre l'usage des données brutes et les encourager à tenir compte des anomalies détectées et à les remplacer par les valeurs ajustées pour des analyses de tendance.

Introduction : l'inévitable imperfection des données de comptage vélo

Les erreurs de mesure constituent un problème inévitable avec les données de capteurs : hors d'un laboratoire, celles-ci contiennent *toujours* des valeurs manquantes, surestimées ou sous-estimées. Cette difficulté est particulièrement prégnante pour les capteurs de comptage vélo, pour une série de raisons propre aux objets décomptés et aux technologies disponibles pour le faire[1]. Ce problème a d'ailleurs déjà été documenté pour des capteurs de comptage vélo dans l'espace public, à Vancouver[2] ou à Zurich[3].

Des problèmes identiques affectent la mesure de fréquentation des services internet. Dans ce domaine, les outils d'analyse de séries chronologiques (rassemblés sous le terme "web analytics") se sont considérablement développés. La plupart des grandes plateformes web ont ainsi mis au point, fait valider dans des revues scientifiques et versé en open source des algorithmes qui rendent très aisée la détection d'anomalies, l'analyse de tendances et la prédiction. Nous appliquons ici l'un de ces algorithmes aux données de comptage vélos pour en améliorer la fiabilité et l'analyse.

*Ce document est un markdown (rmd) : il est généré de manière automatique par un script R, publié conjointement. Cela permet à tout un chacun, à partir du script R, de consulter et vérifier le code et de reproduire le même résultat en l'exécutant depuis son ordinateur.

Boucles de comptage vélo à Nantes

Le comptage des vélos est réalisé au moyen de boucles magnétiques placées sur ou dans la chaussée. Des boucles de comptage ont été installées sur le territoire de Nantes Métropole depuis les années 2000. La carte suivante indique leur localisation. La couleur des marqueurs traduit la date d'installation et la forme des marqueurs correspond au mode de transmission de la donnée (transmission automatique ou relevé manuel).

Localisation des boucles de comptage vélo de Nantes Métropole



Mode de transmission

- Relevé manuel
- Transmission automatique

Année d'établissement

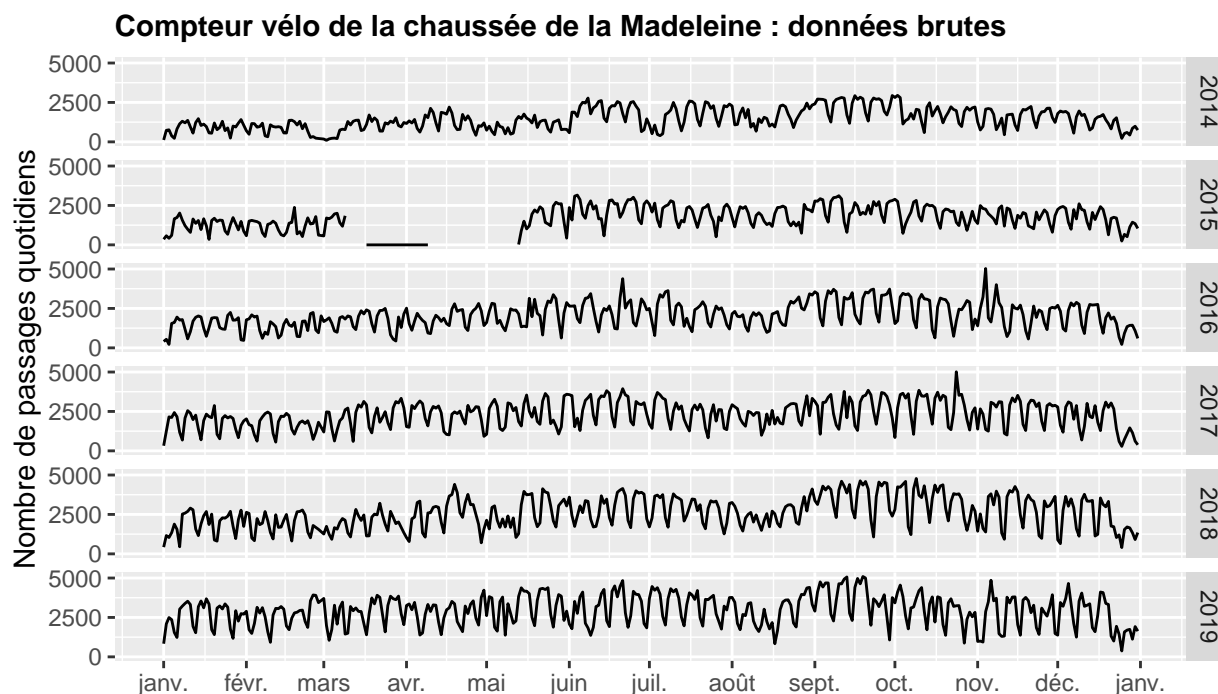
- 2000 (1 point de comptage)
- 2006 (34 points de comptage)
- 2011 (2 points de comptage)
- 2013 (2 points de comptage)
- 2014 (3 points de comptage)
- 2015 (8 points de comptage)
- 2017 (8 points de comptage)
- 2018 (6 points de comptage)
- 2019 (4 points de comptage)

Fond de carte : données OpenStreetMap (ODbL), moteur de rendu graphique Stamen Design (CC-BY)

Cette carte montre que la plupart des boucles de comptage ont été installées sur la commune de Nantes. La plupart l'ont été en 2006, puis d'autres ont été progressivement ajoutés entre 2011 et 2019.

Problème de fiabilité des données brutes de comptage

Comme c'est toujours le cas pour des capteurs placés hors d'un environnement confiné, les compteurs vélos placés sur la voie publique dysfonctionnent ponctuellement. Cela entraîne des valeurs manquantes ou erronées dans les séries de données chronologiques, qui faussent les statistiques agrégées produites à partir des données brutes. Pour illustrer ce point, on représente ci-dessous les relevés effectués sur le compteur localisé Chaussée de la Madeleine, et qui figure parmi les plus fiables.



On observe sur le graphique certaines données manquantes (ex. ici d'avril à mai 2015) et des données apparemment aberrantes, anormalement basses (ex. mars 2014 ou avril 2015, avec des valeurs à 0), ou anormalement hautes (novembre 2016, octobre 2017). Le problème est que de telles variations biaisent les statistiques agrégées que l'on peut produire à partir des données brutes, comme le montre le tableau suivant.

Table 1: Les données manquantes ou aberrantes biaisent les statistiques agrégées et les tendances d'évolution

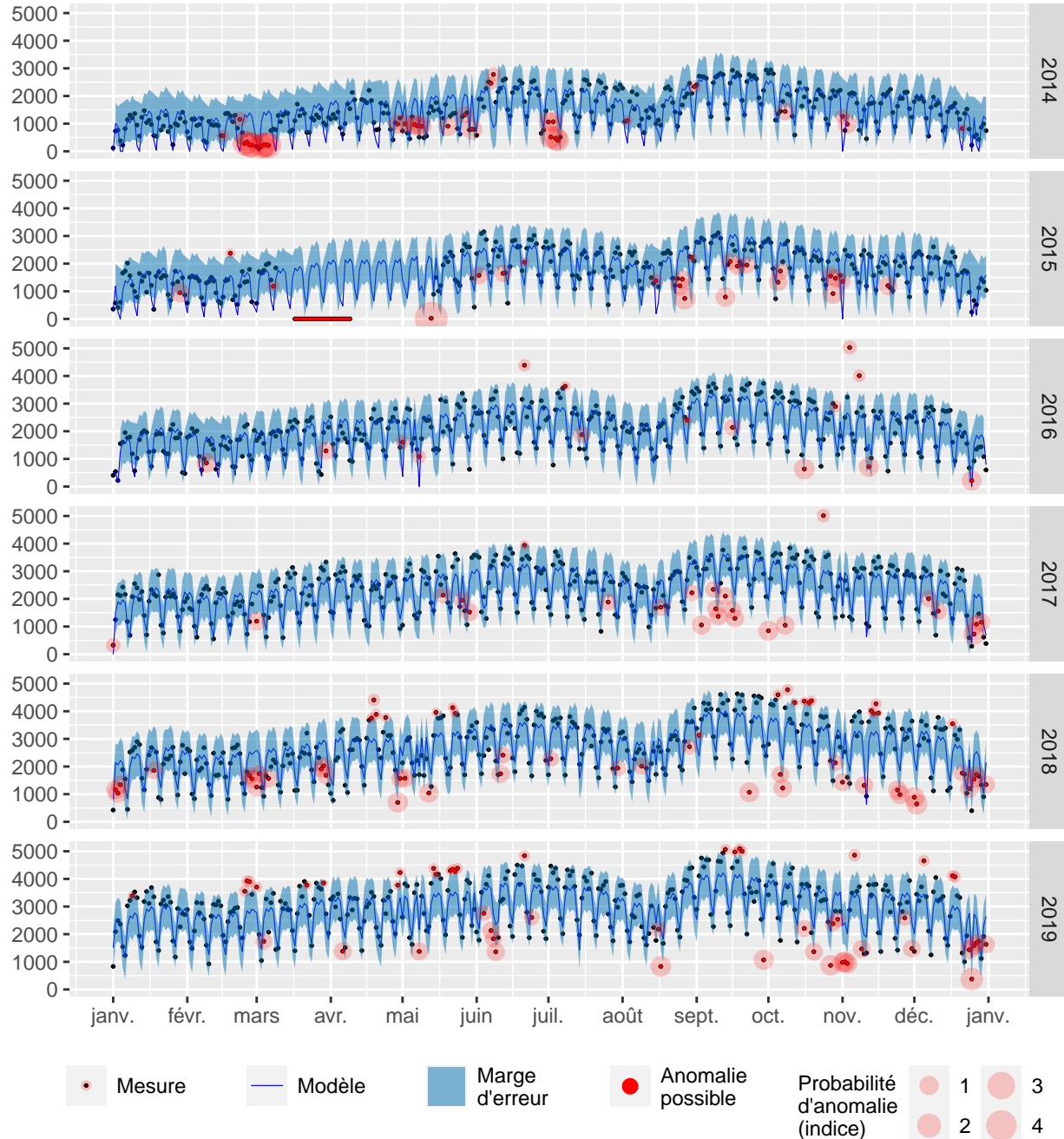
Année	2014	2015	2016	2017	2018	2019
Somme	529408	537747 (+1.6%)	754995 (+40.4%)	847730 (+12.3%)	952853 (+12.4%)	1075569 (+12.9%)
Moyenne	1450	1655 (+14.1%)	2063 (+24.7%)	2323 (+12.6%)	2611 (+12.4%)	2947 (+12.9%)
Médiane	1368	1664 (+21.6%)	2021 (+21.5%)	2421 (+19.8%)	2579 (+6.5%)	3108 (+20.5%)

La somme des passages comptés est une statistique qui est souvent produite par les services de déplacements ou les médias, car elle répond en principe à la question que l'on se pose spontanément avec ce type de données : est-ce que le nombre total de passages augmente dans le temps ? C'est toutefois l'agrégation la plus sensible aux erreurs de mesures ou aux données manquantes. La moyenne est moins affectée par les données manquantes, mais elle demeure faussée par les données aberrantes (par exemple les valeurs à 0 enregistrées en mars 2015 ou anormalement élevées en novembre 2016 et octobre 2017). La médiane est moins sensible aux valeurs extrêmes et donc meilleure indicatrice des tendances de fond. Mais c'est une statistique moins courante, dont le public est peu familier.

Fiabilisation des données et calcul de tendance

Pour résoudre ce problème, nous utilisons un modèle qui repose sur l'identification des variations qui présentent une régularité périodique (ici, les variations hebdomadaires et annuelles) pour en déduire la tendance sous-jacente et détecter les inflexions dans cette tendance[4]. Ce modèle tient compte des jours fériés et des données manquantes. Il a été rendu particulièrement aisé d'application grâce à la création d'une bibliothèque R intitulée *Prophet* et diffusée en licence ouverte[5]. On illustre dans le graphique ci-dessous la manière dont il peut être utilisé pour repérer les anomalies de mesure, en utilisant le même compteur que précédemment, celui placé Chaussée de la Madeleine.

Variations récurrentes et détection des anomalies (Chaussée de la Madeleine)

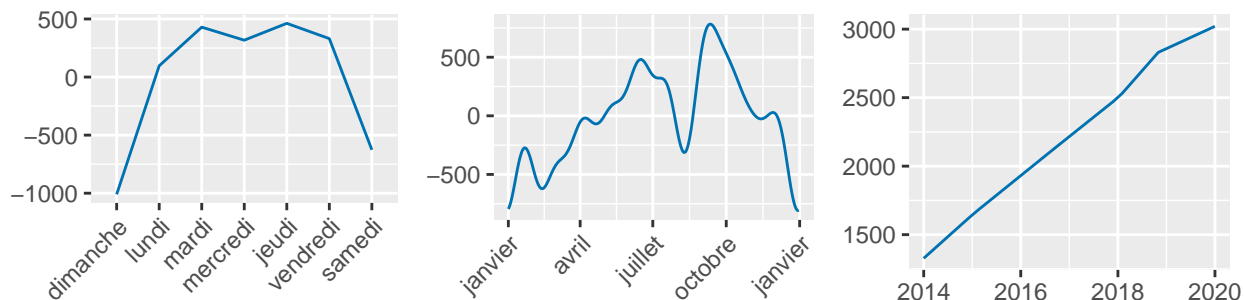


Sur le graphique précédent, la ligne bleue correspond à l'estimation réalisée par le modèle. Pour parvenir à cette estimation, le modèle analyse les mesures de compteurs vélos pour identifier les variations qui se produisent régulièrement d'un jour à l'autre de la semaine et d'un jour à l'autre de l'année, et ainsi déduire la tendance d'évolution sur l'ensemble de la période. Quatre niveaux de variations (hebdomadaire, annuel, tendance de fond et points d'inflexion de la tendance de fond) se combinent pour estimer une valeur théorique pour chaque jour de la période. Une marge d'erreur (zone grisée) est calculée en fonction de la dispersion de l'ensemble des mesures issues des compteurs vélos par rapport à l'ensemble des estimations produites par le modèle. Le seuil de confiance retenu est de 90%, c'est-à-dire que les points situés au-delà de la zone grisée ont 9 chances sur 10 d'être des anomalies. Plus un point est éloigné de la zone grisée, plus la probabilité qu'il s'agisse d'une anomalie est élevée. Si le modèle prédit mal les valeurs observées dans la réalité, la marge d'erreur est très étendue. Ici, elle reste contenue, ce qui indique que le modèle fonctionne assez bien. Inclure des facteurs de météo (pluie, vent, température) permettrait sans doute d'améliorer la fiabilité du modèle.

En plus de la procédure décrite au paragraphe précédent, on a ajouté une étape préalable qui consiste à pointer comme anormale une donnée nulle dès lors qu'on observe trois données nulles successives. En effet, dans certains cas, un compteur en panne renvoie une donnée égale à 0 plutôt qu'une donnée nulle. En cas de pannes prolongées sur plusieurs semaines ou mois, l'algorithme **Prophet** n'interprète plus ces données nulles comme une anomalie ponctuelle, mais comme une tendance de fond. Les experts de la circulation à vélo à Nantes Métropole indique qu'il est hautement improbable de n'avoir absolument aucun cycliste qui emprunte une voie pendant trois jours consécutif. Les tests réalisés montrent que l'algorithme fonctionne mieux avec ce filtre préalable : les calculs des tendances sur 9 compteurs qui déjà interprétables sans ce filtre ne sont pratiquement pas altérés. En revanche, les données de 11 compteurs additionnels deviennent exploitables avec ce filtre.

L'analyse des variations hebdomadaires et annuelles est intéressante, mais c'est surtout la mise en lumière des tendances de fond qui nous intéresse, car elle reflète l'évolution de l'usage du vélo dans la métropole nantaise.

Variations récurrentes intra-hebdomadaires et intra-annuelles et tendance de fond (Chaussée de la Madeleine)



Il est assez surprenant de constater la stabilité de l'augmentation dans le cas du compteur situé Chaussée de la Madeleine. L'algorithme utilisé repère les inflexions dans la tendance de fond (une infime diminution du rythme de croissance peut être décelée en 2018) et elles seraient visibles si on avait des modulations importantes sur la période. Les autres compteurs dont les tendances sont présentées plus bas montrent des inflexions plus marquées entre 2014 et 2018.

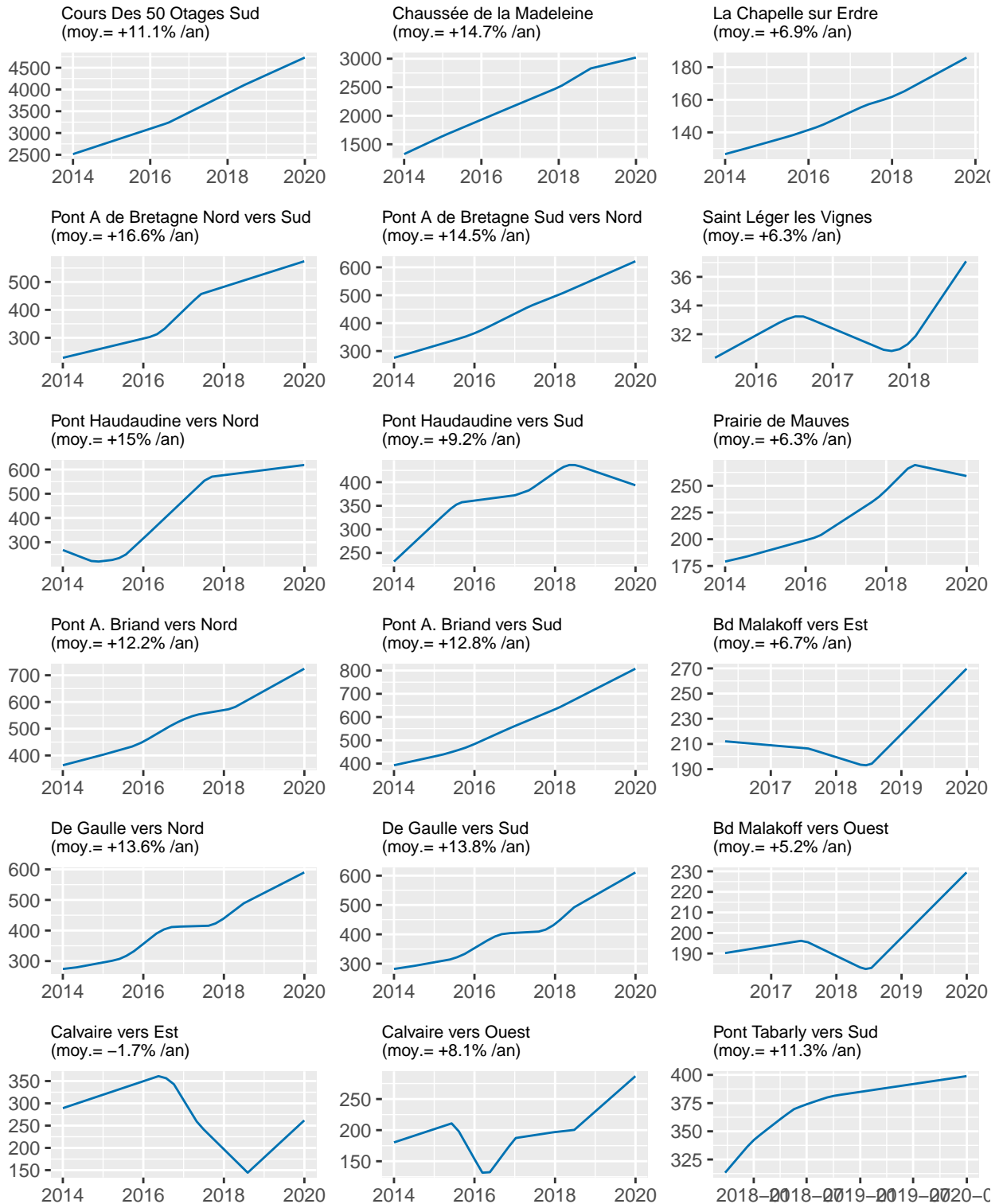
Ainsi, on observe que les comptages vélos effectués à la Madeleine reflètent, une fois corrigé le "bruit" des mesures, une augmentation très importante du vélo : +128% sur 5.9958932 années, soit une croissance de +14.7% par an en moyenne.

Application à l'ensemble des compteurs exploitables à Nantes

Qu'il s'agisse de sommes, moyennes ou médianes, les données en valeur absolue ne sont pas comparables entre elles. Pris isolément, aucun capteur n'est représentatif de l'usage global du vélo de la ville. Le nombre de

passages qu'il capture dépend de circonstances arbitraires : Relève-t'il l'ensemble des passages sur la voie ou seulement ceux qui sont sur la piste cyclable ? S'agit-il d'un passage obligé ou y a-t-il des axes par lesquelles le flux est dispersé (exemple d'un pont par opposition à des rues parallèles) ?, etc. En revanche, l'évolution d'un même capteur au cours du temps reflète l'évolution relative globale de l'usage du vélo sur cet axe, sous réserve que la circulation n'ait pas été structurellement modifiée pendant la période (piétonnisation, fermeture d'autres axes, etc.)

Une part importante des compteurs placés à Nantes a connu des dysfonctionnements trop importants sur des périodes trop prolongées pour être exploitables. On trouve toutefois 20 compteurs qui disposent de séries chronologiques ininterrompues sur plusieurs années, et dont aucune variation invraisemblable n'indique de dysfonctionnement prolongé des dispositifs. Nous présentons ci-dessous les tendances de variations sur la période pour les 18 principaux.

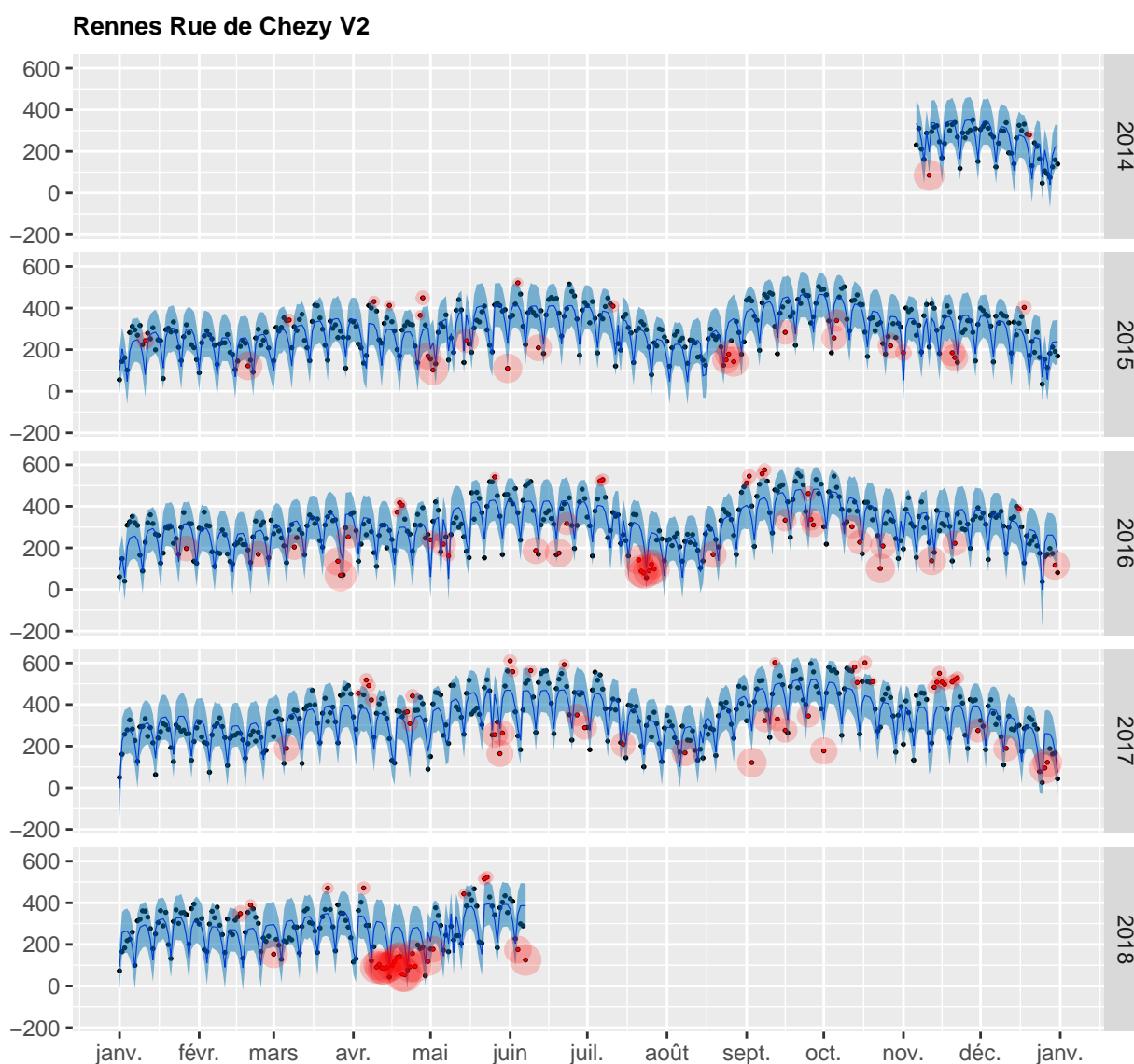


Le présent rapport et le jeu de données complet sont disponibles sur data.nantesmetropole.fr. Sur la même page, se trouve également un fichier html qui permet de visualiser les relevés détaillés et les tendances sous-jacentes (intra-hebdomadaires, intra-annuelles et tendance de fond) des 20 capteurs disposés sur le territoire de la Métropole de Nantes sont disponibles dans l'annexe jointe à cette note.

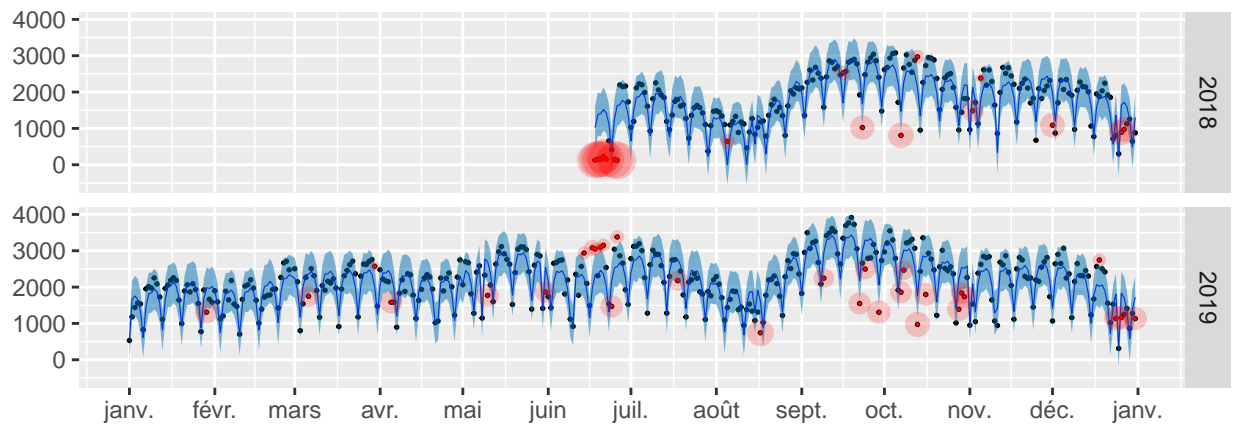
Comparaison avec d'autres métropoles françaises

Rennes

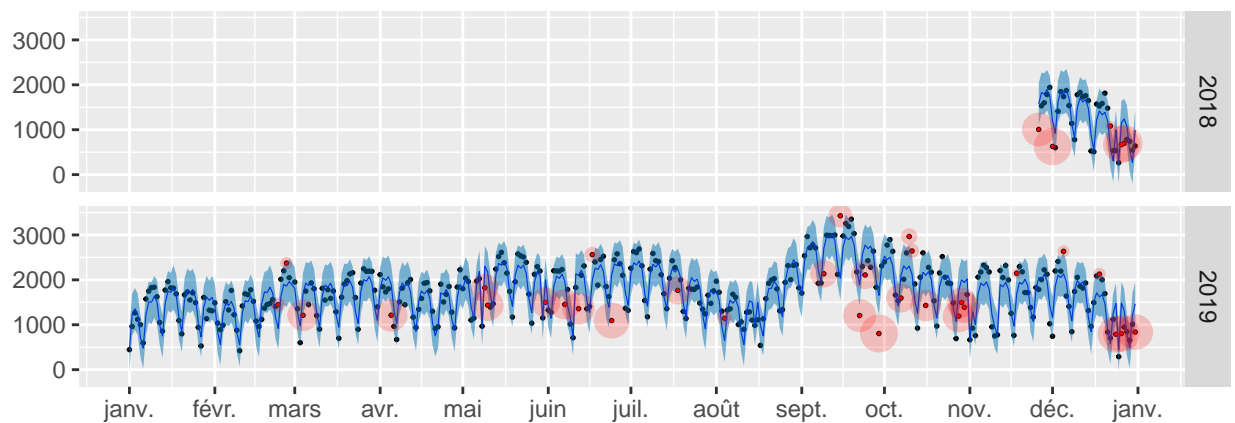
Rennes publie en open data les données brutes de quatre capteurs disposés sur son territoire. Les métadonnées associées à ces comptages indiquent que ces capteurs dénombrent à la fois les passages des vélos et ceux des piétons. Certains articles en ligne indiquent toutefois que l'un des compteurs (place de Bretagne) est dédié exclusivement aux vélos[6]. Deux caractéristiques des données de Rennes gênent les comparaisons avec Nantes. D'une part, elles cumulent comptages piétons et vélos, alors que Nantes ne décompte que les vélos. D'autre part, trois des compteurs (Place de Bretagne, rue d'Isly et rue de Chezyelles) ont été disposés il y a moins d'un an et demi, ce qui limite notre capacité à distinguer des tendances claires. Le capteur placé sur le Boulevard Georges Pompidou contient des données plus anciennes, mais son fonctionnement a visiblement été interrompu mi-2018. Comme à Nantes, les données de Rennes présentent des anomalies, qui semblent toutefois bien détectées par le modèle.



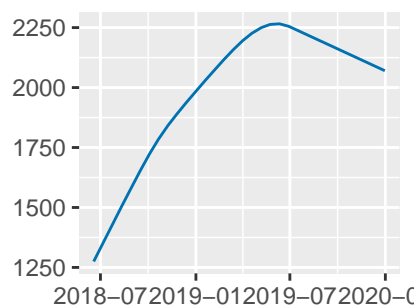
Eco-Display Place de Bretagne



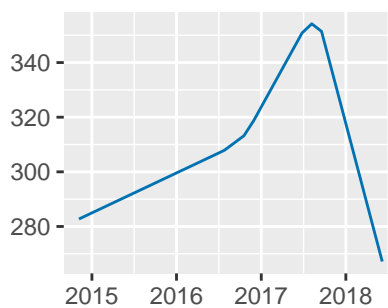
Boulevard Georges Pompidou



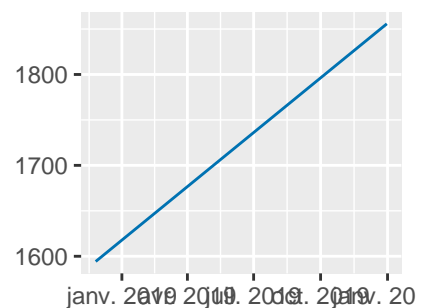
Eco-Display Place de Bretagne
(moy.= +37.1% /an)



Boulevard Georges Pompidou
(moy.= -1.6% /an)

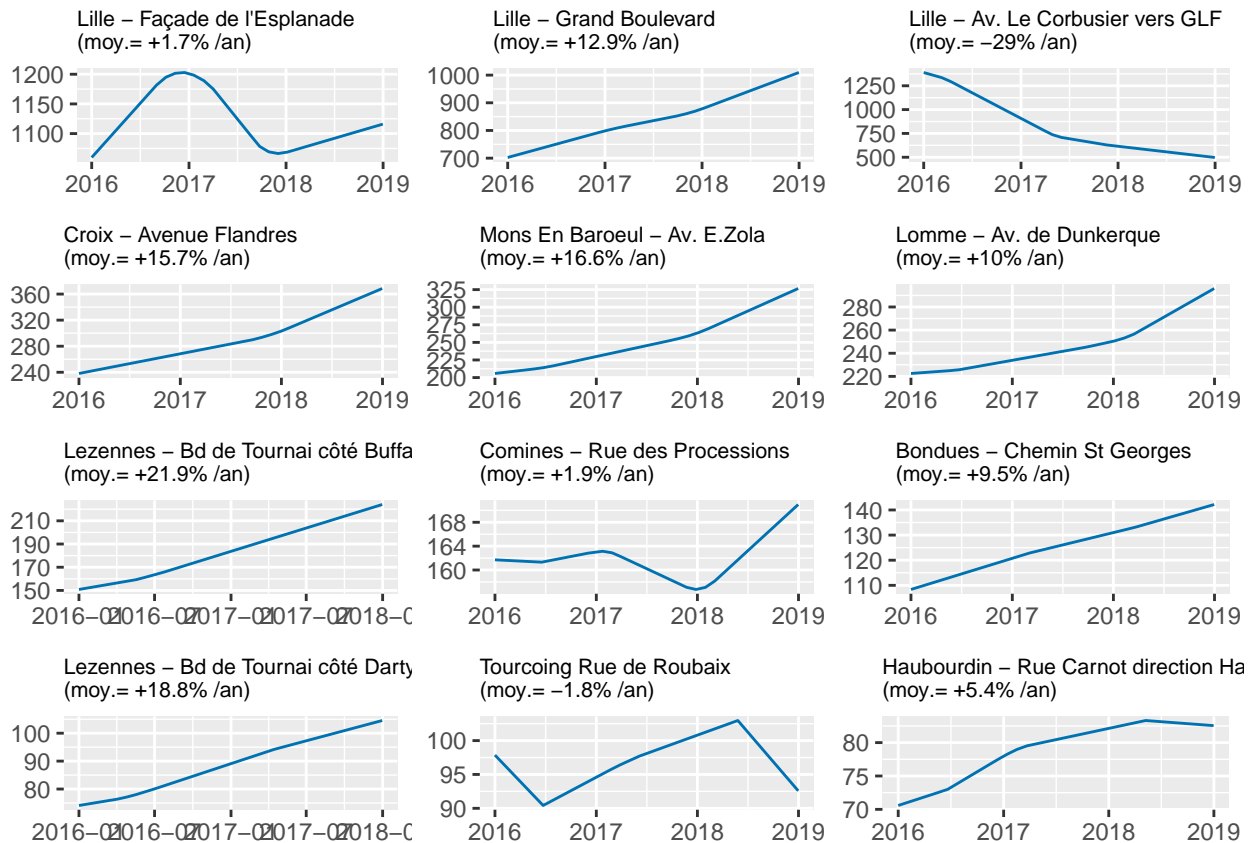


Rennes Rue de Chezy V2
(moy.= +14.9% /an)



Lille

Le site d'open data de Lille Métropole fournit les données de 46 points de comptage, collectées entre 2016 et 2018. Les données de 25 de ces compteurs sont exploitables pour l'analyse. Les autres ne contiennent qu'une année de relevés, ou des données comportant une part trop importante d'erreurs flagrantes pour en tirer une tendance. Nous présentons ci-dessous les données des 12 compteurs exploitables enregistrant le plus de passage.



Ouverture et utilisation des données de comptage à Nantes

Sur la base des éléments mentionnés dans cette note, nous trois actions sont mises en oeuvre :

- **Publier au fil de l'eau les données comptages à partir de 2020** : depuis début 2020, les données de comptage de tous les compteurs fixes à Nantes Métropole sont publiées quotidiennement à mesure qu'elles sont produites. Les données de comptage d'un jour sont ainsi mises en ligne le lendemain sur le portail data.nantesmetropole.fr. Mais ces données ne permettent pas de réaliser des analyses de tendance. Un repérage des erreurs est effectué, mais en utilisant une méthode plus basique que celle présentée ici, qui se limite à signaler une erreur potentielle lorsque plusieurs relevés horaires sont nuls ou manquants.
- **Publier l'historique des données de comptage depuis 2014** pour les 20 compteurs exploitables mentionnés plus haut. Compte tenu des difficultés exposées dans ce document, nous publions ces données avec les variables suivantes : date, compteur, valeur mesurée, anomalie détectée et valeur ajustée. Une note explicative est incluse en en-tête du jeu de donnée afin que les réutilisateurs potentiels en soient conscients des précautions requises pour l'analyse de ces données. Les données au fil de l'eau mentionnées au point précédent seront ajoutées tous les ans à ce jeu de données.
- **Publier cette note technique ainsi que le code associé (R)** qui a servi à la détection des anomalies et la modélisation de valeurs théoriques.

Notes

- [1] P. Ryus *et al.*, “Guidebook on pedestrian and bicycle volume data collection,” Institute of Transportation Studies, UC Berkeley, 2014.
- [2] M. El Esawey, A. I. Mosa, and K. Nasr, “Estimation of daily bicycle traffic volumes using sparse data,” *Computers, Environment and Urban Systems*, vol. 54, pp. 195–203, 2015.
- [3] D. Baehler, D. Marincek, and P. Rérat, “Les comptages vélos dans les villes suisses,” Institut de géographie et durabilité (IGD), Université de Lausanne, 2018.
- [4] S. Taylor and B. Letham, “Forecasting at scale,” *PeerJ Preprints*, vol. 5:e3190v2, 2017.
- [5] S. Taylor and B. Letham, *Prophet: Automatic forecasting procedure*. 2019 [Online]. Available: <https://CRAN.R-project.org/package=prophet>
- [6] J.-C. Collet, “Place de Bretagne, on comptera les vélos !” *Rennes Info Autrement*, Jun. 2018 [Online]. Available: <https://www.rennes-infos-autrement.fr/place-de-bretagne-on-comptera-les-velos/>