

# validation\_baseline

April 16, 2025

## 1 Validation de la simulation de référence

La validation de la simulation de référence repose sur les données ouvertes fournies par Île-de-France Mobilités, correspondant aux deux derniers semestres de 2024.

Il est à noter que certaines lignes ou stations ont pu subir des modifications pendant cette période, ainsi qu'entre le 1er janvier 2025 et le 17 février 2025, date de récupération des données GTFS. Par exemple, la station Villejuif - Gustave Roussy, ouverte en janvier 2025, est présente dans la simulation mais absente des données de référence.

Les données utilisées sont réparties en deux catégories : - les données des voies ferrées, correspondant aux validations à chaque arrêt (hors correspondances), - et les données de surface, concernant les lignes de bus et la majorité des tramways, disponibles sous forme agrégée par ligne

Nous séparerons donc cette analyse en deux, entre le réseau ferré et le réseau de surface.

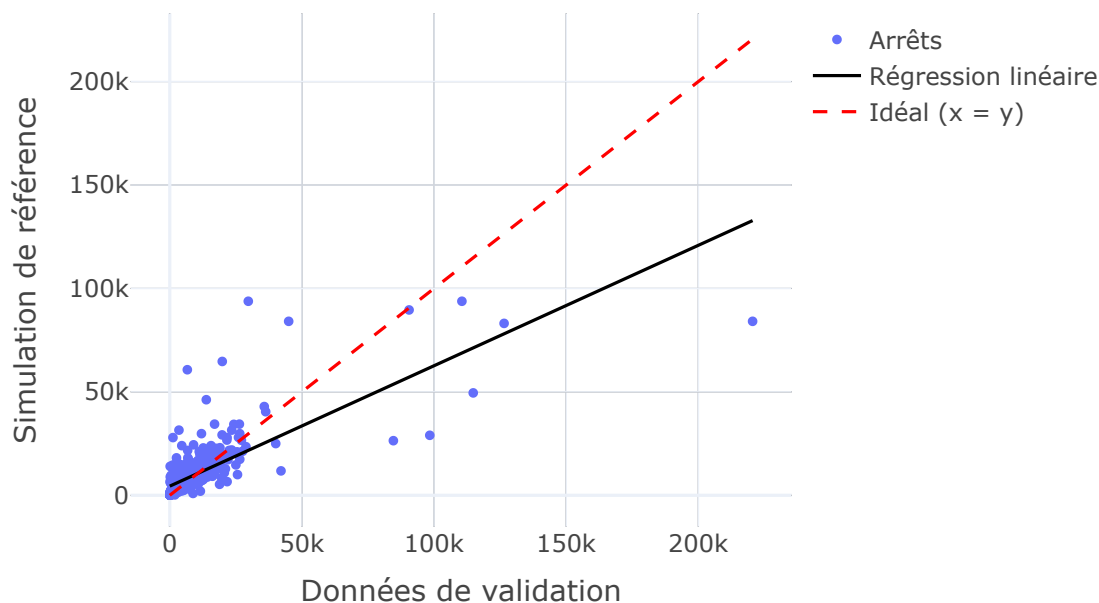
### 1.1 Réseau ferré

Le réseau ferré est composé des lignes demandant une validation à un arrêt physique pour y accéder. Il comporte les lignes de TER, RER, Transilien, Métro ainsi que les tramways 4, 11, 12, 13.

La majorité des correspondances entre ligne ne nécessite pas de validation. Par exemple, à la Motte-Picquet Grenelle, le passage du métro 10 au métro 8 peut se faire sur le même quai. Pour cette raison, les correspondances ne sont pas incluses dans les données ouvertes de l'IDFM. Les résultats de la simulation de référence ont été traités en prenant en compte cette information.

Dans un même trajet, seuls les tronçons de type subway ou rail sont retenus, à condition qu'ils ne soient pas précédés par un autre tronçon de même type, et que les identifiants des arrêts de départ et d'arrivée correspondent.

## Écarts entre les données de validation et la simulation par arrêt

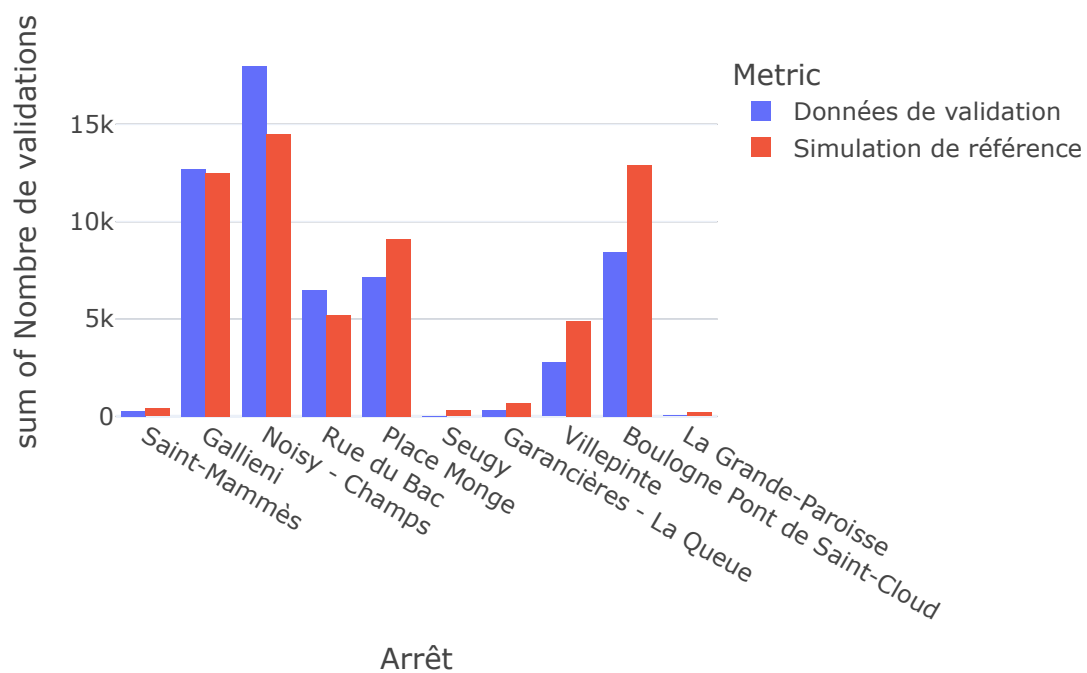


Sur la figure ci-dessus, est représenté la correspondance entre la simulation et les données de validation du nombre de trajets quotidiens pour chaque arrêt du réseau ferré.

Cela montre que la simulation tend à surestimer le nombre de validations pour les arrêts faiblement fréquentés (moins de 15 000 validations), puis à sous-estimer progressivement les volumes plus élevés.

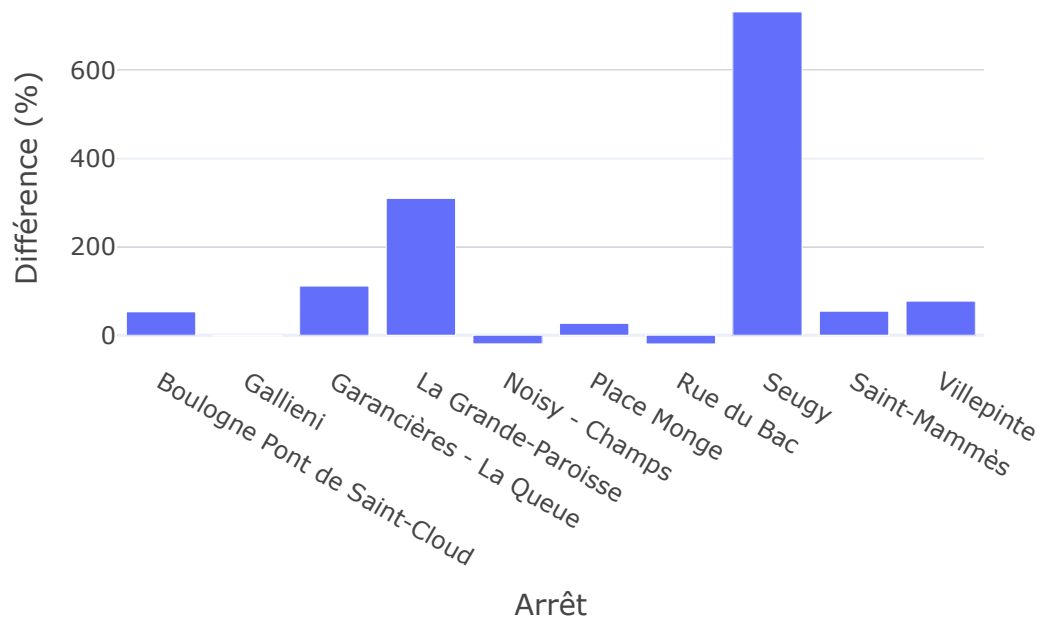
Cette sous-estimation, notamment présente dans les grandes gares (comme Saint-Lazare), peut s'expliquer par le fait que seule la population résidant en Île-de-France est prise en compte dans la simulation. Les voyageurs longue distance (TGV), les touristes ou les usagers des aéroports ne sont donc pas représentés, ce qui conduit logiquement à une sous-estimation sur ces points d'intérêt majeurs.

## Comparaison entre la référence et la simulation



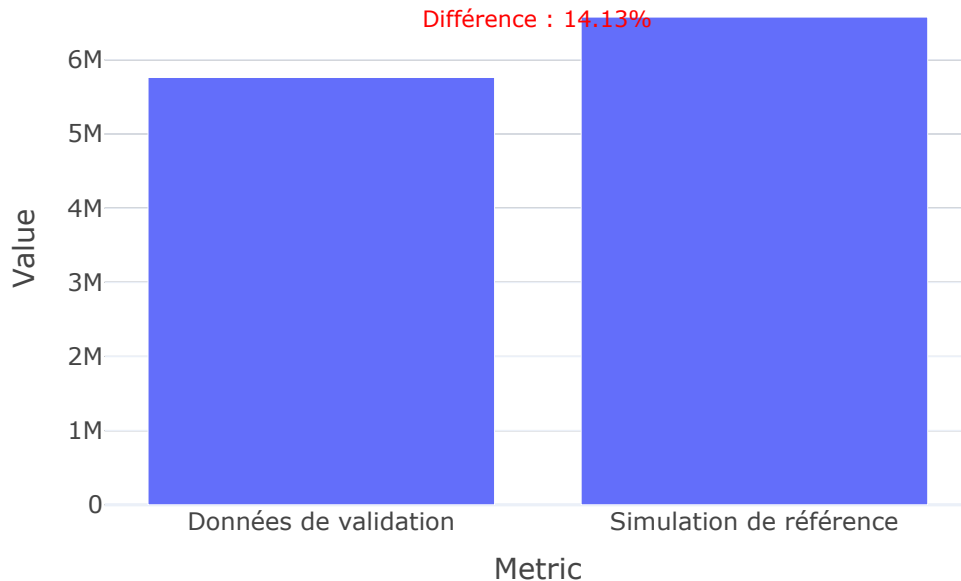
La figure ci-dessus nous permet d'avoir un aperçu plus précis du nombre de validations pour un sample de 10 gares.

## Différence (en %) entre les données de référence et la simulat



Pour ces mêmes arrêts, on peut avoir un aperçu de la différence entre les données de validation et de la simulation de référence en pourcentage.

## Somme totale des valeurs



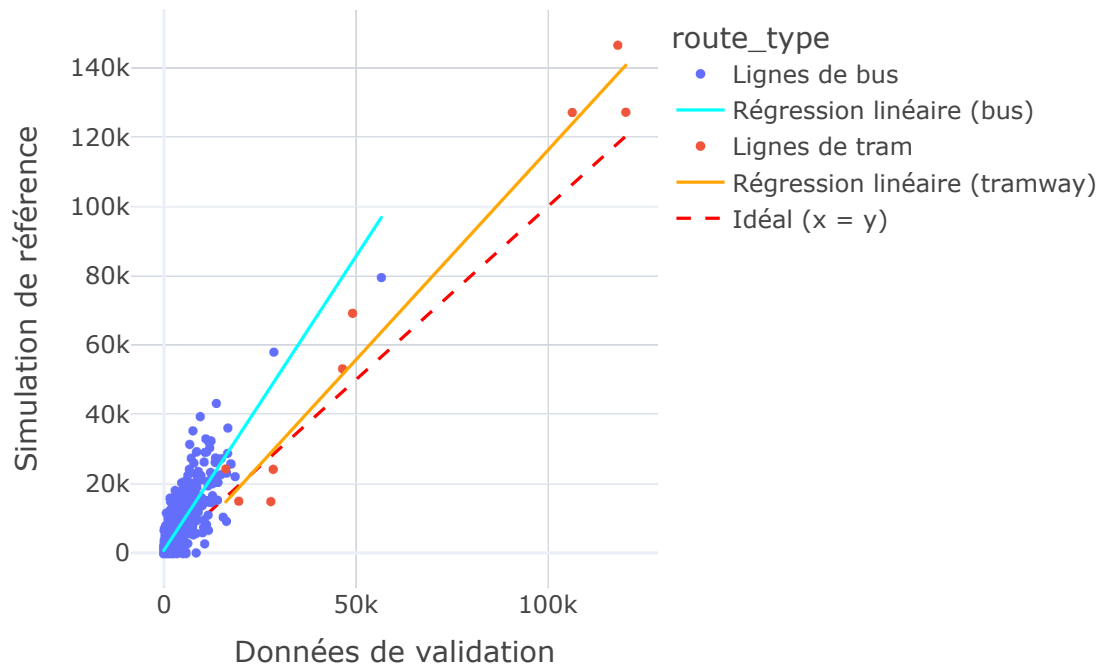
Pour la figure suivante, on agrège les données pour avoir un aperçu du nombre de validation total sur le réseau ferré. Ici on voit que, la simulation a tendance à surestimer les données réelles, à hauteur de 25%

### 1.2 Réseau de surface

Les données sur le réseau de surface contiennent donc les lignes de bus ainsi que la majorité des lignes de tramway.

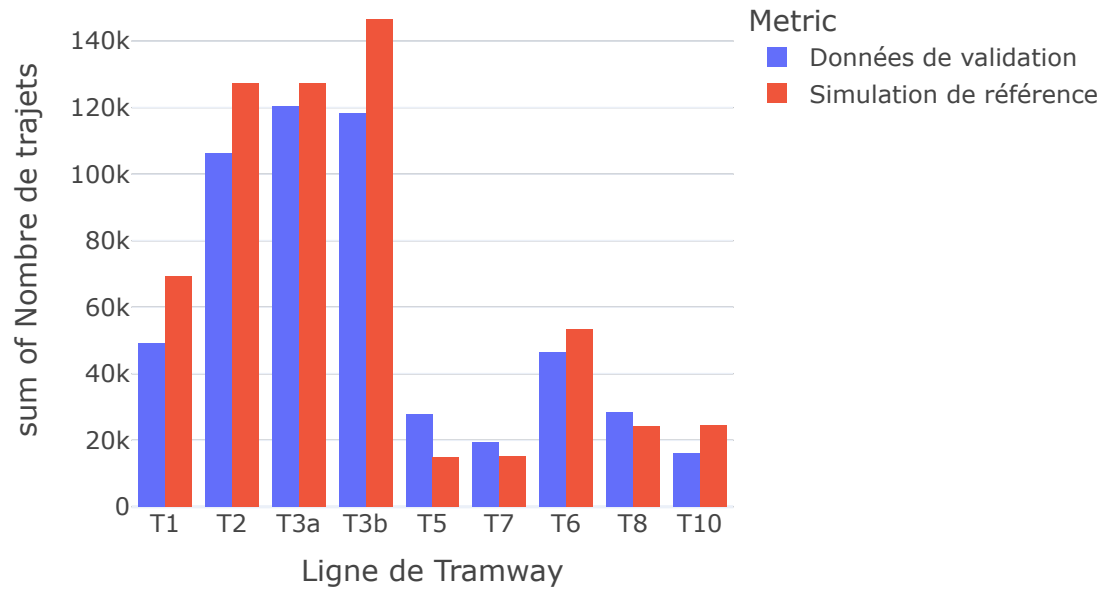
L'identifiant utilisé pour une ligne dans la base de données de validation n'a pas de correspondance directe avec les données GTFS fournies par l'Île-de-France Mobilités. Pour avoir une correspondance, on récupère le référentiel de lignes dans les données ouvertes de l'IDFM.

Une fois la correspondance faite, on peut désormais analyser les résultats.



La figure ci-dessus montre la correspondance du nombre de validations pour un jour, pour chaque ligne du réseau de surface. On remarque que les bus correspondent relativement bien à l'idéal. Quant aux tramways, ceux-ci sont surestimés dans la simulation.

## Comparaison entre les données de référence et la simulation b



Ici on a une vision plus précise du nombre de trajets pour les lignes de tramways.

Différence (en %) entre les données de référence et la simulat

