Construire ensemble le nouveau service de décompte ferroviaire

Compte rendu de l'Atelier de Travail de Concertation sur le décompte ferroviaire

23 novembre 2021 – 10h à 12h

Agenda

1. Introduction et Rappels
2. Méthode d'estimation de la consommation
3. Critères de validation de l'algorithmie d'estimation
4. Processus de qualification, validation et de bascule

Présentation partagée

Retrouvez toutes les informations de la concertation sur le décompte ferroviaire sur le site https://www.concerte.fr/ (GT Données & Comptage)

A l'occasion de cet atelier de travail, SNCF Réseau et RTE ont initié des discussions avec les participants, sur la base de questions qui ont été reprises dans un questionnaire adressé par mail aux participants. Ce questionnaire est disponible sur https://www.concerte.fr/ ou sur demande à RTE-DECOFER-CLIENT@rte-france.com.

RTE et SNCF Réseau rappellent que la présente concertation se déroule en parallèle de la construction du nouveau service, et du développement des nouveaux logiciels associés. Il est nécessaire que les EF et GI participant à la concertation fassent part de leurs retours, avis, questions et besoins le plus tôt possible au fur et à mesure de ce processus de concertation, afin que ces éléments puissent être pris en compte au plus tôt par RTE et SCNF Réseau.

Le présent compte rendu vient en complément des slides présentés lors de l'Atelier de Travail.

Synthèse des échanges :

1. Introduction et Rappels

Principaux éléments présentés

- Le projet de décompte ferroviaire SNCF R – RTE prévoit une transition progressive des rôles et responsabilités des activités de comptage et de décompte
    - Notamment, pour la phase 2, prévoit que SNCF Réseau assure l'estimation de la consommation des engins non équipées de compteur circulant sur son réseau, sur la base des données de circulation.

Diffusable SNCF RESEAU

- Pour ce faire, SNCF Réseau développe et met en œuvre le logiciel ORES (Outil de Rapprochement et d'Estimation de SNCF Réseau), qui vise à réaliser le journal des circulations, et l'estimation de la consommation des engins non équipés de compteurs.

- La chaine d'acquisition et de traitement des données de comptage mise en œuvre par RTE en phase 2 permet d'aboutir à une courbe de consommation pour chaque EF, à la maille de son parc d'engins équipés
  - L'estimation de la consommation des engins non équipés de chaque EF permet de compléter le décompte réalisé sur la base des engins équipés. Le décompte final intègre donc tous les engins
  - Cependant, RTE et SNCF Réseau rappellent que les estimations n'ont pas vocation à remplacer les compteurs. L'équipement en compteurs reste une obligation réglementaire des EF, et un prérequis pour la fiabilisation du décompte.

- RTE et SNCF Réseau rappellent les éléments présentés en AT le 1/07/2021 concernant le journal des circulations :
  - Le journal des circulations est une liste des circulations réalisées, affectées à un identifiant de l'engin moteur (rame ou locomotive), la présence ou non d'un compteur, et complétée d'une estimation de la consommation réalisée sur la circulation.

2. Méthode d'estimation de la consommation

Principaux éléments présentés

- Travaux de construction des algorithmes ORES :
  - Les équipes de Data Scientist SNCF Réseau travaillent sur les algorithmes de l'outil ORES, avec pour objectif d'estimer la consommation des circulations.
  - Les choix méthodologiques, la conception et la validation (statistiques, machine learning, etc) sont réalisés en co-construction avec la Direction de l'Ingénierie de la Traction Electrique, sous le pilotage de la Direction de la Stratégie.
  - Les éléments présentés lors de l'AT du 23/11 constituent un état des travaux à date, l'outil ORES étant en cours d'optimisation et de validation.

- Principes de construction de la courbe de charge des engins non équipés, à la maille d'une EF
  - Pour chaque circulation, l'outil ORES calcule une courbe de consommation estimée, au pas 5 min.
  - L'agrégation des courbes de consommation ainsi estimées permet d'obtenir l'estimation de la consommation de l'EF (à la maille de la partie non équipée de son parc)

- Définition d'un modèle d'estimation
  - Un « modèle » d'estimation est une fonction permettant de calculer l'estimation de la consommation pour une circulation, sur la base d'un certain nombre de variables d'entrées.

Diffusable SNCF RESEAU

- Le modèle d'estimation fonctionne sur un pas de 5 min (pour les variables d'entrée et la conso calculée en sortie)
- Les équipes SNCF Réseau ont construit à date 7 modèles, en se basant sur les « types de circulation » tels que définis dans le DRR

- Choix des variables d'entrée pour les modèles d'estimation
  - Des formules existantes, permettant de calculer une approximation de la puissance de traction, ont permis d'identifier des variables pertinentes retenues comme données d'entrée dans les modèles d'estimation. Le choix de ces variables s'est également appuyé sur la disponibilité de ces données dans le SI industriel de SNCF Réseau :
    - Infocentre Circulation, qui capte la donnée de circulation des trains au long de leur parcours
    - Base GAIA, qui décrit les infrastructures, et notamment la déclivité des voies
    - Base Météo : données météo précises, en fonction de la localisation du train

- Méthode de paramétrage du modèle par entraînement
  - Chaque modèle (1 par type de circulation) est entrainé sur la base
    - des données d'entrée de circulation
    - et des données de consommation mesurées par les compteurs.
  - Le lien entre les données d'entrée (issues des circulations) et la consommation mesurée est possible grâce au rapprochement.

- RTE et SNCF Réseau rappellent deux prérequis indispensables pour le bon fonctionnement des modèles d'estimation :
  - Les modèles d'estimations sont entraînés sur les circulations qui sont réalisées par des engins équipés de compteurs
    - Afin d'assurer la bonne représentativité de ces modèles, les EF doivent donc être équipées de suffisamment de compteurs, pour entrainer les modèles et les maintenir à jour
  - Déclarer avec précision le parc de matériel roulant (Référentiel matériel) est indispensable pour assurer le rapprochement et les estimations.
    - Par ailleurs, la précision des modèles d'estimations dépend également des informations contenues dans les sillons (tonnage, US/UM, type d'engin)

- SNCF Réseau précise la forme des modèles d'estimation, qui utilisent le principe d'arbres de décision, afin de s'adapter aux différentes sous-catégories de circulations, sans avoir à les modéliser manuellement (les catégories de l'arbre de décision étant déterminée par entraînement du modèle).

Echanges en séance

- A la question de savoir si l'estimation se fait au niveau d'une circulation, ou d'un engin, SNCF Réseau répond que la maille choisie est celle de la rame.
- SNCF Réseau précise que la catégorie « tramway » embarque également les tram-trains.

Diffusable SNCF RESEAU

- SNCF Réseau et RTE notent la remarque formulée en séance sur les données « infocentre circulation », qui sont des données déclaratives renseignées par les EF en conformité avec la STI exploitation : ces données ne sont pas saisies aujourd'hui avec la finalité d'estimer la consommation des engins, et les flux informatiques correspondant pourraient nécessiter des modifications ou adaptations, afin de viser une qualité plus importante qu'aujourd'hui.
    - Cependant, SNCF Réseau précise que l'outil d'estimation repose sur des données disponibles dans les gisements de données à sa disposition, et exploite le croisement et le rapprochement de ces données pour être le plus précis possible. C'est à l'EF de renseigner au mieux les données qui seront utilisées par l'outil, mais « Infocentre circulation » n'a pas vocation à devenir un outil d'estimation des consommations. Il est rappelé que la cible visée est que le matériel moteur doit être équipé de compteurs.
- SNCF Réseau note la demande formulée en séance, concernant le besoin de différencier les engins moteur en mode traction ou en mode véhicule afin d'éviter d'estimer ou de profiler (par circulation) les engins en mode véhicule. Ce sujet sera instruit et un retour sera fait aux entreprises de fret.
- SNCF Réseau précise que des travaux sont en cours concernant les différences entre Unités Simples et Unités Multiples, en lien avec les différences de comportement des trains selon le réseau et selon les limitations de puissance tels que définis dans le registre d'infrastructure.
- SNCF Réseau précise que les modèles d'estimation suivent les catégories du DRR (6 types d'activité + tramway). Il n'est pas prévu d'aller à une maille plus fine, par type de matériel moteur par exemple. Ainsi, les circulations de trains régionaux, nationaux et internationaux de voyageurs aptes à la grande vitesse, ne seront représentées que par un seul modèle, ne dépendant pas du type de TGV utilisé. **En effet, l'objectif est de de viser une estimation globale à la maille du parc, et pas de produire une estimation précise à la circulation près. La transition vers un parc 100% équipé sera longue, mais c'est bien cette cible qui est visée.**
- SNCF Réseau précise suite à une question que le critère de longueur du train n'est pas encore pris en compte dans les modèles. Ce paramètre sera étudié par SNCF Réseau.

3. Critères de validation de l'algorithmie d'estimation

Principaux éléments présentés

- Deux critères clés sont utilisés pour mesurer la qualité de l'estimation :
    - Le **biais** est une évaluation de l'écart entre les valeurs d'énergie estimées et les valeurs réelle mesurées.
    - La **dispersion** décrit la variabilité des valeurs d'estimation.
- Il faut viser un biais minimum et une dispersion minimale.
- Le biais peut s'estimer via la moyenne des valeurs par niveau d'agrégation.
- Les dispersions peuvent s'estimer en utilisant le rapport entre l'écart type et la médiane d'une estimation.
- La validation sera réalisée en agrégeant à différents niveaux
    - Sur une série de circulations identiques
    - Au périmètre d'un type de circulation
    - Au périmètre d'une EF, sur un mois.

Diffusable SNCF RESEAU

- Les sensibilités seront aussi vérifiées par rapport aux variables : Masse, Vitesse, temps de parcours
- SNCF Réseau présente en séance un exemple de résultats obtenus sur un ensemble de circulations entre deux villes données, dans des conditions équivalentes.

Echanges en séance

- Sur la question du seuil à partir duquel une estimation sera considérée comme satisfaisante, SNCF Réseau explique que l'observation des consommation mesurées (engins équipés de compteurs) montre, pour une même circulation, même trajet, même type d'engin, même composition, on constate des variations de +/- 10%. La variation de l'estimation de +/- 9,43% est située dans l'intervalle de variation de la consommation mesurée. Ceci implique qu'il n'est pas pertinent de viser une meilleure précision que l'intervalle de variation de la consommation mesurée avec les modèles d'estimation.

4. Processus de qualification, validation et de bascule

Principaux éléments présentés

- RTE et SNCF Réseau présentent les étapes préalables à la mise en œuvre opérationnelle des estimations ORES dans le décompte (*plus de détails dans les slides 26 à 28*) :
    - Premier jalon : Mise en Production DECOFER & ORES interconnectés
        - Observation et évaluation de la performance des modèles d'estimation sur la base des données réelles
        - Les résultats de l'observation seront partagés avant le démarrage de la marche en double.
    - Deuxième jalon : Mise en place de la marche en double
        - Dans cette phase (durée : minimum 3 mois), le décompte officiel publié aux acteurs de marché reste le décompte sans données ORES.
            - En parallèle, RTE et SNCF Réseau mettent à disposition des acteurs qui le souhaitent, le futur décompte, incluant les données ORES
        - La marche en double permettra
            - De confirmer les modèles
            - D'évaluer l'impact du nouveau calcul de la consommation de chaque EF, par rapport au calcul existant
            - D'analyser les éventuels écarts
    - Date à déterminer (S2 2022) – le décompte inclut ORES

- RTE et SNCF Réseau affichent un planning résumant les évolutions des périmètres de services en 2022. La phase 2 telle qu'initialement prévue sera finalement mise en place en deux étapes :
    - Etape 1 en avril 2022 : RTE devient responsable du décompte ferroviaire
        - RTE Télérelève les EMS 2017 et les protocoles RFF
        - RTE calcule la consommation des engins équipés de compteurs

Diffusable SNCF RESEAU

- Etape 2 : bascule de la consommation non équipée sur les estimations ORES, prise en compte du Journal des Circulations. RTE intègre l'estimation ORES dans la consommation de l'EF.

- Puis la phase 3, début 2023, prévoit la bascule du terme bouclant sur le GI.

Echanges en séance

- Concernant les critères de validation des modèles présentés en partie 3, RTE précise que ces critères ont pour but de valider la pertinence et la précision des estimations dans cette phase amont du projet (phase d'observation et évaluation, puis phase de marche en double). Une fois ORES opérationnel, RTE utilisera l'estimation fournie comme donnée d'entrée par le GI sans l'interroger systématiquement.
- SNCF Réseau précise que pendant la marche en double, un des objectifs sera de s'assurer de la cohérence de l'ensemble des courbes de charges des EF, en lien avec la consommation globale du RFN.
- RTE précise que certaines EF, aujourd'hui très largement équipées, pourront faire l'objet d'une bascule assez rapide sur les estimations ORES. En revanche, c'est plus complexe pour le EF moins équipées. Cependant, il est envisageable de comparer la déclaration actuellement réalisée pour ces EF avec l'estimation ORES. Le contrat de prestation annexe proposé par RTE permet de couvrir l'ensemble des cas de figure.

Diffusable SNCF RESEAU