

ORES – DOCUMENTATION TECHNIQUE GLOBALE



SOMMAIRE

1	INTRODUCTION	4
1.1	Contexte.....	4
1.2	Organisation du DOCUMENT.....	4
1.3	Terminologie	5
2	PRESENTATION DE L'OUTIL ORES	6
2.1	Fonctionnement global	6
2.2	Synthèse des fonctionnalités.....	7
2.3	Les différentes chaines de traitement d'ORES.....	8
2.4	Source des données	10
2.4.1	Infocentre Circulation	10
2.4.2	DECOFER	10
2.4.3	GSM-R	11
2.4.4	GAIA	11
2.4.5	Météo	11
3	PRESENTATION DE LA SOLUTION TECHNIQUE.....	12
3.1	interpolation des circulations	12
3.2	Rapprochement des consommations électriques aux circulations	12
3.2.1	Rapprochement GSM-R	12
3.2.2	Rapprochement GPS.....	14
3.3	Estimation de la consommation électrique	15
3.3.1	Modélisation de la circulation	15
3.3.2	Construction de la base d'apprentissage des modèles d'estimation.....	16
3.4	Journal des consommations	17
3.4.1	Traitement des données de circulation.....	17
3.4.2	Ajout des rapprochements	17
3.4.3	Calcul de l'équipement des engins	18
3.4.4	Estimation d'ORES	18
3.4.5	TKE	19
3.4.6	Circulations fantômes.....	19
3.5	Courbe de charge	19
3.5.1	Transmission des données à RTE.....	19

3.5.2	Ajout de la consommation résiduelle (Hors circulation)	19
3.5.3	Génération des livrables avec manque de données source	19
3.6	consommation résiduelle	20
3.6.1	Calcul de la consommation résiduelle	20
3.6.2	Intégration de la consommation résiduelle dans les courbes de charges	20
3.7	Validation	20
3.7.1	Validation du JDC	20
3.7.2	Remplacement	21

1 INTRODUCTION

1.1 CONTEXTE

La facturation de l'énergie électrique de traction dans le secteur ferroviaire présente plusieurs défis majeurs. La principale difficulté réside dans l'attribution correcte de l'énergie consommée aux différentes entreprises ferroviaires (EF), notamment en raison du faible taux d'équipement en compteurs électriques des trains. Cette situation engendre des lacunes dans les données de consommation, rendant difficile une facturation précise. De plus, le réseau d'alimentation subit des pertes électriques entre les sous-stations et le point de captage par les trains, compliquant encore davantage les calculs. L'estimation de la consommation d'énergie devient donc cruciale pour compléter les données de consommation issues des compteurs embarqués et déterminer les pertes attribuées à SNCF Réseau. Le relevé de la consommation électrique des trains équipés de compteurs est réalisé grâce à l'outil DECOFER développé par RTE.

Dans ce contexte, le décompte ferroviaire joue un rôle central. Il permet de relier les circulations ferroviaires et les données d'alimentations issues des sous-stations afin de répartir les consommations électriques entre les différentes EF et le gestionnaire de réseau.

Pour faire face à ces défis, SNCF Réseau a développé un nouvel outil, ORES (Outil de Rapprochement et d'Estimation). ORES repose sur des modèles d'estimation de la consommation électrique, créés à partir de données réelles issues des données de circulation et des compteurs embarqués. Ces données sont utilisées comme une base de données d'apprentissage afin de prédire la consommation électrique des engins non équipés de compteurs.

Le rapprochement entre les données de consommation et les données de circulation qui consiste à associer chaque circulation effectuée à un identifiant d'engin unique est un prérequis à la réalisation des estimations.

1.2 ORGANISATION DU DOCUMENT

L'ordre des sections suit l'ordre des traitements effectués par ORES. La première partie est consacrée aux rapprochements entre les données de circulation et les données de consommation. Puis, la modélisation de la circulation ainsi que l'estimation de la consommation électrique sont développées. Enfin les livrables Journal des Consommations et Courbes de Charge sont décrits.

1.3 TERMINOLOGIE

TERME	DEFINITION
CDC	Courbe de Charge
DECOFER	Nom de la base de données RTE contenant la consommation électrique télérelevée.
EF	Entreprise Ferroviaire
Engin	Un engin est un assemblage d'une caisse munie de ses équipements nécessaires, de bogies et de moteurs électriques et ou thermiques
Fret	Activité ferroviaire de transport de marchandise.
GSM-R	GSM-R désigne le réseau téléphonique GSM installé par SNCF réseau le long des lignes de trains.
HC	Hors Circulation
Id MR	L'identifiant du matériel roulant est un numéro permettant d'identifier tous les caisses et engins formant une rame.
IDFM	Île-de-France Mobilités
JDC	Journal des consommations
NULOC	Numéro d'engin dans la base de données GSMR. Il existe une méthode systématique pour convertir le numéro d'engin qui est dans DECOFER en NULOC.
NUTRA	Numéro de circulation dans la base de données GSMR. Il existe une méthode systématique pour convertir le numéro de circulation qui est dans Infocentre Circulation en Neutra.
ORES	Outil de Rapprochement et d'Estimation de SNCF-Réseau
Point Kilométrique	Le Point Kilométrique (PK) est un repère kilométrique permettant de localiser avec précision : un lieu, une installation ferroviaire ou encore un incident. Le PK est une information donnée par le conducteur au PC, en cas de panne de son train. Ce principe est le même que celui utilisé sur l'autoroute, où l'on retrouve également, le long des voies de circulation, des Points Kilométriques qui permettent de se repérer.
Point Remarquable	Un point remarquable (PR) correspond à un point facilement repérable par un agent de conduite (fiche horaire). Il est identifié par un Code Immuable (CI) et un code Chantier (CH).
Rapprochement	On appelle rapprochement une association entre un numéro de circulation et un numéro d'engin.
TAGV	Train Apte à la Grande Vitesse
TCT	Abréviation pour la catégorie de Type de Convoi-Trafic, qui décrit le type de la circulation (fret, TAGV, etc.).
TKE	Train kilomètre électrique

2 PRESENTATION DE L'OUTIL ORES

2.1 FONCTIONNEMENT GLOBAL

ORES est un outil permettant le rapprochement des données de circulation avec les données de consommations, ainsi que de l'estimation de la consommation des circulations non équipées de compteurs. ORES est structuré autour de deux blocs fonctionnels majeurs :

- Le module de « rapprochement » associe chaque circulation électrique sur le réseau à un identifiant spécifique d'engin moteur,
- Le module « estimation » calcule la consommation estimée pour chaque circulation.

L'outil s'appuie également sur diverses bases de données, telles que :

- Les données de GSM-R, qui constituent une base de données enregistrant l'ensemble des transactions générées par les boîtiers embarqués sur les engins,
- Les données de consommation des engins équipés de compteur télérelévées par DECOFER,
- Les données d'Infocentre Circulation, qui détaillent l'ensemble des circulations sur le réseau ferroviaire,
- Les données de GAIA, issues d'un référentiel décrivant le réseau ferroviaire,
- Les données de la météo.

L'ensemble de ces données permet à ORES de réaliser à la fois le rapprochement et l'estimation de la consommation des circulations, aboutissant à la production de deux livrables principaux :

- Le journal des consommations (JDC) : Un fichier exhaustif regroupant toutes les données de consommation d'énergie de traction par circulation,
- La courbe de charge des consommations (CDC) : Elle est calculée à l'échelle de chaque circulation ou EF, et permet de suivre les consommations énergétiques au pas de 5min.

Ces livrables permettent de réaliser le décompte ferroviaire, tout en suivant un processus d'échange avec RTE, ainsi que le calcul du terme bouclant de SNCF Réseau, qui consiste à soustraire la consommation des entreprises ferroviaires (EF) à la consommation mesurée aux bornes des sous-stations sur le Réseau Ferré National (RFN).

Ci-dessous, un schéma global d'ORES illustrant les données et les modules nécessaires à la réalisation du décompte ferroviaire et au calcul du terme bouclant :

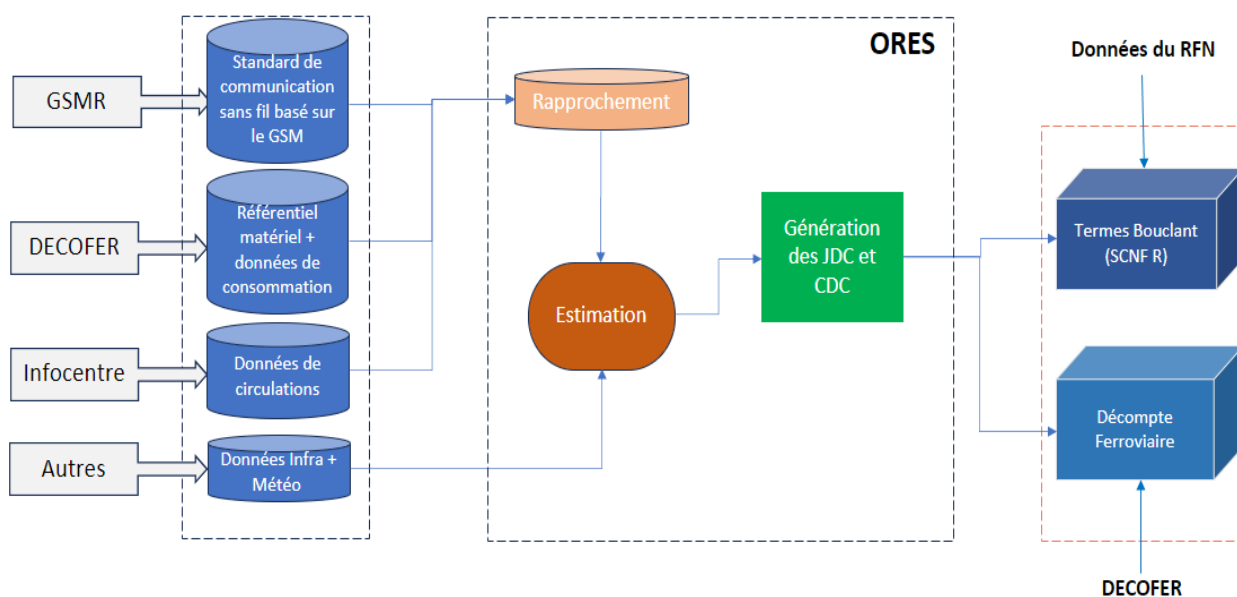


Figure 1 - Fonctionnement globale d'ORES

2.2 SYNTHÈSE DES FONCTIONNALITÉS

Les traitements d'ORES permettent de créer les journaux de consommation qui résument la consommation des circulations et les courbes de charge qui regroupent la consommation des entreprises ferroviaires tout au long de la journée. ORES traite les données d'entrée jour par jour.

Un algorithme ORES, appelé *Pilote*, surveille la disponibilité des données issues d'Infocentre Circulation. Une fois ces données disponibles, il déclenche automatiquement les traitements ORES pour les jours J-4, J-12, J-50 et J-300. Il effectue ensuite une vérification dans la table de suivi pour identifier les traitements qui sont tombés en erreur afin de les relancer.

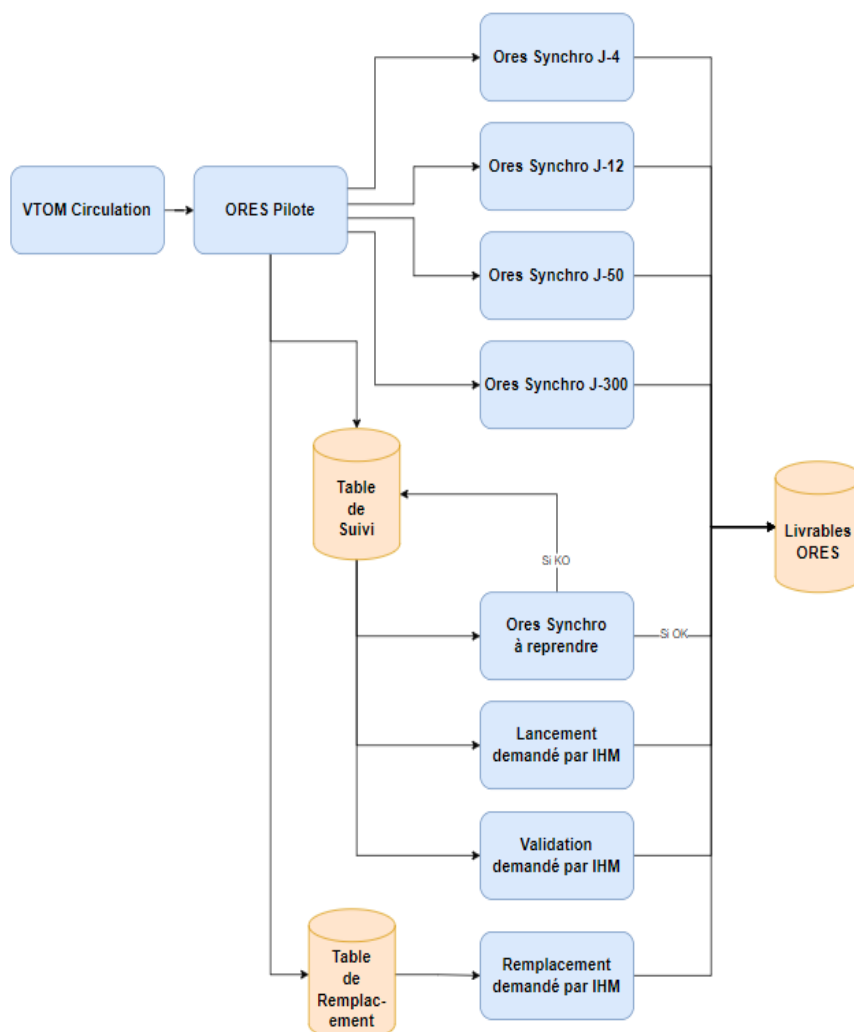


Figure 2 - Fonctionnement du pilote d'ORES

2.3 LES DIFFERENTES CHAINES DE TRAITEMENT D'ORES

Les traitements d'ORES utilisent trois paramètres :

- Le premier est l'environnement d'exécution des scripts. Certains référentiels dépendent de l'environnement.
- Le deuxième paramètre est la date du jour qui sera traitée.
- Le troisième est le mode d'exécution : run (chaîne de traitement classique), validation (chaîne de validation) ou remplacement (chaîne de remplacement).

Les traitements nominaux fonctionnent de la manière suivante (mode = run) :

- La partie 0 consiste à vérifier la complétude des sources de données entrantes d'ORES (Infocentre Circulation, DECOFER et GSM-R),
- La partie 1 consiste à calculer les rapprochements GPS en trois étapes : l'interpolation de la position GPS de la circulation toutes les 5 minutes, puis la création des tables de contingence et enfin le calcul

des rapprochements GPS. La deuxième partie consiste à calculer directement les rapprochements GSM-R,

- La partie 3 modélise la circulation et estime les consommations associées. Les parties 4 et 5 calculent respectivement les journaux de consommation et les courbes de charge. Cette partie 5 inclut aussi le calcul des résidus hors circulation à la journée,
- La partie 6 permet de créer les courbes de charges à la maille circulation,
- La partie 7 intègre et calcule le terme bouclant.

Une fois tous les traitements terminés, ORES met à jour la table de suivi.

Il existe deux autres chaînes de traitements : les validations et les remplacements. L'objectif des validations est de pouvoir valider l'estimation de toutes les circulations pour vérifier qu'ORES envoie des données correctes à RTE via API. Cette partie de validation permet de modifier l'estimation si nécessaire avant de valider.

Un traitement de validation se déroule comme suit :

Dans un premier temps, ORES valide le journal des consommations, validation et modification si nécessaire, ensuite il relance le calcul des courbes de charges, pour impacter les modifications qui ont été réalisées dans le journal des consommations sur les courbes de charges. Finalement la validation des courbes de charges est lancée. Quand les traitements de ces trois scripts sont terminés la table de suivi est mise à jour.

Pour ce qui est des remplacements, le livrable courbe de charge d'une journée est remplacé par celui d'une autre en changeant la date de ce dernier. Pour ce traitement, il faut ajouter des paramètres supplémentaires à ceux des deux autres car il nous faut obligatoirement deux dates : une à remplacer et une date de remplacement. Ainsi en plus des paramètres d'environnements et de mode (qui doit être à 'remplacement'), nous avons besoin des champs : *remplacement_date*, *replaced_date* et *indice_remplacement* qui se remplit automatiquement si jamais il est lancé depuis le pilote.

2.4 SOURCE DES DONNEES

FILIERE	FOURNISSEUR	DONNEES	FREQUENCE DE COLLECTE	LOCALISATION (L, V, PR, PK, X, Y)
Usage du réseau	Infocentre Circulation	Circulation	Quotidienne	L, PR, PK
Usage du réseau	DECOFER	Consommation	Quotidienne	X, Y
Usage du réseau	GSM-R	Appels téléphoniques à partir des motrices	Quotidienne	X, Y
Description du réseau	GAIA	Structure réseau	Trimestrielle	L, PK
Météo	PSS	Météo	Quotidienne	X, Y

2.4.1 Infocentre Circulation

Infocentre Circulation décrit toutes les circulations qui ont eu lieu sur le RFN. La circulation est dite jalonnée puisque l'on a accès à la position des trains seulement quand ils passent à proximité d'un point remarquable. Ainsi, pour chacune des circulations, Infocentre Circulation décrit :

- Le « jal_dci_id », identifiant unique d'une circulation,
- Le numéro de circulation, numéro unique sur la journée,
- La position du train associée au point remarquable, au format Ligne / PK,
- La date de passage au point remarquable, au format CET,
- La distance parcourue entre chaque point remarquable.

2.4.2 DECOFER

DECOFER est une base de données alimentée quotidiennement par RTE. Elle contient toutes les consommations des engins équipés de compteur au cours du temps. Pour chaque engin enregistré, la base de données décrit :

- Le numéro de l'engin qui consomme de l'énergie,
- Le GUID de la sous-famille associée à l'engin,
- L'énergie active positive, négative et nette consommée par l'engin,
- L'entreprise ferroviaire propriétaire de l'engin,
- L'entreprise ferroviaire utilisatrice de l'engin, une entreprise peut louer un engin à une autre,
- Le GUID de la rame associée à l'engin,
- L'EVN de tête associée à l'engin,
- L'EVN de queue associée à l'engin,
- La date de consommation, toute les 5 minutes au format UTC,
- La position GSP de l'engin.

Si la donnée DECOFER n'est pas présente, les livrables d'ORES sont quand même générés, en revanche les rapprochements GPS et GSM-R (3.2) ne sont pas effectués.

2.4.3 GSM-R

La base de données GSM-R enregistre toutes les transactions émises par les boîtiers GSM des engins. Ces données sont télérelevées et transitent par le réseau 2G. ORES utilise les enregistrements relatifs aux communications émises par le personnel de bord ou les différentes applications automatiques. Ces communications sont de plusieurs natures :

- Les communications entre le mécanicien et le dirigeant de service,
- Les communications générales du personnel de bord,
- Les envois de messages du train pour le service de contrôle du train,
- Des applications spécifiques comme la communication dans les tunnels, le transfert et l'analyse des données de service, des données de contrôle...etc.,
- Une communication à chaque début de circulation pour déclarer le numéro de sillon,
- Une communication à chaque fin de circulation pour libérer le numéro de sillon.

Ainsi avec ces différentes communications, la base de données GSM-R lie trois informations entre elles :

- La date de la transaction, au format CET,
- Le NULOC, qui est l'équivalent du numéro de l'engin. Ce numéro est censé être immuable et unique,
- Le NUTRA, qui est l'équivalent du numéro de circulation. Ce numéro est censé être unique pour chaque journée.

Si la donnée GSM-R n'est pas présente, les livrables d'ORES sont quand même générés. Ne seront pris en compte que les rapprochements GPS.

2.4.4 GAIA

L'ensemble des référentiels décrivant le RFN sont générés à partir des données présentées dans GAIA.

2.4.5 Météo

Les données météo sont fournies par Météo France, ORES utilise la température de l'air.

3 PRESENTATION DE LA SOLUTION TECHNIQUE

3.1 INTERPOLATION DES CIRCULATIONS

La circulation fournie par Infocentre Circulation est une circulation décrite par les horaires de passage aux points remarquables. De ce point de vue, nous maîtrisons la localisation des trains mais l'horaire de passage aux points remarquables est soumis aux aléas quotidiens. Nous souhaitons inverser ce paradigme pour maîtriser les horaires de passage (toutes les 5 minutes).

Pour estimer la position des trains toutes les 5 minutes, une interpolation linéaire est faite entre la position du train et l'horaire de passage.

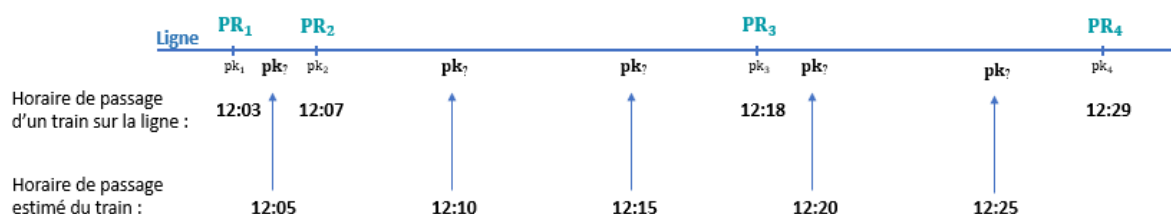


Figure 3 – Schéma linéaire d'une ligne avec exemple de PR et PK

Une fois cette interpolation réalisée, la circulation jalonnée a été transformée en une circulation horodatée toutes les 5 minutes. Il est maintenant possible d'associer la circulation avec la consommation électrique.

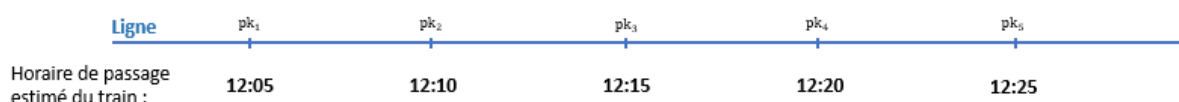


Figure 4 – Schéma linéaire d'une ligne avec regroupement par PK

Cette interpolation est faite en ligne PK alors que pour les rapprochements GPS il s'agissait d'une interpolation par rapport aux coordonnées GPS.

3.2 RAPPROCHEMENT DES CONSOMMATIONS ELECTRIQUES AUX CIRCULATIONS

Dans un premier temps, il est nécessaire de fiabiliser le calcul du décompte ferroviaire et de disposer d'un couplage fiable entre les données de circulation des trains et les données de consommation électrique. Le rapprochement des données de mesure d'énergie télérelevées avec les données de circulation est un prérequis à l'établissement des estimations de consommations. Nous décrivons ici les deux traitements utilisés par ORES.

3.2.1 Rapprochement GSM-R

Le rapprochement GSM-R se fait en deux temps. Premièrement, un référentiel comportant les triplets (Engin, EVN de tête, EVN de queue) est créé à partir des données DECOFER du jour. En parallèle, le référentiel matériel voyageur est généré. Il contient les triplets (EVN, Immatriculation, Id MR) auxquels est ajouté le NULOC à partir de règles de conversions. Ainsi il est possible de créer un référentiel DECOFER comprenant les informations :

- Numéro d'engin,
- EVN,
- NULOC,
- Id MR,
- Immatriculation.

Puis la base de données GSM-R du jour est enrichie avec le référentiel précédemment calculé via le champ en commun : le NULOC. Finalement, il a été construit une table contenant les informations

- Numéro d'engin,
- EVN,
- NULOC,
- Id MR,
- Immatriculation,
- NUTRA.

Dans un second temps, l'ensemble des circulations du jour sont prises en compte. A partir du numéro de circulation, le NUTRA est construit grâce à des règles de conversions. Ainsi pour chaque `jal_dci_id` un NUTRA est calculé. A partir de ce NUTRA, il est possible de remonter au NULOC via la base de données GSM-R, puis au numéro d'engin via le référentiel DECOFER.

Ainsi une table de rapprochement GSM-R est créée et associe un `jal_dci_id` à un numéro d'engin. C'est-à-dire, une circulation à sa consommation.

Il est à noter qu'il est toujours possible de calculer un NUTRA à partir d'un numéro de circulation mais que le NUTRA peut être absent de la base de données GSM-R. De même, il se peut que le NULOC associé au NUTRA soit invalide, multiple ou absent du référentiel DECOFER. Dans tous ces cas cités, une trace est gardée dans la table de rapprochements GSM-R.

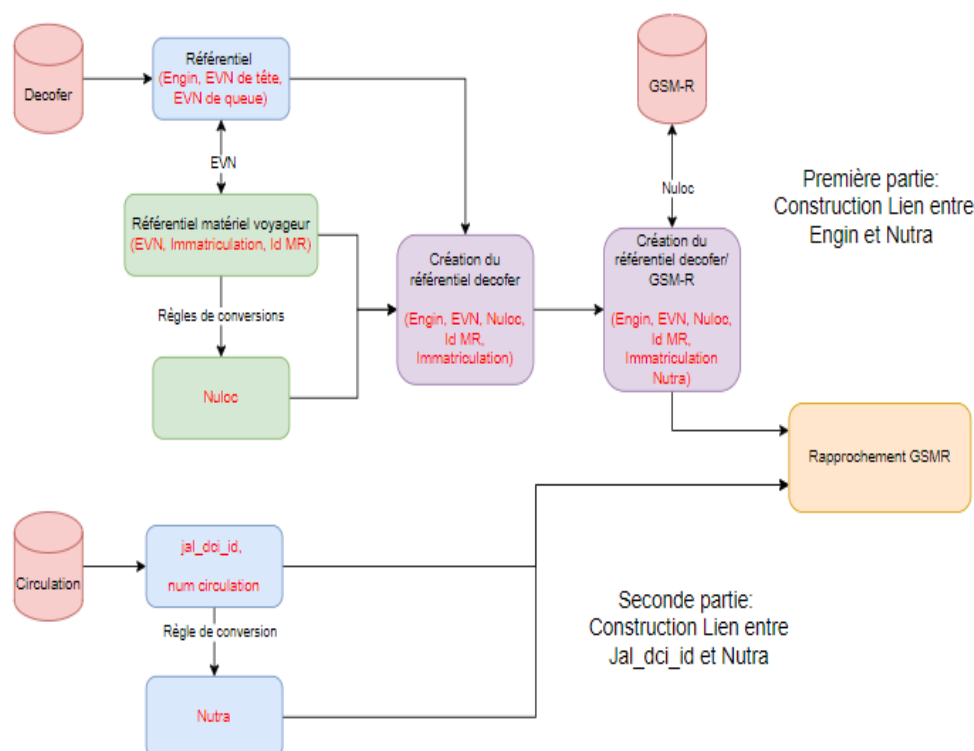


Figure 5 - Fonctionnement des rapprochements GSM-R

Enfin, certains rapprochements seront fiabilisés dans la création du journal des consommations.

3.2.2 Rapprochement GPS

Le rapprochement GPS se fait en trois étapes, séparées en trois scripts distincts. Les étapes sont les suivantes : l'interpolation GPS, la création d'une table de contingence et enfin la création des rapprochements GPS.

Premièrement, la position des trains est donnée aux points remarquables avec une date et heure de passage associée. La position des points remarquables est donnée par les couples lignes, PK. Ces positions sont transformées en coordonnées GPS. Enfin la position GPS des trains est interpolée linéairement toutes les 5 minutes.

Dans un second temps, les tables de contingence sont calculées. La position GPS de chaque engin présent dans la table DECOFER est donnée toutes les 5 minutes. Seules les mesures considérées comme étant « en mouvement » sont gardées. Une mesure est en mouvement si l'engin s'est déplacé de 500 mètres en 5 minutes. Puis les positions des circulations interpolées toutes les 5 minutes sont rapprochées aux positions GPS des engins si elles sont à moins de 3 km l'une de l'autre (2 km pour les transiliens). Les tables de contingence rapprochent les engins aux circulations. Puis les couples engin $\Leftrightarrow$ circulation avec un TCT incohérent avec l'entreprise utilisatrice de l'engin sont supprimés.

Enfin, il faut sélectionner les bons couples engin $\Leftrightarrow$ circulation. En se plaçant du point de vue de la circulation, on garde l'engin le plus probable en faisant un test binomial. Théoriquement, l'engin qui a effectué la circulation devrait apparaître à chaque fois. Malheureusement, avec les aléas de la disponibilité des données ou de l'imprécision des interpolations ce n'est pas toujours le cas. Pour chaque circulation, on compte le nombre de fois qu'un engin a été vu à moins de 3km pour voir si la proportion observée est significativement

différente de la proportion attendue. Dans les cas où il y aurait un doute entre plusieurs engins, seul l'engin le plus probable est retenu. Pour chaque rapprochement la p-valeur est enregistrée et un rapprochement GPS est validé si et seulement la p-valeur est inférieure à 5%.

3.3 ESTIMATION DE LA CONSOMMATION ELECTRIQUE

Afin d'estimer la consommation des circulations, des modèles d'estimation ont été développés dans ORES, avec un modèle spécifique pour chaque type de convoi (TAGV, train régional de voyageurs conventionné hors IDFM, train de voyageurs longue distance, train fret, train régional de voyageurs conventionné IDFM et train de l'infrastructure) fondé sur des caractéristiques propres en termes de vitesse, d'accélération et de distance parcourue. Ceci permet une optimisation maximale des prédictions.

Les modèles d'estimation reposent sur l'algorithme d'apprentissage supervisé "XGBoost", une méthode qui combine plusieurs modèles faibles (arbres de décision) pour créer un modèle global plus précis et robuste.

Pour l'entraînement des modèles, une base d'apprentissage est mise en place. Cette base fournit des données de consommation des engins équipés de compteurs et des caractéristiques des circulations nécessaires pour que les modèles puissent apprendre et s'adapter aux différents scénarios de circulation.

Une fois les modèles créés et entraînés, ils sont intégrés dans ORES, où ils sont utilisés pour faire des prédictions sur les nouvelles circulations. Ces prédictions permettent d'estimer avec précision la consommation d'énergie des trains.

3.3.1 Modélisation de la circulation

L'objectif de la modélisation est de pouvoir associer 5 minutes de circulation avec 5 minutes de consommation. Avec cette correspondance bijective entre la circulation et la consommation, la base d'apprentissage va être construite pour permettre d'entraîner plusieurs modèles d'estimation de la consommation électrique.

Toutes les 5 minutes, plusieurs variables sont calculées. Les grandeurs physiques relatives à la circulation :

- La masse du train théorique, en tonnes, est disponible via Info-Centre,
- La masse du train déclaré, en tonnes, est disponible via Info-Centre,
- La longueur du train, en mètres, est déclarée via Info-Centre,
- La vitesse et l'accélération,
- La vitesse au carré et la vitesse au cube,
- Vitesse maximale et minimale pendant les 5 minutes,
- La longueur de la portion parcourue en mètres,
- Le nombre d'arrêts pendant 5 minutes,
- Le TCT (TAGV, fret...).

Les caractéristiques relatives à la description du réseau et à son environnement :

- La déclivité positive, déclivité négative et déclivité totale pendant 5 minutes,
- Le nombre de mesures de déclivité positive et le nombre de mesures de déclivité négative pendant 5 minutes,
- Le rayon de courbure moyen, en mètres,
- Le pourcentage de la portion parcouru en courbe, compris entre 0 et 1,
- Le type d'alimentation (1,5 kV, 25 kV, 2X25kV),

- La Température de l'air, en °C.

Les calculs d'interaction :

- La vitesse multipliée par le tonnage,
- La vitesse multipliée par la déclivité positive,
- La vitesse multipliée par la déclivité négative,
- La vitesse multipliée par l'accélération,
- La vitesse multipliée par le tonnage multipliée par la déclivité positive,
- La vitesse multipliée par le tonnage multipliée par la déclivité négative,
- La vitesse multipliée par le tonnage multipliée par l'accélération.

Une fois les variables calculées, il convient ensuite d'attribuer la consommation électrique correspondante en utilisant les rapprochements. Ainsi nous disposons d'une base d'apprentissage qui fait correspondre les caractéristiques de la circulation avec la consommation associée. Il est à préciser que toutes les rames ont un unique compteur sauf pour les TAGV qui en ont deux. Ainsi pour les TAGV, il faut que les deux compteurs soient rapprochés pour que la circulation associée soit prise en compte dans l'entraînement.

3.3.2 Construction de la base d'apprentissage des modèles d'estimation

Tous les enregistrements présents dans la base d'apprentissage ne sont pas utilisables tels quel. Certains sont incohérents, d'autres sont clairement des anomalies, d'autres encore comportent des erreurs d'acquisition dans le cheminement de la donnée. De plus, il a été identifié que certains compteurs étaient défectueux : la consommation est soit trop grande, soit trop faible. Pour les identifier nous avons mis en place une méthode statistique. Ainsi, il existe deux types d'anomalies, les anomalies à la maille 5 minutes et les anomalies à la maille circulation.

3.4 JOURNAL DES CONSOMMATIONS

Le journal des consommations est l'un des deux livrables fournis par ORES. Il regroupe les données de consommation à la maille circulation. Les grandes étapes sont :

- L'import et la mise en forme des données de circulation,
- L'ajout des données relatives aux rapprochements,
- L'ajout des données de consommations télérelevées par DECOFER et estimées par ORES,
- La fiabilisation des variables d'intérêt,
- L'ajout des informations des différents référentiels.

3.4.1 Traitement des données de circulation

Les journaux de circulation se basent directement sur la donnée d'infocentre circulation enrichie par quelques informations.

La colonne « gi » pour gestionnaire d'infrastructure est créée. Par défaut, elle vaut la variable « Réseau » pour indiquer la ligne utilisée est gérée par SNCF Réseau. Les lignes '566000', '538315', '570345', '570360', '538311', '570380', '570390', '570385', '570350' sont attribuées à Liséa et la ligne '837000' à la ligne Figueras Perpignan (LFP).

La colonne « CA » est créée à partir d'un référentiel pour indiquer si l'entreprise est un candidat autorisé ou pas.

3.4.2 Ajout des rapprochements

Les rapprochements GSM-R sont importés et enrichis avec les référentiels DECOFER et d'identifiant de matériel roulant. Ainsi une table temporaire des rapprochements GSM-R est créée avec les colonnes suivantes :

- 'jal_dci_id'
- 'ID_RAME_GSMR'
- 'ENGIN'
- 'NUM_ENGIN_GSMR'
- 'NULOC'
- 'CODE_ENGIN_GSM_MATCH'
- 'PREMIER_MATCH_GSMR'
- 'DERNIER_MATCH_GSMR'
- 'FAMILLE_GSMR'
- 'SOUS_FAM_GSMR'
- 'EF_PROPRIE_GSMR'
- 'EF_UTILISA_GSMR'
- 'guid_rame_gsmr'

Les rapprochements GPS sont importés et enrichis avec les référentiels DECOFER et d'identifiant de matériel roulant. Ainsi une table temporaire des rapprochements GPS est créée avec les colonnes suivantes :

- 'jal_dci_id'
- 'ID_RAME_GPS'
- 'ENGIN'
- 'NUM_ENGIN_GPS'
- 'SCORE_FIABILITE'
- 'PROBA'
- 'ratio_fiab'
- 'PREMIER_MATCH_GPS'
- 'DERNIER_MATCH_GPS'
- 'FAMILLE_GPS'
- 'SOUS_FAM_GPS'
- 'EF_PROPRIE_GPS'
- 'EF_UTILISA_GPS'
- 'guid_rame_gps'

Ces deux tables sont jointes avec le couple de clés 'jal_dci_id' et 'ENGIN'. La colonne 'ENGIN' est renommée en 'NUM_ENGIN_CUMUL' pour indiquer que la circulation a été rapprochée à un même engin à la fois par la méthode GSMR et la méthode GPS.

3.4.3 Calcul de l'équipement des engins

Dans le cas où la circulation est rapprochée à un EVN qui existe sur DECOFER, alors la circulation est considérée comme équipée de compteur.

Si la circulation est rapprochée à un engin qui n'existe pas sur DECOFER, deux cas sont envisagés en se basant sur le référentiel matériel :

- Soit la circulation a été rapprochée à un engin qui n'existe pas dans DECOFER, mais appartient à la même rame qu'un engin existant sur DECOFER alors la circulation est considérée équipée de compteur,
- Sinon la circulation a été faite par une rame qui n'a pas de compteur, elle est donc considérée comme non équipée de compteur.

3.4.4 Estimation d'ORES

La consommation fournie par ORES résulte directement du script d'estimation. Cette consommation est validée par l'équipe métier.

Les principes de calcul des consommations estimées sont les suivants :

- Pour les rames électriques, les estimations sont laissées brutes,
- Pour les rames diesel, les estimations sont réduites à zéro,
- Pour les rames bi-modes, les estimations sont calculées au prorata de l'électrification des lignes par rapport aux kilomètres circulés.

3.4.5 TKE

Le nombre de kilomètre électrifié (TKE) est calculé à partir du référentiel d'électrification des lignes de GAIA :

- Pour les rames électriques, les TKE correspondent au kilomètres circulés,
- Pour les rames diesel, les TKE sont réduits à zéro,
- Pour les rames bi-modes, les TKE sont calculé au prorata de l'électrification des lignes par rapport au kilomètres circulés.

3.4.6 Circulations fantômes

Les circulations qui n'ont aucune date de passage observée dans Infocentre Circulation sont considérées comme « fantômes ». La consommation estimée est mise à 0.

3.5 COURBE DE CHARGE

Les courbes de charge forment le deuxième livrable produit par ORES. Elles sont générées à partir des journaux de consommation et des prédictions augmentées du jour J, J-1 et J-2. Ce paragraphe résume les grandes étapes pour générer les courbes de charge.

Dans un premier temps les prédictions augmentées des jours J, J-1 et J-2 sont importées et filtrées :

- De minuit à minuit CET
 - Soit de 23h à 23h l'hiver,
 - Soit de 22h à 22h l'été.
- Réduction du périmètre
 - GI : SNCF Réseau.

Dans un deuxième temps, les données de consommation rapprochées sont ajoutées à la circulation. Les données sont ensuite regroupées par code UI, TCT, date et équipé ou non. Par groupe ainsi créé, le nombre de circulation uniques sont comptés, les consommations estimées, le nombre de kilomètres circulés, le tonnage et la consommation télérelevée sont sommés. Puis les résidus sont ajoutés en prenant en compte les modifications à apporter via les répertoires DEC.

Enfin du formatage est fait pour respecter les noms des colonnes définis dans les US (User Story) et le contrat d'interface.

3.5.1 Transmission des données à RTE

ORES est une application qui estime la consommation de l'ensemble des trains roulants sur le RFN. Cependant seulement une partie de la consommation de ces trains est transmise à RTE. Pour faire la différence entre ce qui est envoyé ou non, un référentiel est utilisé. Ce référentiel permet de déterminer la part qui doit être transmise à RTE pour chaque EF et chaque code UI. Toutes ces informations sont ensuite déposées dans une colonne dédiée à RTE.

3.5.2 Ajout de la consommation résiduelle (Hors circulation)

La consommation hors circulation est ajoutée dans le script des courbes de charges. Elle représente la part de la consommation des trains qui ne sont pas comptabilisée dans une circulation.

3.5.3 Génération des livrables avec manque de données source

En cas d'absence de données GSM-R ou DECOFER, le livrable « courbe de charge » est tout de même généré. Des évolutions spécifiques ont été développées pour permettre sa génération même en l'absence de ces données.

3.6 CONSOMMATION RESIDUELLE

La consommation hors circulation, ou consommation résiduelle représente la part de consommation effectuée en dehors des circulations. Par exemple quand les trains se mettent en conditionnement avant les circulations ou quand ils font des manœuvres en gare entre les circulations. Cette consommation est calculée en multipliant la consommation de la journée par un coefficient hors circulation propre à chaque TCT.

Pour ce faire, on sépare les données de consommation en deux, celles qui sont en circulation et celles qui ne le sont pas. Pour cela, on utilise les 2 types de rapprochements GPS et GSM-R d'ORES. A partir de ces données, on peut calculer la quantité de kWh consommé durant une semaine pour un TCT et trouver le pourcentage que cette consommation représente sur la consommation totale de ce TCT. C'est ce pourcentage de consommation que l'on appelle coefficient de consommation hors circulation.

3.6.1 Calcul de la consommation résiduelle

Une étude a permis d'adopter un calcul du hors circulation à la journée. Concrètement le hors circulation est calculé au réel ce qui signifie que l'on récupère le hors circulation de la journée J pour l'ajouter à cette même journée J.

Ainsi chaque jour, il est généré un fichier de résidu dans lequel il y a le résidu de la journée J pour chaque triplet ('tct', 'uic', 'engin_equipe'). Ce calcul suit les étapes suivantes :

1. Récupération du JDC de la journée J,
2. Récupération des coefficients hors-circulation pour la semaine en cours (Attention il s'agit de semaine RTE donc du samedi au samedi, ainsi si le jour J est un samedi, on prend le numéro de la semaine du J+2 et s'il s'agit d'un dimanche on prend celle du J+1),
3. Multiplication de la consommation de la journée par le coefficient hors circulation tct par tct,
4. Passage au format 5 min par 5 min en divisant la consommation résiduelle par 24×12 (représentant respectivement nombre d'heures * nombre de pas de 5 minutes dans une heure),
5. Puis enregistrement sous format csv, une base de données est basée sur le dossier où l'on dépose les fichiers csv pour qu'il soit plus facile ensuite d'utiliser les informations enregistrées.

3.6.2 Intégration de la consommation résiduelle dans les courbes de charges

Lorsque le script 8 est exécuté, une requête est lancée sur la base des résidus hors circulation, pour récupérer tous les résidus de la journée concernée. Cette consommation est par la suite ajoutée à chaque triplet ('tct', 'code_uic', 'equipe_ou_non') pour tous les pas de temps 5 min de la journée générée. Une colonne de résidu est créée sur la courbe de charge afin d'être ajoutée à la consommation. Cette colonne n'est pas transmise, elle est seulement temporaire. Cependant elle peut être recrée en faisant la différence entre les colonnes `energie_active_soutirage_estimee` et `energie_active_soutirage_estimee_brute` (qui sont respectivement les consommations avec hors circulation et sans hors circulation).

3.7 VALIDATION

3.7.1 Validation du JDC

La fonctionnalité de validation des journaux de consommation (JDC) d'ORES permet d'ajuster la consommation de certaines circulations en utilisant les informations suivantes :

- `jal_dci_id`,
- `date_heure_depart`,
- `date_heure_arrivee`.

Ces ajustements peuvent être effectués par les équipes métiers via des fichiers au format CSV, à déposer dans le répertoire dédié.

D'un point de vue technique, un premier script effectue ces modifications pour les circulations nécessitant une mise à jour et modifie également la colonne *is_validate* à *True* pour indiquer que le JDC a été validé. Ensuite, un second script est exécuté pour régénérer les courbes de charge avec les modifications apportées. Enfin, un dernier script valide ces courbes de charge.

3.7.2 Remplacement

La fonctionnalité de remplacement d'ORES permet de substituer la courbe de charge d'une journée par celle d'une autre. En cas d'indisponibilité des données pour une journée, le remplacement est effectué par la courbe de charge correspondante à J-7.