

Document de travail :

**Étude économétrique sur les facteurs
de disparités géographiques de la
ponctualité des trains de voyageurs en
France sur la période 2012-2019**

Alexandre Barbusse, sous la direction d'Alain Sauvart

FÉVRIER 2022



AQOST
AUTORITÉ
DE LA QUALITÉ DE SERVICE DANS LES TRANSPORTS

SOMMAIRE

1	INTRODUCTION	5
2	METHODOLOGIE	6
2.1	Démarche méthodologique et objectifs	6
2.2	Données utilisées	7
2.2.1	Variable à expliquer	7
2.2.2	Variables explicatives	9
2.2.2.1	TGV	9
2.2.2.2	Intercités	14
2.2.2.3	TER	20
2.2.2.4	RER et Transilien	25
2.3	Forme des modèles statistiques testés	33
3.	RESULTATS	34
3.1	Note de lecture des résultats des tests statistiques	34
3.2	Résultats de la modélisation statistique	37
3.2.1	Résultats pour les TGV	37
3.2.1.1	Résultats détaillés des régressions par variables explicatives	37
3.2.1.2	Synthèse des résultats	37
3.2.1.3	Modèle de régression linéaire multiple	40
3.2.2	Résultats pour les trains Intercités	48
3.2.2.1	Résultats détaillés des régressions par variables explicatives	48
3.2.2.2	Synthèse des résultats	48
3.2.2.3	Modèle de régression linéaire multiple	53
3.2.3	Résultats pour les TER	60
3.2.3.1	Résultats détaillés des régressions par variables explicatives	60
3.2.3.2	Synthèse des résultats	60
3.2.3.3	Modèle de régression linéaire multiple	64
3.2.4	Résultats pour les RER et Transilien	71
3.2.4.1	Résultats détaillés des régressions par variables explicatives	71
3.2.4.2	Synthèse des résultats	71

3.2.4.3	Modèle de régression linéaire multiple retenu pour expliquer le taux de retard par ligne	74
3.2.4.4	Modèles de régressions linéaires multiples retenus pour expliquer le taux de retard par branche	85
4.	CONCLUSION	93
	BIBLIOGRAPHIE	99
	ANNEXES	100
	Annexe 1 : Estimation du taux annuel de retard des TER par ancienne région administrative sur la période 2018-2019	100
	Annexe 2 : Distances ferroviaires par liaison TGV	102
	Annexe 3 : Estimation de l'âge moyen des voies des LGV par axe entre 2012 et 2019	105
	Annexe 4 : Nombre de bifurcations majeures des voies LGV par liaison TGV	110
	Annexe 5 : Estimation du nombre d'arrêts moyen par liaison TGV au cours d'un Jour Ouvrable de Base en 2020	112
	Annexe 6 : Estimation de la longueur de ligne parcourue sur le réseau classique par liaison TGV sur la période 2012-2019	114
	Annexe 7 : Résultats détaillés des régressions pour chaque variable explicative retenue	118
	Ponctualité des TGV par liaisons :	118
	Ponctualité des trains Intercités par liaisons :	139
	Ponctualité des TER par ancienne région :	166
	Ponctualité des RER et Transilien d'Île-de-France par ligne :	188
	Annexe 8 : Tableau de colinéarité entre les variables explicatives du taux de retard des liaisons TGV	203
	Annexe 9 : Première analyse des coefficients associés aux variables explicatives du modèle de régression linéaire multiple retenu pour les services TGV	204
	Annexe 10 : Distances ferroviaires par liaison Intercités	209
	Annexe 11 : Estimation du nombre d'arrêts moyen par liaison Intercités au cours d'un Jour Ouvrable de Base en 2020	212
	Annexe 12 : Réseau Intercités en 2016	214
	Annexe 13 : Tableau de colinéarité entre les variables explicatives du taux de retard des liaisons Intercités	215
	Annexe 14 : Première analyse des coefficients associés aux variables explicatives du modèle de régression linéaire multiple retenu pour les services Intercités	217
	Annexe 15 : Nombre moyen de circulations TER quotidiennes aux heures de pointe par ancienne région en 2016	221

Annexe 16 : Nombre moyen d'arrêts effectués par les trains TER sur les liaisons les plus fréquentées par ancienne région	222
Annexe 17 : Estimation de l'âge moyen du matériel roulant TER par ancienne région	228
Annexe 18 : Tableau de colinéarité entre les variables explicatives du taux de retard des TER	241
Annexe 19 : Estimation de la fréquentation quotidienne par ligne de RER et Transilien sur la période 2013-2019	243
Annexe 20 : Estimation du nombre moyen d'arrêts par circulation par ligne RER et Transilien à l'heure de pointe du matin	244
Annexe 21 : Estimation de l'âge moyen du matériel roulant RER et Transilien par ligne	245
Annexe 22 : Estimation de la moyenne du Trafic Moyen Journalier Annuel par voie et de son niveau sur la section la plus chargée	250
Annexe 23 : Tableau de colinéarité entre les variables explicatives du taux de retard des RER et Transilien	259
Annexe 24 : Effet de l'âge moyen du matériel roulant sur le taux de retard des RER et Transilien par ligne hors ligne K	260
Annexe 25 : Cartes SNCF Réseau – mode d'espacement des trains et lignes équipées de contrôle de vitesse en 2018	261
Annexe 26 : Modèles testant l'effet du trafic par branche pour l'ensemble des lignes RER et Transilien	262
Annexe 27 : Modèle testant l'effet du nombre de voies de service en terminus de branche pour l'ensemble des lignes RER et Transilien	266
Annexe 28 : Comparaison des résultats pour des différents services ferroviaires	268
Comparaison des modèles TGV et Intercités	268
Comparaison des modèles TER et RER et Transilien d'Île-de-France	272

1 INTRODUCTION

La présente étude cherche à identifier des variables pouvant expliquer la variabilité géographique du taux de retard des trains de voyageurs en France par activité ferroviaire (TGV, Intercités, TER, RER et Transilien) sur la période 2012-2019. En effet, d'importantes disparités géographiques du taux de retard des trains par liaison, ligne ou région (selon l'activité ferroviaire) apparaissent pour cette période dans les données de la base de l'AQST.

Des données publiques concernant de nombreuses variables explicatives potentielles (intensité du trafic, volume de voyageurs, caractéristiques des liaisons et des lignes ferroviaires, du matériel roulant...) ont été recueillies afin de constituer des bases de données par activité ferroviaire.

Ensuite, des modèles statistiques ont été implémentés à l'aide du logiciel R pour tester la significativité statistique de l'impact de ces variables sur la variabilité géographique du taux de retard par activité ferroviaire sur la période 2012-2019.

2 METHODOLOGIE

2.1 Démarche méthodologique et objectifs

D'importantes disparités géographiques du taux de retard des trains longue distance par liaison des TGV et des trains régionaux de voyageurs - TER par ancienne région et RER et Transilien d'Île-de-France par ligne - apparaissent pour la période 2012-2019 (TGV et Intercités) ou 2013-2019 (TER et RER et Transilien) dans la base de données de l'AQST, comme le montrent les [bilans de la qualité de service](#) publiés par l'AQST.

Afin d'identifier des variables pouvant expliquer cette variabilité géographique du taux de retard par activité ferroviaire, des bases de données ont été constituées par activité ferroviaire à partir de données publiques provenant de diverses sources (OpenData et sites internet des autorités organisatrices et des opérateurs de transport, rapports d'audit, études techniques, schémas directeurs, comités de ligne, fiches horaires, etc.). Pour certaines variables, les données manquantes ont été estimées via un calcul ou une modélisation simple sur la base de données publiques fiables. Les bases de données ainsi constituées rassemblent des variables de différentes natures (intensité du trafic, volume de voyageurs, caractéristiques des liaisons et des lignes ferroviaires, du matériel roulant...).

Dans un premier temps, l'effet de chacune de ces variables sur le taux de retard des trains de voyageurs par activité ferroviaire a été testé économétriquement à l'aide du logiciel R, afin d'identifier les variables qui présentent l'effet individuel le plus significatif statistiquement. Ensuite, des modèles économétriques simples (régressions linéaires multiples, modèles exponentiels, modèles log-log) ont été implémentés pour décrire et interpréter la corrélation entre la ponctualité des trains par activité et les principales variables permettant d'expliquer sa variabilité géographique sur la période 2012-2019.

2.2 Données utilisées

2.2.1 Variable à expliquer

La variable à expliquer est le taux de retard annuel des trains en France par services ferroviaires. Pour cela, les données de la base AQST ont été utilisées.

Le tableau ci-dessous précise la période des données de la base AQST utilisées par service, le seuil de retard utilisé et la granularité géographique.

Service	Période	Seuil de retard en minutes	Granularité géographique
TGV	2012-2019	Composite 5-10-15	Par liaisons ferroviaires
Intercités	2012-2019	Composite 5-10-15	Par liaisons ferroviaires
TER	2013-2019	5 min	Par ancienne régions administratives*
RER et Transilien d'Île-de-France	2013-2019	Ponctualité voyageurs à 5 min (indicateur Île-de-France Mobilités)	Par lignes
	2013-2018		Par branches

***Pour disposer d'une information plus détaillée, le découpage des régions administratives d'avant la réforme territoriale entrée en vigueur le 1^{er} janvier 2016 a été utilisé. Une estimation du taux de retard des TER par ancienne région administrative pour les années 2017 (pour certaines régions), 2018 et 2019 a été effectuée. Elle est présentée en annexe 1.**

Les données du 1^{er} semestre 2020 n'ont pas été intégrées à l'étude, d'une part pour ne pas intégrer de biais statistique si le taux de retard présentait une certaine saisonnalité semestrielle, mais aussi car les données sur le taux de retard au 1^{er} semestre 2020 pourraient être atypiques du fait des conséquences de la grève nationale contre la réforme des retraites du mois de janvier, puis celles de l'épidémie de Covid-19 en France et les mesures adoptées pour lutter contre la propagation du virus, dont notamment le confinement national du 17 mars au 11 mai 2020.

Le tableau ci-dessous détaille davantage les seuils de retard des données utilisées. Les seuils spécifiques de retard sont précisés ci-dessous. Ils sont stricts et relatifs aux minutes entières, arrondies par défaut (i.e. : « >15min » \Leftrightarrow « ≥ 16 min »).

TGV, Intercités			TER	RER et Transilien
Durée ≤ 1 h30	1h30 < Durée ≤ 3 h			
>5min	>10min	>5min	>5min	> 5 min (ponctualité voyageurs)

2.2.2 Variables explicatives

2.2.2.1 TGV

Le tableau de la page suivante détaille les variables explicatives à caractère géographique dont la significativité statistique de l'effet sur le taux de retard a été testée.

Les données associées à certaines variables ont nécessité de recourir à des sources de données ou des méthodes de calcul particulières. Elles sont listées ci-dessous et présentées plus en détail en annexes.

Variable	Description de la variable	Annexe
DistanceMoy	Distance moyenne parcourue par un train entre les 2 gares sur la liaison	Annexe 2
AgeLGV_axe	Age moyen des voies LGV du premier axe TGV emprunté	Annexe 3
NbBifurcationsLGV	Nombre de bifurcations majeures LGV sur la liaison (y compris en Île-de-France) en 2019	Annexe 4
NbMoyenArrêts_2020	Nombre moyen d'arrêts effectués par les TGV sur la liaison au cours d'un Jour Ouvrable de Base (JOB) de 2020 en dehors des périodes de vacances	Annexe 5
LongueurRéseauClassique	Estimation de la longueur parcourue sur le réseau classique pour une circulation de la liaison par année (en km)	Annexe 6

Nom_Variable	Description de la variable	Source des données ou calcul effectué	Remarques
Liaison	Liaison	Base AQST	
Année	Année		
Service	Service		
Axe	Axe		
TxRetard	Taux de retard		
TxAnnul	Taux d'annulation		
NbCirc_Liaison_an	Nombre de trains ayant circulé sur la liaison		
DuréeTrajet	Durée moyenne du trajet sur la liaison		
DistanceMoy	Distance moyenne parcourue par un train entre les 2 gares sur la liaison	Mémento statistique de la SNCF - 2018 ou Rome2Rio	
VitesseMoyenne	Vitesse moyenne des circulations de trains sur la liaison	Distance moyenne parcourue par un train entre les 2 gares sur la liaison / Durée moyenne du trajet sur la liaison	
AgeLGV_axe	Age moyen des voies LGV du premier axe TGV emprunté	Base ART (2015-2018) estimation AQST autres années	
DépartParis	La liaison part de Paris intra-muros (1 si oui ; 0 sinon)		
ArrivéeParis	La liaison arrive à Paris intra-muros (1 si oui ; 0 sinon)		
Liaison_radiale	La liaison part ou arrive à Paris intra-muros (1 si oui ; 0 sinon)		
DépartMarseille	La liaison part de Marseille (1 si oui ; 0 sinon)		
ArrivéeMarseille	La liaison arrive à Marseille (1 si oui ; 0 sinon)		

DépartLyon	La liaison part de Lyon-Part-Dieu (1 si oui ; 0 sinon)		
ArrivéeLyon	La liaison arrive à Lyon-Part-Dieu (1 si oui ; 0 sinon)		
DépartOuArr_Lyon_Marseille	La liaison part ou arrive à Marseille ou à Lyon-Part-Dieu (1 si oui; 0 sinon)		
NbNoeudsMajeursTraversés ¹	Nombre de nœuds majeurs (Paris, Lyon, Marseille) traversés sur la liaison		Si la liaison passe par les gares de Massy-TGV, Marne-la-Vallée ou Aéroport Charles-de-Gaulles TGV, on considère qu'elle passe par le nœud ferroviaire parisien. Si la liaison passe par les gares de Lyon-Part-Dieu, Lyon-Perrache, ou Lyon-St-Exupéry, on considère qu'elle passe par le nœud ferroviaire lyonnais.
LiaisonPasseParAvignon	La liaison passe par le triangle ferroviaire des Angles (Gard) sur la LGV Méditerranée (simplifié ici comme le nœud ferroviaire d'Avignon) (1 si oui ; 0 sinon)		
LiaisonPasseParBordeaux	La liaison passe par le nœud ferroviaire de Bordeaux (1 si oui ; 0 sinon)		
LiaisonPasseParDijon	La liaison passe par le nœud ferroviaire de Dijon (1 si oui ; 0 sinon)		
LiaisonPasseParLille	La liaison passe par le nœud ferroviaire de Lille (1 si oui ; 0 sinon)		Si la liaison passe par les gares de Lille-Flandres ou Lille-Europe, on considère qu'elle passe par le nœud ferroviaire lillois.

¹ Afin d'intégrer aux variables explicatives le passage par un ou des nœuds ferroviaires répartis en 2 catégories (« primaires » ou « secondaires »), une liste non exhaustive et non définitive de nœuds a été établie en repérant les gares où se croisent plusieurs lignes du réseau ferré d'après l'Atlas SNCF Réseau 2020 : page 50-51 « LE TRAFIC TOTAL - VOYAGEURS ET FRET ». La répartition des nœuds ferroviaires selon les 2 catégories est une proposition de l'auteur basée sur une brève revue de presse et de littérature spécialisée.

LiaisonPasseParRennes	La liaison passe par le nœud ferroviaire de Rennes (1 si oui ; 0 sinon)		
LiaisonPasseParLeMans	La liaison passe par le nœud ferroviaire du Mans (1 si oui ; 0 sinon)		
LiaisonPasseParTours	La liaison passe par le nœud ferroviaire de Tours (1 si oui ; 0 sinon)		
DépartOuArr_NoedSecondaire ¹	La liaison part ou arrive au niveau d'un nœud ferroviaire secondaire (Avignon, Bordeaux, Lille, Le Mans, Tours) (1 si oui ; 0 sinon)		
NbNoeudsSecondaires ¹	Nombre de nœuds secondaires (Avignon, Bordeaux, Lille, Le Mans, Tours) sur la liaison		
NbBifurcationsLGV	Nombre de bifurcations majeures LGV sur la liaison (y compris en Île-de-France)	Atlas SNCF Réseau 2019 : page 42-43 LE TRAFIC DES TRAINS APTES A LA GRANDE VITESSE	Une bifurcation correspond à une séparation des voies LGV rencontrées sur la liaison, qui amène le TGV à emprunter l'une des deux voies. Une séparation des voies offrant plus de 2 chemins possibles est comptée comme une seule bifurcation. Les bifurcations des LGV en Île-de-France sont prises en compte.
NbBifurcationsLGV_horsIDF	Nombre de bifurcations majeures LGV sur la liaison (hors bifurcations en Île-de-France)		Une bifurcation correspond à une séparation des voies LGV rencontrées sur la liaison, qui amène le TGV à emprunter l'une des deux voies. Une séparation des voies offrant plus de 2 chemins possibles est comptée comme une seule bifurcation. Les bifurcations des LGV en Île-de-France ne sont pas prises en compte.
NbBifurcationsRéseauClassique	Nombre de bifurcations majeures du réseau classique sur la liaison	Atlas SNCF Réseau 2019 : page 38-39 LE RÉSEAU FERRÉ	Une bifurcation correspond à une séparation des voies rencontrées sur la liaison, qui amène le TGV à emprunter l'une des deux voies. La première bifurcation prise en compte correspond à la sortie de la LGV vers le réseau classique. Ensuite, toutes les bifurcations qui interviennent sur le réseau classique sont prises en compte. Une séparation des voies offrant plus de 2 chemins possibles est comptée comme une seule bifurcation.

TMJA_max_TGV_Liaison_2018	Trafic moyen journalier annuel (TMJA) des TGV de la section la plus intensément utilisée sur la liaison en 2018	Atlas SNCF Réseau 2019 : page 42-43 LE TRAFIC DES TRAINS APTES A LA GRANDE VITESSE	
NbMoyenArrêts_2020	Nombre moyen d'arrêts effectués par les TGV sur la liaison au cours d'un Jour Ouvrable de Base (JOB) de 2020 en dehors des périodes de vacances	https://www.horaires-de-trains.fr/ (17 novembre 2020) d'après l'OpenData SNCF	Calcul du nombre d'arrêts moyen des trains circulant sur la liaison au cours d'un Jour Ouvrable de Base (JOB) - le 17 novembre 2020 - pour l'amplitude horaire retenue de 6h à 22h. Le départ et l'arrivée ne sont pas pris en compte.
LongueurRéseauClassique	Estimation de la longueur parcourue sur le réseau classique pour une circulation de la liaison par année (en km)	Atlas SNCF Réseau 2019 : page 42-43 LE TRAFIC DES TRAINS APTES A LA GRANDE VITESSE	
TMJA_max_Total_Liaison_2000	Trafic journalier moyen annuel (TJMA) pour l'ensemble des circulations de trains (y compris fret et "haut le pied") sur la section la plus intensément utilisée de la liaison en 2000	Carte SNCF - TRAFIC TOTAL 2000 SNCF / Direction du Contrôle de gestion (2000)	Trafic journalier moyen annuel (TJMA) pour l'ensemble des circulations de trains (y compris fret et "haut le pied") sur la section la plus intensément utilisée de la liaison en l'année 2000. Nous avons utilisé les données de cette carte relativement ancienne car elles fournissent un meilleur niveau de précision que les cartes publiques récentes fournies par la SNCF. En particulier, elle précise le TJMA des sections de lignes au niveau des nœuds ferroviaires du réseau ferré.

2.2.2.2 Intercités

Le tableau de la page suivante détaille les variables explicatives à caractère géographique dont la significativité statistique de l'effet sur le taux de retard a été testée.

Les données associées à certaines variables ont nécessité de recourir à des sources de données ou des méthodes de calcul particulières. Elles sont listées ci-dessous et présentées plus en détail en annexes.

Variable	Description de la variable	Annexe
DistanceMoy	Distance moyenne parcourue par un train entre les 2 gares sur la liaison	Annexe 10
NbMoyenArrêts_2020	Nombre moyen d'arrêts effectués par les Intercités sur la liaison au cours d'un Jour Ouvrable de Base (JOB) de 2020 en dehors des périodes de vacances	Annexe 11

Nom_Variable	Description de la variable	Source des données ou calcul effectué	Remarques
Liaison	Liaison		
Année	Année		
Service	Service		
Axe	Axe		
TpsTrajet	Durée moyenne du trajet sur la liaison		
Trafic_NbProg	Nombre de trains programmés		
TxRetard	Taux de retard		
NbCirc_Liaison_an	Nombre de trains ayant circulé sur la liaison		
DistanceMoy	Distance moyenne entre les 2 gares	Mémento statistique de la SNCF - 2018 ou Rome2Rio	
VitesseMoyenne	Vitesse moyenne des circulations de trains sur la liaison	Distance moyenne parcourue par un train entre les 2 gares sur la liaison / Durée moyenne du trajet sur la liaison	
LiaisonNuit	Liaison de nuit (1 si oui ; 0 sinon)		
DépartParis	La liaison part de Paris intra-muros (1 si oui ; 0 sinon)		
ArrivéeParis	La liaison arrive à Paris intra-muros (1 si oui ; 0 sinon)		
Liaison_radiale	La liaison part ou arrive à Paris intra-muros (1 si oui ; 0 sinon)		
DépartMarseille	La liaison part de Marseille (1 si oui ; 0 sinon)		
ArrivéeMarseille	La liaison arrive à Marseille (1 si oui ; 0 sinon)		

DépartLyon	La liaison part de Lyon (1 si oui ; 0 sinon)		Les gares de Lyon-Part-Dieu et Lyon-Perrache sont prises en compte.
ArrivéeLyon	La liaison arrive à Lyon (1 si oui ; 0 sinon)		Les gares de Lyon-Part-Dieu et Lyon-Perrache sont prises en compte.
NbNoeudsMajeursTraversés ²	Nombre de nœuds majeurs (Paris, Lyon, Marseille) sur la liaison		Si la liaison passe par Massy-TGV, Marne-la-Vallée ou Aéroport Charles-de-Gaulles TGV, on considère qu'elle passe par le nœud ferroviaire parisien. Si la liaison passe par Lyon-Part-Dieu, Lyon-Perrache, ou Lyon-St-Exupéry, on considère qu'elle passe par le nœud ferroviaire lyonnais.
LiaisonTraverseBordeaux	La liaison passe par le nœud de Bordeaux (1 si oui ; 0 sinon)		
LiaisonTraverseNantes	La liaison passe par le nœud de Nantes (1 si oui ; 0 sinon)		
LiaisonTraverseToulouse	La liaison passe par le nœud de Toulouse (1 si oui ; 0 sinon)		
LiaisonTraverseTours	La liaison passe par le nœud de Tours (1 si oui ; 0 sinon)		
LiaisonTraverseOrléans	La liaison passe par le nœud d'Orléans (1 si oui ; 0 sinon)		

² Afin d'intégrer aux variables explicatives le passage par un ou des nœuds ferroviaires répartis en 2 catégories (« primaires » ou « secondaires »), une liste non exhaustive et non définitive de nœuds a été établie en repérant les gares où se croisent plusieurs lignes du réseau d'après la carte du réseau Intercités de 2016 présentée en annexe 12. La répartition des nœuds ferroviaires selon les 2 catégories est une proposition de l'auteur basée sur une brève revue de presse et de littérature spécialisée.

LiaisonTraverseBrive	La liaison passe par le nœud de Brive (1 si oui ; 0 sinon)		
LiaisonTraverseNarbonne	La liaison passe par le nœud de Narbonne (1 si oui ; 0 sinon)		
LiaisonTraverseNevers	La liaison passe par le nœud de Nevers (1 si oui ; 0 sinon)		
LiaisonTraverseDijon	La liaison passe par le nœud de Dijon (1 si oui ; 0 sinon)		
LiaisonTraverseRouen	La liaison passe par le nœud de Rouen (1 si oui ; 0 sinon)		
LiaisonTraverseAmiens	La liaison passe par le nœud d'Amiens (1 si oui ; 0 sinon)		
LiaisonTraverseCaen	La liaison passe par le nœud de Caen (1 si oui ; 0 sinon)		
LiaisonTraverseSaint-Quentin	La liaison passe par le nœud de Saint-Quentin (1 si oui ; 0 sinon)		

DépartOuArr_NoedSecondaire ²	La liaison part ou arrive au niveau d'un nœud secondaire (Bordeaux, Nantes, Toulouse, Orléans, Tours, Brive-la-Gaillarde, Narbonne, Nevers, Dijon, Caen, Amiens, Saint-Quentin) (1 si oui ; 0 sinon)		
NbNoeudsSecondaires ²	Nombre de nœuds secondaires sur la liaison (Bordeaux, Nantes, Toulouse, Orléans, Tours, Brive-la-Gaillarde, Narbonne, Nevers, Dijon, Caen, Amiens, Saint-Quentin)		
Axe_Nord_Normandie	La liaison appartient à l'axe Nord-Normandie de 2016 (1 si oui ; 0 sinon)		
Axe_Med_Atlantique	La liaison appartient à l'axe Méditerranée-Atlantique de 2016 (1 si oui ; 0 sinon)		
Axe_Paris_Centre_SudOuest	La liaison appartient à l'axe Paris-Centre-Sud-Ouest de 2016 (1 si oui ; 0 sinon)		
Axe_Est_Centre	La liaison appartient à l'axe Est-Centre de 2016 (1 si oui ; 0 sinon)		
Est_Ouest	La liaison décrit un trajet de l'Est vers l'Ouest (1 si oui ; 0 sinon)		La liaison est considérée comme « de l'Est vers l'Ouest » si l'angle entre la demi-droite de la gare d'origine vers la gare terminus et le parallèle traversant la gare d'origine est compris entre 135° et 225°
Ouest_Est	La liaison décrit un trajet de l'Ouest vers l'Est (1 si oui ; 0 sinon)		La liaison est considérée comme « de l'Ouest vers l'Est » si l'angle entre la demi-droite de la gare d'origine vers la gare

			terminus et le parallèle traversant de la gare d'origine est compris entre 45° et -45°
Nord_Sud	La liaison décrit un trajet du Nord vers le Sud (1 si oui ; 0 sinon)		La liaison est considérée comme « du Nord vers le Sud » si l'angle entre la demi-droite de la gare d'origine vers la gare terminus et le méridien traversant de la gare d'origine est compris entre 135° et 225°
Sud_Nord	La liaison décrit un trajet du Sud vers le Nord (1 si oui ; 0 sinon)		La liaison est considérée comme « du Sud vers le Nord » si l'angle entre la demi-droite de la gare d'origine vers la gare terminus et le méridien traversant de la gare d'origine est compris entre 45° et -45°
NbMoyenArrêts_2020	Nombre moyen d'arrêts effectués par les trains sur la liaison au cours d'un Jour Ouvrable de Base (JOB) de 2020 en dehors des périodes de vacances	https://www.horaires-de-trains.fr/ (17 novembre 2020) d'après l'OpenData SNCF	Calcul du nombre d'arrêts moyen des trains circulant sur la liaison au cours d'un Jour Ouvrable de Base (JOB) - le 17 novembre 2020 - pour l'amplitude horaire retenue de 6h à 22h. Le départ et l'arrivée ne sont pas pris en compte.
TMJA_max_Total_Liaison_2000	Trafic journalier moyen annuel (TJMA) pour l'ensemble des circulations de trains (y compris fret et "haut le pied") sur la section la plus intensément utilisée de la liaison en 2000 ou en 2018	Atlas SNCF Réseau 2020 : page 50-51 LE TRAFIC TOTAL - VOYAGEURS ET FRET Carte SNCF - TRAFIC TOTAL 2000 SNCF / Direction du Contrôle de gestion (2000)	Pour chaque liaison, est retenue la valeur maximale entre le trafic moyen journalier de la section la plus utilisée d'après la carte de l'atlas de SNCF Réseau - édition 2020, et celle fournie par la carte SNCF - TRAFIC TOTAL 2000 qui détaille le débit du trafic au niveau des nœuds ferroviaires. La valeur retenue comprend les circulations fret et les circulations de trains "haut le pied" (à vide). Pour les liaisons de nuit, la valeur retenue correspond au débit moyen journalier annuel de la liaison de nuit au trafic le plus élevé empruntant les lignes partagées avec la liaison considérée, d'après la base de l'AQST en 2012, 2016 ou 2019 (en fonction de la date d'arrêt du service de la liaison de nuit).
TMJA_max_Total_Liaison_2000_horsParis	Trafic journalier moyen annuel (TJMA) pour l'ensemble des circulations de trains (y compris fret et "haut le pied") sur la	Atlas SNCF Réseau 2020 : page 50-51 LE TRAFIC TOTAL - VOYAGEURS ET FRET	Identique à la variable précédente, mais le débit au niveau des gares dans Paris intra-muros n'est pas pris en compte.

	section la plus intensément utilisée de la liaison en 2000 ou en 2018	Carte SNCF - TRAFIC TOTAL 2000 SNCF / Direction du Contrôle de gestion (2000)	
--	---	---	--

2.2.2.3 TER

Le tableau de la page suivante détaille les variables explicatives à caractère géographique dont la significativité statistique de l'effet sur le taux de retard a été testée.

Les données associées à certaines variables ont nécessité de recourir à des sources de données ou des méthodes de calcul particulières. Elles sont listées ci-dessous et présentées plus en détail en annexes.

Variable	Description de la variable	Annexe
NbCirc_HPM et NbCirc_HPS	Nombre moyen de circulations quotidiennes à l'heure de pointe du matin (HPM) et à l'heure de pointe du soir (HPS) en 2016	Annexe 15
NbMoyArrêts_2020_PpalesLiaisons et TxCircPpalesLiaisonsTraversantNoeudRégional	Nombre moyen d'arrêts effectués en 2020 par les trains TER sur le panel des liaisons les plus fréquentées par ancienne région (panel de l'Autorité de Régulation des Transports (ART)). Estimation du taux de circulations de trains TER traversant le nœud ferroviaire régional retenu sur le panel des liaisons les plus fréquentées par ancienne région (selon l'ART).	Annexe 16
AgeMoyen_MaterielRoulant	Estimation AQST de l'âge moyen du parc de matériel roulant en fin d'année	Annexe 17

Nom_Variable	Description de la variable	Source des données ou calcul effectué	Remarques
Année	Année	Base AQST	
TxRetard	Taux de retard		
NbCirc_an_AQST	Nombre de circulations TER dans l'année selon la base AQST		en milliers de circulations
NbCirc_HPM	Nombre moyen de circulations quotidiennes à l'heure de pointe du matin en 2016	Base ART ferroviaire (2015-2018)	l'heure de pointe du matin prise en compte est [7-8h[
NbCirc_HPS	Nombre moyen de circulations quotidiennes à l'heure de pointe du soir en 2016		l'heure de pointe du soir prise en compte dépend de la région : [17h-18h[ou [18-19h[
DistanceMoy	Distance moyenne parcourue par les trains TER		en trains-km
VoyKm_an	Nombre de passagers-km transportés (en millions) dans l'année		en millions
NbPlacesMoy	Capacité d'emport moyenne des TER en nombre de sièges		en nombre de sièges
TauxOccupationMoy	Taux d'occupation moyen des TER		Voyageurs-km transportés/Sièges-km offerts
SiegesKmParHab	Nombre de sièges-km offerts par habitant		
VoyageurKmParHab	Nombre de voyageurs-km transportés par habitant		
IntensitéUsageTER	Intensité quotidienne d'utilisation du réseau par les services régionaux conventionnés		Train-km TER ou Transilien par km de ligne par jour
LongueurRéseau_HorsLGV	Longueur du réseau hors LGV		Km de lignes
DensitéRéseau_HorsLGV	Densité du réseau hors LGV (kilomètre de lignes pour 100 km ²)	Km de lignes/100km ²	
PartCatUIC_5_6	Part des voies ferroviaires de catégories UIC 5 à 6 dans le réseau hors LGV	% de voies	

PartCatUIC_7_9_AV	Part des voies ferroviaires de catégories UIC 7 à 9 AV dans le réseau hors LGV		% de voies
AgeRéseau_HorsLGV	Âge du réseau hors LGV	ART - Bilan du marché ferroviaire de voyageurs 2017 ; Bilan du marché ferroviaire de voyageurs 2018 volume 2	ans
AgeRéseauCatUIC_5_6	Âge des lignes ferroviaires des catégories UIC 5 à 6)	Base ART ferroviaire (2015-2018)	ans
AgeRéseauCatUIC_7_9_AV	Âge des lignes ferroviaires des catégories UIC 7 à 9 AV)		ans
DensitéRéseau_CatUIC_5_6	Densité du réseau des lignes de catégories UIC 5 à 6 (kilomètre de voies pour 100 km ²)		Km de voies/100km ²
DensitéRéseau_CatUIC_7_9_AV	Densité du réseau des lignes de catégories UIC 7 à 9 AV (kilomètre de voies pour 100 km ²)		Km de voies/100km ²
NbLiaisonsInfrarégionales	Nombre de liaisons infrarégionales desservies par les services TER	ART - Annexes du Bilan du transport ferroviaire de voyageurs 2015/2016 Bilan du marché ferroviaire de voyageurs 2017	Données de l'ART disponibles pour les années 2016 et 2017
NbLiaisonsInterrégionales	Nombre de liaisons interrégionales desservies par les services TER		
NbCommunesDesservies	Nombre de communes desservies au sein de la région par les services TER		
PartPopInf10kmGareTER	Part de la population de chaque région résidant dans une commune située à moins de 10 km d'une gare régionale (TER ou Transilien)		en % de la population

EffectifsPour100000trkm_2015	Ratio des effectifs TER pour 100 000 trains-km réalisés à fin 2015	ART - Bilan du transport ferroviaire de voyageurs 2015/2016	
NbMoyArrêts_NvlleReg_2018	Nombre moyen d'arrêts par ancienne région en 2018 (nouvelles régions administratives après la réforme territoriale de 2016)	ART - Bilan du marché du transport ferroviaire de voyageurs 2018 - vol.2	Données de l'ART fournies uniquement pour les nouvelles régions administratives après la réforme territoriale de 2016. La figure de l'ART dont sont issues les données ne précise pas si le départ et le terminus sont pris en compte dans le calcul du nombre moyen d'arrêts.
TMJA_max_TER_2018	Niveau du Trafic journalier moyen annuel (TJMA) des trains TER sur la section de ligne la plus intensément utilisée dans la région en 2018	Atlas SNCF Réseau 2020 : pages 44-45 LE TRAFIC VOYAGEURS RÉGIONAUX	en nombre de circulations TER/jour
NbMoyArrêts_2020_PpalesLiaisons	Nombre moyen d'arrêts effectués par les trains TER en 2020 sur le panel des liaisons les plus fréquentées par ancienne région (selon l'ART) en 2018	https://www.horaires-de-trains.fr/ (mardi 27 octobre 2020) d'après l'OpenData SNCF	Calcul du nombre moyen d'arrêts effectués par les trains TER - sur les liaisons les plus fréquentées - au cours d'un Jour Ouvrable de Base (JOB) - le mardi 27 octobre 2020 - pour l'amplitude horaire retenue de 6h à 20h. Le panel des liaisons retenues est détaillé en annexe. Le départ et le terminus ne sont pas pris en compte dans le nombre d'arrêts.
TxCircPpalesLiaisonsTraversant-NoeudRégional	Estimation du taux de circulations de trains TER traversant le nœud ferroviaire régional retenu pour le panel des liaisons les plus fréquentées par ancienne région (selon l'ART) en 2018	Le panel des principales liaisons en volume de trafic TER est issu du Bilan du marché du transport ferroviaire de voyageurs 2018 - volume 1 de l'ART	La liste des nœuds ferroviaires régionaux retenus est fournie en annexe.

AgeMoyen_MaterielRoulant	Estimation AQST de l'âge moyen du parc de matériel roulant en fin d'année	<p>Données partielles sur l'âge moyen du matériel roulant en fin d'année :</p> <ul style="list-style-type: none"> - Chambres Régionales des Comptes, rapports d'observations définitives, Enquêtes Transports Express Régionaux, Exercices 2012 et suivants <p>Données sur la composition du parc de matériel roulant :</p> <ul style="list-style-type: none"> - Chambres Régionales des Comptes, rapports d'observations définitives, Enquêtes Transports Express Régionaux, Exercices 2012 et suivants - Fiches Wikipédia sur les TER par ancienne régions, et sur le parc de matériel roulant TER par ancienne régions. 	Des modèles de parc ont été construits sur la base des données détaillées de composition du parc du matériel roulant TER par ancienne régions, et sur les dates de mises en services et de radiations fournies par les différentes sources de données citées dans la colonne adjacente.
--------------------------	---	---	---

2.2.2.4 RER et Transilien

Le tableau de la page suivante détaille les variables explicatives à caractère géographique dont la significativité statistique de l'effet sur le taux de retard par ligne de RER et Transilien a été testée.

Les données associées à certaines variables ont nécessité de recourir à des sources de données ou des méthodes de calcul particulières. Elles sont listées ci-dessous et présentées plus en détail en annexes.

Variable	Description de la variable	Annexe
NbVoyageursJour	Estimation de la fréquentation quotidienne des lignes	Annexe 19
NbMoyArrêtsHPM_2019 NbMoyArrêtsHPMdansParis_2019 TpsParcoursMoyHPM_2019 DuréeMoyInterstation_2019	Nombre d'arrêts moyen à l'heure de pointe du matin durant un Jour Ouvrable de Base (JOB) Nombre d'arrêts moyen dans Paris intra-muros à l'heure de pointe du matin durant un JOB Durée moyenne du parcours total des trains à l'heure de pointe du matin durant un JOB Durée moyenne entre 2 arrêts à l'heure de pointe du matin durant un JOB	Annexe 20
AgeMoyen_MaterielRoulant_1erJanvier	Estimation AQST de l'âge moyen du parc de matériel roulant en début d'année	Annexe 21
TMJA_Tot_ParVoie_Moy et TMJA_Tot_ParVoie_Max	Estimation de la moyenne du Trafic Moyen Journalier Annuel (TMJA) par voie pour l'ensemble des circulations de trains (y compris fret et "haut-le-pied") Estimation du Trafic Moyen Journalier Annuel (TMJA) par voie sur la section la plus chargée de la ligne pour l'ensemble des circulations de trains (y compris fret et "haut-le-pied")	Annexe 22

Nom_Variable	Description de la variable	Source des données ou calcul effectué	Remarques
Année	Année	Base AQST	
TxRetard	Taux de retard		
LongueurLigne	Longueur totale de la ligne	OpenData SNCF "présentation du réseau Transilien"	[en km] Toutes branches comprises
NbArrêts	Nombre total de gares/stations sur la ligne		
DistanceMoyEntreArrêts	Distance moyenne entre 2 gares de la ligne	Longueur totale de ligne/Nombre d'arrêts	[en km]
NbCircJour_2019	Nombre de trains par jour en 2019 de la ligne	OpenData SNCF "présentation du réseau Transilien", comités de ligne RER et Transilien, sites officiels des lignes	D'après les services SNCF Transilien : "Année de référence des dates de comptage des voyageurs RER A 2014 ; RER B 2014 ; RER C 2017 ; RER D 2015 ; RER E 2016 ; H 2017 ; J 2016 ; K 2017 ; L 2014 ; L Grande Ceinture Ouest 2016 ; N 2013 ; P 2015 ; R 2017 ; U 2016" Il s'agit des trains de la ligne, mais pas de ceux qui circulent sur la voie support en cas de partage de voies entre plusieurs lignes.
NbVoyageursJour	Estimation de la fréquentation quotidienne des lignes	OpenData SNCF, comités de ligne RER et Transilien, sites officiels des lignes	[en milliers de voyageurs par jour]
NbBranches	Nombre de branches de la ligne	Plans des lignes	Si un "tronçon central" existe, il n'est pas considéré comme une branche supplémentaire. Une "boucle" compte comme 2 branches. Il s'agit des branches de la ligne et pas des branches des voies support.
TrKm	Offre en trains-km réalisés	OMNIL, 2020	[en millions de trains-km réalisés]
VoyKm	Nombre de voy-km transportés	$(NbVoyageursJour/NbCircJour) * TrKm$ $*(DistanceMoyVoy/DistanceMoyTrain)$	[en milliards de voy-km transportés]
DistanceMoyTrain_2019	Distance moyenne parcourue par les trains de la ligne en 2019	$TrKm_{2019}/(330*NbCircJour_{2019})$	[en km]

IntervalleMin_2019	Intervalle minimal à l'hyperpointe du matin en 2019	Comités de ligne RER et Transilien, sites officiels des lignes, fiches Wikipédia des lignes, documentation technique des transporteurs	[en secondes] En cas de section de ligne partagée (uniquement avec d'autres lignes de RER et Transilien), l'intervalle minimal retenu prend en compte les circulations de trains des autres lignes sur ces sections. L'hyperpointe désigne la période durant laquelle la demande de transport est la plus forte. Elle peut être infra-horaire, par exemple 7h45-8h15, à la différence de l'heure de pointe, qui représente une heure complète.
NbTrainsHPM_2019	Nombre de trains par sens à l'Heure de Pointe du Matin (HPM) à l'interstation la plus chargée		En cas de section de ligne partagée (uniquement avec d'autres lignes de RER et Transilien), les trains des autres lignes sont intégrés au nombre de trains circulant à l'HPM sur ces sections.
VoiesPartagées	Présence de sections de ligne partagées		[0 = non ; 1 = oui] Ne sont prises en compte que les voies partagées avec d'autres lignes de RER et Transilien. Autrement dit, si le partage des voies ne concerne que des lignes de TER ou Grandes Lignes SNCF, la variable prend la valeur 0.
AgeMoyen_MaterielRoulant-1erJanvier	Estimation de l'âge moyen du matériel roulant en début d'année	- Fiches Wikipédia sur les lignes de Transilien et sur les différentes séries du matériel roulant composant leurs parcs. - articles de la presse spécialisée sur les mises en service et radiations du matériel roulant	Des modèles de parc ont été construits sur la base des données détaillées de composition du parc du matériel roulant Transilien par ligne, et sur les dates de mises en services et de radiations fournies par les différentes sources de données citées.
NbMoyArrêtsHPM_2019	Nombre d'arrêts moyen à l'heure de pointe du matin durant un JOB	Fiches horaires des lignes	Le départ et le terminus ne sont pas intégrés au nombre d'arrêts
NbMoyArrêtsHPMdansParis-2019	Nombre d'arrêts moyen <u>dans Paris intra-muros</u> à l'heure de pointe du matin durant un JOB		Le départ et le terminus ne sont pas intégrés au nombre d'arrêts
TpsParcoursMoyHPM_2019	Durée moyenne du parcours total des trains à l'heure de pointe du matin durant un JOB		[en minutes] y compris les temps d'arrêts (échange voyageurs)
DuréeMoyInterstation_2019	Durée moyenne entre 2 arrêts à l'heure de pointe du matin durant un JOB		[en minutes] La durée moyenne d'interstation comprend la durée moyenne d'un arrêt (échange voyageurs)

TypeContrôleVitesse	Type de contrôle de vitesse utilisé depuis 2015	<p>- Rapport de l'Audit de la performance du système Transilien, PANOS TZIEROPOULOS, YVES PUTALLAZ, EPFL, 2015 [6]</p> <p>- Normes de tracé horaire sur les lignes nationales pour le SA 2021, Document opérationnel, SNCF Réseau, 2019</p>	<p>Type de contrôle de vitesse utilisé :</p> <p>- "2" : système "Favorable à la fluidification du trafic" sur l'ensemble de la ligne d'après l'Audit EPFL : contrôle de vitesse à réouverture (KVBP, KCVB) ou système de pilotage automatique (SACEM, etc.)</p> <p>- "1" : système "Favorable à la fluidification du trafic" sur une partie de la ligne</p> <p>- "0" : système "simple" au sens de l'Audit EPFL = contrôle de vitesse par balise (KVB)</p>
GareDuNord	La ligne contient la station "Gare du Nord".	Plans des lignes	
GareDeLyon	La ligne contient la station "Gare de Lyon".		
GareStLazare	La ligne contient la station "Gare Saint-Lazare".		La gare d'Hausmann – Saint-Lazare sur la ligne E n'est pas intégrée à cette variable, car les voies de cette gare ne sont pas reliées aux voies de surface de la gare Saint-Lazare avant le prolongement de la ligne E à l'Ouest (projet Eole dont la mise en service du premier tronçon est prévu pour 2022).
GareParisEst	La ligne contient la station "Gare de l'Est".		La gare de Magenta sur la ligne E est intégrée à cette variable car certaines fonctionnalités ferroviaires sont interdépendantes. En particulier « les lignes E et P partagent la même infrastructure au départ de la gare de l'Est » [6]
ChâteletLesHalles	La ligne contient la station "Châtelet - Les Halles".		
TMJA_Tot_ParVoie_Moy	Estimation de la moyenne du Trafic Moyen Journalier Annuel (TMJA) par voie pour l'ensemble des circulations de trains (y compris fret et "haut le pied")	<p>- Trafic Transilien : Atlas SNCF Réseau 2020 Carte "Trafic Transilien" (page 46-47)</p> <p>- Trafic Grandes Lignes + TER + fret : Carte SNCF - TRAFIC TOTAL 2000 SNCF / Direction du Contrôle de gestion (2000)</p>	<p>Moyenne pondérée du nombre de km de ligne par tronçon ayant un débit journalier constant.</p> <p>Pour le trafic Grandes Lignes + TER + fret, des données assez anciennes (2000) ont été utilisées car les cartes plus récentes à disposition n'offrent pas un aussi bon niveau de précision : le découpage du réseau par tronçons n'est pas aussi fin.</p>

TMJA_Tot_ParVoie_Max	Estimation du Trafic Moyen Journalier Annuel (TMJA) par voie sur la section la plus chargée de la ligne pour l'ensemble des circulations de trains (y compris fret et "haut le pied")	<p>- Trafic Transilien : Atlas SNCF Réseau 2020 Carte "Trafic Transilien" (page 46-47)</p> <p>- Trafic Grandes Lignes + TER + fret : Carte SNCF - TRAFIC TOTAL 2000 SNCF / Direction du Contrôle de gestion (2000)</p>	<p>Niveau du TMJA toutes circulations ferroviaires prises en compte sur le tronçon le plus chargé de la ligne.</p> <p>Pour le trafic Grandes Lignes + TER + fret, des données assez anciennes (2000) ont été utilisées car les cartes plus récentes à disposition n'offrent pas un aussi bon niveau de précision : le découpage du réseau par tronçons n'est pas aussi fin.</p>
----------------------	---	--	---

La variable portant sur le passage de la ligne par la gare Montparnasse n'est pas étudiée car seule la ligne N la dessert. De fait, si nous obtenions un résultat significatif quant à l'impact de cette variable sur la ponctualité, nous ne pourrions pas savoir s'il est dû à des problèmes rencontrés au niveau de la gare Montparnasse ou à d'autres caractéristiques de la ligne N qui dégradent la ponctualité.

Le tableau de la page suivante détaille les variables explicatives à caractère géographique dont la significativité statistique de l'effet sur le taux de retard **par branche** de RER et Transilien a été testée.

Nom_Variable	Description de la variable	Source des données ou calcul effectué	Remarques
Année	Année	Base AQST	
TxRetard	Taux de retard		
Branche	Branche	OMNIL	
Année	Année		
TxPonctu	Taux de ponctualité voyageurs pour les voyageurs de la branche		
TxRetard	Taux de retard voyageurs pour les voyageurs de la branche		
TpsParcoursMoy	Temps de parcours total des trains de la branche	Fiches horaires des lignes de RER et Transilien	Temps de parcours total moyen des trains des missions qui desservent la branche à l'HPM [en min]
NbArrêts	Nombre d'arrêts effectués par les trains de la branche		Les arrêts de départ et de terminus ne sont pas pris en compte.
VoiesPartagées_Transilien	Présence de sections partagées entre plusieurs lignes de RER ou Transilien sur la branche		[0 = non ; 1 = oui] Ne sont prises en compte que les voies partagées avec d'autres lignes de RER et Transilien. Autrement dit, si le partage des

			voies ne concerne que des lignes de TER ou Grandes Lignes SNCF, la variable prend la valeur 0.
VoiesPartagées-AutresActivitésFer	Présence de sections partagées avec d'autres activités ferroviaires (voyageurs TER, Intercités, TGV, ou fret) sur la branche		[0 = non ; 1 = oui] Ne sont prises en compte que les voies partagées avec d'autres activités ferroviaires de voyageurs ou fret. Autrement dit, si le partage des voies ne concerne que des lignes de RER et Transilien, la variable prend la valeur 0.
AgeMoyen_MaterielRoulant-1erJanvier	Âge moyen du matériel roulant de la ligne en début d'année	- Fiches Wikipedia sur les lignes de Transilien et sur les différentes séries du matériel roulant composant leurs parcs. - articles de la presse spécialisée sur les mises en service et radiations du matériel roulant	Des modèles de parc ont été construits sur la base des données détaillées de composition du parc du matériel roulant Transilien par ligne, et sur les dates de mises en services et de radiations fournies par les différentes sources de données citées.
GareDuNord	Les trains de la branche passent par "Gare du Nord".		
GareDeLyon	Les trains de la branche passent par la station "Gare de Lyon".		
GareStLazare	Les trains de la branche passent par la station "Gare Saint-Lazare".		
ChâteletLesHalles	Les trains de la branche passent par la station "Châtelet - Les Halles".		
GareParisEst	Les trains de la branche passent par la station "Gare de l'Est".		La station "Magenta" sur la ligne E est intégrée car certaines infrastructures ferroviaires entre les 2 gares sont interdépendantes
TMJA_ParVoie_Max_ParcoursCompleat	Estimation de la moyenne du Trafic Moyen Journalier Annuel (TMJA) par voie pour l'ensemble des circulations de trains (y compris fret et "haut le pied") sur la section la plus chargée du parcours total des trains de la branche	Trafic Transilien : Atlas SNCF Réseau 2020 Carte "Trafic Transilien" (page 46-47) Trafic Grandes Lignes + TER + fret : Carte SNCF - TRAFIC TOTAL 2000	Niveau du TMJA (toutes circulations ferroviaires prises en compte) sur le tronçon le plus chargé emprunté par les trains de la branche

TMJA_ParVoie_Max_Branche	Estimation de la moyenne du Trafic Moyen Journalier Annuel (TMJA) par voie pour l'ensemble des circulations de trains (y compris fret et "haut le pied") sur la section la plus chargée de la branche	SNCF / Direction du Contrôle de gestion (2000)	Niveau du TMJA (toutes circulations ferroviaires prises en compte) sur le tronçon le plus chargé de la branche
TMJA_ParVoie_Moy_Branche	Estimation de la moyenne du Trafic Moyen Journalier Annuel (TMJA) par voie pour l'ensemble des circulations de trains (y compris fret et "haut le pied") sur l'ensemble de la branche		Niveau du TMJA (toutes circulations ferroviaires prises en compte) moyen pondéré de la longueur des tronçons de la branche
ContrôleVitesseAvancé	Utilisation d'un contrôle de vitesse plus avancé que le KVP (contrôle de vitesse par balises) sur une partie de la branche ou sur la totalité de la branche	<ul style="list-style-type: none"> - Rapport de l'Audit de la performance du système Transilien, PANOS TZIEROPOULOS, YVES PUTALLAZ, EPFL, 2015 - Normes de tracé horaire sur les lignes nationales pour le SA 2021, Document opérationnel, SNCF Réseau, 2019 	Type de contrôle de vitesse utilisé : <ul style="list-style-type: none"> - "1" : contrôle de vitesse à réouverture (KVBP, KCVB, KCVP) ou système de pilotage automatique (SACEM, etc.) - "0" : contrôle de vitesse par balise (KVB)
NbVoiesGarageTerminus	Nombre de voies de garage de rames en bout de branche	<ul style="list-style-type: none"> - Carte du Réseau Express Régional d'Île-de-France, © Franklin JARRIER, 2020 en ligne : http://cartometro.com/cartes/rer-idf/ <ul style="list-style-type: none"> - Fiches Wikipedia des lignes de RER et Transilien 	Les voies principales sur lesquels partent ou arrivent les trains de la branche sont comptées comme des voies de garage lorsque le terminus de la branche représente un terminus physique de ces voies. Si la branche est un tronçon de ligne mais pas un terminus de ligne, la variable prend la valeur NA.
NbVoiesGarageTerminus_100trains	Nombre de voies de garage de rames en bout de branche pour 100 circulations de trains quotidiennes par voie de la branche	Calcul effectué : $(NbVoiesGarageTerminus * 100) / TMJA_ParVoie_Moy_Branche$	

2.3 Forme des modèles statistiques testés

Des modèles économétriques de régressions linéaires simples et multiples ont été implémentés à l'aide du logiciel R pour décrire et interpréter la corrélation éventuelle entre une variable explicative à caractère géographique retenue et le taux de retard des trains par service ferroviaire.

Deux formes de modèles ont été utilisées :

1. Des régressions linéaires simples dont la variable expliquée est le taux de retard et la variable explicative est l'une des différentes variables à caractère géographique retenues. Ces modèles prennent la forme suivante :

$$y_{l,t} = a + b \times x_{l,t} + \varepsilon_{l,t}$$

Avec

$y_{l,t}$ la valeur du taux de retard pour la liaison l lors de l'année

a la constante de la régression linéaire

b le coefficient associé à la variable explicative

$x_{l,t}$ la valeur de la variable explicative pour la liaison l et l'année t

$\varepsilon_{l,t}$ le résidu (ou terme d'erreur) de l'estimation du taux de retard pour la liaison l et l'année t

2. Des régressions linéaires multiples dont la variable expliquée est le taux de retard et les variables explicatives sont les variables explicatives à caractère géographique retenues. Ces modèles prennent la forme suivante :

$$y_{l,t} = a_0 + \sum_{i=1}^n a_i \times x_{i,l,t} + \varepsilon_{l,t}$$

Avec

$y_{l,t}$ la valeur du taux de retard pour la liaison l lors de l'année

a_0 la constante de la régression linéaire

$x_{i,l,t}$ la valeur de la i-ème variable explicative pour la liaison l et l'année t

a_i la constante (indépendante du temps et de la liaison) associée à la i-ème variable explicative

$\varepsilon_{l,t}$ le résidu (ou terme d'erreur) de l'estimation du taux de retard pour la liaison l et l'année t

3. RESULTATS

3.1 Note de lecture des résultats des tests statistiques

Pour chaque modèle de régression retenu, des résultats portant sur la significativité du modèle et de l'impact des déterminants supposés de la ponctualité sur cette ponctualité sont fournis dans la section suivante.

Un test de Student est utilisé pour décrire la significativité de l'influence d'un déterminant « supposé de la ponctualité » sur cette ponctualité. Il porte sur la nullité du coefficient a_i associé au i -ème déterminant « supposé de la ponctualité » $X_{i,t}$, selon l'hypothèse nulle H_0 et l'hypothèse alternative H_1 suivantes :

$$H_0: a_i = 0$$

$$H_1: a_i \neq 0$$

Les résultats du test de Student permettent de savoir si l'on peut rejeter ou non l'hypothèse nulle H_0 . Si on choisit de rejeter l'hypothèse nulle H_0 et d'accepter l'hypothèse alternative H_1 , on considère que l'effet du i -ème déterminant « supposé de la ponctualité » sur la ponctualité est significatif. Si on choisit d'accepter l'hypothèse nulle H_0 , on considère qu'on ne peut pas conclure qu'il y a un effet du i -ème déterminant « supposé de la ponctualité » sur cette ponctualité d'après le modèle retenu.

La p-value associée au test de Student appliqué à un coefficient représente la probabilité de rejeter l'hypothèse nulle H_0 de nullité de ce coefficient, alors qu'elle vraie. Ici, plus la p-value est petite, plus la probabilité de faire une erreur en concluant qu'il y a un effet significatif du i -ème déterminant « supposé de la ponctualité » sur la ponctualité est faible.

En pratique, pour savoir si l'on doit ou non rejeter l'hypothèse nulle, on peut comparer la p-value à un seuil de significativité. Si la p-value est inférieure au seuil, on conclut qu'il y a un effet du i -ème déterminant « supposé de la ponctualité » sur la ponctualité. Dans le cas contraire, on ne peut pas conclure avec certitude à l'existence d'un effet (même très faible) du i -ème déterminant « supposé de la ponctualité » sur la ponctualité. En pratique, un seuil de significativité de 0,05 est souvent utilisé.

Dans le logiciel R, les résultats de la modélisation portant sur les coefficients des différentes variables explicatives du modèle sont présentés sous forme de tableau avec :

- En colonne "Estimate" la valeur estimée du coefficient ;
- En colonne "Std. Error" l'écart-type de cette estimation ;
- En colonne "t value" la statistique du test de Student ;
- En colonne "Pr(>|t|)" la p-value associée au test de Student.

Dans le tableau de résultats de la régression modélisée, sont précisés à droite des p-value des codes indiquant le niveau de significativité atteint :

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Pour chaque modèle, la significativité du modèle dans son ensemble est caractérisée à l'aide d'un test de Fisher et du calcul du coefficient de détermination R^2 ajusté, dont les résultats sont également fournis dans la section suivante :

- Le coefficient de détermination ajusté « R^2 ajusté » quantifie la part de la variance expliquée par le modèle. Le coefficient de détermination non ajusté « R^2 » augmente généralement à mesure que l'on ajoute des variables dans le modèle – donc des degrés de liberté. Pour s'affranchir de cet effet, on utilise le coefficient de détermination ajusté, qui est corrigé du nombre de degrés de liberté ;
- Le test de Fisher caractérise la significativité globale du modèle et porte sur la nullité de l'ensemble des coefficients des variables explicatives du modèle, selon l'hypothèse nulle H_0 et l'hypothèse alternative H_1 suivantes :

H_0 : tous les coefficients sont nuls

H_1 : un des coefficients au moins est non nul

Les résultats du test de Fisher permettent de savoir si l'on peut rejeter ou non l'hypothèse nulle H_0 . Si on choisit de la rejeter, on considère que le modèle permet d'expliquer une part significative de l'évolution de la ponctualité des trains en France. Si on choisit d'accepter l'hypothèse nulle H_0 et de rejeter l'hypothèse alternative H_1 , on considère qu'on ne peut pas conclure avec certitude que le modèle retenu permet d'expliquer une part significative de l'évolution de la ponctualité des trains en France.

Pour savoir si l'on doit ou non rejeter l'hypothèse nulle, on peut alors comparer la p-value à un seuil de significativité. Si la p-value est inférieure au seuil, on rejette l'hypothèse nulle et on accepte l'hypothèse alternative.

Dans le logiciel R, les résultats portant sur la significativité du modèle dans son ensemble sont fournis en-dessous du tableau des coefficients :

- "Residual standard error" est la somme des carrés résiduels correspondant à la variabilité non-expliquée par le modèle ;
- "Multiple R-squared" est le coefficient de détermination R^2 qui quantifie la part de la variance expliquée par le modèle. Cet indicateur augmente généralement à mesure que l'on ajoute des variables dans le modèle – donc des degrés de liberté. Pour s'affranchir de cet effet, on utilise le coefficient de détermination ajusté, présenté ci-après ;
- "Adjusted R-squared" est le coefficient de détermination ajusté ou R^2 ajusté, il s'agit du coefficient de détermination R^2 corrigé du nombre de degrés de liberté ;

- "F-statistic" est la statistique du test de Fisher ;
- "p-value" est la p-value associée au test de Fisher.

3.2 Résultats de la modélisation statistique

3.2.1 Résultats pour les TGV

3.2.1.1 Résultats détaillés des régressions par variables explicatives

Les résultats détaillés des modèles de régressions testés sont présentés en Annexe 7.

3.2.1.2 Synthèse des résultats

Le tableau ci-dessous énumère les variables explicatives pour lesquelles le modèle de régression linéaire utilisé n'est pas statistiquement fiable au sens où il ne passe pas le test de Fisher au seuil de significativité de 5%. On ne peut donc pas conclure sur l'effet de ces variables sur le taux de retard avec les données à notre disposition.

Variable explicative	Description de la variable	R ² ajusté régression	p-value test de Fisher de la régression	p-value test de Student variables explicatives
NbCirc_Liaison_an	Nombre de trains ayant circulé sur la liaison	0,3%	5% < p < 10%	5% < p < 10%
AgeLGV_axe	Age moyen des voies LGV du premier axe TGV emprunté	< 0,1%	> 10%	> 10%
ArrivéeParis	La liaison arrive à Paris intra-muros	< 0,1%	> 10%	> 10%
LiaisonPasseParBordeaux	La liaison passe par le nœud ferroviaire de Bordeaux	< 0,1%	> 10%	> 10%
LiaisonPasseParLeMans	La liaison passe par le nœud ferroviaire du Mans	< 0,1%	> 10%	> 10%
LiaisonPasseParTours	La liaison passe par le nœud ferroviaire de Tours	< 0,1%	> 10%	> 10%

Le tableau ci-dessous énumère les variables explicatives pour lesquelles le modèle de régression linéaire est significatif – il passe le test de Fisher au seuil de significativité de 5% - mais pour lequel le signe de l'impact de la variable explicative sur le taux de retard est contraire à l'intuition. On ne peut donc pas trouver d'explication évidente à la corrélation observée avec les données à notre disposition.

Variable explicative	Description de la variable	R ² ajusté régression	p-value test de Fisher de la régression	p-value test de Student variables explicatives	Impact sur le taux de retard
NbBifurcationsRéseauClassique	Nombre de bifurcations majeures du réseau classique sur la liaison	3,6%	< 0,1%	< 0,1%	-0,9 pts / bifurcation
LongueurRéseauClassique	Estimation de la longueur parcourue sur le réseau classique pour une circulation de la liaison (en km)	6,1%	< 0,1%	< 0,1%	-1,2 pts / 100 km

Le tableau ci-dessous énumère les variables explicatives pour lesquelles le modèle de régression linéaire est significatif – il passe le test de Fisher au seuil de significativité de 5% - et pour lesquelles le signe de l'impact de la variable explicative sur le taux de retard est conforme à l'intuition, ou tout du moins ne s'oppose pas à l'intuition.

Variable explicative	Description de la variable	R ² ajusté régression	p-value test de Fisher de la régression	p-value test de Student variables explicatives	Impact sur le taux de retard
DuréeTrajet	Durée moyenne du trajet sur la liaison	0,8%	< 1%	< 1%	+0,42 pts / heure de trajet
DistanceMoy	Distance moyenne parcourue par un train entre les 2 gares sur la liaison	4%	< 0,1%	< 0,1%	+0,53 pts / 100 km
VitesseMoyenne	Vitesse moyenne théorique des circulations de trains sur la liaison	3,5%	< 0,1%	< 0,1%	+0,33 pts / 10 km/h
DépartParis	La liaison part de Paris intra-muros	8,8%	< 0,1%	< 0,1%	-3,3 pts

Liaison_radiale	La liaison part ou arrive à Paris intra-muros	16%	< 0,1%	< 0,1%	-6,7 pts
DépartMarseille	La liaison part de Marseille	3,3%	< 0,1%	< 0,1%	+5,7 pts
ArrivéeMarseille	La liaison arrive à Marseille	2,1%	< 0,1%	< 0,1%	+4,7 pts
DépartLyon	La liaison part de Lyon-Part-Dieu	8%	< 0,1%	< 0,1%	+7,2 pts
ArrivéeLyon	La liaison arrive à Lyon-Part-Dieu	0,5%	< 5%	< 5%	+1,8 pts
DépartOuArr_Lyon_Marseille	La liaison part ou arrive à Marseille ou à Lyon-Part-Dieu	9%	< 0,1%	< 0,1%	+4,8 pts
NbNoeudsMajeursTraversés	Nombre de nœuds majeurs (Paris, Lyon, Marseille) traversés sur la liaison	7,5%	< 0,1%	< 0,1%	+2,4 pts
LiaisonPasseParAvignon	La liaison passe par le nœud ferroviaire des Angles (Gard) sur la LGV Méditerranée (nœud ferroviaire simplifié ici comme Avignon)	11%	< 0,1%	< 0,1%	+4,5 pts
LiaisonPasseParDijon	La liaison passe par le nœud ferroviaire de Dijon	4%	< 0,1%	< 0,1%	-4,7 pts
LiaisonPasseParLille	La liaison passe par le nœud ferroviaire de Lille	2,6%	< 0,1%	< 0,1%	+3,3 pts
LiaisonPasseParRennes	La liaison passe par le nœud ferroviaire de Rennes	3%	< 0,1%	< 0,1%	-2,8 pts
DépartOuArr_NoedSecondaire	La liaison part ou arrive au niveau d'un nœud ferroviaire secondaire (Avignon, Bordeaux, Lille, Le Mans, Tours)	4%	< 0,1%	< 0,1%	+2,7 pts
NbNoeudsSecondaires	Nombre de nœuds secondaires (Avignon, Bordeaux, Lille, Le Mans, Tours) sur la liaison	2,8%	< 0,1%	< 0,1%	+1,6 pts

NbBifurcationsLGV	Nombre de bifurcations majeures LGV sur la liaison (y compris en Île-de-France)	11%	< 0,1%	< 0,1%	+3 pts
NbBifurcationsLGV_horsIDF	Nombre de bifurcations majeures LGV sur la liaison (hors bifurcations en Île-de-France)	2,9%	< 0,1%	< 0,1%	+1,9 pts
TMJA_max_TGV_Liaison_2018	Trafic moyen journalier annuel (TMJA) des TGV de la section la plus intensément utilisée sur la liaison en 2018	0,5%	< 5%	< 5%	+1,2 pts / 100 circulations TGV/jour
NbMoyenArrêts_2020	Nombre moyen d'arrêts effectués par les TGV sur la liaison au cours d'un « Jour Ouvrable de Base » (JOB) de 2020 en dehors des périodes de vacances	1,4%	< 0,1%	< 0,1%	+0,4 pts / arrêt
TMJA_max_Total_Liaison_2000	Trafic journalier moyen annuel (TJMA) pour l'ensemble des circulations de trains (y compris fret et "haut le pied") sur la section la plus intensément utilisée de la liaison en 2000	3,6%	< 0,1%	< 0,1%	+1,1 pts / 100 circulations de trains/jour

3.2.1.3 Modèle de régression linéaire multiple

A partir de l'analyse des résultats des régressions linéaires par variables explicatives, un modèle de régression linéaire multiple a été construit. Les variables explicatives pour lesquelles le modèle de régression linéaire simple ne passe pas le test de Fisher au seuil de significativité de 5%, ainsi que celles pour lesquelles le signe de l'impact de la variable explicative sur le taux de retard semble contraire à l'intuition, ont été exclues de la sélection des variables explicatives du modèle.

Certaines variables explicatives testées dans la section 3.2.1.2 présentent des colinéarités. Un tableau de synthèse des relations de colinéarité entre ces variables est fourni en Annexe 8. Pour s'affranchir des effets de colinéarité, les variables colinéaires avec l'une des variables explicatives retenues ont été exclues des modèles testés.

Différents modèles ont été testés à l'aide du logiciel R. Le modèle retenu est celui qui minimise le critère d'information d'Akaike (AIC), défini comme :

$$AIC = 2k - 2 \ln(L)$$

avec k le nombre de paramètres à estimer et L le maximum de la fonction de vraisemblance du modèle.

Ce critère permet de retenir le modèle qui offre le meilleur compromis entre la qualité de l'ajustement (notamment la part de la variance du taux de retard expliquée par le modèle) et le nombre de variables explicatives retenues dans le modèle. Il pénalise en effet les modèles constitués d'un grand nombre de variables explicatives et limite ainsi l'effet d'amélioration de la qualité de l'ajustement à mesure que l'on ajoute des variables explicatives. En pratique, entre plusieurs modèles, on retient celui qui minimise le critère d'information d'Akaike.

Le modèle de régression linéaire multiple suivant a été retenu :

```
Call:
lm(formula = TxRetard ~ DépartParis + DépartLyon + LiaisonPasseParAvignon +
    LiaisonPasseParRennes + LiaisonPasseParDijon + DépartOuArr_NoedSecondaire +
    NbMoyenArrêts_2019 + TMJA_max_Total_Liaison_2000, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-13.1148  -3.2723  -0.1854   2.8244  18.7606

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   10.986504   0.669113   16.420 < 2e-16 ***
DépartParis   -2.291971   0.335944   -6.822 1.76e-11 ***
DépartLyon     4.890913   0.754562    6.482 1.58e-10 ***
LiaisonPasseParAvignon  2.460692   0.489295    5.029 6.08e-07 ***
LiaisonPasseParRennes -2.849098   0.566522   -5.029 6.08e-07 ***
LiaisonPasseParDijon -4.838904   0.743421   -6.509 1.33e-10 ***
DépartOuArr_NoedSecondaire  3.113007   0.402450    7.735 3.10e-14 ***
NbMoyenArrêts_2019    0.314262   0.104626    3.004 0.00275 **
TMJA_max_Total_Liaison_2000  0.005308   0.002261    2.348 0.01912 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

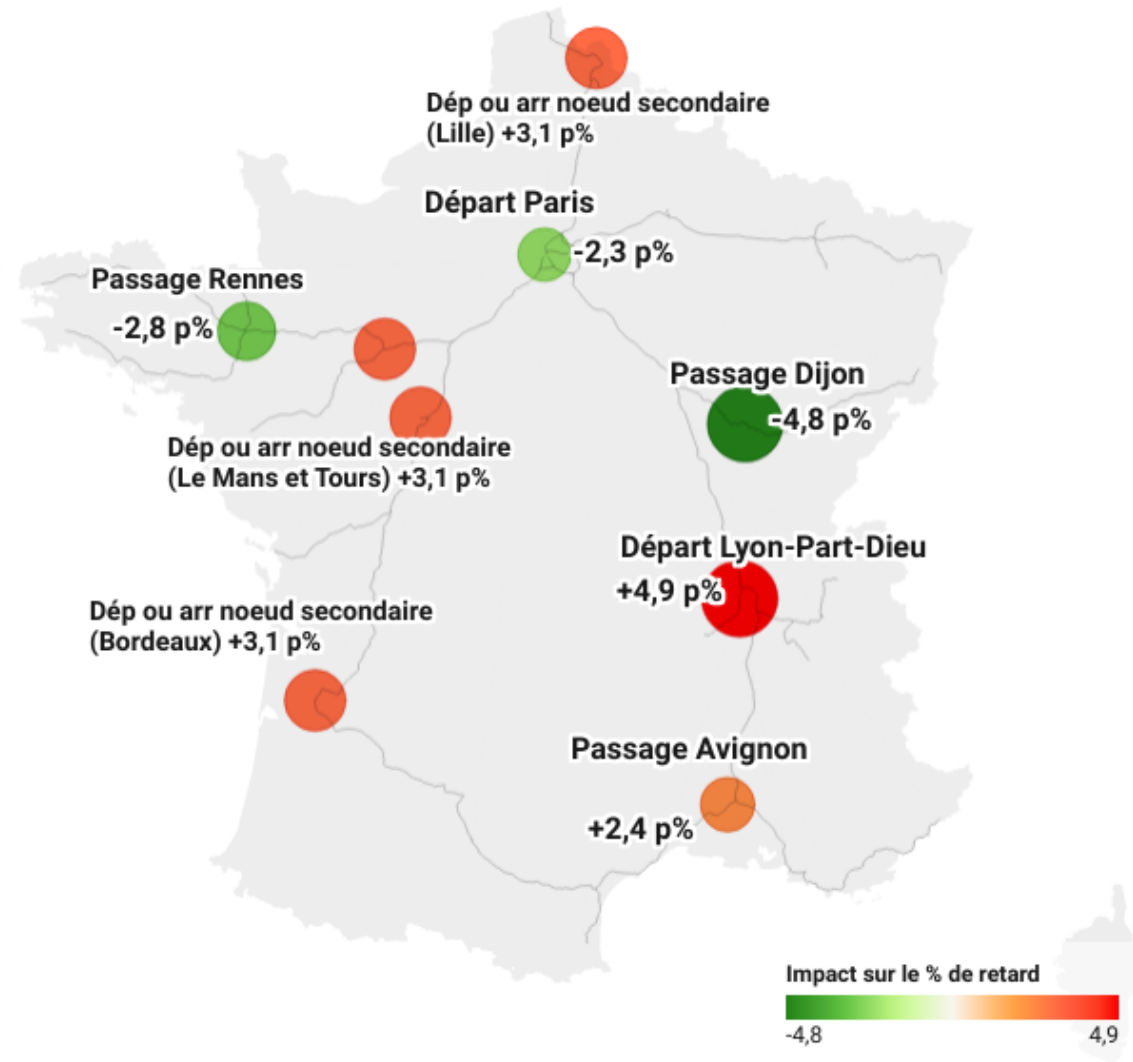
Residual standard error: 4.583 on 803 degrees of freedom
Multiple R-squared:  0.3244,    Adjusted R-squared:  0.3177
F-statistic: 48.2 on 8 and 803 DF,  p-value: < 2.2e-16
```

D'après la p-value associée au test de Fisher sur la régression, le modèle retenu est significatif au seuil de 0,1%. D'après les p-values associées aux différentes variables explicatives du modèle, l'effet du nombre moyen d'arrêts des circulations des TGV sur le taux de retard est significatif au seuil 1% et celui du niveau du trafic au cours de l'année 2000 de l'ensemble des circulations de trains (y compris fret et « haut-le-pied ») sur la section la plus intensément utilisée du réseau empruntée par la liaison TGV est significatif au seuil 2% mais pas au seuil 1%. L'effet des autres variables explicatives sur le taux de retard est significatif au seuil 0,1%.

Une première analyse des coefficients des différentes variables explicatives retenues dans le modèle est détaillée en annexe 9.

La carte ci-dessous permet de visualiser l'effet sur la ponctualité des TGV du départ, du passage ou de l'arrivée au niveau de certains nœuds ferroviaires du réseau TGV d'après le modèle retenu :

Estimation économétrique de l'effet sur la ponctualité du passage des TGV par certains nœuds ferroviaires



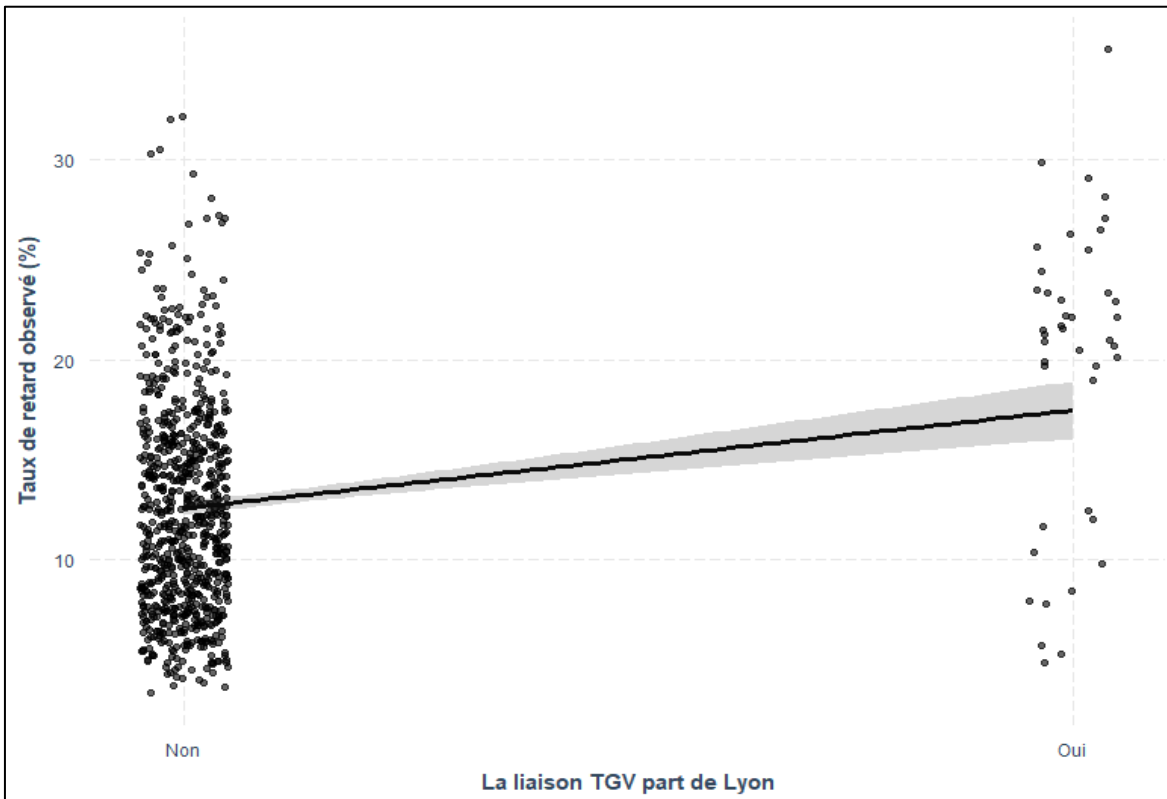
Carte: Alexandre Barbusse/AQST - Créé avec Datawrapper

Des graphiques illustrant les relations de corrélations identifiées avec le modèle retenu ont été construits. Certains d'entre eux, pour lesquels ces relations semblent assez apparentes graphiquement, sont présentés ci-après. Les autres graphiques, moins parlant, sont disponibles en annexe 9.

D'après le modèle retenu, une liaison TGV ayant pour départ Paris présente en moyenne un taux de retard plus faible d'environ 2,3 points, alors qu'une liaison TGV au départ de Lyon Part-Dieu présente en moyenne un taux de retard plus élevé de 4,9 points. Une explication possible à ce dernier point pourrait être qu'il y a une saturation du réseau ferroviaire au niveau du nœud ferroviaire

lyonnais, ce qui rejoint le constat effectué par plusieurs experts ferroviaires dans la littérature spécialisée. [\[1\]](#)[\[2\]](#)[\[3\]](#)

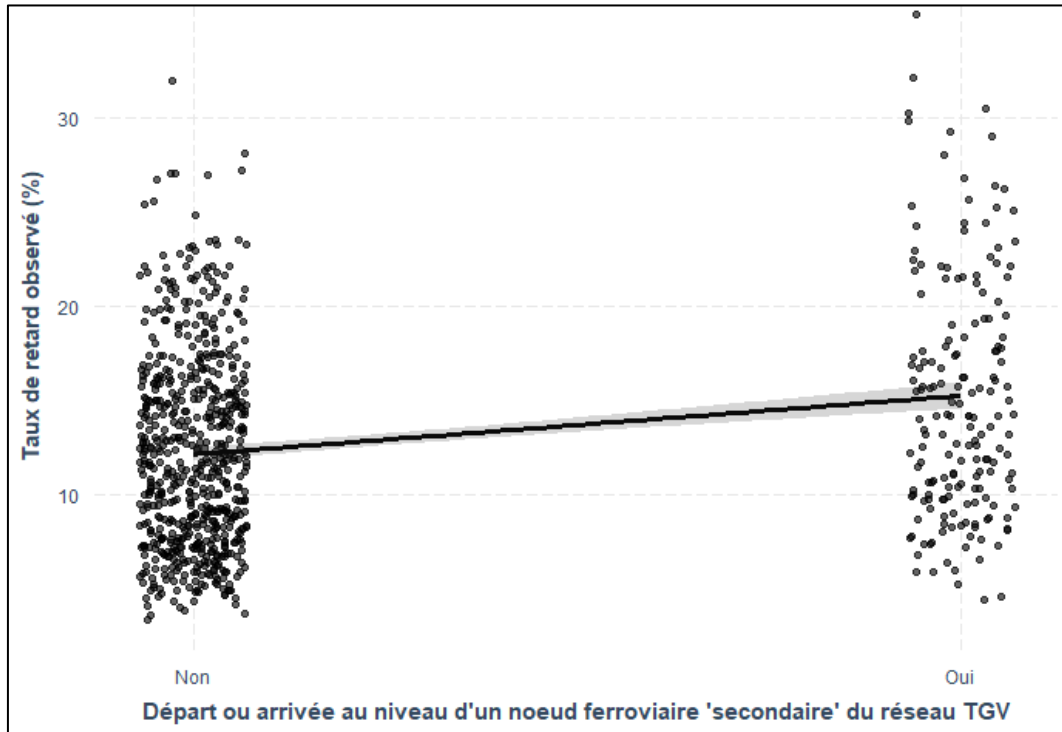
Le graphique ci-après montre qu'en moyenne les liaisons TGV partant de Lyon-Part-Dieu sont moins ponctuelles que celles qui ne partent pas de Lyon-Part-Dieu, même si la distribution des taux de retard observés pour les liaisons TGV partant de Lyon-Part-Dieu semble divisée en 2 groupes bien distincts. La zone ombrée autour de la droite de régression correspond à la plage d'incertitude statistique sur cette droite.



Le modèle indique également que les liaisons TGV qui passent par Avignon sur la LGV Méditerranée et desservent ou bien partent du littoral méditerranéen sont moins ponctuelles

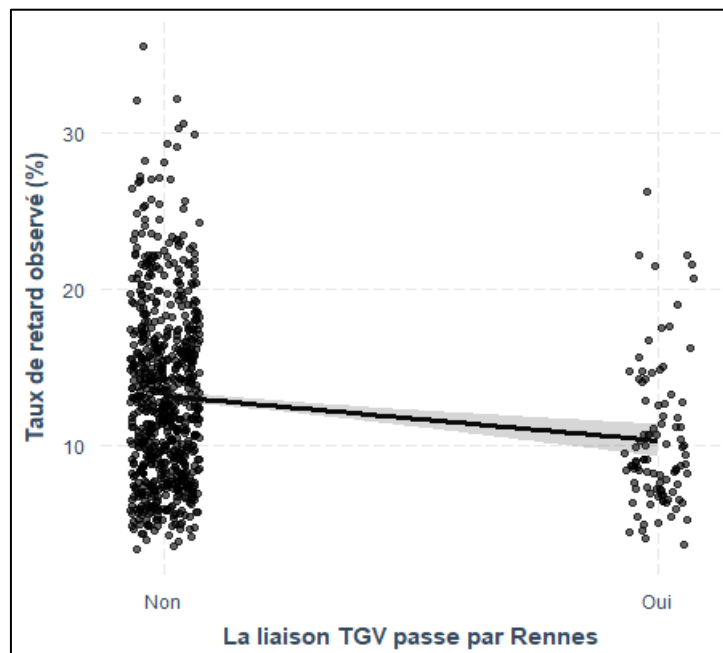
Le modèle indique également que le taux de retard des TGV partant ou arrivant au niveau d'un des nœuds ferroviaires secondaires du réseau ferré retenus (Avignon, Bordeaux, Le Mans, Lille, Tours) est supérieur de près de 3,1 points à celui des autres TGV. Du fait de la saturation du réseau ferré observée au niveau de ces nœuds, il est possible qu'un plus grand nombre de TGV soient impactés si un retard s'y produit. Les difficultés de gestion du trafic au niveau d'un nœud ferroviaire pourraient également accroître le risque de retard.

Le graphique ci-dessous illustre la moins bonne ponctualité des TGV partant de ou arrivant au niveau d'un nœud ferroviaire secondaire du réseau TGV :



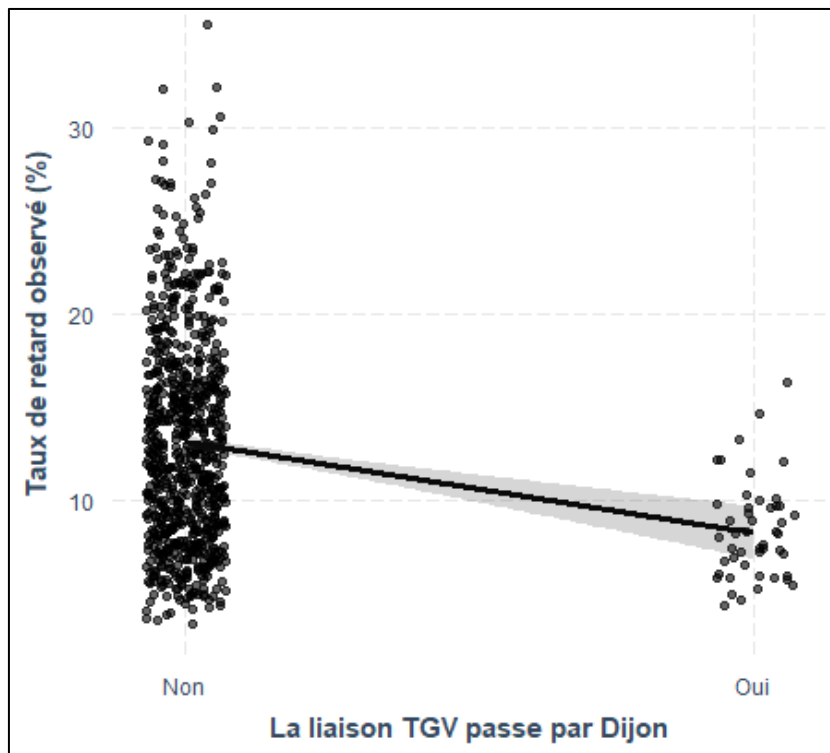
La valeur du coefficient associé au passage de la liaison TGV par Rennes indique qu'en moyenne, le taux de retard des liaisons concernées est moins élevé de 2,8 points. Les liaisons TGV qui traversent Rennes empruntent toutes la LGV Bretagne-Pays de la Loire. Le nœud ferroviaire de Rennes a fait l'objet d'investissements importants de capacité lors de la construction de la LGV Bretagne-Pays de la Loire (mise en service en 2017), ce qui a probablement contribué à la bonne ponctualité des liaisons TGV qui le traversent.

Le graphique ci-dessous illustre la meilleure ponctualité en moyenne des liaisons TGV passant par Rennes :



Un raisonnement similaire pourrait être invoqué pour expliquer en partie la nettement meilleure ponctualité des liaisons TGV traversant l'agglomération de Dijon (-4,8 points de retard en moyenne). En effet, ces liaisons TGV empruntent la LGV Rhin-Rhône. Cette LGV est en effet récente (mise en service en 2011) et était relativement peu intensément utilisée en 2018 d'après l'atlas du réseau ferré en France – édition 2020, publié par SNCF Réseau.

Le graphique ci-dessous illustre le fait qu'en moyenne les liaisons TGV passant par Dijon sont plus ponctuelles que celles qui ne partent pas de Dijon :



Selon le modèle retenu, plus une circulation TGV effectue d'arrêts, plus elle est susceptible de connaître des retards, ce qui est conforme à l'intuition relativement à l'effet de cette variable. En revanche, on peut noter que l'impact du nombre d'arrêts sur la ponctualité n'est pas très prononcé d'après le modèle (+0,3 points par arrêt supplémentaire).

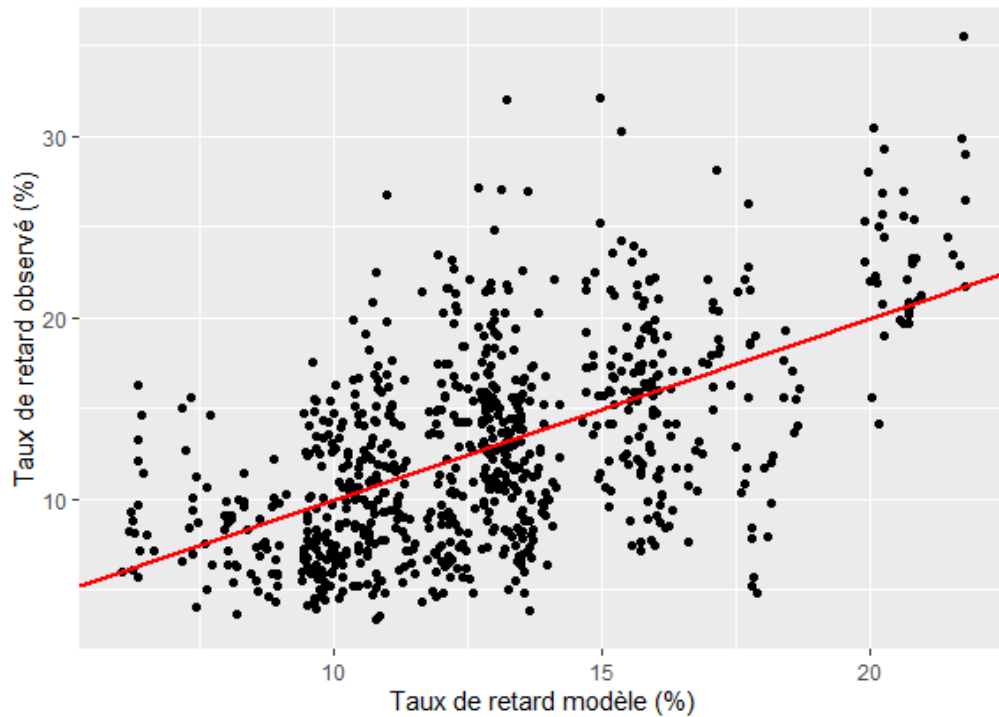
La valeur du coefficient associé au niveau du trafic moyen journalier pour l'année 2000 de l'ensemble des circulations de trains (y compris fret et « haut-le-pied ») sur le tronçon dimensionnant du réseau ferré emprunté par la liaison TGV indique qu'en moyenne, si ce tronçon est emprunté par 100 trains de plus par jour, il présentera un taux de retard plus élevé de 0,5 points. Un élément partiel d'explication pourrait être que lorsqu'une circulation TGV emprunte des lignes très fréquentées, elle soit davantage exposée au risque de subir un retard qui se propagerait sur le réseau d'une circulation de train à la suivante, et cela sans que ce retard ne puisse nécessairement être « rattrapé » au cours du trajet, puisque le débit de circulations de trains sur la ligne est déjà très élevé.

Enfin, le coefficient de détermination ajusté ou R^2 ajusté du modèle retenu – qui minimise le critère d'information d'Akaike – est de 0,318, ce qui signifie que seulement 31,8% de la variance du taux de retard moyen annuel par liaisons TGV sur la période 2012-2019 est expliquée par le modèle. Autrement dit, 68,2% de la variabilité du taux de retard ne peuvent pas être expliqués par les variables explicatives retenues dans le modèle.

En effet, le taux de retard des trains peut dépendre d'un très grand nombre de paramètres, et de nombreux évènements ayant un impact significatif sur le taux de retard ne peuvent pas être intégrés au modèle :

- Les évènements météorologiques majeurs (inondations et pluies intenses, tempêtes, épisodes neigeux et de froid, canicule et fortes chaleurs, etc.) ;
- Les mouvements sociaux ;
- Les autres évènements externes au transport, dont notamment les actes de malveillance, les accidents de personne, les collisions des trains avec des véhicules, ou avec des animaux, ...)
- Les incidents liés au dysfonctionnements des infrastructures ferroviaires, par exemple les ruptures de caténaire ou les pannes de signalisation ;
- Les dysfonctionnements du matériel roulant, avec notamment les pannes du matériel moteur ;
- Les incidents voyageurs (malaise d'un voyageur, blocage involontaire des portes par un voyageur, etc.)

Le graphique ci-après, croisant le taux de retard prédit par le modèle (en abscisse) et le taux de retard observé (en ordonnée), permet d'apprécier d'une autre manière la capacité du modèle à décrire la variabilité du taux de retard par liaison TGV. La ligne rouge correspond à la droite $y = x$ sur laquelle tous les points doivent s'aligner dans un modèle parfait théorique. On constate que la distribution des points présente une légère tendance à suivre cette droite à mesure que le taux de retard prédit augmente, mais que de nombreux points s'écartent fortement de la droite. Cela confirme le caractère limité du pouvoir descriptif du modèle retenu face à la forte variabilité du taux de retard des TGV par liaison.



3.2.2 Résultats pour les trains Intercités

3.2.2.1 Résultats détaillés des régressions par variables explicatives

Les résultats détaillés des modèles de régressions testés sont présentés en Annexe 7.

3.2.2.2 Synthèse des résultats

Le tableau ci-dessous énumère les variables explicatives pour lesquelles le modèle de régression linéaire utilisé n'est pas statistiquement fiable au sens où il ne passe pas le test de Fisher au seuil de significativité de 5%. On ne peut donc pas conclure sur l'effet de ces variables sur le taux de retard avec les données à notre disposition.

Variable explicative	Description de la variable	R² ajusté régression	p-value test de Fisher de la régression	p-value test de Student variables explicatives
NbCirc_Liaison_an	Nombre de trains ayant circulé sur la liaison	< 0,1%	> 10%	> 10%
VitesseMoyenne	Vitesse moyenne des circulations de trains sur la liaison	< 0,1%	> 10%	> 10%
Liaison_radiale	La liaison part ou arrive à Paris	< 0,1%	> 10%	> 10%
DépartLyon	La liaison part de Lyon	< 0,1%	> 10%	> 10%
ArrivéeLyon	La liaison arrive à Lyon	< 0,1%	> 10%	> 10%
LiaisonTraverseTours	La liaison passe par le nœud ferroviaire de Tours/Saint-Pierre-des-Corps	< 0,1%	> 10%	> 10%
LiaisonTraverseBrive	La liaison passe par le nœud ferroviaire de Brive-la-Gaillarde	< 0,1%	> 10%	> 10%
LiaisonTraverseNevers	La liaison passe par le nœud ferroviaire de Nevers	< 0,5%	> 10%	> 10%
LiaisonTraverseCaen	La liaison passe par le nœud ferroviaire de Caen	< 0,5%	> 5%	> 5%
LiaisonTraverseAmiens	La liaison passe par Amiens	0,7%	> 5%	> 5%
LiaisonTraverseSaint-Quentin	La liaison passe par Saint-Quentin (Aisne)	< 0,1%	> 10%	> 10%
Sens_Est_Ouest	Le trajet sur la liaison s'effectue en sens "Est -> Ouest"	< 0,1%	> 10%	> 10%
Sens_Sud_Nord	Le trajet sur la liaison s'effectue en sens "Sud -> Nord"	< 0,1%	> 10%	> 10%

TMJA_max_Total_Liaison_2000	Trafic journalier moyen annuel (TJMA) pour l'ensemble des circulations de trains (y compris fret et "haut le pied") sur la section la plus intensément utilisée de la liaison en 2000	< 0,1%	> 10%	> 10%
TMJA_max_Total_Liaison_2000_horsParis	Trafic journalier moyen annuel (TJMA) pour l'ensemble des circulations de trains (y compris fret et "haut le pied") sur la section la plus intensément utilisée de la liaison en 2000 hors de Paris intra-muros	< 0,1%	> 10%	> 10%

Le tableau ci-dessous énumère les variables explicatives pour lesquelles le modèle de régression linéaire est significatif – il passe le test de Fisher au seuil de significativité de 5% - et pour lesquelles le signe de l'impact de la variable explicative sur le taux de retard est conforme à l'intuition, ou tout du moins ne s'oppose pas à l'intuition.

Variable explicative	Description de la variable	R ² ajusté régression	p-value test de Fisher de la régression	p-value test de Student variables explicatives	Impact sur le taux de retard
DuréeTrajet	Durée moyenne du trajet sur la liaison	5,1%	< 0,1%	< 0,1%	+0,45 pts / heure de trajet
DistanceMoy	Distance moyenne parcourue par un train entre les 2 gares sur la liaison	8,4%	< 0,1%	< 0,1%	+0,44 pts / 100 km
LiaisonNuit	La liaison est une liaison Intercités de nuit	1,9%	< 1%	< 1%	+2,5 pts
DépartParis	La liaison part de Paris	1,4%	< 1%	< 1%	+1,7 pts
ArrivéeParis	La liaison arrive à Paris	1,6%	< 1%	< 1%	-1,8 pts

DépartMarseille	La liaison part de Marseille	3,9%	< 0,1%	< 0,1%	+9,8 pts
ArrivéeMarseille	La liaison arrive à Marseille	1,3%	< 0,1%	< 0,1%	+4,7 pts
NbNoeudsMajeursTraversés	Nombre de nœuds majeurs (Paris, Lyon, Marseille) traversés sur la liaison	6,7%	< 0,1%	< 0,1%	+3,2 pts/nœud majeur traversé
LiaisonTraverseBordeaux	La liaison passe par le nœud ferroviaire de Bordeaux	10%	< 0,1%	< 0,1%	+6,5 pts
LiaisonTraverseNantes	La liaison passe par le nœud ferroviaire de Nantes	1,3%	< 0,1%	< 0,1%	+3,4 pts
LiaisonTraverseToulouse	La liaison passe par le nœud ferroviaire de Toulouse Matabiau	2,8%	< 0,1%	< 0,1%	+3,0 pts
LiaisonTraverseOrléans	La liaison passe par le nœud ferroviaire d'Orléans	1,1%	< 5%	< 5%	+1,7 pts
LiaisonTraverseNarbonne	La liaison passe par le nœud ferroviaire de Narbonne	7,2%	< 0,1%	< 0,1%	+6,2 pts
LiaisonTraverseDijon	La liaison passe par le nœud ferroviaire de Dijon	3,6%	< 0,1%	< 0,1%	< 0,1%
LiaisonTraverseRouen	La liaison passe par le nœud ferroviaire de Rouen	1,9%	< 1%	< 1%	-3,6 pts
DépartOuArr_NoedSecondaire	La liaison part ou arrive au niveau d'un nœud ferroviaire secondaire (Bordeaux, Nantes, Toulouse, Orléans, Tours, Brive-la-Gaillarde, Narbonne, Nevers, Dijon, Caen, Amiens, Saint-Quentin)	3,0%	< 0,1%	< 0,1%	+2,3 pts
NbNoeudsSecondaires	Nombre de nœuds secondaires (Bordeaux, Nantes, Toulouse, Orléans, Tours, Brive-la-Gaillarde, Narbonne, Nevers, Dijon, Caen, Amiens, Saint-Quentin) sur la liaison	8,2%	< 0,1%	< 0,1%	+1,7 pts/nœud secondaire traversé

Axe_Nord_Normandie	La liaison appartient à l'axe "Nord-Normandie" du réseau Intercités de 2016	1,7%	< 1%	< 1%	-1,9 pts
Axe_Med_Atlantique	La liaison appartient à l'axe "Méditerranée-Atlantique" du réseau Intercités de 2016	5,5%	< 0,1%	< 0,1%	+4,7 pts
Axe_Paris_Centre_SudOuest	La liaison appartient à l'axe "Paris – Centre - Sud-Ouest" du réseau Intercités de 2016	1,1%	< 5%	< 5%	+1,7 pts
Axe_Est_Centre	La liaison appartient à l'axe "Est – Centre" du réseau Intercités de 2016	1,7%	< 1%	< 1%	-2,0 pts
Ouest_Est	Le trajet sur la liaison s'effectue en sens "Ouest -> Est"	1,1%	< 5%	< 5%	-1,8 pts
Nord_Sud	Le trajet sur la liaison s'effectue en sens "Nord -> Sud"	3,7%	< 0,1%	< 0,1%	+2,8 pts
NbMoyenArrêts_2020	Nombre moyen d'arrêts effectués par les trains Intercités sur la liaison au cours d'un Jour Ouvrable de Base (JOB) de 2020 en dehors des périodes de vacances	0,9%	< 5%	< 5%	-0,15 pts/arrêt

3.2.2.3 Modèle de régression linéaire multiple

A partir de l'analyse des résultats des régressions linéaires par variables explicatives, un modèle de régression linéaire multiple a été construit. Les variables explicatives pour lesquelles le modèle de régression linéaire simple ne passe pas le test de Fisher au seuil de significativité de 5% ont été exclus de la sélection des variables explicatives du modèle.

Certaines variables explicatives testées dans la section 3.2.2.2 présentent des colinéarités. Un tableau de synthèse des relations de colinéarité entre ces variables est fourni en Annexe 13. Pour s'affranchir des effets de colinéarité, les variables colinéaires avec l'une des variables explicatives retenues ont été exclues des modèles testés.

Différents modèles ont été testés à l'aide du logiciel R. Le modèle retenu est celui qui minimise le critère d'information d'Akaike (AIC), présenté précédemment dans la section 3.2.1.3. Il s'agit du modèle suivant :

```
Call:
lm(formula = TxRetard ~ LiaisonNuit + NbNoeudsMajeursTraversés +
    LiaisonTraverseBordeaux + LiaisonTraverseToulouse + LiaisonTraverseDijon +
    LiaisonTraverseRouen + DépartOuArr_NoeudSecondaire + Axe_Est_Centre +
    Nord_Sud, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-15.0183  -3.2128  -0.3527   2.6899  22.4246

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)         6.9484     0.8298   8.374 8.36e-16 ***
LiaisonNuit         -3.9108     1.0465  -3.737 0.000212 ***
NbNoeudsMajeursTraversés
                    5.2014     0.7097   7.329 1.19e-12 ***
LiaisonTraverseBordeaux
                    7.2998     0.8684   8.406 6.59e-16 ***
LiaisonTraverseToulouse
                    2.8881     0.8259   3.497 0.000521 ***
LiaisonTraverseDijon
                    4.2044     1.7489   2.404 0.016645 *
LiaisonTraverseRouen
                   -2.9840     1.0266  -2.907 0.003848 **
DépartOuArr_NoeudSecondaire
                    1.6491     0.6238   2.644 0.008505 **
Axe_Est_Centre     -2.2518     0.7699  -2.925 0.003633 **
Nord_Sud            2.4554     0.5921   4.147 4.08e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

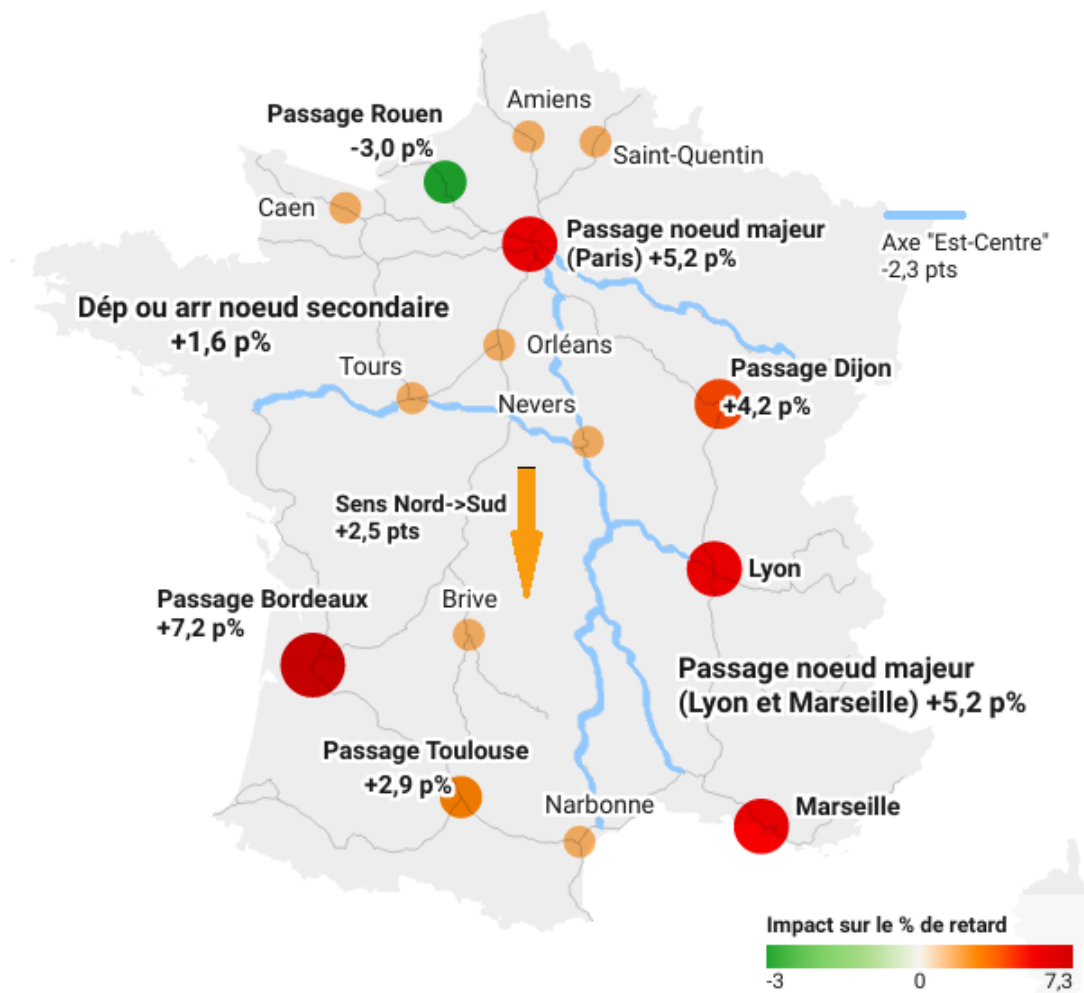
Residual standard error: 5.288 on 421 degrees of freedom
Multiple R-squared:  0.3482,    Adjusted R-squared:  0.3343
F-statistic: 24.99 on 9 and 421 DF,  p-value: < 2.2e-16
```

D'après la p-value associée au test de Fisher sur la régression, le modèle retenu est significatif au seuil de 0,1%. D'après les p-values associées aux différentes variables explicatives du modèle, l'effet de la traversée du nœud ferroviaire dijonnais sur le taux de retard est significatif au seuil 2%, mais pas au seuil 1%. L'effet du passage par Rouen, celui d'un départ ou d'une arrivée au niveau d'un nœud ferroviaire secondaire du panel retenu (Bordeaux, Nantes, Toulouse, Orléans, Tours, Brive-la-Gaillarde, Narbonne, Nevers, Dijon, Caen, Amiens, Saint-Quentin), et celui d'une liaison appartenant à l'axe « Est – Centre » du réseau Intercités de 2016, présenté en annexe 12, sont significatifs au seuil 1%. L'effet des autres variables explicatives sur le taux de retard est significatif au seuil 0,1%.

Une première analyse des coefficients des différentes variables explicatives retenues dans le modèle est détaillée en annexe 14.

La carte ci-dessous permet de visualiser l'effet sur la ponctualité des Intercités du départ, du passage ou de l'arrivée au niveau de certains nœuds ferroviaires du réseau Intercités (période 2012-2019) d'après le modèle retenu. De nombreuses liaisons sont désormais dans le périmètre des services régionaux TER. Le diagnostic de l'effet sur la ponctualité du passage par ces nœuds ferroviaires reste toutefois instructif malgré les changements de périmètres.

Estimation économétrique de l'effet sur la ponctualité du passage des trains Intercités par certains nœuds ferroviaires



Carte: Alexandre Barbusse/AQST - Créé avec Datawrapper

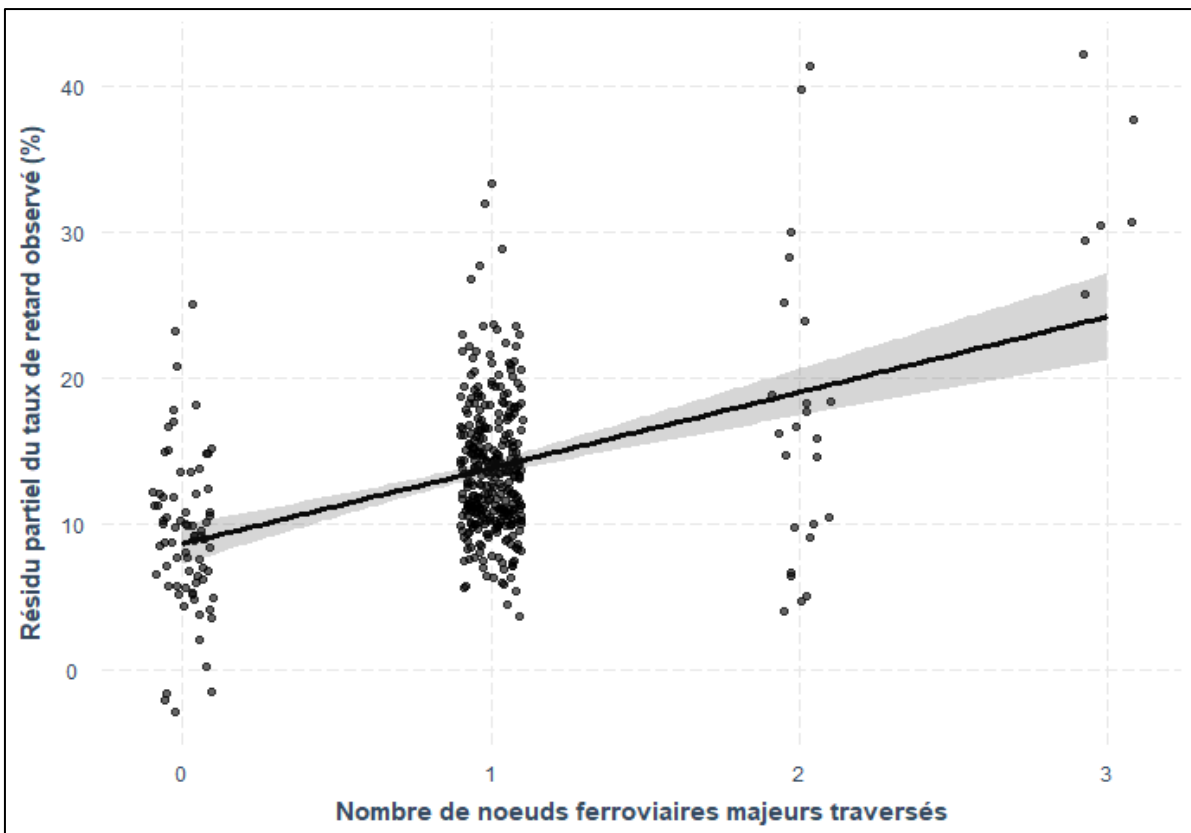
Des graphiques illustrant les relations de corrélations identifiées avec le modèle retenu ont été construits. Certains d'entre eux, pour lesquels ces relations semblent assez apparentes

graphiquement, sont présentés ci-après. Les autres graphiques, moins parlant, sont disponibles en annexe 14.

D'après le modèle retenu, une liaison Intercités de nuit présente en moyenne un taux de retard plus faible d'environ 3,9 points. Le plus faible niveau de trafic des liaisons de nuit relativement aux autres liaisons Intercités pourrait expliquer pour partie ce meilleur niveau de ponctualité. Les retards liés à la gestion du trafic et aux effets de réseau – propagation d'un retard d'une circulation de train à la suivante – ont en effet moins de chance de survenir lorsque la fréquence de circulation de trains est faible.

Le modèle suggère qu'en moyenne, chaque nœud majeur supplémentaire traversé par une liaison Intercités se traduit par une forte hausse du taux de retard (+5,2 points/ nœud majeur traversé).

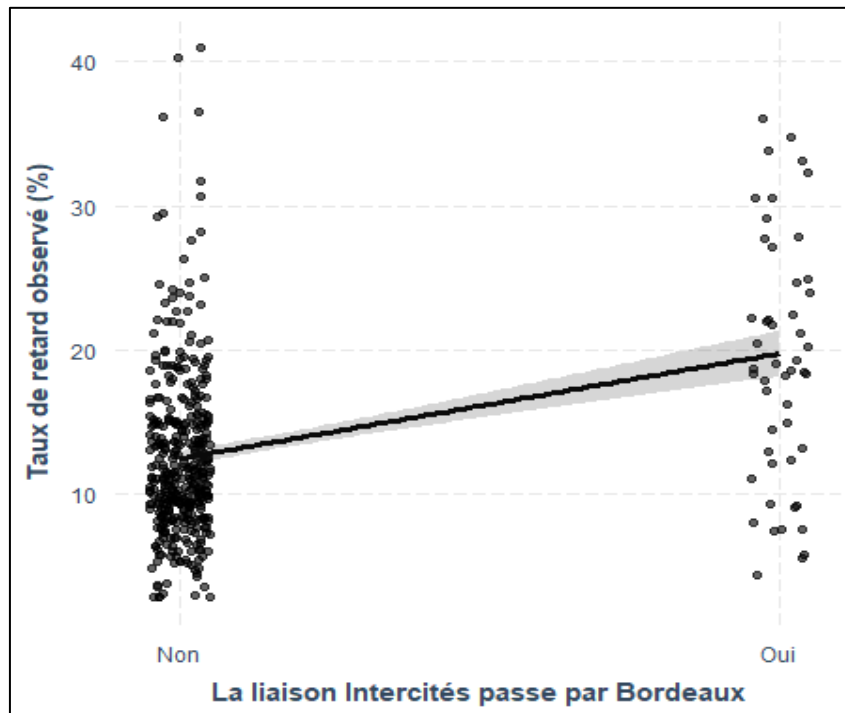
Le graphique ci-dessous illustre cet effet. La zone ombrée autour de la droite de régression correspond à la plage d'incertitude statistique sur cette droite. En axe des ordonnées, le résidu partiel du taux de retard observé est représenté. Il s'agit en fait du taux de retard observé auquel on retranche l'effet des autres variables de la régression, afin d'identifier plus facilement la relation de corrélation graphiquement. Une relation linéaire semble convenir, même si, à nombre de nœuds majeurs traversés fixé, la dispersion du résidu autour de la droite est importante.



Le modèle indique également que les liaisons Intercités traversant les nœuds ferroviaires de Bordeaux, Toulouse et Dijon sont moins ponctuelles. Au-delà de l'effet d'engorgement qui peut se

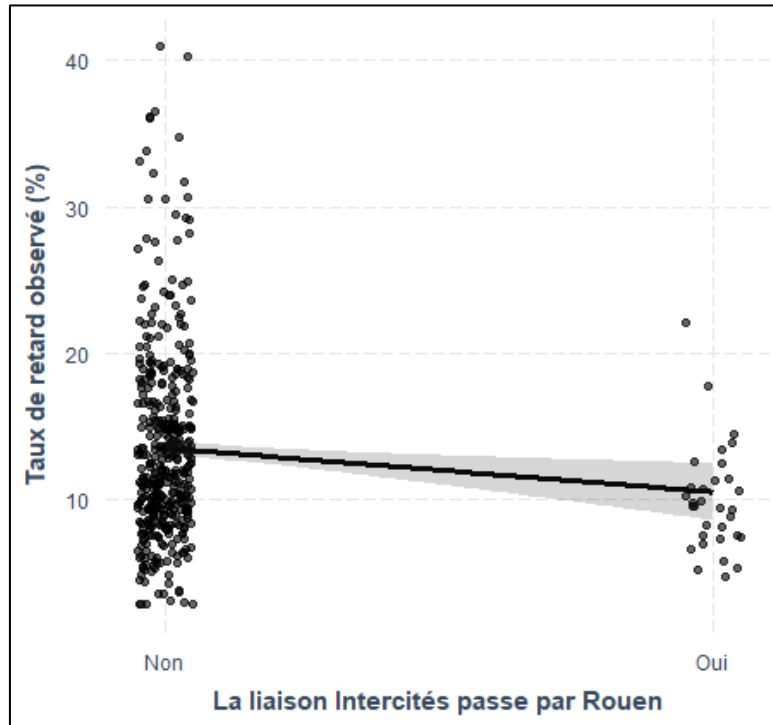
produire au niveau du nœud ferroviaire à proprement parler, on peut noter que le moins bon niveau de ponctualité des liaisons passant par Bordeaux est influencé par les mauvaises performances de ponctualité de la liaison Bordeaux – Marseille, pour laquelle il n’y a pas d’explication évidente *a priori*. La moins bonne ponctualité des liaisons Intercités passant par Toulouse-Matabiau semble en accord avec la nécessité d’augmenter la capacité d’accueil des trains en gare de Toulouse-Matabiau soulevée par SNCF Réseau dans son rapport de concertation préalable au projet de modernisation du nœud ferroviaire de Toulouse [5].

Le graphique ci-dessous illustre le fait qu’en moyenne les liaisons Intercités passant par Bordeaux sont moins ponctuelles que celles qui n’y passent pas :

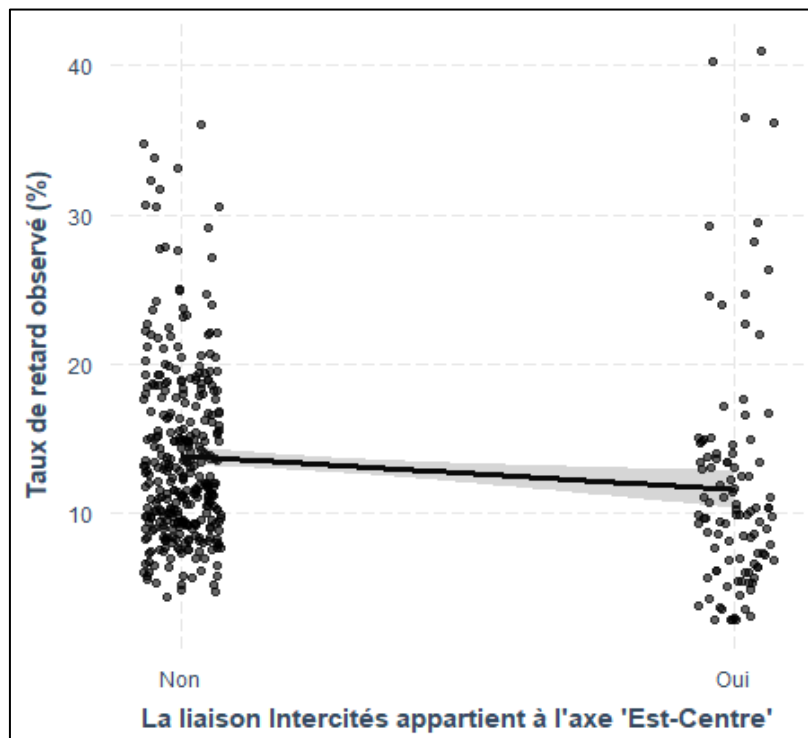


Selon le modèle retenu, les trains Intercités dont le départ ou l’arrivée s’effectue au niveau de l’un des nœuds ferroviaires secondaires retenus (Bordeaux, Nantes, Toulouse, Orléans, Tours, Brive-la-Gaillarde, Narbonne, Nevers, Dijon, Caen, Amiens, Saint-Quentin) sont également moins ponctuels que les autres trains Intercités (+1,6 points de retard). Du fait de la saturation du réseau ferré observée au niveau de certains de ces nœuds, un plus grand nombre de trains Intercités sont impactés lorsqu’un retard s’y produit. Les difficultés de gestion du trafic au niveau d’un nœud ferroviaire tendent également à accroître le risque de retard.

Le graphique ci-dessous illustre le fait qu’en moyenne les liaisons Intercités passant par Rouen sont davantage ponctuelles que celles qui n’y passent pas :



Le graphique ci-dessous montre qu'en moyenne les liaisons de l'ancien axe Intercités « Est-Centre » étaient plus ponctuelles que celles des autres axes Intercités :



Enfin, le modèle retenu suggère que les liaisons Intercités de l'axe « Est – Centre » du réseau Intercités de 2016 (présenté en Annexe 12) sont plus ponctuelles que les autres (-2,3 points de retard), et que les trains Intercités empruntant un trajet dans la direction Nord -> Sud sont moins ponctuels

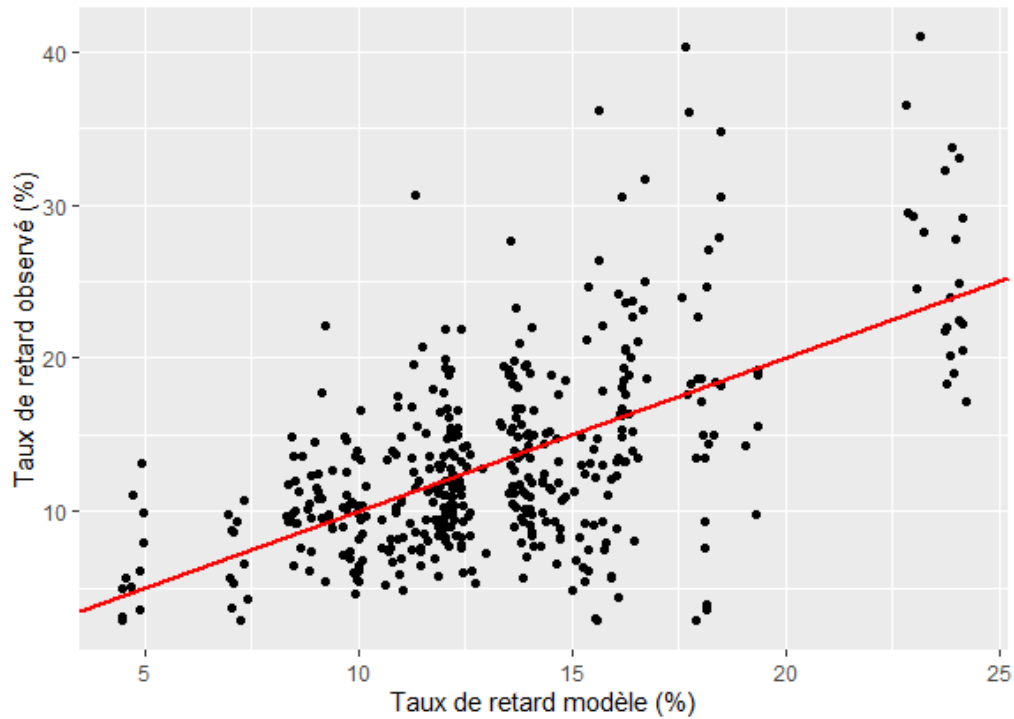
que les autres (+2,5 points de retard). Nous n'avons pas identifié à ce stade d'explication évidente à l'effet de ces variables sur la ponctualité des trains Intercités.

Enfin, le coefficient de détermination ajusté ou R^2 ajusté du modèle retenu est de 0,348, ce qui signifie que seulement 34,8% de la variance du taux de retard moyen annuel par liaisons Intercités sur la période 2012-2019 est expliquée par le modèle. Autrement dit, 65,2% de la variabilité du taux de retard ne peuvent pas être expliqués par les variables explicatives retenues dans le modèle.

En effet, le taux de retard des trains peut dépendre d'un très grand nombre de paramètres, et de nombreux événements ayant un impact significatif sur le taux de retard ne peuvent pas être intégrés au modèle :

- Les événements météorologiques majeurs (inondations et pluies intenses, tempêtes, épisodes neigeux et de froid, canicule et fortes chaleurs, etc.) ;
- Les mouvements sociaux ;
- Les autres événements externes au transport, dont notamment les actes de malveillance, les accidents de personne, les collisions des trains avec des véhicules, ou avec des animaux, ...)
- Les incidents liés au dysfonctionnements des infrastructures ferroviaires, par exemple les ruptures de caténaire ou les pannes de signalisation ;
- Les dysfonctionnements du matériel roulant, avec notamment les pannes du matériel moteur ;
- Les incidents voyageurs (malaise d'un voyageur, blocage involontaire des portes par un voyageur, etc.)

Le graphique croisant le taux de retard prédit par le modèle (en abscisse) et le taux de retard observé (en ordonnée) permet d'apprécier d'une autre manière la capacité du modèle à décrire la variabilité du taux de retard par liaison Intercités. La ligne rouge correspond à la droite $y = x$ sur laquelle tous les points doivent s'aligner dans un modèle parfait théorique. On constate que la distribution des points présente une légère tendance à suivre cette droite à mesure que le taux de retard prédit augmente, mais que de nombreux points s'écartent fortement de la droite. Cela confirme le caractère limité du pouvoir descriptif du modèle retenu face à la forte variabilité géographique et temporelle du taux de retard des liaisons Intercités.



3.2.3 Résultats pour les TER

3.2.3.1 Résultats détaillés des régressions par variables explicatives

Les résultats détaillés des modèles de régressions testés sont présentés en Annexe 7.

3.2.3.2 Synthèse des résultats

Le tableau ci-dessous énumère les variables explicatives pour lesquelles le modèle de régression linéaire utilisé n'est pas statistiquement fiable au sens où il ne passe pas le test de Fisher au seuil de significativité de 5%. On ne peut donc pas conclure sur l'effet de ces variables sur le taux de retard avec les données à notre disposition.

Variable explicative	Description de la variable	R ² ajusté régression	p-value test de Fisher de la régression	p-value test de Student variables explicatives
SiegesKmParHab	Nombre de sièges-km offerts par habitant	< 0,1%	> 10%	> 10%
VoyageurKmParHab	Nombre de voyageurs-km par habitant	< 0,1%	> 10%	> 10%
IntensitéUsageTER	Intensité quotidienne d'utilisation du réseau par les services régionaux conventionnés	< 0,1%	> 10%	> 10%
PartCatUIC_7_9_AV	Part des voies ferroviaires de catégories UIC 7 à 9 avec voyageurs (AV) dans le réseau hors LGV	1,6%	> 5%	> 5%
AgeRéseau_HorsLGV	Âge du réseau hors LGV	0,6%	> 10%	> 10%
AgeRéseauCatUIC_5_6	Âge des lignes ferroviaires des catégories UIC 5 à 6	0,9%	> 10%	> 10%
AgeRéseauCatUIC_7_9_AV	Âge des lignes ferroviaires des catégories UIC 7 à 9 AV	0,7%	> 10%	> 10%
PartPopInf10kmGareTER	Part de la population de chaque région résidant dans une commune située à moins de 10 km d'une gare régionale (TER ou Transilien) [en %]	1,7%	> 5%	> 5%
EffectifsPour100000trkm_2015	Ratio des effectifs TER pour 100 000 trains-km réalisés à fin 2015	< 0,1%	> 10%	> 10%
TxCircPpalesLiaisonsTraversant-NoeudRégional	Estimation du taux de circulations de trains TER traversant le nœud ferroviaire régional retenu sur le panel des liaisons les plus fréquentées par ancienne région (selon l'ART) en 2018	< 0,1%	> 10%	> 10%
AgeMoyen_MaterielRoulant	Estimation AQST de l'âge moyen du parc de matériel roulant en fin d'année	1,7%	> 5%	> 5%

Le tableau ci-dessous énumère les variables explicatives pour lesquelles le modèle de régression linéaire est significatif – il passe le test de Fisher au seuil de significativité de 5% - et pour lesquelles le signe de l'impact de la variable explicative sur le taux de retard est conforme à l'intuition, ou tout du moins ne s'oppose pas à l'intuition.

Variable explicative	Description de la variable	R ² ajusté régression	p-value test de Fisher de la régression	p-value test de Student variables explicatives	Impact sur le taux de retard
NbCirc_an_AQST	Nombre de circulations TER dans l'année selon la base AQST	2,7%	< 5%	< 5%	+0,008 pts / 1000 circulations TER
NbCirc_HPM	Nombre moyen de circulations quotidiennes à l'heure de pointe du matin (HPM)	4,1%	< 1%	< 1%	+3,4 pts / 100 circulations TER à l'HPM
NbCirc_HPS	Nombre moyen de circulations quotidiennes à l'heure de pointe du soir (HPS)	2,4%	< 5%	< 5%	+2,6 pts / 100 circulations TER à l'HPS
DistanceMoy	Distance moyenne parcourue par les trains TER	11%	< 0,1%	< 0,1%	+0,53 pts / 10 km
VoyKm_an	Nombre de passagers-km transportés (en millions) dans l'année	16%	< 0,1%	< 0,1%	+2,1 pts / milliard voy-km
NbPlacesMoy	Capacité d'emport moyenne des TER en nombre de sièges	2,6%	< 5%	< 5%	+0,62 pts / 100 sièges
TauxOccupationMoy	Taux d'occupation moyen des rames TER	6,3%	< 1%	< 1%	+1,8 pts / 10%
LongueurRéseau_HorsLGV	Longueur du réseau hors LGV	16%	< 0,1%	< 0,1%	+3,0 pts/ 1000 km de ligne
DensitéRéseau_HorsLGV	Densité du réseau hors LGV (kilomètre de ligne pour 100 km ²)	11%	< 0,1%	< 0,1%	- 0,66 pts/ 1km de ligne/100 km ² de superficie
PartCatUIC_5_6	Part des voies de catégories UIC 5 à 6 dans le réseau hors LGV	9,1%	< 0,1%	< 0,1%	- 0,66 pts / 10%
DensitéRéseau_CatUIC_5_6	Densité du réseau des lignes de catégories UIC 5 à 6 (kilomètre de voie pour 100 km ²)	12%	< 0,1%	< 0,1%	- 0,76 pts/ 1km de voie/100 km ² de superficie

DensitéRéseau_CatUIC_7_9_AV	Densité du réseau des lignes de catégories UIC 7 à 9 AV (kilomètre de voie pour 100 km ²)	18%	< 0,1%	< 0,1%	- 1,0 pt/ 1km de voie/100 km ² de superficie
NbLiaisonsInfrarégionales	Nombre de liaisons infrarégionales desservies par les services TER	14%	< 0,1%	< 0,1%	+2,1 pts/ 1000 liaisons
NbLiaisonsInterrégionales	Nombre de liaisons interrégionales desservies par les services TER	16%	< 0,1%	< 0,1%	+4,7 pts/ 1000 liaisons
NbCommunesDesservies	Nombre de communes desservies au sein de la Région par les services TER	16%	< 0,1%	< 0,1%	+2,2 pts/ 100 communes
NbMoyArrêts_NvllleReg_2018	Nombre moyen d'arrêts par ancienne région en 2018 (nouvelles régions administratives après la réforme territoriale de 2016)	51%	< 0,1%	< 0,1%	+1,5 pts/arrêt
NbMoyArrêts_2020_PpalesLiaisons	Nombre moyen d'arrêts effectués par les trains TER en 2020 sur le panel des liaisons les plus fréquentées par ancienne région (selon l'ART) en 2018	24%	< 0,1%	< 0,1%	+0,70 pts/arrêt
TMJA_max_TER_2018	Niveau du Trafic journalier moyen annuel (TJMA) des trains TER sur la section de ligne la plus intensément utilisée dans la région en 2018	3,9%	< 5%	< 5%	+2,1 pts / 100 circulations TER

3.2.3.3 Modèle de régression linéaire multiple

A partir de l'analyse des résultats des régressions linéaires par variable explicative, un modèle de régression linéaire multiple a été construit. Les variables explicatives pour lesquelles le modèle de régression linéaire simple ne passe pas le test de Fisher au seuil de significativité de 5% ont été exclus de la sélection des variables explicatives du modèle.

Certaines variables explicatives testées dans la section 3.2.3.2 présentent des colinéarités. Un tableau de synthèse des relations de colinéarité entre ces variables est fourni en Annexe 18. Pour s'affranchir des effets de colinéarité, les variables colinéaires avec l'une des variables explicatives retenues ont été exclues des modèles testés.

Différents modèles ont été testés à l'aide du logiciel R. Le modèle retenu est celui qui optimise un ensemble de critères statistiques de sélection – en particulier, le modèle minimise le critère d'information d'Akaike (AIC) présenté précédemment dans la section 3.2.1.3 ainsi que le critère d'information bayésien (BIC) – et pour lequel aucun signe des coefficients associés aux variables explicatives ne semble contre-intuitif ou inexplicable. Il s'agit du modèle suivant :

```
Call:
lm(formula = TxRetard ~ NbMoyArrêts_Nv11eReg_2018 + DistanceMoy +
    VoyKm_an + DensitéRéseau_CatUIC_7_9_AV, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-3.6934 -1.5131  0.0175  1.1919  5.5013

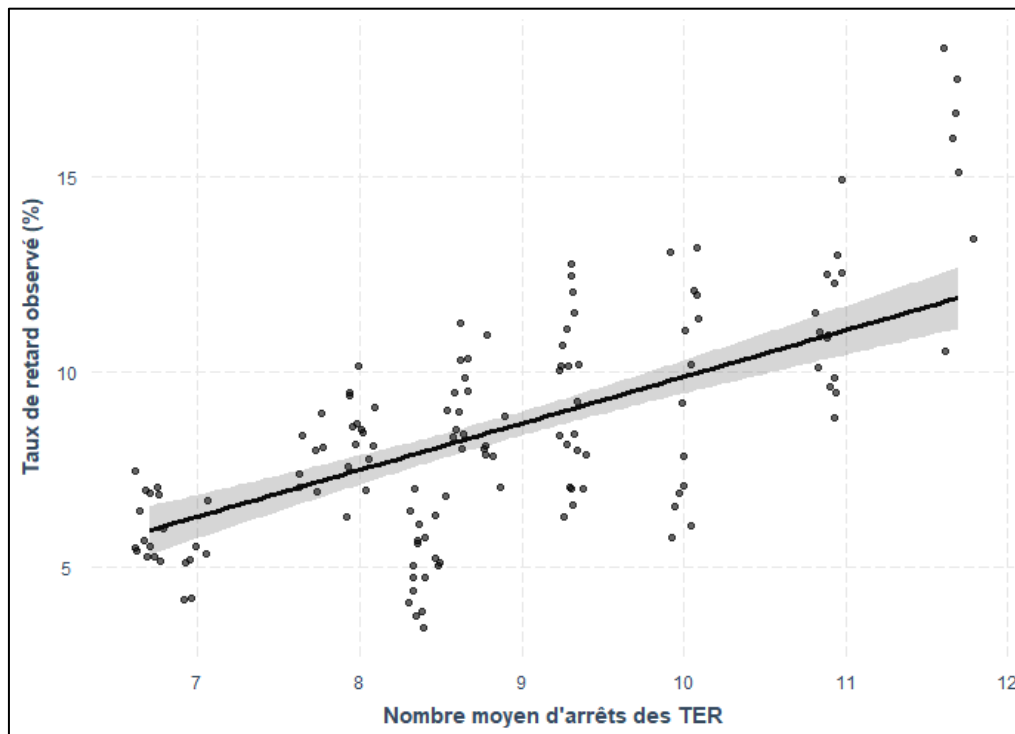
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -3.7310704   1.3581200   -2.747  0.006857 **
NbMoyArrêts_Nv11eReg_2018
  1.1909839   0.1278642    9.314 3.81e-16 ***
DistanceMoy
  0.0247197   0.0096825    2.553 0.011825 *
VoyKm_an
  0.0012298   0.0003152    3.902 0.000152 ***
DensitéRéseau_CatUIC_7_9_AV
 -0.4642732   0.1384980   -3.352 0.001048 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.809 on 131 degrees of freedom
Multiple R-squared:  0.6286,    Adjusted R-squared:  0.6173
F-statistic: 55.44 on 4 and 131 DF,  p-value: < 2.2e-16
```

D'après la p-value associée au test de Fisher sur la régression, le modèle retenu est significatif au seuil de 0,1%. D'après les p-values associées aux différentes variables explicatives du modèle, l'effet de la distance moyenne parcourue par les TER sur le taux de retard est significatif au seuil 2%, mais pas au seuil 1%. L'effet de la densité des voies UIC 7 à 9 avec voyageurs (AV) du réseau ferré régional est significatif au seuil 1%. L'effet sur le taux de retard des 2 autres variables explicatives - le nombre moyen d'arrêts effectués par les circulations TER en 2018, et le nombre voyageurs-km transportés dans l'année (en millions) - est significatif au seuil 0,1%.

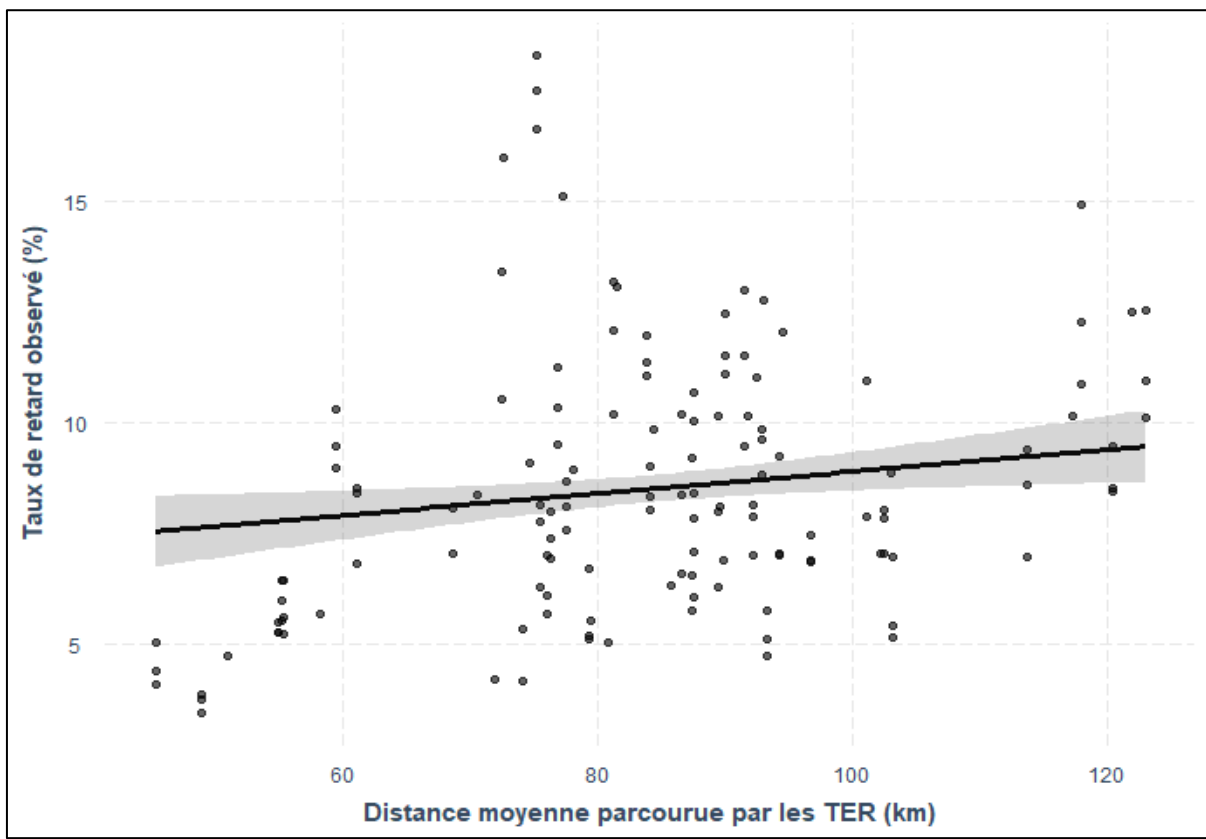
D'après le modèle retenu, chaque arrêt supplémentaire effectué par une circulation TER s'accompagne en moyenne d'un taux de retard plus élevé de 1,2 points. Pour expliquer cette corrélation, on peut supposer que le risque d'occurrence de retards est important lors d'un arrêt. En effet, certains événements peuvent survenir lors d'un arrêt en gare et retarder le départ du train : par exemple, malaise voyageur, blocage (volontaire ou non) des portes par un voyageur retardant le train, dysfonctionnement du matériel roulant lors de son redémarrage, ou parfois des infrastructures ferroviaires en sortie de gare (par exemple la signalisation), etc.

Le graphique ci-dessous permet d'observer l'effet du nombre d'arrêts des TER par ancienne région sur le taux de retard observé. La zone ombrée autour de la droite de régression correspond à la plage d'incertitude statistique sur cette droite. On note que la modélisation linéaire s'ajuste bien avec les moyennes des distributions régionales observées.



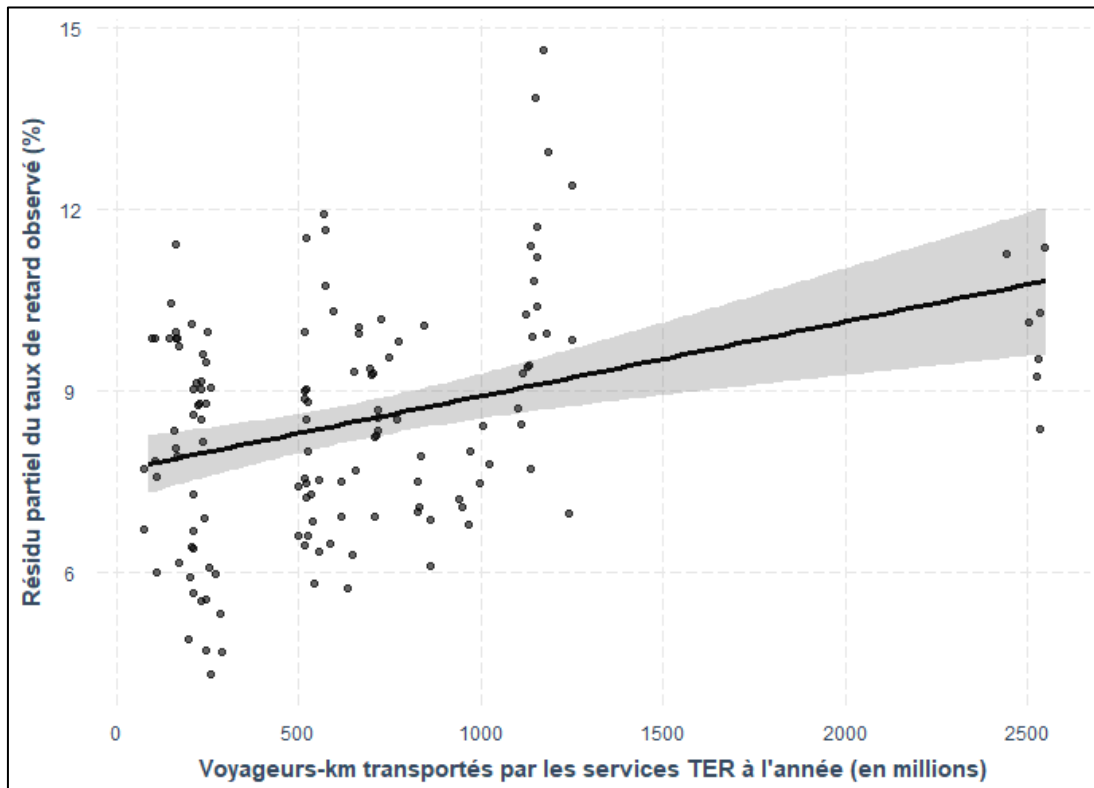
Le modèle suggère qu'en moyenne, un trajet en TER plus long de 10 km présente un taux de retard plus élevé de 0,25 points. Sur l'année 2016, l'étendue entre les régions au sein de la série de la distance moyenne parcourue par les TER est de 73 km, ce qui se traduit d'après le modèle par un impact d'environ 1,8 points de retard des TER entre la région où les TER parcourent le plus de distance (Languedoc-Roussillon) et celle où ils en parcourent le moins (Alsace). La corrélation observée s'explique probablement par le fait que plus un trajet en TER dure longtemps (à vitesse moyenne inchangée entre les trajets) plus il est soumis longtemps au risque d'occurrence d'un incident entraînant un retard.

Sur le graphique suivant, on note une tendance de hausse du taux de retard lorsque la distance moyenne qu'ils parcourent augmente. Cependant, la distribution du taux de retard est très dispersée autour de la moyenne à valeur donnée de la distance parcourue, de sorte qu'une modélisation par une droite est très imparfaite pour cette variable. Ceci explique pourquoi la statistique de Student pour cette variable ne passe pas le test au seuil 1% mais seulement au seuil 5%.



Le modèle indique également qu’une hausse de 1 milliard de voyageurs-km TER entraîne une hausse du taux de retard de 1,2 points environ. Sur l’année 2016, l’étendue entre les régions est de 2,46 milliards de voy-km transportés, ce qui se traduit d’après le modèle par une différence d’environ 3,0 points de retard des TER entre la région qui en comptabilise le plus (Rhône-Alpes) et celle qui en comptabilise le moins (Limousin). Au sein de la série, l’écart-type est de 568 millions de voy-km transportés, se traduisant par une différence de taux de retard de 0,70 points en moyenne d’après le modèle.

Le fait que la moyenne de la distribution du taux de retard observé augmente à mesure que le volume de voyageurs-km transportés augmente est visible sur le graphique ci-après. En axe des ordonnées, le résidu partiel du taux de retard observé est représenté. Il s’agit en fait du taux de retard observé auquel on retranche l’effet des autres variables de la régression, afin d’identifier plus facilement graphiquement la relation de corrélation pour cette variable spécifiquement. On observe une forte dispersion du résidu partiel autour de sa valeur moyenne. La zone ombrée autour de la droite de régression correspond à la plage d’incertitude statistique qui résulte de cette dispersion. On remarque que cette incertitude est importante pour les valeurs les plus élevées du volume de voyageurs-km transportés. Néanmoins, les moyennes du taux de retard des « paquets » observés au sein de la distribution s’ajustent assez bien au modèle linéaire.



Pour expliquer l'impact sur le taux de retard du nombre de voyageurs-km effectués à l'année par les TER, on peut commencer par décomposer le nombre de voyageurs-km comme le produit du nombre de circulations de TER à l'année par le nombre moyen de voyageurs par train, que multiplie la distance moyenne parcourue par les TER.

L'analyse de colinéarité entre les variables, effectuée à l'aide du logiciel R, montre un niveau élevé de colinéarité avec le nombre de circulations de TER, que ce soit à l'année, par jour au niveau de la section dimensionnant du réseau en termes de circulations de TER, ou à l'heure de pointe du matin ou du soir. Le niveau de colinéarité est en revanche bien plus faible avec le taux d'occupation des trains, comme avec la distance moyenne parcourue par les TER. En conséquence, il n'y a pas de problème de colinéarité entre les variables explicatives du modèle que sont le nombre de voyageurs-km réalisés par les TER et la distance moyenne parcourue par les TER.

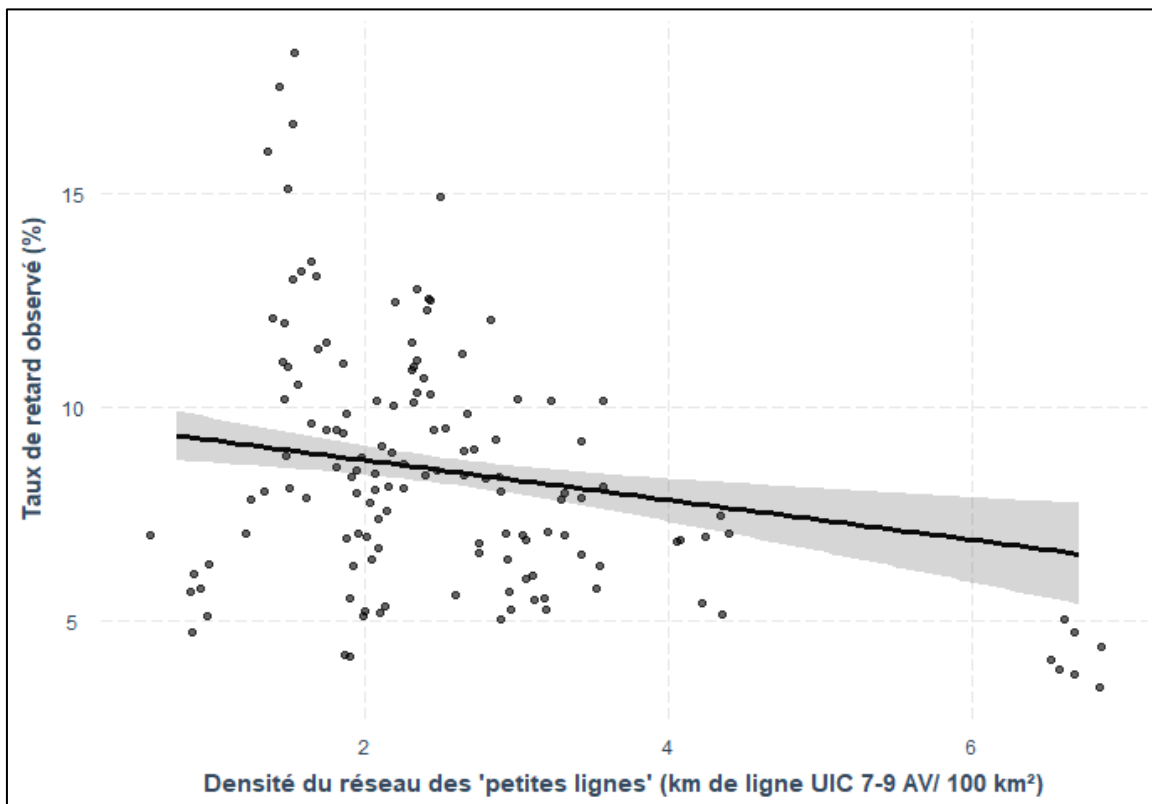
D'après l'analyse de la modélisation par variables explicatives, les 3 composantes intervenant dans la décomposition du nombre de voyageurs-km à l'année ont un impact de même signe sur le taux de retard :

- Le nombre de circulations de TER dans la région à l'année : on peut supposer que cette variable puisse engendrer des retards pour diverses raisons. Lorsqu'une circulation TER emprunte des lignes plus fréquentées, elle est davantage exposée au risque de subir un retard qui se propage sur le réseau d'une circulation de train à l'autre. Les infrastructures ferroviaires (notamment les voies et la signalisation) et le matériel roulant se détériorent à mesure qu'ils sont intensément utilisés, et leur usure augmente le risque de panne engendrant des retards ;
- Le taux d'occupation des trains : on peut supposer que lorsque des rames TER transportent plus de voyageurs, elles sont davantage susceptibles de connaître des retards liés aux

voyageurs - par exemple, malaise voyageur, blocage (volontaire ou non) des portes par un voyageur retardant le train, etc. ;

- La distance moyenne parcourue : plus un trajet en TER dure longtemps (à vitesse moyenne inchangée entre les trajets), plus il est soumis longtemps au risque d'occurrence d'un incident entraînant un retard.

Enfin, le modèle retenu suggère qu'une hausse de 1 km de voie des catégories UIC 7 à 9 (AV)/100 km² de superficie - variable caractérisant la densité des « petites lignes » sur la superficie régionale, s'accompagne d'une baisse du taux de retard de 0,46 point environ. Sur l'année 2016, l'étendue de la densité de ces lignes au sein de la série est de 5,2 km de voie UIC 7-9 (AV)/100 km². Cela se traduit, selon le modèle, par une différence de taux de retard de 2,4 points entre la région Centre qui présente le réseau le moins dense en « petites lignes », et la région Alsace, qui présente le réseau le plus dense pour cette catégorie de lignes ferroviaires. Le graphique ci-dessous illustre l'effet identifié de la densité des « petites lignes » sur le taux de retard des TER. Malgré une importante dispersion du taux de retard autour de sa moyenne, l'effet est statistiquement significatif au seuil 1%.



Le signe de la corrélation indique que selon le modèle, plus le réseau ferroviaire des « petites lignes » est dense, meilleure est la ponctualité des TER. Un élément d'explication pourrait être qu'un réseau ferroviaire plus dense s'accompagnerait d'une répartition du trafic plus homogène sur les différentes « petites lignes » du réseau, et permettrait peut-être de limiter en partie la congestion des circulations de TER au niveau des sections du réseau les plus fréquentées. Or cette congestion accroît le risque d'incidents générant des retards. On pourrait de la même façon imaginer qu'un réseau de « petites lignes » moins dense concentre davantage le trafic des TER sur quelques lignes ferroviaires. On pourrait alors supposer que les conséquences de la propagation d'un éventuel « gros retard » d'un

TER à d'autres circulations TER soient moins lourdes en présence d'un réseau de « petites lignes » dense qui offrirait éventuellement des possibilités de déviation et des trajets alternatifs entre plusieurs gares.

Enfin, le coefficient de détermination ajusté ou R^2 ajusté du modèle retenu est de 0,617, ce qui signifie que 61,7% de la variance du taux de retard moyen annuel des TER par ancienne région administrative sur la période 2013-2019³ sont expliqués par le modèle.

On peut remarquer que ce chiffre est nettement plus élevé (près du double) que celui obtenu avec les modèles retenus pour les trains longue distance (TGV et Intercités). On peut émettre différentes hypothèses permettant d'expliquer la différence constatée :

- Des variables explicatives importantes n'auraient pas pu être prises en compte dans la modélisation du taux de retard des TGV et des trains Intercités ;
- Le modèle de régression linéaire multiple serait mieux adapté pour décrire la distribution du taux de retard des trains régionaux que celle du taux de retard des trains longue distance ;
- Le modèle de régression linéaire multiple serait mieux adapté pour décrire la distribution du taux de retard des trains d'un ensemble de liaisons agrégées (les régions dans le cas des TER) que celle du taux de retard des trains par liaison.

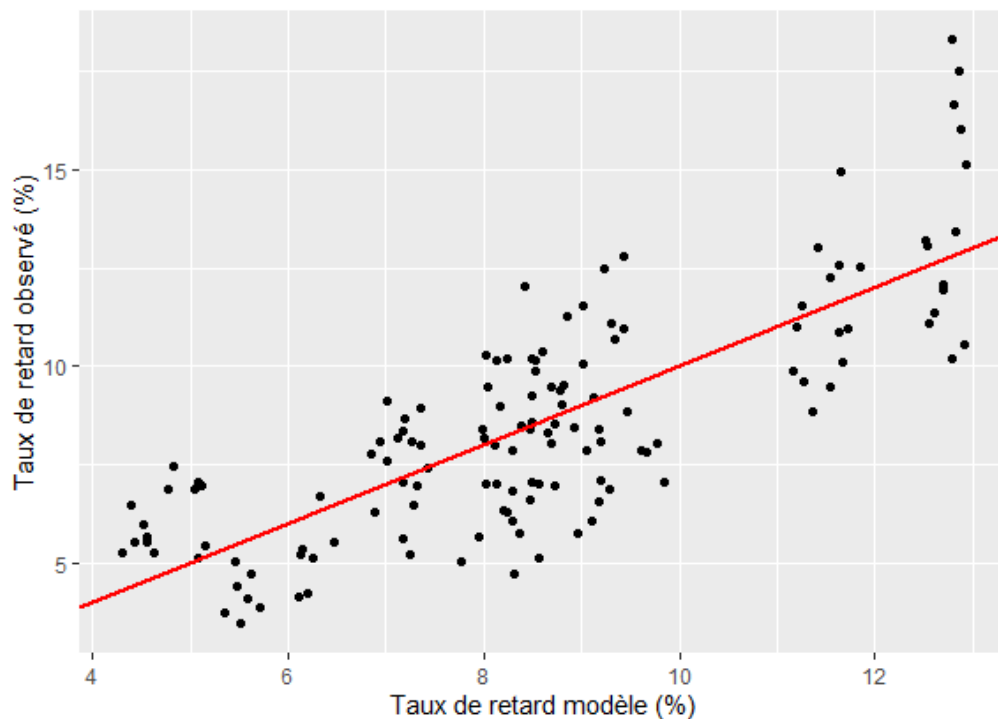
On rappelle que le taux de retard des trains peut dépendre d'un très grand nombre de paramètres, et que de nombreux événements ayant un impact significatif sur le taux de retard ne sont pas intégrés au modèle :

- Les événements météorologiques majeurs (inondations et pluies intenses, tempêtes, épisodes neigeux et de froid, canicule et fortes chaleurs, etc.) ;
- Les mouvements sociaux ;
- Les autres événements externes au transport, dont notamment les actes de malveillance, les accidents de personne, les collisions des trains avec des véhicules, ou avec des animaux, ...)
- Les incidents liés aux dysfonctionnements des infrastructures ferroviaires, par exemple les ruptures de caténaire ou les pannes de signalisation. Le risque d'occurrence de ces incidents est cependant pour partie lié à l'âge des lignes ferroviaire du réseau ferré, dont l'effet sur le taux de retard des TER n'est pas apparu comme statistiquement significatif sur la période d'étude à l'issue de la modélisation ;
- Les dysfonctionnements du matériel roulant, avec notamment les pannes du matériel moteur. Le risque d'occurrence de ces incidents est toutefois partiellement lié à l'âge du parc de matériel roulant des TER, dont l'effet sur le taux de retard des TER n'est pas apparu comme statistiquement significatif sur la période d'étude à l'issue de la modélisation ;
- Les incidents voyageurs (malaise d'un voyageur, blocage involontaire des portes par un voyageur, etc.), dont le risque d'occurrence augmente possiblement avec le taux

³ Pour disposer d'une information plus détaillée, le découpage des régions administratives d'avant la réforme territoriale entrée en vigueur le 1er janvier 2016 a été utilisé. Une estimation du taux de retard des TER par ancienne région administrative pour les années 2017 (pour certaines régions), 2018 et 2019 a été effectuée. Elle est présentée en annexe 1.

d'occupation des trains. Cette variable présente un effet significatif sur le taux de retard des TER sur la période d'étude d'après l'analyse par variable explicative.

Le graphique croisant le taux de retard prédit par le modèle (en abscisse) et le taux de retard observé (en ordonnée) permet d'apprécier d'une autre manière la capacité du modèle à décrire la variabilité du taux de retard des TER par ancienne région. La ligne rouge correspond à la droite $y = x$ sur laquelle tous les points doivent s'aligner dans un modèle parfait théorique. On constate que la distribution des points s'aligne assez bien sur cette droite, mais que les points s'écartent de la droite pour les valeurs les plus élevées du taux de retard. Le graphique illustre le plus grand pouvoir descriptif du modèle retenu pour les TER en comparaison à celui des modèles retenus pour les TGV et les services Intercités. Cette différence notable se traduit par une valeur du R^2 ajusté bien plus élevée.



3.2.4 Résultats pour les RER et Transilien

3.2.4.1 Résultats détaillés des régressions par variables explicatives

Les résultats détaillés des modèles de régressions testés pour expliquer le taux de retard des lignes de RER et Transilien **par ligne** sont présentés en Annexe 7.

3.2.4.2 Synthèse des résultats

Le tableau ci-dessous énumère les variables explicatives pour lesquelles le modèle de régression linéaire utilisé n'est pas statistiquement fiable au sens où il ne passe pas le test de Fisher au seuil de significativité de 5%. On ne peut donc pas conclure sur l'effet de ces variables sur le taux de retard **par ligne** avec les données à notre disposition.

Variable explicative	Description de la variable	R ² ajusté régression	p-value test de Fisher de la régression	p-value test de Student variables explicatives
NbGares	Nombre total de points de gares/stations sur la ligne	3%	> 5%	> 5%
DistanceMoyEntreArrêts	Distance moyenne entre 2 points d'arrêts de la ligne	2,1%	> 5%	> 5%
NbCircJour_2019	Nombre de trains par jour en 2019	2,8%	> 5%	> 5%
NbBranches	Nombre de branches	1,2%	> 10%	> 10%
IntervalleMin_2019	Intervalle minimal à l'hyperpointe du matin en 2019	< 0,1%	> 10%	> 10%
VoiesPartagées	Présence de sections de ligne partagées	< 0,1%	> 10%	> 10%
NbMoyArrêtsHPMdansParis-2019	Nombre d'arrêts moyen dans Paris intra-muros à l'heure de pointe du matin durant un JOB	2,2%	> 5%	> 5%
DuréeMoyInterstation_2019	Durée moyenne entre 2 arrêts à l'heure de pointe du matin durant un JOB	2,5%	> 5%	> 5%
GareDuNord	La ligne contient la station "Gare du Nord"	< 0,1%	> 10%	> 10%
GareStLazare	La ligne contient la station "Gare Saint-Lazare"	< 0,1%	> 10%	> 10%
TMJA_Tot_ParVoie_Moy	Estimation de la moyenne du Trafic Moyen Journalier Annuel (TMJA) par voie pour l'ensemble des circulations de trains (y compris fret et "haut le pied")	< 0,1%	> 10%	> 10%

Le tableau ci-dessous énumère la variable explicative pour laquelle le modèle de régression linéaire est significatif – il passe le test de Fisher au seuil de significativité de 5% - mais pour lequel le signe de l'impact de la variable explicative sur le taux de retard **par ligne** est contraire à l'intuition.

Variable explicative	Description de la variable	R ² ajusté régression	p-value test de Fisher de la régression	p-value test de Student variables explicatives	Impact sur le taux de retard
TypeContrôleVitesse	Type de contrôle de vitesse utilisé depuis 2015	15,3%	< 0,1%	< 0,1%	+1,9 pts / amélioration du contrôle de vitesse

Une explication probable est que la modernisation du contrôle de vitesse a été rendue prioritaire sur les lignes les moins ponctuelles, sans que cette amélioration ne parvienne à faire de ces lignes les plus ponctuelles du réseau puisqu'il s'agit des lignes RER sur lesquelles pèsent le plus de contraintes d'exploitation.

Le tableau ci-dessous énumère les variables explicatives pour lesquelles le modèle de régression linéaire est significatif – il passe le test de Fisher au seuil de significativité de 5% - et pour lesquelles le signe de l'impact de la variable explicative sur le taux de retard **par ligne** est conforme à l'intuition, ou tout du moins ne s'oppose pas à l'intuition.

Variable explicative	Description de la variable	R ² ajusté régression	p-value test de Fisher de la régression	p-value test de Student variables explicatives	Impact sur le taux de retard
LongueurLigne	Longueur totale de la ligne	6,4%	< 1%	< 1%	+1,3 pts / 100 km
NbVoyageursJour	Estimation de la fréquentation quotidienne des lignes	7,1%	< 1%	< 1%	+2,9 pts / million voy quotidiens
TrKm	Offre en trains-km réalisés	13,3%	< 0,1%	< 0,1%	+0,42 pt / million trains-km
VoyKm	Nombre de voy-km transportés	10,3%	< 1%	< 1%	+0,57 pt / milliard voy-km
NbTrainsHPM_2019	Nombre de trains par sens à l'Heure de Pointe du Matin (HPM) à l'interstation la plus chargée	5,4%	< 5%	< 5%	+1,1 pts / 10 circulations RER à l'HPM/sens
AgeMoyen_MaterielRoulant-1erJanvier	Estimation de l'âge moyen du matériel roulant en début d'année	15%	< 0,1%	< 0,1%	+1,5 pts / 10 ans
NbMoyArrêtsHPM_2019	Nombre d'arrêts moyen à l'heure de pointe du matin durant un JOB	7,1%	< 1%	< 1%	+0,18 pts/arrêt
TpsParcoursMoyHPM_2019	Durée moyenne du parcours total des trains à l'heure de pointe du matin durant un JOB	39,6%	< 0,1%	< 0,1%	+1,8 pts / 10 minutes
GareDeLyon	La ligne contient la station "Gare de Lyon"	24,4%	< 0,1%	< 0,1%	+4,1 pts
ChâteletLesHalles	La ligne contient la station "Châtelet - Les Halles"	18,9%	< 0,1%	< 0,1%	+3,7 pts
GareParisEst	La ligne contient la station "Gare de l'Est"	5,2%	< 5%	< 5%	-2,4 pts
DistanceMoyTrain_2019	Distance moyenne parcourue par les trains de la ligne en 2019	18,4%	< 0,1%	< 0,1%	+1,0 pts / 10 km
TMJA_Tot_ParVoie_Max	Estimation du Trafic Moyen Journalier Annuel (TMJA) par voie sur la section la plus chargée de la ligne pour l'ensemble des circulations de trains (y compris fret et "haut le pied")	13,8%	< 0,1%	< 0,1%	+1,7 pts / 100 trains/jour

3.2.4.3 Modèle de régression linéaire multiple retenu pour expliquer le taux de retard par ligne

A partir de l'analyse des résultats des régressions linéaires par variable explicative, un modèle de régression linéaire multiple a été construit pour expliquer le taux de retard des RER et Transilien **par ligne**. Les variables explicatives pour lesquelles le modèle de régression linéaire simple ne passe pas le test de Fisher au seuil de significativité de 10% ont été exclus de la sélection des variables explicatives du modèle.

Certaines variables explicatives testées dans la section 3.2.4.2 présentent des colinéarités. Un tableau de synthèse des relations de colinéarité entre ces variables est fourni en Annexe 23. Pour s'affranchir des effets de colinéarité, les variables colinéaires avec l'une des variables explicatives retenues ont été exclues des modèles testés.

Différents modèles ont été testés à l'aide du logiciel R. Le modèle retenu est celui qui optimise un ensemble de critères statistiques de sélection et pour lequel aucun signe des coefficients associés aux variables explicatives ne semble contre-intuitif ou inexplicable. Il s'agit du modèle suivant :

```
Call:
lm(formula = TxRetard ~ +TMJA_Tot_ParVoie_Max + DistanceMoyTrain_2019 +
    AgeMoyen_MaterielRoulant_1erJanvier + GareDeLyon + GareStLazare +
    GareDuNord + GareParisEst, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-4.7175 -1.4897 -0.1622  1.0475  6.2767

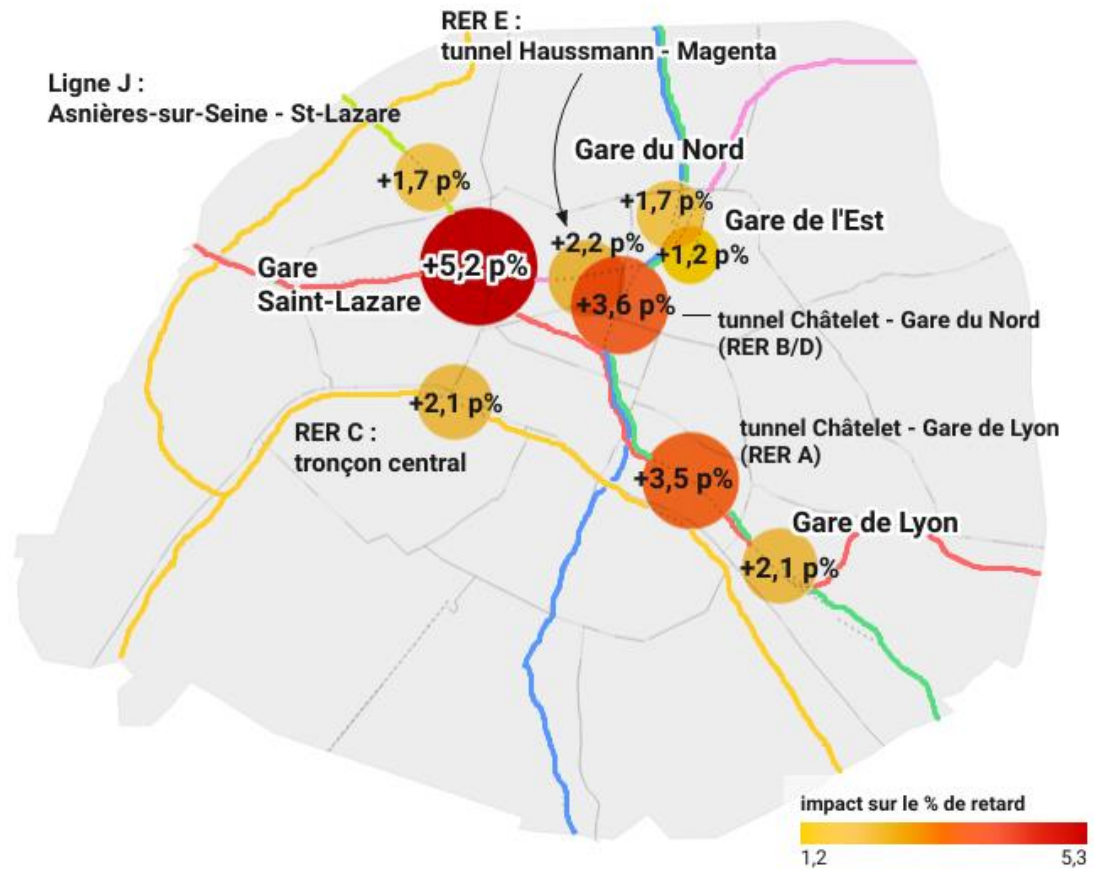
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.744333   1.273050  -0.585  0.56035
TMJA_Tot_ParVoie_Max
0.010989   0.003781   2.907  0.00468 **
DistanceMoyTrain_2019
0.111010   0.021098   5.262 1.10e-06 ***
AgeMoyen_MaterielRoulant_1erJanvier
0.123266   0.025393   4.854 5.60e-06 ***
GareDeLyon     2.139459   0.862388   2.481  0.01513 *
GareStLazare   5.317582   0.793394   6.702 2.31e-09 ***
GareDuNord     1.684449   0.558527   3.016  0.00340 **
GareParisEst   1.209996   0.755191   1.602  0.11290
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.12 on 83 degrees of freedom
Multiple R-squared:  0.6587,    Adjusted R-squared:  0.6299
F-statistic: 22.88 on 7 and 83 DF,  p-value: < 2.2e-16
```

D'après la p-value associée au test de Fisher sur la régression, le modèle retenu est significatif au seuil de 0,1%. D'après les p-values associées aux différentes variables explicatives du modèle, le fait que la ligne passe ou non par la gare de l'Est n'est pas significatif au seuil de 10%. Le fait que la ligne passe ou non par la gare de Lyon est significatif au seuil de 5%. Le niveau du trafic moyen annuel journalier par voie sur la section de plus fort trafic, et le fait qu'elle passe ou non par la gare du Nord sont significatifs au seuil de 1%. L'effet des autres variables explicatives retenues sur le taux de retard est significatif au seuil de 0,1%.

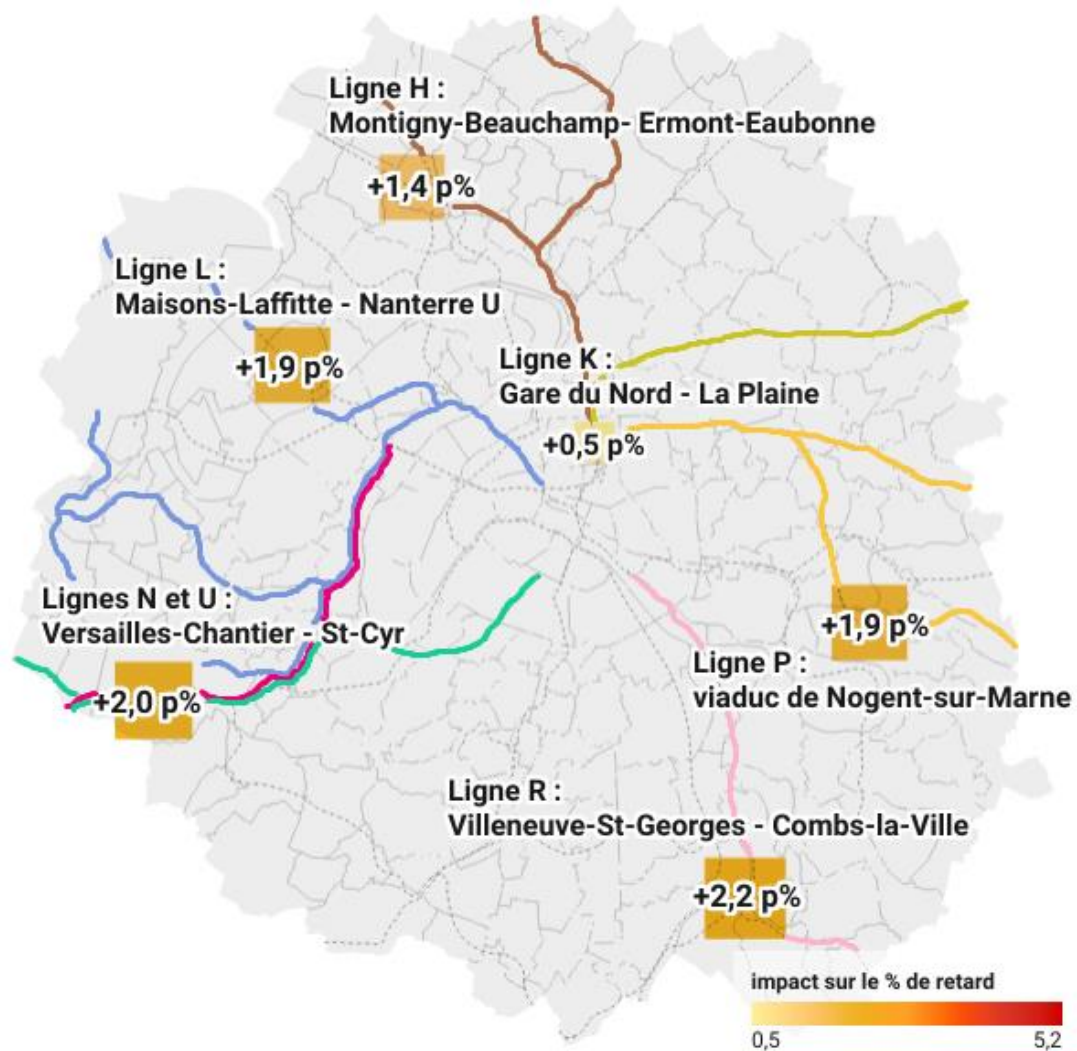
Les cartes ci-dessous illustrent l'effet sur la ponctualité des RER et Transilien du passage par certaines grandes gares parisiennes et par le tronçon où le trafic (toutes activités ferroviaires confondues) est le plus intense pour chaque ligne.

Estimation économétrique de l'effet du passage par certaines grandes gares parisiennes et certains tronçons sur la ponctualité des RER et Transilien



Carte: Alexandre Barbusse/AQST • Créé avec Datawrapper

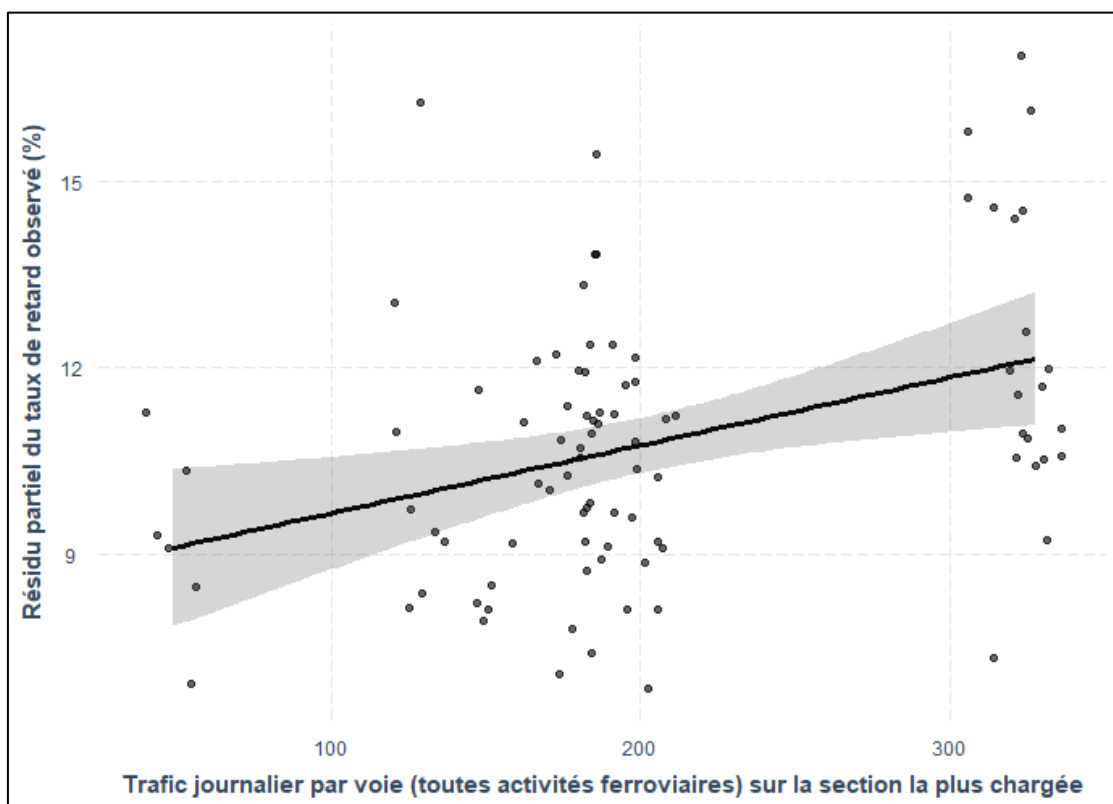
Estimation économétrique de l'effet du passage par certains tronçons sur la ponctualité des Transilien



Carte: Alexandre Barbusse/AQST • Créé avec Datawrapper

D'après le modèle retenu, la circulation de 100 trains/jour/voie supplémentaires – toutes circulations ferroviaires prises en compte (y compris fret et « haut-le-pied ») - au niveau de la section la plus chargée de la ligne, se traduit en moyenne par un taux de retard plus élevé de 1,1 point. L'effet de cette variable sur le taux de retard est cohérent avec l'intuition selon laquelle lorsqu'une voie connaît un trafic plus intense, la ponctualité des lignes qui l'utilisent se dégrade car elles sont davantage exposées au risque de subir un retard se propageant sur cette voie d'une circulation de train à l'autre. On observe par ailleurs que les sections les plus saturées identifiées par ligne (annexe 22) correspondent à celles désignées dans les documents publics comme posant d'importantes difficultés d'exploitation pour la majorité des lignes : la section centrale du RER A entre Gare de Lyon et La Défense, le tunnel entre Châtelet – Les Halles et la gare du Nord pour les lignes RER B et D [6],

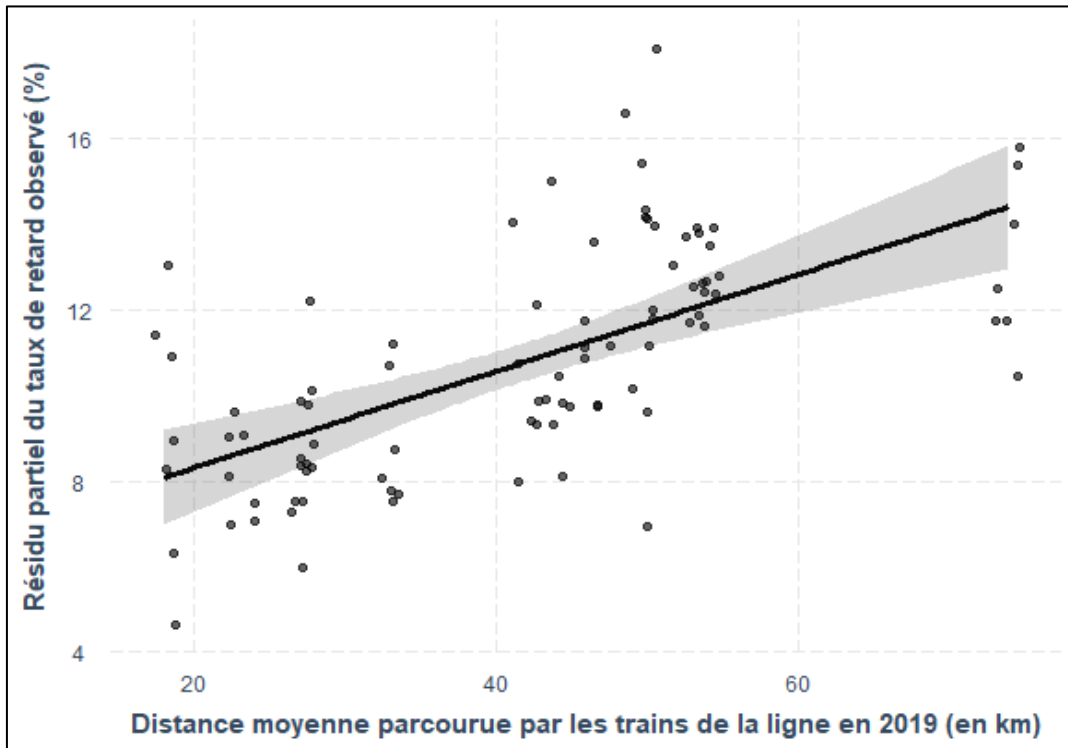
le tronçon central du RER C entre Champ de Mars-Tour Eiffel et la gare d'Austerlitz, le tunnel entre les gares d'Hausmann – Saint-Lazare et Magenta sur la ligne RER E, les deux voies du groupe V en sortie de la gare Saint-Lazare, utilisées par les lignes Transilien L et J ainsi que des trains Grandes Lignes [10], les deux voies de la ligne R entre la gare de Lyon et Créteil-Pompadour - également empruntées par des TER Bourgogne-Franche-Comté, des trains Intercités ainsi que trains de fret – la section à deux voies entre Saint-Cloud et La Défense exploitée par les Transilien L et U ou encore le rétrécissement de quatre à deux voies au niveau du viaduc de Nogent-sur-Marne pour le RER E et la ligne P.



Le graphique ci-dessus permet d’observer l’effet du trafic par voie au niveau de la section la plus chargée de la ligne RER ou Transilien sur le taux de retard observé. En axe des ordonnées, le résidu partiel du taux de retard observé est représenté. Il s’agit en fait du taux de retard observé auquel on retranche l’effet des autres variables de la régression, afin d’identifier plus facilement graphiquement la relation de corrélation pour cette variable spécifiquement. On note une forte dispersion des points autour de la droite de régression du taux de retard. Il en résulte une incertitude statistique importante sur les paramètres de cette droite, matérialisée par la zone ombrée autour de la droite.

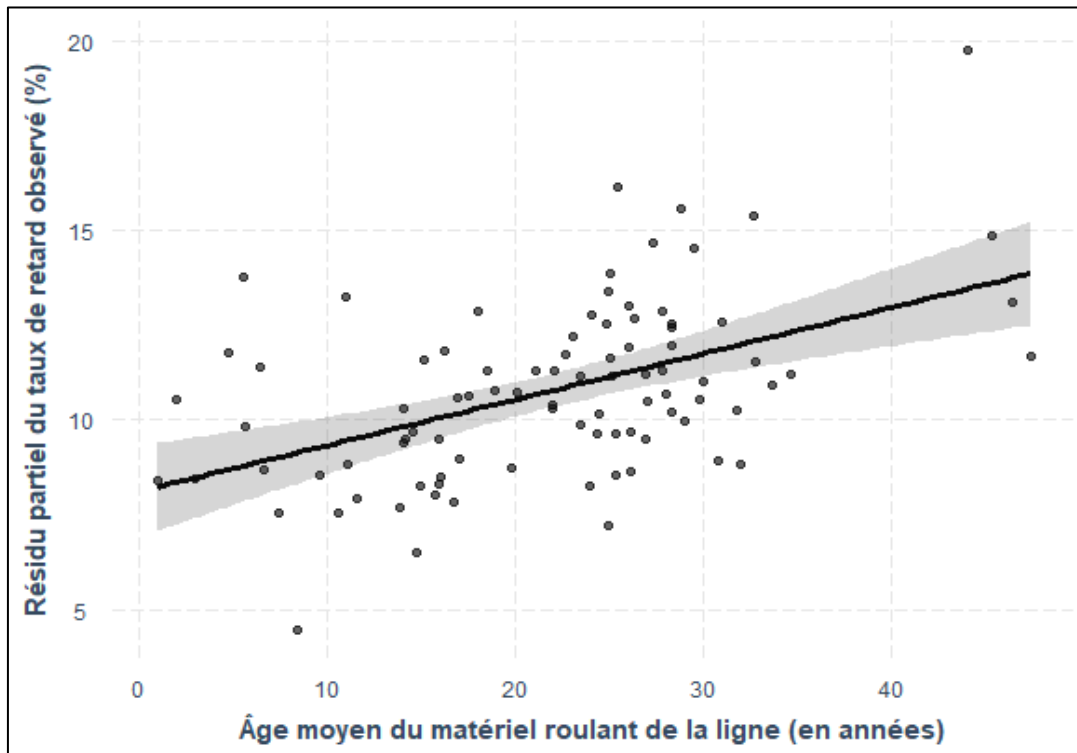
Le modèle suggère qu’en moyenne, un trajet en RER ou Transilien plus long de 10 km présente un taux de retard plus élevé de 1,2 points. Sur l’année 2019, l’étendue entre lignes de la distance moyenne parcourue par les RER ou Transilien vaut 56 km, ce qui se traduit d’après le modèle par un différence d’environ 6,6 points de retard entre la ligne RER ou Transilien pour laquelle les trajets sont les plus longs en moyenne (R) et celle pour laquelle ils sont les plus courts en moyenne (L). Au sein de la série, l’écart-type est d’environ 16 km, se traduisant par une différence de taux de retard de 1,8 points en moyenne d’après le modèle. La corrélation observée s’explique probablement par le fait que plus un trajet en RER ou Transilien dure longtemps, plus il est soumis longtemps au risque

d'occurrence d'un incident entraînant un retard. Le graphique suivant permet de constater que la distribution du résidu partiel du taux de retard observé s'ajuste assez bien avec un modèle linéaire.



Le modèle indique également qu'en moyenne, un parc de matériel roulant sur la ligne plus âgé de 10 ans s'accompagne d'un taux de retard annuel plus élevé de 1,2 points, ce qui est conforme à l'intuition. En effet, un matériel roulant âgé présente un risque de panne ou d'incident matériel plus important. Il est donc susceptible de connaître davantage de retard en moyenne. L'annexe 21 présente les données utilisées pour la variable de l'âge moyen du matériel roulant en début d'année par ligne RER ou Transilien. On peut remarquer une nette rupture entre 2016 et 2017 pour la ligne K. En effet, l'intégralité du matériel roulant de cette ligne a été remplacé le 2 septembre 2016 (remplacement de rames réversibles type RIB par des rames Francilien Z 50000). On pourrait alors imaginer que l'effet de l'âge du matériel roulant sur la ponctualité par ligne soit biaisé par cette rupture, et qu'il soit non significatif en retirant la ligne K de l'échantillon étudié. Nous avons testé cette éventualité, et le test statistique montre que l'effet de l'âge du matériel roulant reste significatif au seuil 1% et de même signe (résultats en annexe 24). Il est en revanche légèrement atténué en retirant la ligne K (+1,0 point de retard/vieillesse de 10 ans du parc de matériel roulant, contre +1,2 points en intégrant la ligne K).

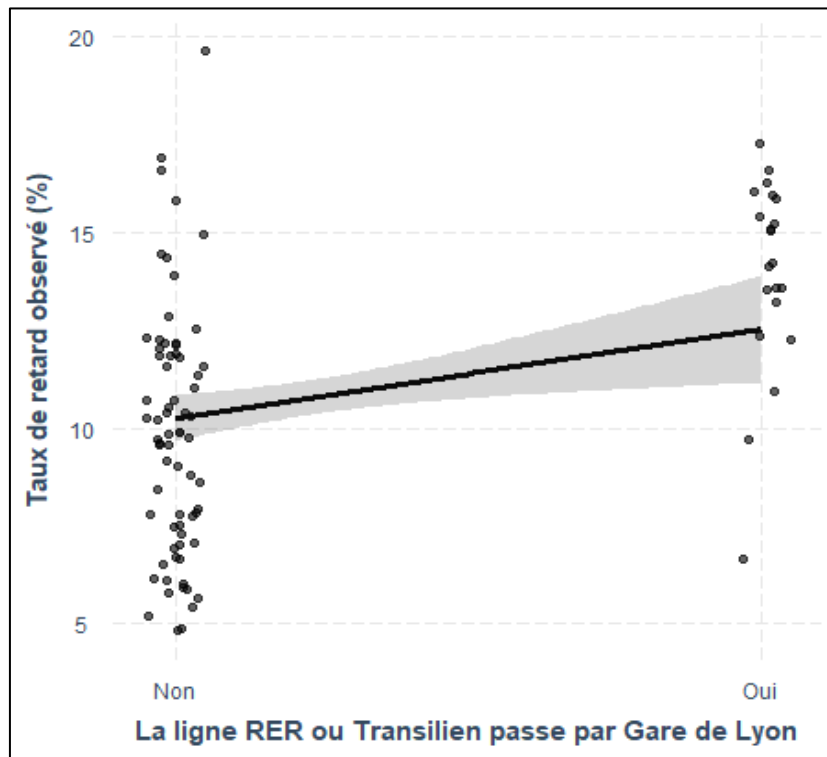
On note en visualisant le graphique ci-après que la dispersion du résidu partiel du taux de retard observé autour de la droite des moindres carrés ordinaires est relativement limitée par rapport aux variables précédentes, ce qui se traduit par une significativité statistique élevée du coefficient associé à l'âge moyen du matériel roulant dans la régression linéaire multiple.



Enfin, le modèle retenu suggère que les grandes gares parisiennes concentrent une grande part des causes de non-punctualité. Ce résultat rejoint le constat effectué par l'EPFL lors de l'analyse de base de données BRÉHAT (Base des Résultats de l'Exploitation Habiles à d'Autres Tâches) - la base utilisée par la Direction des Circulations Ferroviaires de la SNCF pour recenser et documenter tous les événements circulations. L'EPFL constate que « Les grandes gares parisiennes concentrent un nombre très important d'Incidents Origine » [6], incidents cause du retard ou cause du retard ayant provoqué la cascade de retards. L'IAU Île-de-France indique également que 4 des 6 grandes gares de surface parisiennes sont saturées ou bien fonctionnent en limite de capacité [10].

Ainsi, en moyenne, le taux de retard des lignes RER ou Transilien empruntant la Gare de Lyon est supérieur de 2,1 points à celui des autres lignes. La moins bonne punctualité de ces lignes semble en accord avec la qualification de « nœud ferroviaire majeur » donnée à la gare de Lyon par le CGEDD [7], le phénomène de saturation constaté par le CGEDD au niveau du secteur de la gare de Lyon [8], et les contraintes d'exploitation des lignes RER D et Transilien R entraînées par le partage des voies avec des trains TER Bourgogne-Franche-Comté, Intercités de la ligne Paris - Clermont-Ferrand, ainsi que des trains de fret, évoquées par RFF et la SNCF dans le document de préparation au schéma directeur de la ligne Transilien R [9]. Depuis des lignes ferroviaires en provenance du Sud, la ligne R du Transilien dessert la gare de surface et le RER D une gare souterraine également desservie par le RER A au sein d'un tronçon central en souterrain entre Vincennes et la Défense. L'IAU Île-de-France indique que « la gare souterraine de passage fonctionne correctement » [10]. Néanmoins, la saturation de la gare de surface impacte probablement la punctualité des lignes D et R au niveau des voies partagées en direction du Sud en sortie de la gare. Dans son étude des grandes gares parisiennes, l'IAU Île-de-France évoque deux pistes pouvant contribuer à la désaturation de la gare de Lyon : « La gare de Lyon a la possibilité de créer encore deux voies à quai supplémentaires, pour faire face au développement du TGV Sud-est, et de reporter des trains sur la gare d'Austerlitz lorsque le sextuplement des voies sera réalisé entre Paris et Juvisy » [10].

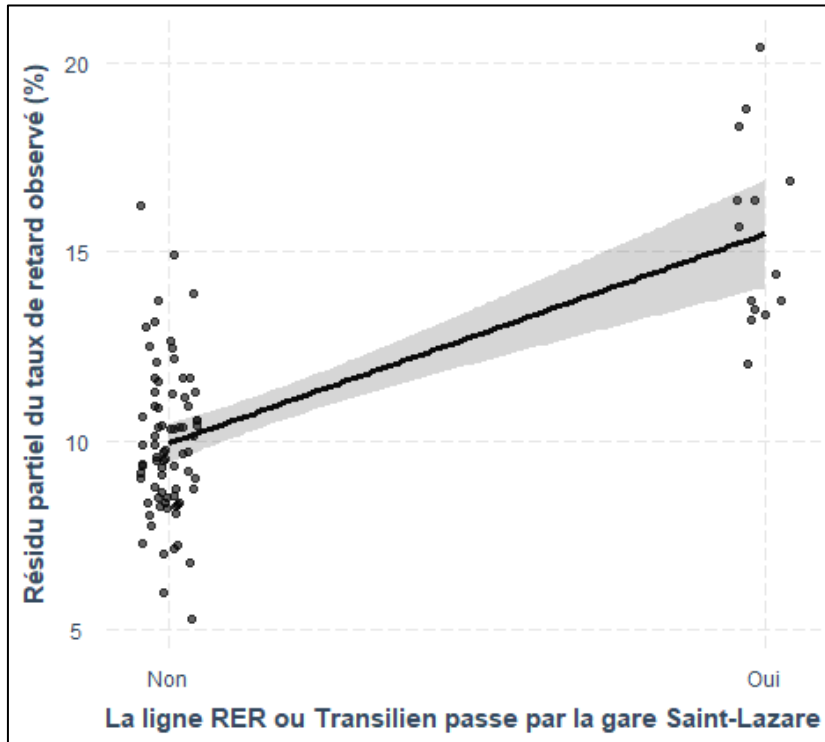
Le graphique ci-dessous illustre la divergence entre les moyennes des taux de retard observés selon que la ligne RER ou Transilien passe ou non par la gare de Lyon.



D'après le modèle, l'impact du passage par la gare Saint-Lazare est très important (+5,3 points de retard). Cela s'explique probablement en grande partie par la densité du trafic sur le secteur de la gare. D'après la SNCF⁴, la gare Saint-Lazare connaît un mouvement de train toutes les 28 secondes à l'heure de pointe. Dans le détail, l'IAU Île-de-France indique que les voies des groupes II et III utilisées par la ligne L « sont particulièrement sollicitées » aux heures de pointe et que « les 10 voies Grandes lignes [...] fonctionnent en limite de capacité » [10]. L'institut constate également « la saturation à la pointe des deux voies du groupe V (Mantes – Paris via Poissy) » utilisées par les lignes Transilien L et J et des trains Grandes Lignes [10] dont l'effet est pris en compte dans le modèle par la variable sur le trafic moyen journalier annuel de la section la plus chargée. Enfin, les insuffisances en matière de fiabilité des équipements d'alimentation des lignes électriques de la gare Saint-Lazare, soulignées en 2015 par l'EPFL [6] pourrait être un autre élément d'explication de l'effet identifié par l'analyse économétrique. L'IAU Île-de-France note également « la gare Saint-Lazare [...] va être soulagée par le transfert des trains Transilien du groupe V dans la gare Hausmann-Saint Lazare du RER E, lorsque ce dernier sera prolongé jusqu'à Mantes » [10] (projet EOLE).

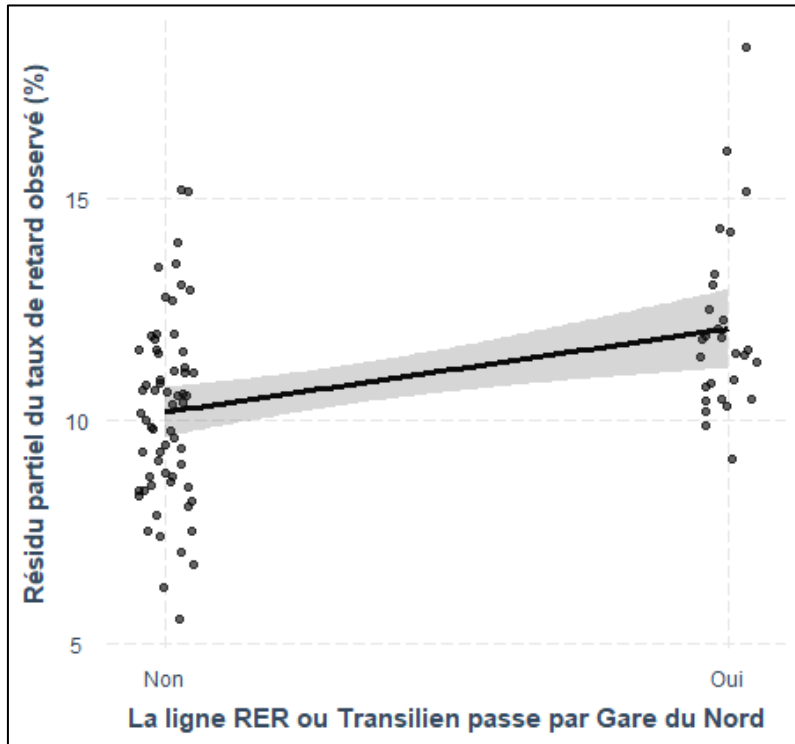
Le graphique ci-après montre que le résidu du taux de retard observé prend en moyenne des valeurs sensiblement plus élevées lorsque la ligne passe par la gare Saint-Lazare.

⁴ D'après la présentation de la SNCF pour le comité de ligne J du 28 juin 2013

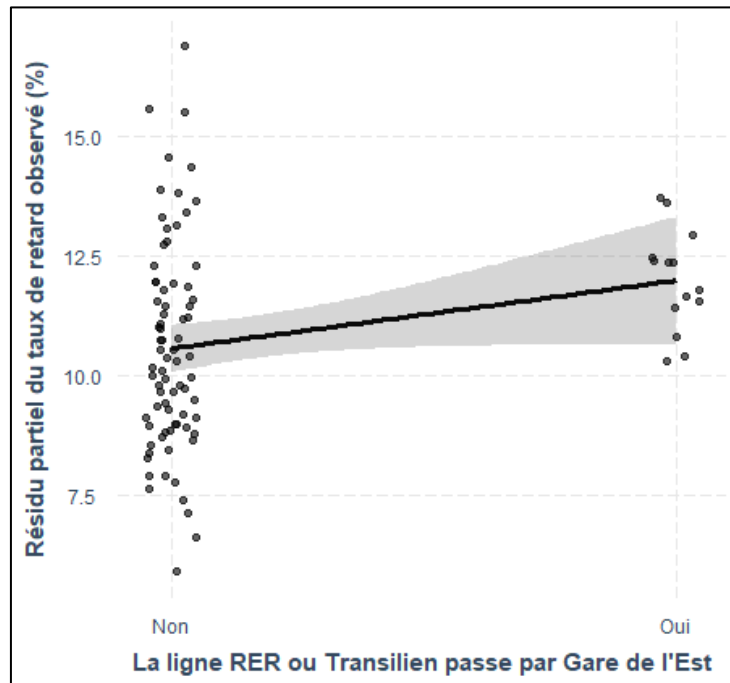


Le passage par la gare du Nord est également un facteur de non-punctualité (+1,7 points de retard). La densité du trafic que connaît la gare du Nord – près de 1370 trains par jour en 2012 [\[10\]](#) – et le niveau de saturation de la gare de surface aux heures de pointe sont des sources d’incidents potentiellement à l’origine de retards. L’IAU Île-de-France précise que la saturation affecte l’ensemble des voies de la gare de surface exceptées les voies de l’Eurostar, et également la gare souterraine – desservie par les RER B et D – aux heures de pointe. L’institut note également la saturation en amont de la gare du Nord des voies partagées entre le RER D, les TER des Hauts-de-France et les trains Grandes Lignes sur la ligne de Creil [\[10\]](#). En outre, les insuffisances en matière de fiabilité des équipements d’alimentation des lignes électriques de la gare, ainsi que des installations de sécurité, étaient en 2015 des facteurs de non-punctualité soulevés par l’EPFL [\[6\]](#). Dans son étude des grandes gares parisiennes, l’IAU Île-de-France notait en 2013 l’absence de grand projet susceptible de décharger la gare du Nord, et formule la proposition d’une liaison souterraine entre la gare Montparnasse et la gare du Nord qui serait destinée aux TGV [\[10\]](#).

On peut constater sur le graphique suivant que le résidu du taux de retard observé prend des valeurs plus élevées en moyenne lorsque la ligne passe par la gare du Nord.



Selon le modèle retenu, le passage par la gare de l'Est entraîne une dégradation de la ponctualité à hauteur de 1,2 points de retard. Cependant, l'effet de cette variable sur le taux de retard n'est pas significatif au seuil de 10%. On observe en effet une certaine incertitude statistique sur l'effet du passage par la gare de l'Est, matérialisée par la zone ombrée dans le graphique suivant :



Nous rappelons que l'effet du passage par la gare Montparnasse n'est pas étudié car seule la ligne N la dessert. De fait, si nous obtenions un résultat significatif quant à l'impact de cette variable sur la ponctualité, nous ne pourrions pas savoir s'il est dû à des problèmes rencontrés au niveau de la gare Montparnasse ou à d'autres caractéristiques de la ligne N qui dégradent la ponctualité. Toutefois, l'IAU Île-de-France constate que cette gare fonctionne également en limite de capacité et que les voies de la gare destinées aux services Transilien sont les plus chargées [10]. La saturation de cette gare a donc probablement un impact sur la ponctualité de la ligne N.

Concernant les technologies de signalisation ferroviaire, il est à noter que le mode d'espacement et la longueur des cantons pourrait influencer sur les performances de ponctualité. Cependant nous n'avons pas identifié de données publiques plus détaillées que la carte sur le mode d'espacement des trains citée dans le Document de Référence du Réseau (DRR). Or, cette carte (fournie en annexe 25) se limite à la présence d'un block automatique lumineux (BAL) qui équipe la quasi-totalité des lignes de RER et Transilien, ne permettant pas d'intégrer cette variable dans notre modèle économétrique.

Pareillement, on pourrait s'attendre à une influence du type de contrôle de vitesse utilisé sur telle ou telle ligne RER ou Transilien sur sa ponctualité. Le DRR fournit là-aussi une carte (annexe 25), mais celle-ci se limite à la présence du KVB (contrôle de vitesse par balises) qui équipe la quasi-totalité des lignes de RER et Transilien. D'autres sources publiques – en particulier le rapport de l'Audit de la performance du système Transilien par l'EPFL – fournissent des informations plus fines ayant permis de constituer la variable « TypeContrôleVitesse », dont nous avons vu plus haut que son effet sur la ponctualité des lignes RER ou Transilien ressort statistiquement comme contre-intuitif du fait que les lignes équipées des systèmes de contrôle de vitesse les plus avancées sont celles qui souffrent structurellement de mauvaises performances de ponctualité.

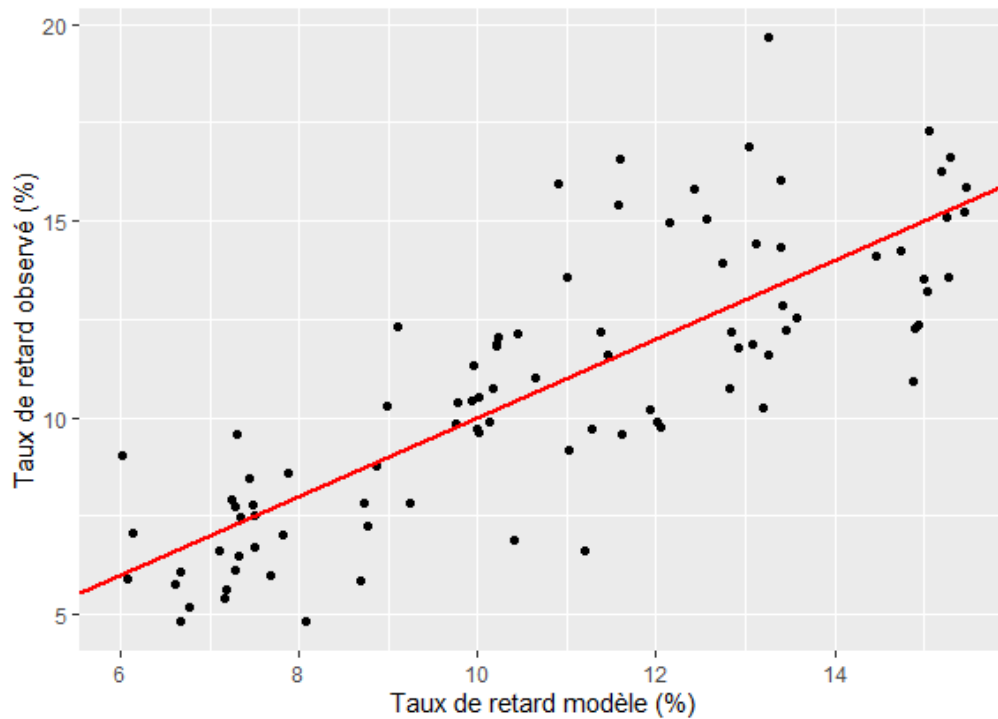
Enfin, le coefficient de détermination ajusté ou R^2 ajusté du modèle retenu est de 0,63, ce qui signifie que 63% de la variance du taux de retard moyen annuel des RER et Transilien par ligne sur la période 2013-2019 sont expliqués par le modèle.

On peut remarquer que ce chiffre est nettement plus élevé que celui obtenu avec les modèles retenus pour les trains longue distance (TGV et Intercités), et relativement proche (légèrement supérieur) de celui obtenu avec le modèle TER retenu.

On rappelle que le taux de retard des trains peut dépendre d'un très grand nombre de paramètres, et que de nombreux événements ayant un impact significatif sur le taux de retard ne sont pas intégrés au modèle :

- Les événements météorologiques majeurs (inondations et pluies intenses, tempêtes, épisodes neigeux et de froid, canicule et fortes chaleurs, etc.) ;
- Les mouvements sociaux ;
- Les autres événements externes au transport, dont notamment les actes de malveillance, les accidents de personne, les collisions des trains avec des véhicules, ou avec des animaux, ... ;
- Les incidents liés au dysfonctionnements des infrastructures ferroviaires, par exemple les ruptures de caténaire ou les pannes de signalisation ;
- Les incidents voyageurs (malaise d'un voyageur, blocage involontaire des portes par un voyageur, etc.), dont le risque d'occurrence augmente possiblement avec le taux d'occupation des trains.

Le graphique croisant le taux de retard prédit par le modèle (en abscisse) et le taux de retard observé (en ordonnée) permet d'apprécier d'une autre manière la capacité du modèle à décrire la variabilité géographique et temporelle du taux de retard des RER et Transilien par ligne. La ligne rouge correspond à la droite $y = x$ sur laquelle tous les points doivent s'aligner dans un modèle parfait théorique. On constate que la majorité des points s'alignent assez bien sur cette droite – notamment pour les taux de retard prédits les plus faibles et les plus élevés - même si certains points s'écartent fortement de la droite. Comme pour les services TER, le modèle retenu présente un plus grand pouvoir descriptif que les modèles retenus pour les TGV et les services Intercités, ce qui se traduit par un meilleur alignement de la distribution du taux de retard sur la droite $y = x$.



3.2.4.4 Modèles de régressions linéaires multiples retenus pour expliquer le taux de retard par branche

Afin de tester l'influence du niveau de trafic par branche sur le taux de retard des RER et Transilien par branche sur la période 2013-2018⁵, des modèles de régressions linéaires multiples ont été implémentés. Les données de l'OMNIL utilisées correspondent au taux de ponctualité voyageurs Île-de-France Mobilités annuel pour les voyageurs qui descendent à l'une des gares de la branche.

Les données utilisées pour les différentes variables explicatives testées sont précisées à la section 2.2.2.4.

Des modèles portant sur le taux de retard des RER et Transilien par branche pour l'ensemble des lignes ont d'abord été construits. Les 4 « meilleurs modèles » de ce type – au sens où ils optimisent un ensemble de critères statistiques – sont présentés en annexe 26.

Cependant, ces modèles indiquent que le niveau de trafic par voie au sein de la branche présente un impact sur son taux de retard contraire à l'intuition : en moyenne, les branches qui connaissent le trafic le plus faible ne sont pas les plus ponctuelles. Le résultat est significatif au seuil de 0,1%, aussi bien en prenant en compte le niveau maximal ou le niveau moyen du trafic au sein de la branche.

D'autres modèles ont alors été construits en se restreignant à certaines lignes de RER et Transilien. Pour certaines lignes, on peut alors observer qu'un trafic par voie plus élevé est corrélé avec une moins bonne ponctualité de la branche. 2 modèles basés sur des groupes de lignes ont ainsi été construits : le premier portant sur le niveau moyen du trafic par voie de la branche, et le second sur le niveau maximal du trafic par voie de la branche.

```
Call:
lm(formula = TxRetard ~ TMJA_ParVoie_Moy_Branche + TMJA_ParVoie_Max_ParcoursComple +
  AgeMoyen_MaterielRoulant_1erJanvier + NbVoiesGarageTerminus,
  data = data_PonctuTransilien_branches_Ligne_A_H_P)

Residuals:
    Min       1Q   Median       3Q      Max
-15.746  -2.498  -0.061   2.775   8.889

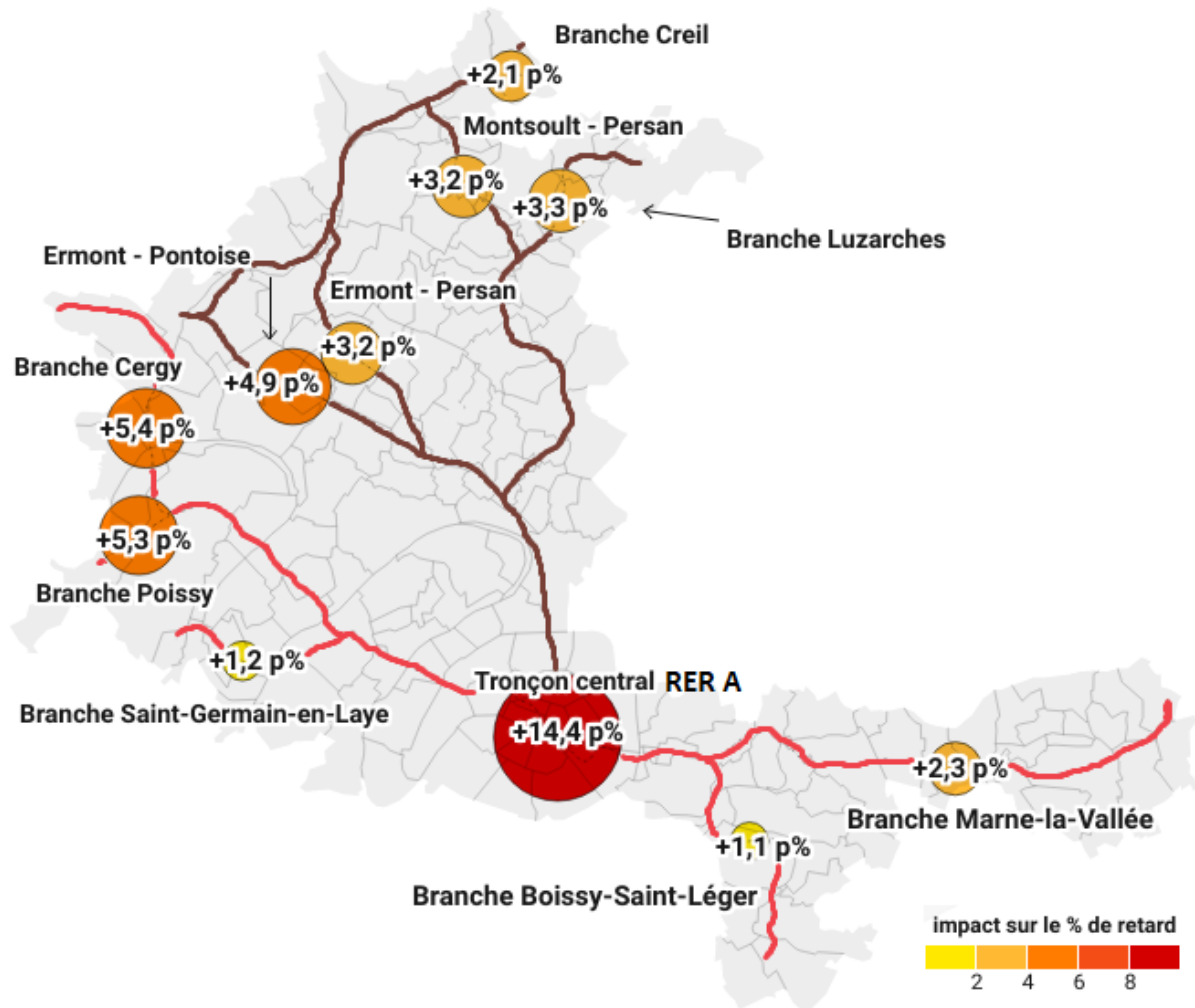
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2.540331   1.925453  -1.319  0.19060
TMJA_ParVoie_Moy_Branche  0.045781   0.015778   2.901  0.00473 **
TMJA_ParVoie_Max_ParcoursComple  0.054745   0.004871  11.239 < 2e-16 ***
AgeMoyen_MaterielRoulant_1erJanvier  0.302066   0.092656   3.260  0.00160 **
NbVoiesGarageTerminus   -0.204434   0.072258  -2.829  0.00582 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.47 on 85 degrees of freedom
Multiple R-squared:  0.6827,    Adjusted R-squared:  0.6678
F-statistic: 45.73 on 4 and 85 DF,  p-value: < 2.2e-16
```

⁵ Pour l'année 2019, les données publiées par l'OMNIL ne couvrent que la période de janvier à juillet 2019. En cas de saisonnalité du taux de ponctualité, il est possible que les disparités entre branches de l'effet de saisonnalité influencent la ponctualité, puis la validité des résultats obtenus avec la modélisation statistique. Nous n'avons donc pas intégré les données par branches de 2019 dans la modélisation.

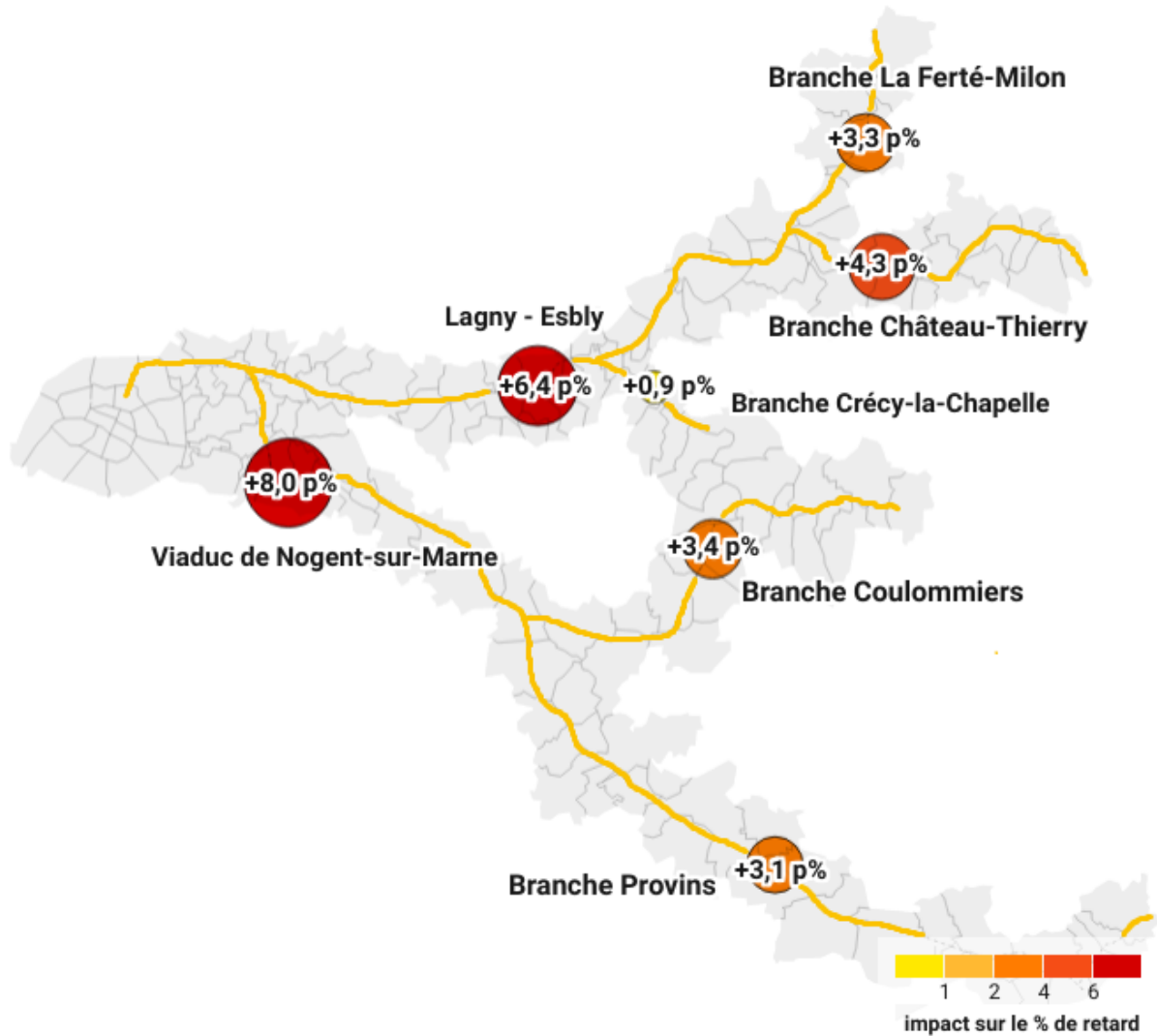
Les cartes ci-dessous illustrent l'effet selon le modèle du trafic journalier moyen par branche sur la ponctualité des lignes A, H et P :

Estimation économétrique de l'effet du trafic journalier moyen par branche sur la ponctualité des lignes A et H



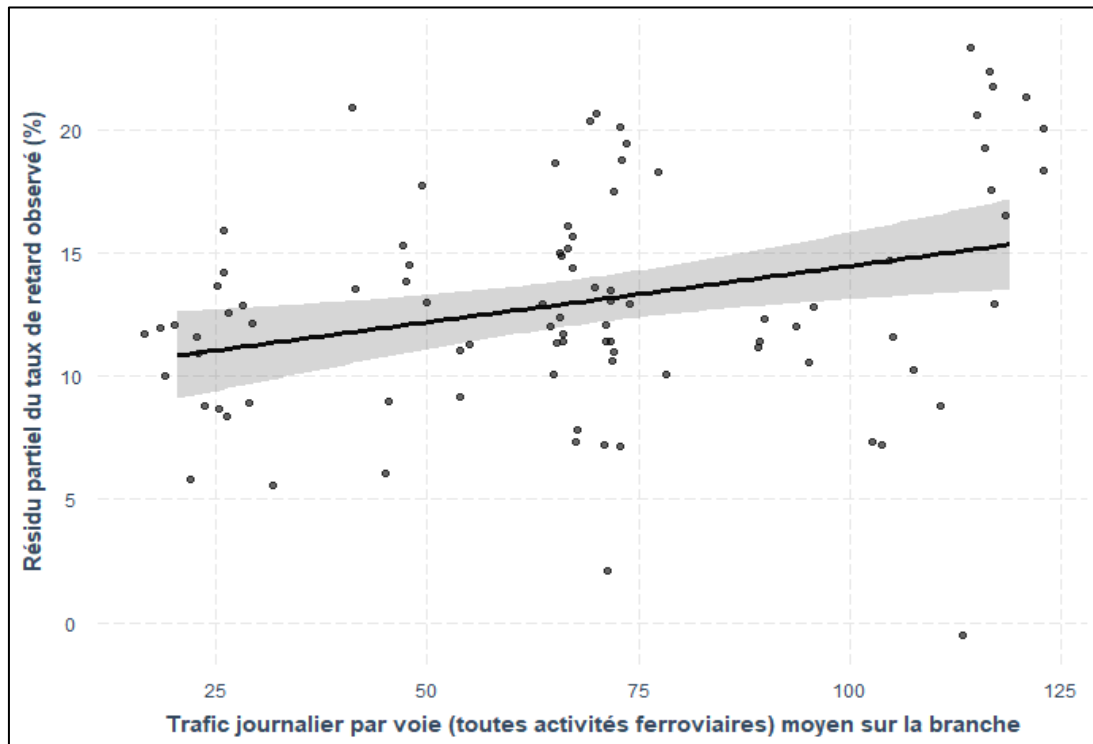
Carte: Alexandre Barbusse/AQST · Créé avec Datawrapper

Estimation économétrique de l'effet du trafic journalier moyen par branche sur la ponctualité de la ligne P



Carte: Alexandre Barbusse/AQST • Créé avec Datawrapper

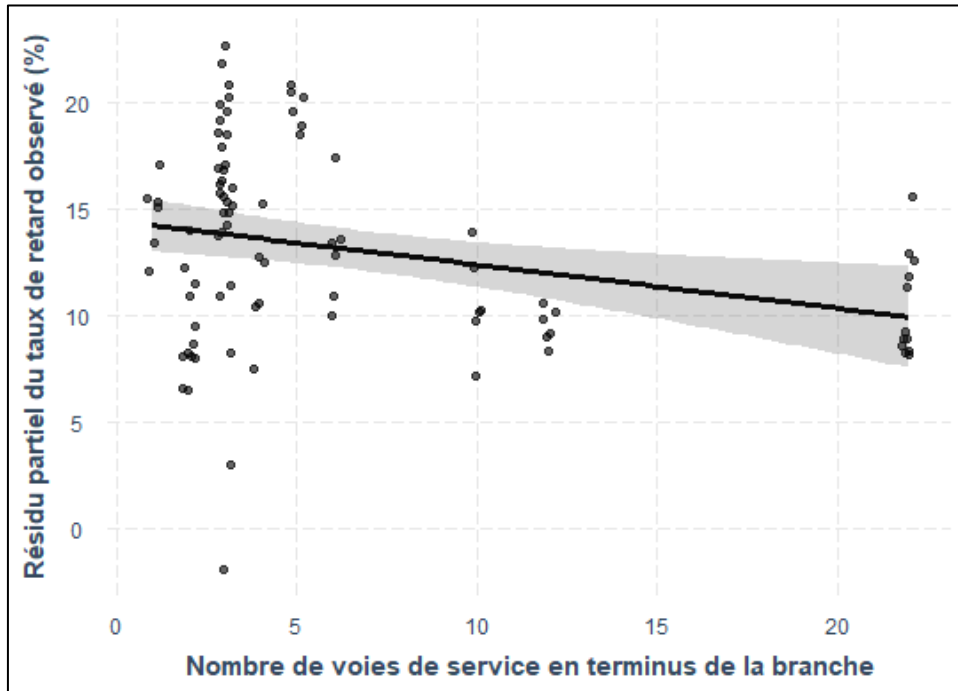
D'après le premier modèle, la circulation de 100 trains quotidiens par voie supplémentaires sur l'ensemble d'une branche des lignes RER A ou Transilien H ou P se traduit par une hausse du taux de retard sur cette branche de 4,5 points en moyenne.



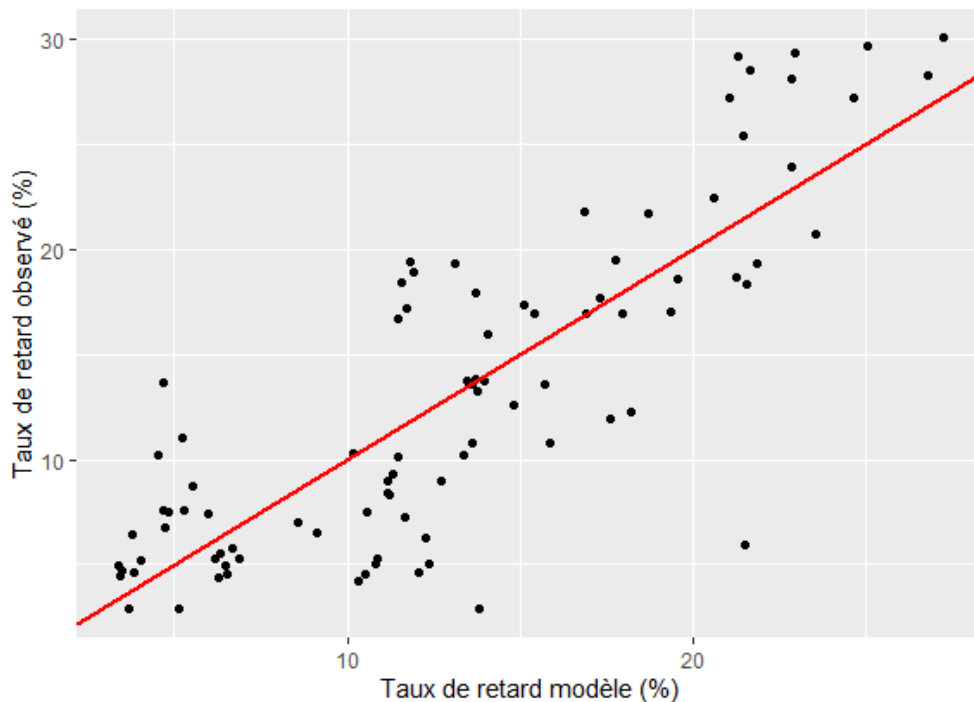
Le graphique ci-dessus montre l'effet modélisé du trafic journalier par voie moyen de la branche sur son taux de retard en se restreignant aux lignes A, H et P. La zone ombrée autour de la droite de régression correspond à la plage d'incertitude statistique sur cette droite. En axe des ordonnées, le résidu partiel du taux de retard observé est représenté. Il s'agit en fait du taux de retard observé auquel on retranche l'effet des autres variables de la régression, afin d'identifier plus facilement graphiquement la relation de corrélation avec la variable considérée. Malgré une dispersion très importante autour de la droite de régression, l'effet de cette variable est significatif au seuil 1%.

Le modèle suggère également qu'un nombre de voies de garage en bout de branche plus élevé s'accompagne d'une meilleure ponctualité, ce qui est également conforme à l'intuition. En effet, si le terminus/départ de la branche dispose de nombreuses voies de service (voies de garage notamment), il sera plus aisé pour l'exploitant d'adapter le plan de circulation en cas d'incident par l'ajout ou la suppression de trains puisqu'il y aura davantage de rames en réserve dans le sens du départ et de voies disponibles pour accueillir des rames dans le sens de l'arrivée. En moyenne sur l'année, cette souplesse améliore la ponctualité de la branche pour les lignes A, H et P. En revanche, nous n'observons pas d'effet significatif de cette variable sur la ponctualité par branche pour l'ensemble des lignes (voir annexe 27).

Le graphique ci-après présente la droite de régression modélisant cet effet, significatif au seuil 1%. On observe toutefois une dispersion importante du résidu partiel du taux de retard pour les branches des lignes A, H et P dont le terminus présente 2 ou 3 voies de service seulement.



Le graphique croisant le taux de retard prédit par le modèle (en abscisse) et le taux de retard observé (en ordonnée) montre une dispersion relativement limitée de la distribution autour de la droite $y = x$, hormis pour quelques points. Cela illustre le plutôt bon pouvoir descriptif du modèle par branche retenu pour les lignes A, H et P.



D'après le second modèle, la circulation de 100 trains quotidiens par voie supplémentaires sur la section la plus chargée d'une branche des lignes RER A ou E ou Transilien N ou R se traduit par une hausse du taux de retard sur cette branche de 2,0 points en moyenne.

```

Call:
lm(formula = TxRetard ~ TMJA_ParVoie_Max_Branche + NbArrêts +
    AgeMoyen_MaterielRoulant_1erJanvier, data = data_PonctuTransilien_branches_Ligne_A_E_R_N)

Residuals:
    Min       1Q   Median       3Q      Max
-13.2662  -2.7112   0.0569   2.3131   8.7641

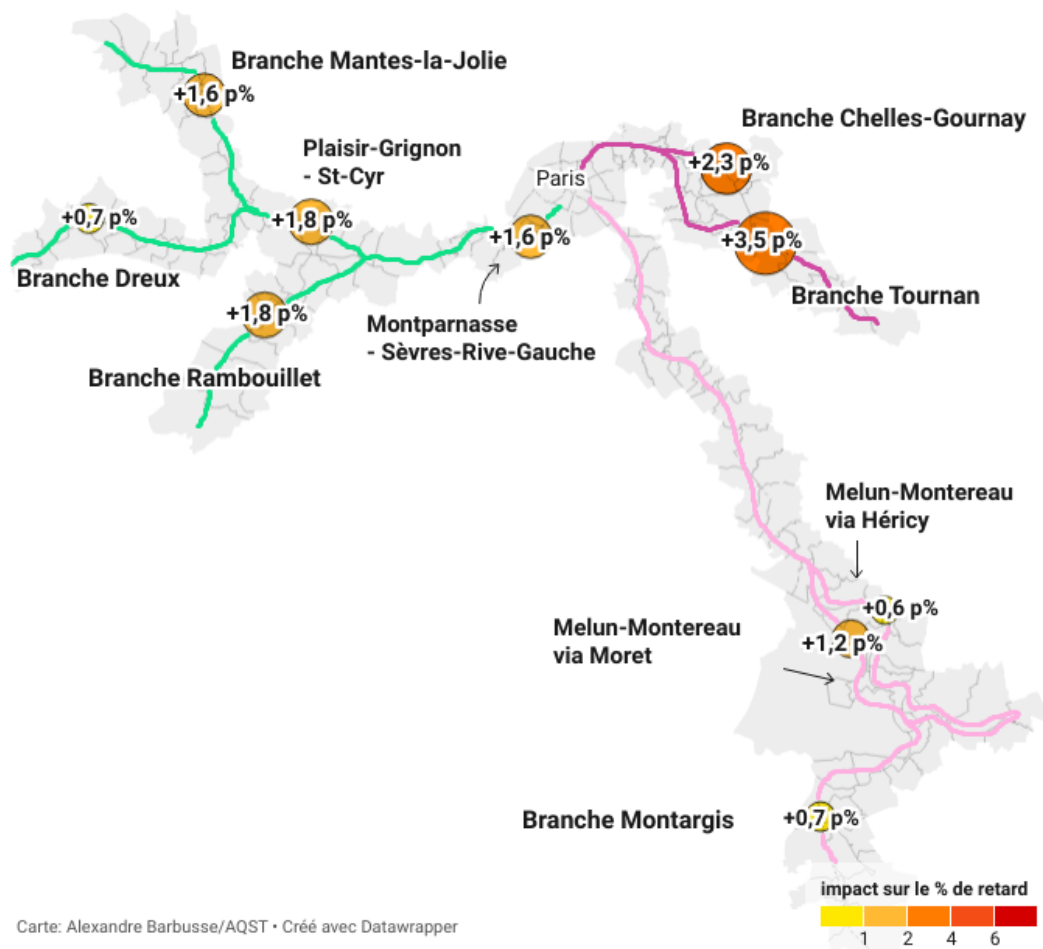
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -13.113672   2.741159  -4.784 7.06e-06 ***
TMJA_ParVoie_Max_Branche
                0.019871   0.007738   2.568  0.012 *
NbArrêts        1.176207   0.084793  13.871 < 2e-16 ***
AgeMoyen_MaterielRoulant_1erJanvier
                0.388447   0.074618   5.206 1.30e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.751 on 86 degrees of freedom
Multiple R-squared:  0.7316,    Adjusted R-squared:  0.7222
F-statistic: 78.14 on 3 and 86 DF,  p-value: < 2.2e-16

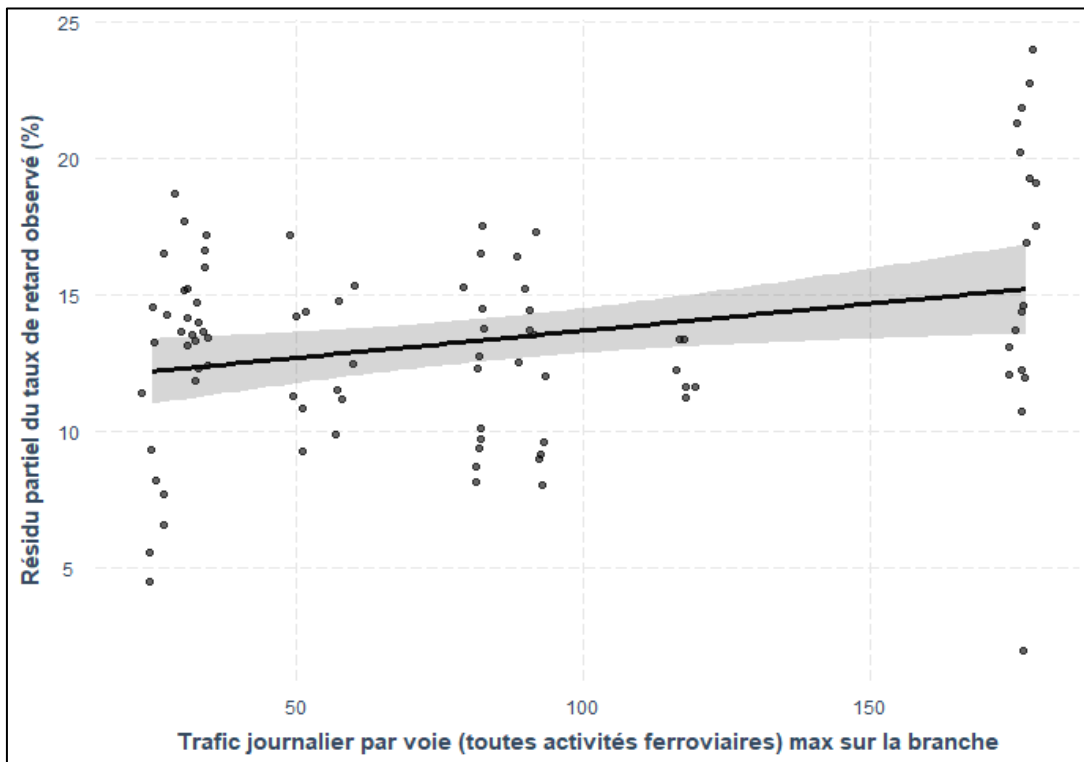
```

La carte ci-dessous illustre l'effet selon le modèle du trafic journalier maximal par branche sur la ponctualité des lignes E, N et R :

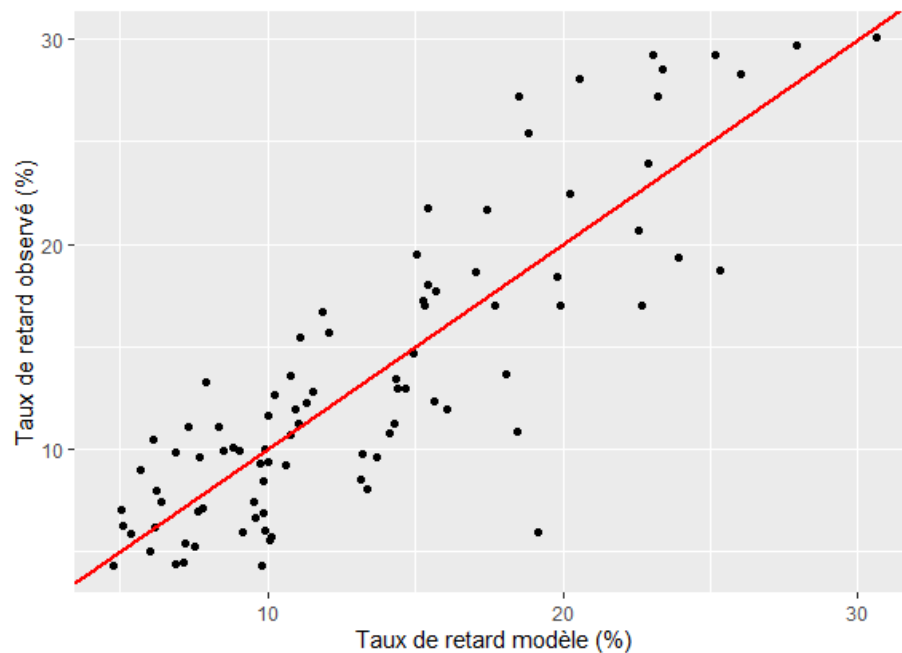
Estimation économétrique de l'effet du trafic journalier maximal par branche sur la ponctualité des lignes E, N et R



Le graphique ci-dessous montre l'effet modélisé – pour les lignes A, E, N et R - du trafic journalier par voie (toutes activités ferroviaires prises en compte) de la section la plus chargée de la branche sur son taux de retard. L'étendue de la zone ombrée autour de la droite de régression matérialise l'incertitude sur ses paramètres. Elle s'explique par la forte dispersion du résidu partiel du taux de retard autour de sa valeur moyenne, à niveau de trafic donné. Cette dispersion se traduit également par une significativité statistique qui passe le test de Student au seuil 5% mais pas au seuil 1%, contrairement aux autres variables du modèle.



Le graphique suivant, croisant le taux de retard prédit par le modèle (en abscisse) et le taux de retard observé (en ordonnée), illustre l'assez bon niveau du coefficient de déterminé R^2 ajusté obtenu avec le modèle par branche retenu pour les lignes A, E, N et R.



Nous avons donc identifié un effet du trafic par branche conforme à l'intuition sur les lignes A, E, H, N, P et R : un trafic par voie plus intense engendre une hausse des effets de propagation des retards d'un train au suivant sur la même voie, et signifie que les voies sont davantage encombrées et donc que les plans de circulation adoptés offrent moins de souplesse à l'exploitant, pour lequel il devient plus difficile de « rattraper un retard ».

En revanche, nous n'observons pas d'effet significatif du niveau de trafic par voie et par branche pour les autres lignes B, C, D, J et L (voir résultats sur l'ensemble des lignes en annexe 26). Il n'y a pas de typologie évidente rassemblant ces lignes et pouvant expliquer l'absence de relation logique.

4. CONCLUSION

Afin d'identifier des déterminants (notamment géographiques) de la ponctualité des trains de voyageurs en France par activité ferroviaire (TGV, Intercités, TER, RER et Transilien d'Île-de-France), une étude économétrique a été réalisée. Les données de la variable à expliquer (taux de retard annuel par liaison ferroviaire, par ancienne région, ou par ligne) proviennent de la base de l'AQST et couvrent la période 2012-2019, et de l'OMNIL pour les données par branche sur les RER et Transilien en Île-de-France. Les variables explicatives testées économétriquement proviennent de documents ou bases de données publiques, ou éventuellement d'estimations issues d'une modélisation basée sur ces données.

Des modèles économétriques simples (régressions linéaires simples puis multiples, modèles exponentiels, modèles log-log) ont été implémentés à l'aide du logiciel R pour décrire et interpréter la corrélation éventuelle entre la ponctualité des trains par activité et l'ensemble des variables explicatives recueillies.

Les tableaux des pages suivantes fournissent une synthèse des modèles retenus par activité.

Pour les modèles retenus, la part de la variance du taux de retard sur la période 2012-2019 expliquée par le modèle est de l'ordre d'1/3 pour les services longue distance et de l'ordre de 60% pour les activités ferroviaires régionales. En outre, d'après les résultats des tests de Fisher implémentés, les modèles retenus sont tous statistiquement significatifs au seuil 0,1%.

Pour les activités longue distance, la modélisation montre que le passage par les nœuds ferroviaires majeurs - particulièrement congestionnés [7] - que sont Paris, la gare de Lyon-Part-Dieu, et celle de Marseille Saint-Charles, engendre en moyenne une dégradation de la ponctualité du train. Concernant le nœud parisien, la dégradation de la ponctualité liée à la saturation de la Gare de Lyon est également visible dans le modèle retenu pour les RER et Transilien. Le départ ou l'arrivée au niveau de certains nœuds « secondaires » du réseau TGV ou Intercités, et parfois leur traversée, s'accompagne également d'une hausse de la probabilité d'occurrence d'un retard d'après les modèles retenus. L'effet est davantage prononcé pour les nœuds du réseau classique (Bordeaux, Toulouse et Dijon) que pour les nœuds du réseau TGV, à l'exception du passage par le nœud d'Avignon.

Concernant les trains régionaux de voyageurs, les modèles indiquent que plus la distance moyenne parcourue est longue, plus le taux de retard est élevé, aussi bien par ligne en Île-de-France, que pour l'ensemble des lignes par ancienne région dans le cas des TER. Cette corrélation semble logique car plus un trajet dure longtemps plus il est soumis longtemps au risque d'occurrence d'un incident entraînant un retard.

Une hausse du niveau de trafic engendrerait également une hausse du taux de retard d'après les modèles. Cette corrélation est également visible pour le niveau de trafic sur la section dimensionnante de la liaison empruntée dans le cas du modèle TGV retenu. Une explication possible pourrait être que, lorsqu'une liaison ferroviaire circule sur des lignes ferroviaires connaissant un trafic intense, elle est davantage exposée au risque de subir un retard qui se propage d'une circulation de train à l'autre. De plus, les infrastructures ferroviaires (notamment les voies et la signalisation) se

détériorent à mesure qu'elles sont intensément utilisées, et cette usure augmente le risque d'incident engendrant des retards.

Activité ou service ferroviaire	Granularité des données utilisées	Période des données utilisées	R ² ajusté régression (p-value test de Fisher)	Variables explicatives	p-value test de Student variables explicatives	Impact sur le taux de retard
TGV	Par liaisons ferroviaires	2012-2019	0,32 ($< 0,001$)	La liaison part de Paris	$< 0,1\%$	-2,3 pts de retard
				La liaison part de Lyon-Part-Dieu	$< 0,1\%$	+4,9 pts
				La liaison traverse le nœud ferroviaire d'Avignon	$< 0,1\%$	+2,4 pts
				La liaison traverse Rennes	$< 0,1\%$	-2,8 pts
				La liaison traverse Dijon	$< 0,1\%$	-4,8 pts
				La liaison part ou arrive au niveau d'un nœud ferroviaire secondaire du réseau TGV	$< 0,1\%$	+3,1 pts
				Nombre moyen d'arrêts effectués par les TGV sur la liaison	$< 1\%$	+0,3 pts / arrêt
				Trafic moyen journalier annuel (TMJA) des TGV sur la section la plus intensément utilisée de la liaison en 2019	$< 2\%$	+0,5 pts / 100 circulations TGV/jour
Intercités	Par liaisons ferroviaires	2012-2019	0,33 ($< 0,001$)	Liaison de nuit	$< 0,1\%$	-3,9 pts de retard
				Nombre de nœuds majeurs (Paris, Lyon, Marseille) traversés sur la liaison	$< 0,1\%$	+5,2 pts / nœud majeur traversé
				La liaison traverse le nœud ferroviaire de Bordeaux	$< 0,1\%$	+7,3 pts
				La liaison traverse le nœud ferroviaire de Toulouse Matabiau	$< 0,1\%$	+2,9 pts
				La liaison traverse le nœud ferroviaire de Dijon	$< 2\%$	+4,2 pts
				La liaison traverse Rouen	$< 1\%$	-3,0 pts
				La liaison part ou arrive au niveau d'un nœud ferroviaire secondaire du réseau classique	$< 1\%$	+1,6 pts
				La liaison appartient à l'axe "Est – Centre" du réseau Intercités de 2016	$< 1\%$	-2,3 pts
				Le trajet sur la liaison s'effectue en sens "Nord -> Sud"	$< 0,1\%$	+2,5 pts

Activité ou service ferroviaire	Granularité des données utilisées	Période des données utilisées	R ² ajusté régression (p-value test de Fisher)	Variables explicatives	p-value test de Student variables explicatives	Impact sur le taux de retard
TER (trains régionaux de province)	Par anciennes régions administratives	2013-2019	0,62 ($< 0,001$)	Nombre moyen d'arrêts par ancienne région en 2018 (nouvelles régions administratives)	$< 0,1\%$	+1,2 pts / arrêt
				Distance moyenne parcourue par les trains TER	$< 2\%$	+0,25 pts / 10 km
				Nombre de voyageurs-km transportés (en millions) dans l'année	$< 0,1\%$	+1,2 pts / milliard voy-km
				Densité des « petites lignes » (catégories UIC 7 à 9 AV)	$< 1\%$	-0,46 pt/ 1km de voie/100 km ² de superficie
RER et Transilien (trains régionaux d'Île-de-France)	Par lignes	2013-2019	0,63 ($< 0,001$)	Nombre de circulations quotidiennes de trains (toutes activités ferroviaires) par voie sur la section la plus chargée de la ligne	$< 1\%$	+1,1 pts / 100 circulations/jour/voie
				Distance moyenne parcourue par les trains RER ou Transilien	$< 0,1\%$	+1,1 pts / 10 km
				Estimation de l'âge moyen du matériel roulant en début d'année	$< 0,1\%$	+1,2 pts / 10 ans
				La ligne comprend la station gare de Lyon	$< 5\%$	+2,1 pts
				La ligne passe par la gare Saint-Lazare	$< 0,1\%$	+5,3 pts
				La ligne passe par la gare du Nord	$< 1\%$	+1,7 pts
				La ligne passe par la gare de l'Est	$> 10\%$	+1,2 pts

Activité ou service ferroviaire	Granularité des données utilisées	Période des données utilisées	R ² ajusté régression (p-value test de Fisher)	Variables explicatives	p-value test de Student variables explicatives	Impact sur le taux de retard
RER et Transilien (trains régionaux d'Île-de-France)	Par branches (lignes A, H et P)	2013-2018	0,67 (< 0,001)	Nombre de circulations quotidiennes de trains (toutes activités ferroviaires) par voie moyen sur la branche	< 1%	+4,5 pts / 100 circulations/jour/voie
				Nombre de circulations quotidiennes de trains (toutes activités ferroviaires) par voie sur la section la plus chargée de la ligne	< 0,1%	+5,4 pts / 100 circulations/jour/voie
				Estimation de l'âge moyen du matériel roulant en début d'année	< 1%	+3,0 pts / 10 ans
				Nombre de voie de service en terminus de branche	< 1%	-0,2 pt/ voie de service
RER et Transilien (trains régionaux d'Île-de-France)	Par branches (lignes A, E, N et R)	2013-2018	0,72 (< 0,001)	Nombre de circulations quotidiennes de trains (toutes activités ferroviaires) par voie sur la section la plus chargée de la branche	< 2%	+2,0 pts / 100 circulations/jour/voie
				Nombre moyen d'arrêts effectués par les RER ou Transilien desservant la branche	< 0,1%	+1,2 pts / arrêt
				Estimation de l'âge moyen du matériel roulant en début d'année	< 0,1%	+3,9 pts / 10 ans

La modélisation suggère également que chaque arrêt supplémentaire sur une liaison TGV ou une liaison TER s'accompagne d'une hausse du risque de retard. Pour l'expliquer, on peut supposer que le risque d'occurrence de retards est important lors d'un arrêt. En effet, certains événements peuvent survenir lors d'un arrêt en gare et retarder le départ du train : par exemple, malaise voyageur, blocage des portes par un voyageur retardant le train, dysfonctionnement du matériel roulant lors de son redémarrage, ou parfois des infrastructures ferroviaires en sortie de gare (par exemple la signalisation), etc.

Les modèles retenus pour les trains RER et Transilien en Île-de-France indiquent qu'un vieillissement du parc de matériel roulant s'accompagne en moyenne d'une baisse de ponctualité, ce qui est conforme à l'intuition puisqu'un matériel roulant âgé présente un risque de panne ou d'incident matériel plus important. En revanche, cette corrélation n'est pas apparue comme statistiquement significative au seuil de 5% pour les TER de province.

Le modèle retenu pour les trains régionaux d'Île-de-France par ligne montre également l'effet sur la ponctualité de la saturation des grandes parisiennes – en particulier la gare du Nord, la gare Saint-Lazare et la gare de Lyon - qui en font des sources majeures d'incidents causant des retards sur le réseau Transilien, en accord avec les constats de l'audit du réseau Transilien [6], et de l'étude de l'IAU Île-de-France consacrée aux grandes gares parisiennes [10].

L'analyse des déterminants du taux de retard par branche des RER et Transilien suggère par ailleurs que - pour le RER A et certaines lignes de Transilien au moins - l'intensité du trafic par voie (toutes activités ferroviaires prises en compte) au sein de la branche affecte son niveau de ponctualité, et qu'un nombre important de voies de service en terminus de branche contribue à l'amélioration de la ponctualité de la branche, peut-être parce que cela offre davantage de souplesse à l'exploitation et à l'adaptation du plan de circulation en situation dégradée.

Enfin, les résultats de la modélisation montrent qu'une part importante de la variabilité temporelle et géographique du taux de retard des services de transport ferroviaire de voyageurs sur la période d'étude reste inexpliquée par les modèles retenus. En effet, le taux de retard des trains peut dépendre d'un très grand nombre de paramètres, et de nombreux événements pouvant impacter significativement le taux de retard ne peuvent pas être intégrés au modèle en l'absence de données précises et structurées sur les dates d'occurrence et les périmètres géographiques impactés par ces événements. On peut notamment citer :

- Les événements météorologiques majeurs (inondations et pluies intenses, tempêtes, épisodes neigeux et de froid, canicule et fortes chaleurs, etc.) ;
- Les mouvements sociaux ;
- Les autres événements externes au transport, dont notamment les actes de malveillance, les accidents de personne, les collisions des trains avec des véhicules, ou avec des animaux, etc.) ;
- Les incidents liés aux dysfonctionnements des infrastructures ferroviaires, par exemple les ruptures de caténaire ou les pannes de signalisation ;
- Les dysfonctionnements du matériel roulant, avec notamment les pannes du matériel moteur ;
- Les incidents voyageurs (malaise d'un voyageur, blocage involontaire des portes par un voyageur, etc.)

BIBLIOGRAPHIE

- [1] Nœud ferroviaire lyonnais – rapport d'étape, Rapport n°006680-01 du CGEDD, Marie-Line MEAUX, 2009
- [2] Coordination des réflexions prospectives sur l'évolution du nœud ferroviaire lyonnais, Rapport n°006680-02 du CGEDD, Marie-Line MEAUX, 2011
- [3] Mobilité 21, « Pour un schéma national de mobilité durable », Rapport au ministre chargé des transports, de la mer et de la pêche, 2013
- [4] LNOBPL – Débat Public – D.2 Nœud de Nantes, Réseau Ferré de France, 2014
- [5] Modernisation du nœud ferroviaire de Toulouse, Optimisation du pôle industriel ferroviaire – Concertation Préalable du 20 mars au 19 mai 2017, SNCF Réseau, 2017
- [6] Rapport de l'Audit stratégique de la performance du système, Panos Tzieropoulos, Yves Putallaz, EPFL, 2015
- [7] Les nœuds ferroviaires et leur prise en charge opérationnelle, Rapport n° 008354-03 du CGEDD, Marie-Anne BACOT, 2015
- [8] Recommandations du conseil scientifique des observatoires de la saturation ferroviaire, Rapport n° 009772-01 du CGEDD, Jean-Paul OURLIAC, 2015
- [9] Préparation du Schéma Directeur de la ligne R – Schéma de secteur, RFF et SNCF, 2015
- [10] Paris, Londres, Berlin, Madrid, Le fonctionnement des grandes gares ferroviaires, IAU Île-de-France, 2013

ANNEXES

Annexe 1 : Estimation du taux annuel de retard des TER par ancienne région administrative sur la période 2018-2019

Les données sur le taux de retard annuel des TER par ancienne région de la base de l'AQST concernent sur la période 2018-2019 les régions selon le découpage administratif postérieur à la réforme du 1^{er} janvier 2016. Or la base de données sur laquelle repose l'étude économétrique concerne le découpage régional antérieure à cette réforme, afin de disposer d'un plus grand nombre d'observations (nombre de lignes) dans la base. Il a donc été nécessaire d'estimer le taux de retard des TER des anciennes régions pour la période 2018-2019 sur la base du taux de retard des nouvelles régions d'après la base de données de l'AQST.

Le calcul effectué suit les hypothèses et le principe suivant :

- Hypothèse que la part du nombre annuel de circulations TER par ancienne région dans le total du nombre de circulations TER de la nouvelle région est égal à la moyenne de cette proportion sur la période 2013-2016. Cela permet d'en déduire le nombre annuel de circulations TER par ancienne région en 2018 et 2019.
- Hypothèse que la part du nombre annuel de circulations TER en retard par ancienne région dans le total du nombre de circulations TER en retard de la nouvelle région est égal à la moyenne de cette proportion sur la période 2013-2016. Cela permet d'en déduire le nombre annuel de circulations TER en retard par ancienne région en 2018 et 2019.
- Le rapport pour chaque ancienne région de l'estimation du nombre annuel de circulations TER en retard sur l'estimation du nombre annuel de circulations TER fournit alors une estimation du taux de retard par ancienne région administrative en 2018 et 2019.

Le tableau suivant détaille les taux de retard obtenus :

Région	Année	TxRetard (Estimation AQST)
Alsace	2017	4,4
Alsace	2018	5,1
Alsace	2019	4,1
Aquitaine	2018	10,7
Aquitaine	2019	8,4
Auvergne	2018	7,7
Auvergne	2019	6,0
Basse-Normandie	2018	7,5
Basse-Normandie	2019	6,9
Bourgogne	2018	8,6
Bourgogne	2019	7,0
Champagne Ardenne	2017	6,1
Champagne Ardenne	2018	7,0
Champagne Ardenne	2019	5,7

Franche-Comté	2018	7,8
Franche-Comté	2019	6,3
Haute-Normandie	2018	6,5
Haute-Normandie	2019	6,0
Languedoc-Roussillon	2018	14,9
Languedoc-Roussillon	2019	10,9
Limousin	2018	8,0
Limousin	2019	6,3
Lorraine	2017	5,6
Lorraine	2018	6,5
Lorraine	2019	5,2
Midi Pyrénées	2018	13,0
Midi Pyrénées	2019	9,5
Nord Pas de Calais	2018	10,3
Nord Pas de Calais	2019	9,5
Picardie	2018	11,3
Picardie	2019	10,4
Poitou-Charentes	2018	8,4
Poitou-Charentes	2019	6,6
Rhône Alpes	2018	13,2
Rhône Alpes	2019	10,2

Annexe 2 : Distances ferroviaires par liaison TGV

Les données sont issues du Mémento statistique 2018 de la SNCF pour les 18 premières relations, et de la base de données Rome2Rio pour les 36 suivantes :

Relation	Distance (km)	Liaison AQST sens_1	Liaison AQST sens_2
Source	Mémento statistique de la SNCF - 2018		
Paris - Strasbourg	452	PARIS EST => STRASBOURG	STRASBOURG => PARIS EST
Paris - Nancy	331	PARIS EST => NANCY	NANCY => PARIS EST
Paris - Lille	226	PARIS NORD => LILLE	LILLE => PARIS NORD
Paris - Le Mans	202	PARIS MONTPARNASSE => LE MANS	LE MANS => PARIS MONTPARNASSE
Paris - Nantes	387	PARIS MONTPARNASSE => NANTES	NANTES => PARIS MONTPARNASSE
Paris - Rennes	365	PARIS MONTPARNASSE => RENNES	RENNES => PARIS MONTPARNASSE
Paris - La Rochelle	468	PARIS MONTPARNASSE => LA ROCHELLE VILLE	LA ROCHELLE VILLE => PARIS MONTPARNASSE
Paris - Tours	223	PARIS MONTPARNASSE => TOURS	TOURS => PARIS MONTPARNASSE
Paris - Bordeaux	569	PARIS MONTPARNASSE => BORDEAUX ST JEAN	BORDEAUX ST JEAN => PARIS MONTPARNASSE
Paris - Poitiers	322	PARIS MONTPARNASSE => POITIERS	POITIERS => PARIS MONTPARNASSE
Paris - Toulouse	826	PARIS MONTPARNASSE => TOULOUSE MATABIAU	TOULOUSE MATABIAU => PARIS MONTPARNASSE
Paris - Dijon	285	PARIS-LYON => DIJON VILLE	DIJON VILLE => PARIS-LYON
Paris - Lyon-Part-Dieu	427	PARIS-LYON => LYON PART DIEU	LYON PART DIEU => PARIS-LYON
Paris - Montpellier	739	PARIS-LYON => MONTPELLIER	MONTPELLIER => PARIS-LYON
Paris - Marseille	749	PARIS-LYON => MARSEILLE ST CHARLES	MARSEILLE ST CHARLES => PARIS-LYON
Paris - Nice	973	PARIS-LYON => NICE VILLE	NICE VILLE => PARIS-LYON
Paris - Grenoble	554	PARIS-LYON => GRENOBLE	GRENOBLE => PARIS-LYON
Lyon-Part-Dieu - Lille	695	LYON PART DIEU => LILLE	LILLE => LYON PART DIEU

Source	Rome2Rio	Liaison AQST sens_1	Liaison AQST sens_2
Paris - Aix-en-Provence	725	PARIS-LYON => AIX EN PROVENCE TGV	AIX EN PROVENCE TGV => PARIS-LYON
Paris - Angers	297	PARIS MONTPARNASSE => ANGERS SAINT LAUD	ANGERS SAINT LAUD => PARIS MONTPARNASSE
Paris - Angoulême	430	PARIS MONTPARNASSE => ANGOULEME	ANGOULEME => PARIS MONTPARNASSE
Paris - Annecy	544	PARIS-LYON => ANNECY	ANNECY => PARIS-LYON
Paris - Arras	179	PARIS NORD => ARRAS	ARRAS => PARIS NORD
Paris - Avignon	652	PARIS-LYON => AVIGNON TGV	AVIGNON TGV => PARIS-LYON
Paris - Bellegarde	466	PARIS-LYON => BELLEGARDE (AIN)	BELLEGARDE (AIN) => PARIS-LYON
Paris - Besançon	361	PARIS-LYON => BESANCON-FRANCHE-COMTE-TGV	BESANCON-FRANCHE-COMTE-TGV => PARIS-LYON
Paris - Brest	604	PARIS MONTPARNASSE => BREST	BREST => PARIS MONTPARNASSE
Paris - Chambéry	519	PARIS-LYON => CHAMBERY CHALLES LES EAUX	CHAMBERY CHALLES LES EAUX => PARIS-LYON
Paris - Douai	205	PARIS NORD => DOUAI	DOUAI => PARIS NORD
Paris - Dunkerque	293	PARIS NORD => DUNKERQUE	DUNKERQUE => PARIS NORD
Paris - Laval	289	PARIS MONTPARNASSE => LAVAL	LAVAL => PARIS MONTPARNASSE
Paris - Montchanin	303	PARIS-LYON => LE CREUSOT-MONTCEAU-MONTCHANIN	LE CREUSOT-MONTCEAU-MONTCHANIN => PARIS-LYON
Lille - Marseille	962	LILLE => MARSEILLE ST CHARLES	MARSEILLE ST CHARLES => LILLE
Lyon - Marseille	318	LYON PART DIEU => MARSEILLE ST CHARLES	MARSEILLE ST CHARLES => LYON PART DIEU
Lyon - Montpellier	314	LYON PART DIEU => MONTPELLIER	MONTPELLIER => LYON PART DIEU
Lyon - Rennes	781	LYON PART DIEU => RENNES	RENNES => LYON PART DIEU
Lyon - Marne La Vallée	418	LYON PART DIEU => MARNE LA VALLEE	MARNE LA VALLEE => LYON PART DIEU
Paris - Mâcon Loché	363	PARIS-LYON => MACON LOCHE	MACON LOCHE => PARIS-LYON
Paris - Metz	313	PARIS EST => METZ	METZ => PARIS EST
Paris - Mulhouse	488	PARIS-LYON => MULHOUSE VILLE	MULHOUSE VILLE => PARIS-LYON
Nantes - Strasbourg	438	NANTES => STRASBOURG	STRASBOURG => NANTES
Paris - Nîmes	686	PARIS-LYON => NIMES	NIMES => PARIS-LYON
Paris - Perpignan	895	PARIS-LYON => PERPIGNAN	PERPIGNAN => PARIS-LYON
Paris - Saint-Etienne	485	PARIS-LYON => SAINT ETIENNE CHATEAUCREUX	SAINT ETIENNE CHATEAUCREUX => PARIS-LYON

Paris - Toulon	811	PARIS-LYON => TOULON	TOULON => PARIS-LYON
Paris - Valence	524	PARIS-LYON => VALENCE-ALIXAN TGV	VALENCE-ALIXAN TGV => PARIS-LYON
Paris - Reims	146	PARIS EST => REIMS	REIMS => PARIS EST
Paris - Quimper	600	PARIS MONTPARNASSE => QUIMPER	QUIMPER => PARIS MONTPARNASSE
Paris - Saint-Malo	437	PARIS MONTPARNASSE => ST MALO	ST MALO => PARIS MONTPARNASSE
Paris - Saint-Pierre-des-Corps	220	PARIS MONTPARNASSE => ST PIERRE DES CORPS	ST PIERRE DES CORPS => PARIS MONTPARNASSE
Paris - Vannes	481	PARIS MONTPARNASSE => VANNES	VANNES => PARIS MONTPARNASSE
Bordeaux - Tourcoing	830	BORDEAUX ST JEAN => TOURCOING	TOURCOING => BORDEAUX ST JEAN
Marseille - Marne La Vallée	736	MARSEILLE ST CHARLES => MARNE LA VALLEE	MARNE LA VALLEE => MARSEILLE ST CHARLES
Marseille - Tourcoing	978	MARSEILLE ST CHARLES => TOURCOING	TOURCOING => MARSEILLE ST CHARLES

Annexe 3 : Estimation de l'âge moyen des voies des LGV par axe entre 2012 et 2019

La base de données sur les services de transport ferroviaire de voyageurs de l'Autorité de Régulation des Transports (ART) a permis d'estimer l'évolution de l'âge moyen des voies des LGV par axe entre 2012 et 2019.

Les données utilisées de la base AQST sont les longueurs de voie LGV (en km) et l'âge de ces voies par ancienne région administrative. Elles couvrent la période 2015-2018. La majorité des régions ne sont traversées que par un seul axe TGV (Atlantique, Nord, Est, Sud-Est), ce qui nous a permis d'estimer l'âge moyen des voies LGV par axe TGV en calculant une moyenne par axe pondérée de la longueur de voie LGV par ancienne région. En Île-de-France, des voies LGV traversent la région pour chacun des axes, et nous ne disposons pas du détail de la longueur de voie LGV par axe pour cette région. Elle a donc été retirée du calcul. Le même problème se présente pour l'ancienne région Picardie – traversée par les axes TGV Nord et Est - que nous avons écartée également. En revanche, les voies LGV traversant l'ancienne région Nord-Pas-de-Calais appartiennent toutes à l'axe TGV Nord. Cette région a donc été intégrée au calcul.

Les tableaux suivants présentent les calculs détaillés pour les années 2015-2019. Pour l'année 2019, seule la longueur de voie du contournement ferroviaire entre Nîmes et Montpellier (LGV) a été ajoutée par rapport à l'année 2018.

Source : ART 2019		Longueur du réseau LGV en km de voie				Âge du réseau LGV
2015	Région/Axe	Atlantique	Nord	Est	Sud-Est	ans
	Auvergne Rhône-Alpes	-	-	-	564	21
	Bourgogne-Franche-Comté	-	-	-	888	21
	Bretagne	-	-	-	-	N/D
	Centre-Val de Loire	395	-	-	-	24
	Grand Est	-	-	467	-	8
	Hauts-de-France (complexe)					21
	dont Nord-Pas-de-Calais	-	446	-	-	23
	dont Picardie (non pris en compte)					19
	Normandie	-	-	-	-	N/D
	Nouvelle-Aquitaine	-	-	-	-	N/D
	Occitanie	-	-	-	104	15
	Pays de la Loire	66	-	-	-	26
	Provence-Alpes-Côte d'Azur	-	-	-	252	15
	Île-de-France (non pris en compte)					21
	Âge du réseau LGV par axe	24,2	23,0	7,9	19,8	Moyenne pondérée de la longueur de voie par région

Source : ART 2019		Longueur du réseau LGV en km de voie				Âge du réseau LGV
2016	Région/Axe	Atlantique	Nord	Est	Sud-Est	ans
	Auvergne Rhône-Alpes	-	-	-	566	22
	Bourgogne-Franche-Comté	-	-	-	928	21
	Bretagne	-	-	-	-	N/D
	Centre-Val de Loire	397	-	-	-	25
	Grand Est	-	-	712	-	8
	Hauts-de-France (complexe)					23
	dont Nord-Pas-de-Calais	-	441	-	-	24
	dont Picardie (non pris en compte)					20
	Normandie	-	-	-	-	N/D
	Nouvelle-Aquitaine	-	-	-	-	N/D
	Occitanie	-	-	-	105	16
	Pays de la Loire	64	-	-	-	27
	Provence-Alpes-Côte d'Azur	-	-	-	257	16
	Île-de-France (non pris en compte)					22
	Âge du réseau LGV par axe	25,0	24,0	8,0	20,4	Moyenne pondérée de la longueur de voie par région

Source : ART 2019		Longueur du réseau LGV en km de voie				Âge du réseau LGV
2017	Région/Axe	Atlantique	Nord	Est	Sud-Est	ans
	Auvergne Rhône-Alpes	-	-	-	562	23
	Bourgogne-Franche-Comté	-	-	-	940	22
	Bretagne	-	-	-	-	N/D
	Centre-Val de Loire	391	-	-	-	25
	Grand Est	-	-	696	-	8
	Hauts-de-France (complexe)					24
	dont Nord-Pas-de-Calais	-	445	-	-	25
	dont Picardie (non pris en compte)					21
	Normandie	-	-	-	-	N/D
	Nouvelle-Aquitaine	-	-	-	-	N/D
	Occitanie	-	-	-	106	17
	Pays de la Loire	64	-	-	-	26
	Provence-Alpes-Côte d'Azur	-	-	-	251	17
	Île-de-France (non pris en compte)					23
	LGV Bretagne - Pays de la Loire (BPL)	363	-	-	-	1
	LGV Sud Europe Atlantique (SEA)	603	-	-	-	1
	Contournement Nîmes Montpellier (CNM)				-	N/D
	Âge du réseau LGV	8,9	25,0	7,8	21,3	Moyenne pondérée de la longueur de voie par région

2018	Région/Axe	Longueur du réseau LGV en km de voie				Âge du réseau LGV ans
		Atlantique	Nord	Est	Sud-Est	
	Auvergne Rhône-Alpes	-	-	-	560	23
	Bourgogne-Franche-Comté	-	-	-	929	23
	Bretagne	-	-	-	-	N/D
	Centre-Val de Loire	392	-	-	-	26
	Grand Est	-	-	699	-	9
	Hauts-de-France (complexe)	-	-	-	-	24
	dont Nord-Pas-de-Calais	-	442	-	-	26
	dont Picardie (non pris en compte)	-	-	-	-	20
	Normandie	-	-	-	-	N/D
	Nouvelle-Aquitaine	-	-	-	-	4
	Occitanie	-	-	-	108	18
	Pays de la Loire	64	-	-	-	27
	Provence-Alpes-Côte d'Azur	-	-	-	253	18
	Île-de-France (non pris en compte)	-	-	-	-	23
	LGV Bretagne - Pays de la Loire (BPL)	363	-	-	-	2
	LGV Sud Europe Atlantique (SEA)	603	-	-	-	2
	Contournement Nîmes Montpellier (CNM)	-	-	-	-	N/D
	Âge du réseau LGV	9,7	26,0	8,8	22,1	Moyenne pondérée de la longueur de voie par région

2019	Source : ART 2019	Longueur du réseau LGV en km de voie				Âge du réseau LGV ans
	Région/Axe	Atlantique	Nord	Est	Sud-Est	
	Auvergne Rhône-Alpes	-	-	-	560	24
	Bourgogne-Franche-Comté	-	-	-	929	24
	Bretagne	-	-	-	-	N/D
	Centre-Val de Loire	392	-	-	-	27
	Grand Est	-	-	699	-	10
	Hauts-de-France (complexe)	-	-	-	-	25
	dont Nord-Pas-de-Calais	-	442	-	-	27
	dont Picardie	-	-	-	-	21
	Normandie	-	-	-	-	N/D
	Nouvelle-Aquitaine	-	-	-	-	5
	Occitanie	-	-	-	108	19
	Pays de la Loire	64	-	-	-	28
	Provence-Alpes-Côte d'Azur	-	-	-	253	19
	Île-de-France (non pris en compte)	-	-	-	-	24
	LGV Bretagne - Pays de la Loire (BPL)	363	-	-	-	3
	LGV Sud Europe Atlantique (SEA)	603	-	-	-	3
	Contournement Nîmes Montpellier (CNM)	-	-	-	122	2
	Âge du réseau LGV	10,7	27,0	9,8	21,8	Moyenne pondérée de la longueur de voie par région

Le tableau présente l'évolution de l'âge moyen des voies LGV par axe TGV entre 2012 et 2019, reprenant les résultats des tableaux précédents pour la période 2015-2019. L'estimation présentée pour la période 2012-2014 est issue d'un prolongement par continuité sur la base du taux d'évolution relative constaté entre 2015 et 2016.

Année	Axe TGV	Âge moyen des voies LGV	Méthode de calcul
Source : ART 2019			
2012	Atlantique	21,6	Prolongement par continuité par rapport à l'année suivante
	Nord	20,0	
	Est	4,9	
	Sud-Est	17,4	
2013	Atlantique	22,4	Prolongement par continuité par rapport à l'année suivante
	Nord	21,0	
	Est	5,9	
	Sud-Est	18,2	
2014	Atlantique	23,3	Prolongement par continuité par rapport à l'année suivante
	Nord	22,0	
	Est	6,9	
	Sud-Est	19,0	
2015	Atlantique	24,2	Moyenne pondérée de la longueur de voie par ancienne région
	Nord	23,0	
	Est	7,9	
	Sud-Est	19,8	
2016	Atlantique	25,0	Moyenne pondérée de la longueur de voie par ancienne région
	Nord	24,0	
	Est	8,0	
	Sud-Est	20,4	
2017	Atlantique	8,9	Moyenne pondérée de la longueur de voie par ancienne région
	Nord	25,0	
	Est	7,8	
	Sud-Est	21,3	
2018	Atlantique	9,7	Moyenne pondérée de la longueur de voie par ancienne région
	Nord	26,0	
	Est	8,8	
	Sud-Est	22,1	

2019	Atlantique	10,7	Moyenne pondérée de la longueur de voie par ancienne région
	Nord	27,0	
	Est	9,8	
	Sud-Est	21,8	

Annexe 4 : Nombre de bifurcations majeures des voies LGV par liaison TGV

Une bifurcation correspond à une séparation des voies LGV rencontrées sur la liaison, qui amène le TGV à emprunter l'une des deux voies. Une séparation des voies offrant plus de 2 chemins possibles est comptée comme une seule bifurcation. Les bifurcations des LGV en Île-de-France sont prises en compte.

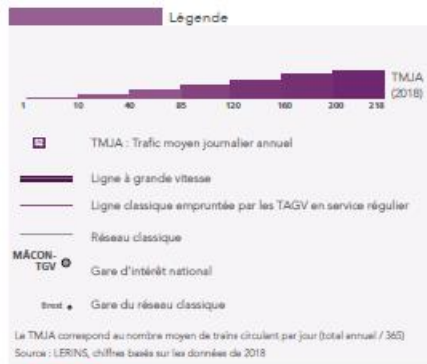
Le nombre de bifurcations LGV par liaison a été compté à l'aide de la carte « Le trafic des trains aptes à la Grande Vitesse » (pages 42-43) de l'atlas du réseau ferré en France – édition 2020 publié par le gestionnaire d'infrastructure SNCF Réseau⁶, présentée ci-dessous.

Les 3 bifurcations LGV hors Île-de-France sont :

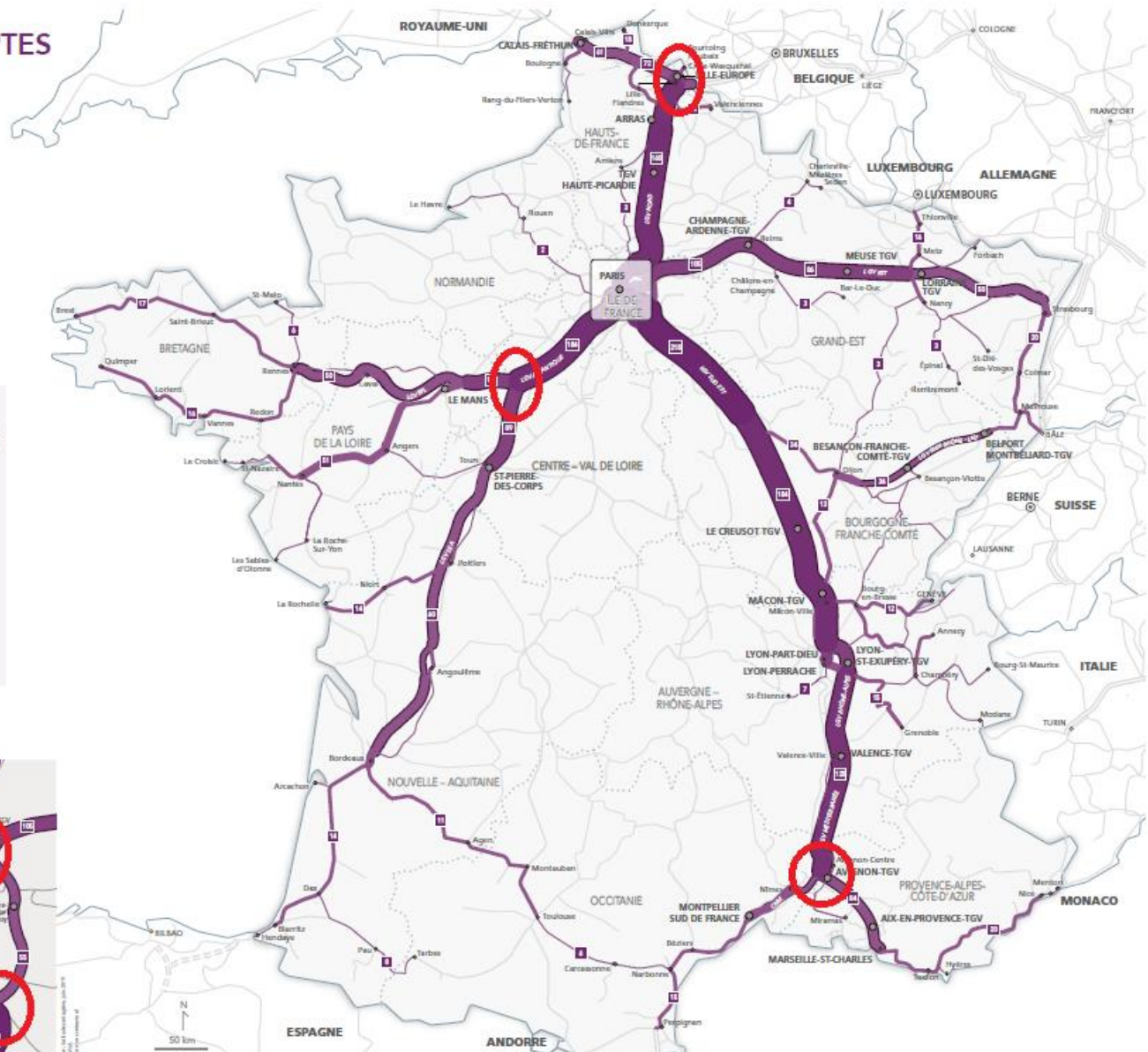
1. La bifurcation de la LGV Méditerranée au sud de la commune des Angles (Gard) en deux branches : la branche « Marseille » ou PACA au sud-est, et la branche « Languedoc » au sud-ouest ;
2. La bifurcation de la LGV Atlantique au sud de la commune de Courtalain (Eure-et-Loir) en deux branches : la branche « ouest » qui dessert la Bretagne et le Pays de la Loire, en partie via la LGV Bretagne-Pays de la Loire ; et la branche « sud-ouest » qui dessert notamment Bordeaux via la LGV Sud Europe Atlantique ;
3. La bifurcation de la LGV Nord au niveau du triangle de Frétil, simplifié ici en deux branches : la branche « Calais » vers l'ouest, qui conduit notamment au tunnel sous la manche, emprunté par les Eurostar, et la branche « est » vers la Belgique qui conduit à la LGV 1 puis à Bruxelles.

⁶ Atlas disponible en ligne : <https://www.sncf-reseau.com/fr/carte/atlas-reseau-ferre-francais>

LE TRAFIC DES TRAINS APTES À LA GRANDE VITESSE



bifurcations LGV retenues



Annexe 5 : Estimation du nombre d'arrêts moyen par liaison TGV au cours d'un Jour Ouvrable de Base en 2020

Une estimation du nombre d'arrêts moyen des trains circulant sur chaque liaison au cours d'un Jour Ouvrable de Base (JOB) - le 17 octobre 2020 – a été calculé pour l'amplitude horaire retenue de 6h à 22h. Les données utilisées ont été recueillies sur le site <https://www.horaires-de-trains.fr/> et proviennent de l'OpenData de la SNCF. Pour simplifier l'étude, nous avons considéré que le nombre d'arrêts moyen ne dépend pas en première approche du sens de circulation sur la liaison TGV. Si cette hypothèse semble raisonnable pour un grand nombre de liaisons, elle n'est pas nécessairement exacte sur la totalité des liaisons.

Liaison Sens 1	Liaison Sens 2	NbArretsMoyen
AIX EN PROVENCE TGV => PARIS-LYON	PARIS-LYON => AIX EN PROVENCE TGV	1,0
ANGERS SAINT LAUD => PARIS MONTPARNASSE	PARIS MONTPARNASSE => ANGERS SAINT LAUD	0,7
ANGOULEME => PARIS MONTPARNASSE	PARIS MONTPARNASSE => ANGOULEME	1,0
ANNECY => PARIS-LYON	PARIS-LYON => ANNECY	2,3
ARRAS => PARIS NORD	PARIS NORD => ARRAS	0
AVIGNON TGV => PARIS-LYON	PARIS-LYON => AVIGNON TGV	0
BELLEGARDE (AIN) => PARIS-LYON	PARIS-LYON => BELLEGARDE (AIN)	1,3
BESANCON-FRANCHE-COMTE-TGV => PARIS-LYON	PARIS-LYON => BESANCON-FRANCHE-COMTE-TGV	1,4
BORDEAUX ST JEAN => PARIS MONTPARNASSE	PARIS MONTPARNASSE => BORDEAUX ST JEAN	0
BREST => PARIS MONTPARNASSE	PARIS MONTPARNASSE => BREST	4,0
CHAMBERY CHALLES LES EAUX => PARIS-LYON	PARIS-LYON => CHAMBERY CHALLES LES EAUX	1,0
DIJON VILLE => PARIS-LYON	PARIS-LYON => DIJON VILLE	0,3
DOUAI => PARIS NORD	PARIS NORD => DOUAI	1,0
DUNKERQUE => PARIS NORD	PARIS NORD => DUNKERQUE	3,0
GRENOBLE => PARIS-LYON	PARIS-LYON => GRENOBLE	1,0
LA ROCHELLE VILLE => PARIS MONTPARNASSE	PARIS MONTPARNASSE => LA ROCHELLE VILLE	2,3
LAVAL => PARIS MONTPARNASSE	PARIS MONTPARNASSE => LAVAL	0,7
LE CREUSOT-MONTCEAU-MONTCHANIN => PARIS-LYON	PARIS-LYON => LE CREUSOT-MONTCEAU-MONTCHANIN	0
LE MANS => PARIS MONTPARNASSE	PARIS MONTPARNASSE => LE MANS	0
LILLE => LYON PART DIEU	LYON PART DIEU => LILLE	2,2
LILLE => PARIS NORD	PARIS NORD => LILLE	0,4
LYON PART DIEU => MARNE LA VALLEE	MARNE LA VALLEE => LYON PART DIEU	0
LYON PART DIEU => PARIS-LYON	PARIS-LYON => LYON PART DIEU	0,6
MACON LOCHE => PARIS-LYON	PARIS-LYON => MACON LOCHE	0,4
MARNE LA VALLEE => MARSEILLE ST CHARLES	MARSEILLE ST CHARLES => MARNE LA VALLEE	2,7
MARSEILLE ST CHARLES => LILLE	LILLE => MARSEILLE ST CHARLES	5,0
MARSEILLE ST CHARLES => LYON PART DIEU	LYON PART DIEU => MARSEILLE ST CHARLES	2,1
MARSEILLE ST CHARLES => PARIS-LYON	PARIS-LYON => MARSEILLE ST CHARLES	1,5

METZ => PARIS EST	PARIS EST => METZ	0,3
MONTPELLIER => LYON PART DIEU	LYON PART DIEU => MONTPELLIER	1,8
MONTPELLIER => PARIS-LYON	PARIS-LYON => MONTPELLIER	2,0
MULHOUSE VILLE => PARIS-LYON	PARIS-LYON => MULHOUSE VILLE	2,3
NANCY => PARIS EST	PARIS EST => NANCY	0,6
NANTES => PARIS MONTPARNASSE	PARIS MONTPARNASSE => NANTES	1,3
NICE VILLE => PARIS-LYON	PARIS-LYON => NICE VILLE	5,3
NIMES => PARIS-LYON	PARIS-LYON => NIMES	0,8
PERPIGNAN => PARIS-LYON	PARIS-LYON => PERPIGNAN	7,0
POITIERS => PARIS MONTPARNASSE	PARIS MONTPARNASSE => POITIERS	1,3
QUIMPER => PARIS MONTPARNASSE	PARIS MONTPARNASSE => QUIMPER	4,8
REIMS => PARIS EST	PARIS EST => REIMS	0
RENNES => LYON PART DIEU	LYON PART DIEU => RENNES	2,0
RENNES => PARIS MONTPARNASSE	PARIS MONTPARNASSE => RENNES	0,7
SAINT ETIENNE CHATEAUCREUX => PARIS-LYON	PARIS-LYON => SAINT ETIENNE CHATEAUCREUX	1,0
ST MALO => PARIS MONTPARNASSE	PARIS MONTPARNASSE => ST MALO	2,7
ST PIERRE DES CORPS => PARIS MONTPARNASSE	PARIS MONTPARNASSE => ST PIERRE DES CORPS	0,6
STRASBOURG => NANTES	NANTES => STRASBOURG	7,0
STRASBOURG => PARIS EST	PARIS EST => STRASBOURG	0,1
TOULON => PARIS-LYON	PARIS-LYON => TOULON	1,4
TOULOUSE MATABIAU => PARIS MONTPARNASSE	PARIS MONTPARNASSE => TOULOUSE MATABIAU	3,0
TOURCOING => BORDEAUX ST JEAN	BORDEAUX ST JEAN => TOURCOING	7,0
TOURCOING => MARSEILLE ST CHARLES	MARSEILLE ST CHARLES => TOURCOING	4,0
TOURS => PARIS MONTPARNASSE	PARIS MONTPARNASSE => TOURS	2,0
VALENCE-ALIXAN TGV => PARIS-LYON	PARIS-LYON => VALENCE-ALIXAN TGV	0
VANNES => PARIS MONTPARNASSE	PARIS MONTPARNASSE => VANNES	1,4

Annexe 6 : Estimation de la longueur de ligne parcourue sur le réseau classique par liaison TGV sur la période 2012-2019

Afin d'estimer la longueur de ligne en km parcourue sur le réseau classique (hors LGV) pour chaque liaison TGV de la base de l'AQST, nous avons utilisé la carte « Le trafic des trains aptes à la Grande Vitesse » (pages 42-43) de l'atlas du réseau ferré en France – édition 2020 publié par le gestionnaire d'infrastructure SNCF Réseau⁷. Elle présente notamment la répartition sur le territoire métropolitain des LGV et des lignes classiques empruntées par les TGV en service régulier. Nous avons essayé d'en estimer la longueur par une lecture de la carte., mais il y a une part d'incertitude inévitable dans l'estimation effectuée ici.

La longueur de ligne parcourue sur le réseau classique a évolué pour un certain nombre de liaisons entre 2012 et 2019, du fait de la mise en service de nouvelles LGV en 2016, 2017 et 2018. La liste des nouvelles LGV mises en service sur cette période est fournie dans le tableau ci-dessous. Nous avons intégré ces évolutions dans notre base de données en appliquant le changement de longueur de ligne parcourue sur le réseau classique à partir de la 1^{ère} année complète d'utilisation de la nouvelle LGV par la liaison après sa mise en service.

Ligne	Axe	Longueur totale de LGV	Source longueur de Ligne	Première année complète d'utilisation après la mise en service
LGV Est Européenne – tronçon	Est	106		2017
LGV Sud Europe Atlantique	Atlantique	302	Wikipédia	2017
LGV Bretagne-Pays de la Loire	Atlantique	182		2017
Contournement Nîmes - Montpellier	Sud-Est	61	WikiSara	2018

L'estimation des longueurs de ligne parcourues sur le réseau classique par liaison au cours des périodes 2012-2016, 2017 et 2018-2019 sont présentées dans le tableau ci-dessous :

⁷ Atlas disponible en ligne : <https://www.sncf-reseau.com/fr/carte/atlas-reseau-ferre-francais>

Liaison\ Longueur parcourue sur le réseau classique (en km)	période 2012-2016	En 2017	depuis 2018
AIX EN PROVENCE TGV => PARIS-LYON	0	0	0
ANGERS SAINT LAUD => PARIS MONTPARNASSE	91	91	91
ANGOULEME => PARIS MONTPARNASSE	190	17	17
ANNECY => PARIS-LYON	153	153	153
ARRAS => PARIS NORD	0	0	0
AVIGNON TGV => PARIS-LYON	0	0	0
BELLEGARDE (AIN) => PARIS-LYON	68	68	68
BESANCON-FRANCHE-COMTE-TGV => PARIS-LYON	74	74	74
BORDEAUX ST JEAN => PARIS MONTPARNASSE	302	0	0
BORDEAUX ST JEAN => TOURCOING	319	17	17
BREST => PARIS MONTPARNASSE	403	221	221
CHAMBERY CHALLES LES EAUX => PARIS-LYON	113	113	113
DIJON VILLE => PARIS-LYON	45	45	45
DOUAI => PARIS NORD	0	0	0
DUNKERQUE => PARIS NORD	17	17	17
GRENOBLE => PARIS-LYON	85	85	85
LA ROCHELLE VILLE => PARIS MONTPARNASSE	202	108	108
LAVAL => PARIS MONTPARNASSE	128	0	0
LE CREUSOT-MONTCEAU-MONTCHANIN => PARIS-LYON	0	0	0
LE MANS => PARIS MONTPARNASSE	11	11	11
LILLE => LYON PART DIEU	0	0	0
LILLE => MARSEILLE ST CHARLES	0	0	0
LILLE => PARIS NORD	0	0	0
LYON PART DIEU => LILLE	0	0	0
LYON PART DIEU => MARNE LA VALLEE	0	0	0
LYON PART DIEU => MARSEILLE ST CHARLES	0	0	0
LYON PART DIEU => MONTPELLIER	51	51	0
LYON PART DIEU => PARIS-LYON	0	0	0
LYON PART DIEU => RENNES	191	9	9
MACON LOCHE => PARIS-LYON	0	0	0
MARNE LA VALLEE => LYON PART DIEU	0	0	0
MARNE LA VALLEE => MARSEILLE ST CHARLES	0	0	0
MARSEILLE ST CHARLES => LILLE	0	0	0
MARSEILLE ST CHARLES => LYON PART DIEU	0	0	0
MARSEILLE ST CHARLES => MARNE LA VALLEE	0	0	0
MARSEILLE ST CHARLES => PARIS-LYON	0	0	0
MARSEILLE ST CHARLES => TOURCOING	17	17	17
METZ => PARIS EST	11	11	11
MONTPELLIER => LYON PART DIEU	51	51	0
MONTPELLIER => PARIS-LYON	51	51	0
MULHOUSE VILLE => PARIS-LYON	111	111	111
NANCY => PARIS EST	23	23	23

NANTES => PARIS MONTPARNASSE	170	170	170
NANTES => STRASBOURG	285	179	179
NICE VILLE => PARIS-LYON	170	170	170
NIMES => PARIS-LYON	0	0	0
PARIS EST => METZ	11	11	11
PARIS EST => NANCY	23	23	23
PARIS EST => REIMS	0	0	0
PARIS EST => STRASBOURG	106	0	0
PARIS MONTPARNASSE => ANGERS SAINT LAUD	91	91	91
PARIS MONTPARNASSE => ANGOULEME	190	17	17
PARIS MONTPARNASSE => BORDEAUX ST JEAN	302	0	0
PARIS MONTPARNASSE => BREST	403	221	221
PARIS MONTPARNASSE => LA ROCHELLE VILLE	202	108	108
PARIS MONTPARNASSE => LAVAL	128	0	0
PARIS MONTPARNASSE => LE MANS	11	11	11
PARIS MONTPARNASSE => NANTES	170	170	170
PARIS MONTPARNASSE => POITIERS	94	0	0
PARIS MONTPARNASSE => QUIMPER	395	213	213
PARIS MONTPARNASSE => RENNES	182	0	0
PARIS MONTPARNASSE => ST MALO	239	57	57
PARIS MONTPARNASSE => ST PIERRE DES CORPS	0	0	0
PARIS MONTPARNASSE => TOULOUSE MATABIAU	517	215	215
PARIS MONTPARNASSE => TOURS	0	0	0
PARIS MONTPARNASSE => VANNES	301	119	119
PARIS NORD => ARRAS	0	0	0
PARIS NORD => DOUAI	0	0	0
PARIS NORD => DUNKERQUE	17	17	17
PARIS NORD => LILLE	0	0	0
PARIS-LYON => AIX EN PROVENCE TGV	0	0	0
PARIS-LYON => ANNECY	153	153	153
PARIS-LYON => AVIGNON TGV	0	0	0
PARIS-LYON => BELLEGARDE (AIN)	68	68	68
PARIS-LYON => BESANCON-FRANCHE-COMTE-TGV	74	74	74
PARIS-LYON => CHAMBERY CHALLES LES EAUX	113	113	113
PARIS-LYON => DIJON VILLE	45	45	45
PARIS-LYON => GRENOBLE	85	85	85
PARIS-LYON => LE CREUSOT-MONTCEAU-MONTCHANIN	0	0	0
PARIS-LYON => LYON PART DIEU	0	0	0
PARIS-LYON => MACON LOCHE	0	0	0
PARIS-LYON => MARSEILLE ST CHARLES	0	0	0
PARIS-LYON => MONTPELLIER	51	51	0
PARIS-LYON => MULHOUSE VILLE	111	111	111
PARIS-LYON => NICE VILLE	170	170	170
PARIS-LYON => NIMES	0	0	0

PARIS-LYON => PERPIGNAN	176	176	125
PARIS-LYON => SAINT ETIENNE CHATEAUCREUX	51	51	51
PARIS-LYON => TOULON	48	48	48
PARIS-LYON => VALENCE-ALIXAN TGV	0	0	0
PERPIGNAN => PARIS-LYON	176	176	125
POITIERS => PARIS MONTPARNASSE	94	0	0
QUIMPER => PARIS MONTPARNASSE	395	213	213
REIMS => PARIS EST	0	0	0
RENNES => LYON PART DIEU	191	9	9
RENNES => PARIS MONTPARNASSE	182	0	0
SAINTE ETIENNE CHATEAUCREUX => PARIS-LYON	51	51	51
ST MALO => PARIS MONTPARNASSE	239	57	57
ST PIERRE DES CORPS => PARIS MONTPARNASSE	0	0	0
STRASBOURG => NANTES	285	179	179
STRASBOURG => PARIS EST	106	0	0
TOULON => PARIS-LYON	48	48	48
TOULOUSE MATABIAU => PARIS MONTPARNASSE	517	215	215
TOURCOING => BORDEAUX ST JEAN	319	17	17
TOURCOING => MARSEILLE ST CHARLES	17	17	17
TOURS => PARIS MONTPARNASSE	0	0	0
VALENCE-ALIXAN TGV => PARIS-LYON	0	0	0
VANNES => PARIS MONTPARNASSE	301	119	119

Annexe 7 : Résultats détaillés des régressions pour chaque variable explicative retenue

Ponctualité des TGV par liaisons :

- **Le nombre de circulations ferroviaires à l'année sur la liaison :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbCirc_Liaison_an, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.1192  -4.4288  -0.6948   3.3356  22.6554

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   13.5055129   0.4263476   31.677  <2e-16 ***
NbCirc_Liaison_an -0.0001917  0.0001075   -1.783   0.075 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.541 on 810 degrees of freedom
Multiple R-squared:  0.003909, Adjusted R-squared:  0.00268
F-statistic: 3.179 on 1 and 810 DF, p-value: 0.07496
```

Le nombre de circulations de trains à l'année sur la liaison a un effet significatif au seuil de 10% mais pas au seuil 5%. La valeur du coefficient montre que cet effet est très faible : un ajout de 1000 circulations de trains à l'année n'entraîne une baisse du taux de retard que de 0,2 points environ. De plus, le signe de la corrélation est dans le sens opposé de l'effet attendu, qui voudrait qu'un trafic plus intense entraîne une hausse des retards sur la liaison.

On note également que le test de Fisher indique que la régression est significative au seuil 10% mais pas au seuil 5%, donc les résultats observés ne sont pas totalement validés.

- **La durée moyenne du trajet (en minutes) sur la liaison au cours de l'année considérée :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ DuréeTrajet, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-9.9978 -4.4817 -0.7322  3.2442 22.4643

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.716068   0.450058   26.03 < 2e-16 ***
DuréeTrajet  0.006954   0.002538    2.74  0.00627 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.526 on 810 degrees of freedom
Multiple R-squared:  0.009186, Adjusted R-squared:  0.007962
F-statistic: 7.509 on 1 and 810 DF, p-value: 0.006273

```

D'après le modèle, plus le temps de trajet sur la liaison (en minutes) est long, plus le taux de retard de cette liaison est élevé, ce qui est conforme à l'intuition. L'effet de la durée du trajet sur le taux de retard est significatif au seuil de 1%.

La valeur du coefficient indique qu'en moyenne, une heure de trajet supplémentaire engendre une hausse du taux de retard de 0,42 points.

- **La distance moyenne du trajet (en km) sur la liaison :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ DistanceMoy, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-9.7767 -4.3368 -0.8473  3.1113 21.5228

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.286788   0.470394   21.868 < 2e-16 ***
DistanceMoy  0.005280   0.000893    5.912 4.97e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.436 on 810 degrees of freedom
Multiple R-squared:  0.04137, Adjusted R-squared:  0.04019
F-statistic: 34.96 on 1 and 810 DF, p-value: 4.968e-09

```

D'après le modèle, plus la distance kilométrique ferroviaire entre les 2 gares est grande, plus le taux de retard de cette liaison est élevé, ce qui est conforme à l'intuition, et avec le précédent modèle

portant sur la durée du trajet. L'effet de la distance kilométrique ferroviaire sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, un trajet en TGV plus long de 100 km présente un taux de retard plus élevé de 0,53 points.

- **La vitesse moyenne du trajet (en km/h) des circulations de trains sur la liaison :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ VitesseMoyenne, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.0994  -4.1981  -0.7553   3.1120  21.4559

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.717825   1.123917   5.977 3.40e-09 ***
VitesseMoyenne 0.032715   0.005929   5.518 4.62e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.45 on 810 degrees of freedom
Multiple R-squared:  0.03623,    Adjusted R-squared:  0.03504
F-statistic: 30.45 on 1 and 810 DF,  p-value: 4.619e-08
```

D'après le modèle, plus la vitesse moyenne des trains circulant sur la liaison est grande, plus le taux de retard de cette liaison est élevé. L'effet de la vitesse moyenne sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, une vitesse des trains plus élevée de 10 km/h se traduit par un taux de retard plus élevé de 0,33 points.

- **L'âge moyen des voies LGV du 1^{er} axe TGV sur lequel circule les TGV :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :


```

Call:
lm(formula = TxRetard ~ AgeLGV_axe, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-9.4863 -4.5002 -0.7774  3.2815 22.5052

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.40925   0.66040  18.790  <2e-16 ***
AgeLGV_axe   0.02258   0.03395   0.665   0.506
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.55 on 810 degrees of freedom
Multiple R-squared:  0.000546, Adjusted R-squared:  -0.0006879
F-statistic: 0.4425 on 1 and 810 DF,  p-value: 0.5061

```

D'après le modèle, l'âge moyen des voies LGV du 1er axe TGV sur lequel circule les TGV qui emprunte la liaison considérée n'a pas d'effet significatif sur le taux de retard. Il serait probablement nécessaire de disposer de données sur l'âge moyen des voies LGV empruntées par la liaison, et non celui de l'axe TGV complet, pour observer une éventuelle corrélation. Malheureusement, de telles données ne sont pas à la disposition de l'AQST à ce stade.

On note également que le test de Fisher indique que la régression n'est pas significative au seuil 10% donc le modèle de régression linéaire n'est pas fiable pour cette variable.

- **Le fait que le TGV parte ou non de Paris intra-muros :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ DépartParis, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.4010 -3.9572 -0.6164  3.3853 21.2020

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.2770   0.2470  57.791  <2e-16 ***
DépartParis -3.3403   0.3752  -8.902  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.298 on 810 degrees of freedom
Multiple R-squared:  0.08912, Adjusted R-squared:  0.088
F-statistic: 79.25 on 1 and 810 DF,  p-value: < 2.2e-16

```

D'après le modèle, les TGV partant de Paris sont plus ponctuels que les TGV qui partent d'ailleurs en France. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, si l'on compare deux échantillons de 100 circulations de TGV chacun, le premier constitué uniquement de trains partant de Paris, et le second de trains ne partant pas de Paris, 3 trains de plus seront en retard dans le deuxième échantillon. L'impact de cette variable est donc important.

- **Le fait que la TGV arrive ou non à Paris intra-muros :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ ArrivéeParis, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-9.3918 -4.4702 -0.7806  3.2983 22.7352

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.7438     0.2588   49.24  <2e-16 ***
ArrivéeParis  0.1966     0.3931    0.50   0.617
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.551 on 810 degrees of freedom
Multiple R-squared:  0.0003088, Adjusted R-squared:  -0.0009253
F-statistic: 0.2502 on 1 and 810 DF, p-value: 0.617
```

D'après le modèle, le fait que le TGV arrive à Paris n'a pas d'effet significatif à la hausse ou à la baisse sur la probabilité qu'il soit en retard à l'arrivée.

On note également que le test de Fisher indique que la régression n'est pas significative au seuil 10% donc le modèle de régression linéaire n'est pas fiable pour cette variable.

- **Le fait que la liaison TGV parte de Paris ou arrive à Paris intra-muros (liaison radiale) :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ Liaison_radiale, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-12.1994  -4.0036  -0.4471   3.4131  20.2174

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    18.6334     0.4872   38.24 <2e-16 ***
Liaison_radiale -6.6948     0.5233  -12.79 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.064 on 810 degrees of freedom
Multiple R-squared:  0.1681,    Adjusted R-squared:  0.1671
F-statistic: 163.7 on 1 and 810 DF,  p-value: < 2.2e-16

```

D'après le modèle, les TGV qui circulent sur des liaisons radiales – i. e. qui partent ou arrivent à Paris – sont plus ponctuels que les TGV qui circulent sur des liaisons transversales reliant 2 gares de province. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, si l'on compare deux échantillons de 100 circulations de TGV chacun, le premier constitué uniquement de trains circulant sur des liaisons radiales, et le second de TGV circulant sur des liaisons transversales, plus de 6 trains de plus seront en retard à l'arrivée dans le deuxième échantillon. La différence de ponctualité entre les liaisons radiales et les liaisons transversales est donc prononcée.

Sachant que les liaisons transversales effectuées en TGV empruntent en quasi-totalité les nœuds ferroviaires de Lyon Part-Dieu et Marseille Saint-Charles, nous étudierons plus loin l'impact de ces nœuds sur le taux de retard de façon plus détaillée.

- **Le fait que le TGV parte ou non de Marseille-Saint-Charles :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ DépartMarseille, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-11.2385  -4.3290  -0.7422   3.3360  22.8408

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.6382     0.1947  64.909 < 2e-16 ***
DépartMarseille  5.7373     1.0678   5.373 1.01e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.455 on 810 degrees of freedom
Multiple R-squared:  0.03442,    Adjusted R-squared:  0.03322
F-statistic: 28.87 on 1 and 810 DF,  p-value: 1.012e-07

```

D'après le modèle, les TGV partant de Marseille Saint-Charles sont moins ponctuels que les TGV qui partent d'ailleurs en France. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des TGV partant de Marseille Saint-Charles est supérieur de plus de 5,7 points à celui des TGV partant d'ailleurs en France. Une explication possible à ce constat pourrait être qu'il y a une saturation du réseau ferroviaire au niveau du nœud ferroviaire de Marseille Saint-Charles. Une autre explication possible pourrait être que la gare de Marseille Saint-Charles soit davantage affectée par des problèmes de gestion du trafic en gare.

- **Le fait que le TGV arrive ou non à Marseille-Saint-Charles :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ ArrivéeMarseille, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-11.3559  -4.3441  -0.7653   3.4037  22.8047

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.6743     0.1959   64.70 < 2e-16 ***
ArrivéeMarseille  4.6516     1.0743    4.33 1.68e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.488 on 810 degrees of freedom
Multiple R-squared:  0.02262,    Adjusted R-squared:  0.02142
F-statistic: 18.75 on 1 and 810 DF,  p-value: 1.678e-05
```

D'après le modèle, les TGV arrivant en gare de Marseille Saint-Charles sont moins ponctuels que les TGV qui arrivent à une gare située ailleurs en France. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des TGV arrivant à Marseille est supérieur de plus de 4,6 points à celui des TGV à destination d'une autre gare en France. Une explication possible à ce constat pourrait être qu'il y a une saturation du réseau ferroviaire au niveau du nœud ferroviaire de Marseille Saint-Charles. Une autre explication possible pourrait être que la gare de Marseille Saint-Charles soit davantage affectée par des problèmes de gestion du trafic en gare.

- **Le fait que le TGV parte ou non de Lyon-Part-Dieu :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ DépartLyon, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-14.7488  -4.1610  -0.4918   3.3325  19.6972

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.4588     0.1917  64.985 <2e-16 ***
DépartLyon    7.1580     0.8430   8.491 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.32 on 810 degrees of freedom
Multiple R-squared:  0.08174, Adjusted R-squared:  0.08061
F-statistic: 72.1 on 1 and 810 DF, p-value: < 2.2e-16

```

D'après le modèle, les TGV partant de la gare de Lyon Part-Dieu sont moins ponctuels que les TGV qui partent d'ailleurs en France. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des TGV partant de Lyon Part-Dieu est supérieur de plus de 7,2 points à celui des TGV partant d'ailleurs en France. Une explication possible à ce constat pourrait être qu'il y a une saturation du réseau ferroviaire au niveau du nœud ferroviaire lyonnais. Or la question du nœud ferroviaire lyonnais a souvent été soulevée par les experts ferroviaires, ce qui appuie cette hypothèse. Une autre explication possible pourrait être que la gare de Lyon Part-Dieu soit davantage affectée par des problèmes de gestion du trafic en gare.

- **Le fait que le TGV arrive ou non à Lyon-Part-Dieu :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ ArrivéeLyon, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.9173  -4.3857  -0.8265   3.2388  22.7435

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.7355     0.1995  63.82 <2e-16 ***
ArrivéeLyon   1.8078     0.8774   2.06  0.0397 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.537 on 810 degrees of freedom
Multiple R-squared:  0.005214, Adjusted R-squared:  0.003986
F-statistic: 4.246 on 1 and 810 DF, p-value: 0.03967

```

D'après le modèle, les TGV arrivant en gare de Lyon Part-Dieu sont moins ponctuels que les TGV à destination d'une autre gare en France. L'effet sur le taux de retard est significatif au seuil de 5%, mais pas au seuil 1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des TGV arrivant en gare de Lyon Part-Dieu est supérieur de plus de 1,8 points à celui des TGV arrivant ailleurs en France.

En comparant les 2 modèles précédents, on peut remarquer que les TGV qui partent de la gare de Lyon Part-Dieu affichent en moyenne des taux de retard significativement plus élevés que ceux qui arrivent à cette gare. Cela tend à appuyer l'hypothèse de l'existence de problèmes de gestion du trafic à cette gare, même s'il est impossible de vérifier cette hypothèse sans effectuer un travail de terrain.

- **Le fait que le TGV parte de ou arrive à Lyon-Part-Dieu ou Marseille-Saint-Charles :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ DépartOuArr_Lyon_Marseille, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-13.2487  -4.0164  -0.4177   3.6046  20.0423

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)         12.1137    0.2012  60.221  <2e-16 ***
DépartOuArr_Lyon_Marseille  4.7611    0.5190   9.174  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.284 on 810 degrees of freedom
Multiple R-squared:  0.09413,    Adjusted R-squared:  0.09301
F-statistic: 84.17 on 1 and 810 DF,  p-value: < 2.2e-16
```

D'après le modèle, les TGV qui partent ou arrivent en gare de Lyon Part-Dieu ou Marseille Saint-Charles sont moins ponctuels que les autres TGV. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard de ces TGV est plus élevé de 4,8 points.

Ce résultat est une généralisation des résultats relatifs aux taux de retard observés plus élevés pour les liaisons concernées, constatés avec les 4 modèles précédents.

- **Le nombre de nœuds ferroviaires majeurs (Paris, Lyon, Marseille) traversés par la liaison TGV :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ NbNoeudsMajeursTraversés, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.937  -4.010  -0.900   3.146  21.190

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)         9.4601    0.4554  20.773 < 2e-16 ***
NbNoeudsMajeursTraversés  2.4144    0.2975   8.116 1.78e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.339 on 810 degrees of freedom
Multiple R-squared:  0.07521,    Adjusted R-squared:  0.07407
F-statistic: 65.88 on 1 and 810 DF,  p-value: 1.779e-15

```

D'après le modèle, plus une circulation TGV traverse les nœuds ferroviaires majeurs (Paris, Lyon, Marseille) et plus la probabilité d'occurrence d'un retard est élevée, ce qui est conforme à l'intuition relativement au niveau de saturation de ces nœuds ferroviaires majeurs sur le réseau ferré français. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, chaque nœud majeur supplémentaire traversé par une liaison se traduit par une hausse du taux de retard de 2,4 points. La traversée d'un nœud ferroviaire majeur augmente donc significativement la probabilité d'occurrence d'un retard pour une circulation TGV.

- **Le fait que le TGV traverse ou non le nœud ferroviaire d'Avignon :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LiaisonPasseParAvignon, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.3465  -4.0590  -0.7195   3.0478  23.6575

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)         11.8215    0.2082  56.79 <2e-16 ***
LiaisonPasseParAvignon  4.4950    0.4397  10.22 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.225 on 810 degrees of freedom
Multiple R-squared:  0.1143,    Adjusted R-squared:  0.1132
F-statistic: 104.5 on 1 and 810 DF,  p-value: < 2.2e-16

```

D'après le modèle, les TGV passant par le triangle ferroviaire des Angles (Gard) sur la LGV Méditerranée (nœud ferroviaire simplifié ici comme Avignon) sont moins ponctuels que les TGV qui n'y passent pas. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des TGV passant par les Angles (Gard) est supérieur de plus de 4,5 points à celui des autres TGV.

Au Sud des Angles (Gard), la LGV Méditerranée se sépare en deux branches : la branche « Marseille » ou PACA au Sud-Est, et la branche « Languedoc » au Sud-Ouest. Les liaisons TGV qui desservent ou bien partent du littoral méditerranéen en empruntant la LGV Méditerranée sont donc moins ponctuelles.

Au-delà de l'impact du nœud ferroviaire majeur de Marseille Saint-Charles, une explication partielle de la moins bonne-punctualité de ces liaisons pourraient être l'impact des événements météorologiques qui surviennent dans ces régions et détériorent les infrastructures ferroviaires. C'est notamment le cas des fortes chaleurs en été, et des épisodes méditerranéens (pluies intenses engendrant fréquemment des inondations).

- **Le fait que le TGV traverse ou non le nœud ferroviaire bordelais :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ LiaisonPasseParBordeaux, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-9.4349 -4.4149 -0.8199  3.3176 22.6921

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.7869     0.1989  64.286 <2e-16 ***
LiaisonPasseParBordeaux  1.0062     0.9720   1.035  0.301
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.548 on 810 degrees of freedom
Multiple R-squared:  0.001321, Adjusted R-squared:  8.831e-05
F-statistic: 1.072 on 1 and 810 DF, p-value: 0.3009
```

L'effet sur le taux de retard du passage par le nœud ferroviaire de Bordeaux n'est pas significatif au seuil de 10%.

- **Le fait que le TGV traverse ou non le nœud ferroviaire dijonnais :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :


```

Call:
lm(formula = TxRetard ~ LiaisonPasseParDijon, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-9.757 -4.294 -0.610  3.122 22.370

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.1085     0.1967  66.625 < 2e-16 ***
LiaisonPasseParDijon -4.7275     0.8092  -5.842 7.47e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.438 on 810 degrees of freedom
Multiple R-squared:  0.04043, Adjusted R-squared:  0.03925
F-statistic: 34.13 on 1 and 810 DF, p-value: 7.466e-09

```

D'après le modèle, les TGV traversant Dijon sont plus ponctuels que les TGV qui n'y passent pas. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des TGV traversant Dijon est inférieur de plus de 4,7 points à celui des autres TGV. La différence est donc marquée, et le passage par le nœud ferroviaire dijonnais n'a pas d'impact négatif sur la ponctualité des TGV.

Les liaisons TGV qui traversent Dijon empruntent toutes la LGV Rhin-Rhône, qui dessert les villes de Dijon, Besançon, Belfort, Montbéliard et Mulhouse. Cette LGV est récente (mise en service en 2011) et était relativement peu intensément utilisée en 2018 d'après la carte « Le trafic des trains aptes à la Grande Vitesse » (pages 42-43) de l'atlas du réseau ferré en France – édition 2020 publié par le gestionnaire d'infrastructure SNCF Réseau⁸. En effet, le nombre moyen de trains circulant sur la section dimensionnante de cette LGV était de 34 circulations TGV par jour en 2018 (en trafic moyen journalier annuel). Cela en fait la LGV la moins intensément utilisée pour cette année, les suivantes étant le second tronçon de la LGV Est Européenne, entre Baudrecourt et Vendenheim, avec 50 circulations TGV par jour, ex aequo avec la LGV Bretagne-Pays de la Loire. Ces éléments expliquent peut-être en partie la bonne performance en termes de ponctualité des liaisons passant par Dijon.

- **Le fait que le TGV traverse ou non le nœud ferroviaire lillois :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

⁸ Atlas disponible en ligne : <https://www.sncf-reseau.com/fr/carte/atlas-reseau-ferre-francais>

```

Call:
lm(formula = TxRetard ~ LiaisonPasseParLille, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-11.9984  -4.3214  -0.7397   3.4856  19.6053

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.5507     0.2007  62.532 < 2e-16 ***
LiaisonPasseParLille  3.3238     0.6936   4.792 1.96e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.475 on 810 degrees of freedom
Multiple R-squared:  0.02757, Adjusted R-squared:  0.02637
F-statistic: 22.97 on 1 and 810 DF, p-value: 1.961e-06

```

D'après le modèle, les TGV traversant le nœud ferroviaire lillois (axe TGV Nord) sont moins ponctuelles que les autres TGV. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des liaisons concernées est plus élevé de 3,3 points. L'impact de cette variable est donc important.

Au-delà de l'effet d'engorgement qui peut se produire au niveau du nœud ferroviaire à proprement parler, le moins bon niveau de ponctualité des liaisons passant par Lille peut être en partie s'expliquer par une plus grande probabilité d'occurrence et une plus grande intensité des épisodes de froids et des épisodes neigeux qui se produisent en région Hauts-de-France. Ces événements météorologiques peuvent entraîner des retards de trains, notamment parce qu'ils dégradent les infrastructures ferroviaires.

- **Le fait que le TGV traverse ou non le nœud ferroviaire rennais :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LiaisonPasseParRennes, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
 -9.8111  -4.1771  -0.7336   3.1599  22.3159

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.1631     0.2046  64.321 < 2e-16 ***
LiaisonPasseParRennes  -2.8256     0.5952  -4.747 2.44e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.476 on 810 degrees of freedom
Multiple R-squared:  0.02707, Adjusted R-squared:  0.02587
F-statistic: 22.54 on 1 and 810 DF, p-value: 2.436e-06

```

D'après le modèle, les TGV traversant le nœud ferroviaire rennais sont plus ponctuelles que les autres TGV. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des liaisons concernées est moins élevé de 2,8 points. L'impact de cette variable est donc important.

Les liaisons TGV qui traversent Rennes empruntent toutes la LGV Bretagne-Pays de la Loire. Cette LGV est très récente (mise en service en 2017) et était relativement peu intensément utilisée en 2018 d'après la carte « Le trafic des trains aptes à la Grande Vitesse » (pages 42-43) de l'atlas du réseau ferré en France – édition 2020 publié par SNCF Réseau². En effet, le nombre moyen de trains circulant sur la section dimensionnante de cette LGV était de 50 circulations TGV par jour en 2018 (en trafic moyen journalier annuel, soit la deuxième LGV la moins intensément utilisée pour cette année. Les voies ferroviaires et les autres infrastructures ferroviaires de la LGV Bretagne-Pays de la Loire sont donc en très bon état pour le moment, ce qui peut expliquer pour partie la bonne ponctualité des trains sur les liaisons passant par Rennes.

- **Le fait que le TGV traverse ou non le nœud ferroviaire du Mans :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ LiaisonPasseParLeMans, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-9.5234 -4.4789 -0.7514  3.3348 22.6036

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.8754     0.2029   63.455 <2e-16 ***
LiaisonPasseParLeMans -0.5890     0.7227  -0.815   0.415
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.549 on 810 degrees of freedom
Multiple R-squared:  0.0008191, Adjusted R-squared:  -0.0004144
F-statistic: 0.664 on 1 and 810 DF, p-value: 0.4154
```

L'effet sur le taux de retard du passage par le nœud ferroviaire du Mans – où s'effectue notamment la jonction entre la LGV Atlantique et la LGV Bretagne-Pays de la Loire - n'est pas significatif au seuil de 10%.

- **Le fait que le TGV traverse ou non le nœud ferroviaire de Tours :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LiaisonPasseParTours, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-9.4039 -4.4257 -0.8134  3.2935 22.7231

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.7559     0.2100  60.736 <2e-16 ***
LiaisonPasseParTours  0.5204     0.5605   0.928  0.353
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.549 on 810 degrees of freedom
Multiple R-squared:  0.001063, Adjusted R-squared:  -0.0001701
F-statistic: 0.8621 on 1 and 810 DF, p-value: 0.3534

```

L'effet sur le taux de retard du passage par le nœud ferroviaire de Tours (que l'on confond ici volontairement avec Saint-Pierre-des-Corps, sachant que les gares de ces 2 communes limitrophes ne sont séparées que de 2,7 km) – où s'effectue notamment la jonction entre la LGV Atlantique et la LGV Sud Europe Atlantique - n'est pas significatif au seuil de 10%.

- **Le fait que le TGV parte ou bien arrive ou non au niveau d'un des nœuds ferroviaires secondaires retenus :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ DépartOuArr_NoeudSecondaire, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.5626 -4.5233 -0.7263  3.3581 20.5304

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.2339     0.2159  56.661 < 2e-16 ***
DépartOuArr_NoeudSecondaire  2.7147     0.4612   5.887 5.77e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.437 on 810 degrees of freedom
Multiple R-squared:  0.04103, Adjusted R-squared:  0.03984
F-statistic: 34.65 on 1 and 810 DF, p-value: 5.767e-09

```

D'après le modèle, les TGV dont le départ ou l'arrivée s'effectue au niveau de l'un des nœuds ferroviaires secondaires retenus (Avignon, Bordeaux, Le Mans, Lille, Tours) sont moins ponctuels que les autres TGV. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard de ces TGV est supérieur de près de 2,7 points à celui des autres TGV.

Néanmoins, les résultats des modèles de régression testés précédemment pour chacun de ces nœuds ferroviaires secondaires montrent des disparités importantes quant à l'effet du passage par l'un de ces nœuds sur le taux de retard. Ainsi, les TGV traversant les nœuds ferroviaires des Andes (Gard) près d'Avignon ou de Lille sont moins ponctuels que les autres TGV, tandis que ceux qui traversent Dijon ou Rennes sont plus ponctuels. L'absence de significativité statistique de la régression ne permet pas de conclure quant à un éventuel impact du passage d'un TGV par les nœuds ferroviaires de Bordeaux, Le Mans ou Tours sur la probabilité qu'il subisse un retard.

- **Le nombre de nœuds ferroviaires secondaires (Avignon, Bordeaux, Le Mans, Lille, Tours) traversés par la liaison TGV :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbNoeudsSecondaires, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-9.5483 -4.3906 -0.6564  3.3806 22.2317

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.5984    0.3156  36.753 < 2e-16 ***
NbNoeudsSecondaires  1.6489    0.3356   4.913 1.08e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.471 on 810 degrees of freedom
Multiple R-squared:  0.02894, Adjusted R-squared:  0.02774
F-statistic: 24.14 on 1 and 810 DF, p-value: 1.084e-06
```

D'après le modèle, plus une circulation TGV traverse de nœuds ferroviaires secondaires retenus (Avignon, Bordeaux, Le Mans, Lille, Tours), plus la probabilité d'occurrence d'un retard est élevée, ce qui est conforme à l'intuition relativement à la situation d'engorgement que connaissent les circulations ferroviaires au niveau de ces nœuds ferroviaires secondaires du réseau ferré français. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, chaque nœud secondaire supplémentaire traversé par une liaison se traduit par une hausse du taux de retard de 1,6 points. On peut observer que la hausse du taux de retard est légèrement moins forte que celle provoquée par le passage par un nœud majeur (Paris, Lyon, Marseille), qui était de 2,4 points.

- **Le nombre de bifurcations majeures LGV sur la liaison TGV :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbBifurcationsLGV, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-9.4303 -4.0293 -0.6773  3.1489 20.0787

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.4264    0.3869  24.367  <2e-16 ***
NbBifurcationsLGV  2.9870    0.2988   9.996  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.238 on 810 degrees of freedom
Multiple R-squared:  0.1098,    Adjusted R-squared:  0.1087
F-statistic: 99.92 on 1 and 810 DF,  p-value: < 2.2e-16
```

D’après le modèle, plus une circulation TGV emprunte de bifurcations majeures des voies LGV, plus elle est susceptible de connaître des retards. L’effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu’en moyenne, chaque bifurcation supplémentaire des voies sur le réseau LGV emprunté par la liaison engendre une hausse du taux de retard de 3 points, ce qui montre un impact important de cette variable.

La corrélation observée peut peut-être s’expliquer en partie par l’occurrence de retards supplémentaires liés à une gestion plus complexe du trafic en amont de la bifurcation lorsque plusieurs branches LGV se rejoignent. Il devient alors nécessaire de synchroniser le plan de circulations des TGV des 2 branches, en particulier en mode dégradé lorsqu’une perturbation du trafic se produit. Une explication inverse à la corrélation observée pourrait être l’effet du mauvais niveau de ponctualité des liaisons TGV reliant les branches « Languedoc » et « PACA » qui constituent la majeure partie des liaisons sur lesquels les TGV effectuent le plus de bifurcations sur le réseau LGV (2 bifurcations). Il apparait donc ici difficile de savoir dans quel sens s’effectue l’éventuel lien de causalité entre un nombre élevé de bifurcations sur le réseau LGV emprunté par la liaison TGV et un niveau de ponctualité relativement mauvais.

- **Le nombre de bifurcations majeures LGV (hors Île-de-France) sur la liaison TGV :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ NbBifurcationsLGV_horsIDF, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-9.7850 -4.4387 -0.6758  3.2795 21.9950

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.6178    0.3079  37.727 < 2e-16 ***
NbBifurcationsLGV_horsIDF  1.8662    0.3711   5.028 6.1e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.467 on 810 degrees of freedom
Multiple R-squared:  0.03027, Adjusted R-squared:  0.02907
F-statistic: 25.28 on 1 and 810 DF, p-value: 6.1e-07

```

D'après le modèle, plus une circulation TGV emprunte de bifurcations majeures des voies LGV hors de l'Île-de-France, plus elle est susceptible de connaître des retards. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, chaque bifurcation supplémentaire des voies sur le réseau LGV emprunté par la liaison engendre une hausse du taux de retard de 1,9 points, soit une hausse moins importante que pour les bifurcations des voies sur le réseau LGV Île-de-France comprise.

Les explications et les remarques sur la corrélation observée sont identiques à celles formulées pour le modèle précédent concernant les bifurcations des voies LGV y compris en Île-de-France.

- **Le nombre estimé de bifurcations sur le réseau classique (hors LGV) sur la liaison TGV :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ NbBifurcationsRéseauClassique, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.2348 -4.2024 -0.7298  3.4386 21.7818

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.8608    0.2653  52.243 < 2e-16 ***
NbBifurcationsRéseauClassique -0.9127    0.1627  -5.608 2.81e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.447 on 810 degrees of freedom
Multiple R-squared:  0.03737, Adjusted R-squared:  0.03618
F-statistic: 31.45 on 1 and 810 DF, p-value: 2.811e-08

```

D'après le modèle, plus une circulation TGV emprunte de bifurcations sur le réseau classique (hors LGV), moins elle est susceptible de connaître des retards, ce qui est contraire à l'intuition relativement à l'effet de cette variable. L'effet sur le taux de retard est significatif au seuil de 0,1%.

Il n'y a pas d'explication évidente à la corrélation observée ici.

- **Le niveau du trafic sur la section la plus intensément utilisée du réseau LGV emprunté par la liaison TGV :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ TMJA_max_Liaison_2019, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-9.8501 -4.4040 -0.9364  3.3150 22.2769

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.506660   1.022145  10.279  <2e-16 ***
TMJA_max_Liaison_2019  0.012364   0.005343   2.314  0.0209 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.533 on 810 degrees of freedom
Multiple R-squared:  0.006568, Adjusted R-squared:  0.005342
F-statistic: 5.355 on 1 and 810 DF, p-value: 0.02091
```

D'après le modèle, plus la section dimensionnante du réseau LGV emprunté par la liaison TGV connaît un trafic intense (en nombre moyen annuel de circulations de TGV par jour), moins la liaison est ponctuelle, ce qui est conforme à l'intuition. L'effet sur le taux de retard est significatif au seuil de 5% mais pas au seuil 1%.

La valeur du coefficient indique qu'en moyenne, une liaison qui traverse un tronçon dimensionnant du réseau LGV emprunté par 100 TGV de plus présentera un taux de retard plus élevé de 1,2 points.

D'après la carte « Le trafic des trains aptes à la Grande Vitesse » (pages 42-43) de l'atlas du réseau ferré en France – édition 2020 publié par SNCF Réseau, l'intensité d'utilisation des différents tronçons du réseau LGV varie de 50 circulations TGV/jour - pour les LGV Rhin-Rhône, Bretagne-Pays de la Loire, et le second tronçon de la LGV Est Européenne, entre Baudrecourt et Vendenheim - à 218 – pour la section de LGV Sud-Est située au nord de la bifurcation vers Dijon. D'après le modèle, cette différence d'intensité d'utilisation des LGV entraînerait une différence de près de 2 points dans le taux de retard.

- **Le nombre estimé d'arrêts moyen par liaison TGV au cours d'un Jour Ouvrable de Base :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :


```

Call:
lm(formula = TxRetard ~ NbMoyenArrêts_2019, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.0594  -4.4977  -0.6741   3.4373  22.4347

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.1715    0.2699  45.101 < 2e-16 ***
NbMoyenArrêts_2019  0.3967    0.1136   3.493 0.000503 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.51 on 810 degrees of freedom
Multiple R-squared:  0.01484,    Adjusted R-squared:  0.01362
F-statistic: 12.2 on 1 and 810 DF,  p-value: 0.0005034

```

D'après le modèle, plus une circulation TGV effectue d'arrêts, plus elle est susceptible de connaître des retards, ce qui est conforme à l'intuition relativement à l'effet de cette variable. En effet, certains événements peuvent survenir lors d'un arrêt en gare et retarder le départ du train : par exemple, malaise voyageur, blocage (volontaire ou non) des portes par un voyageur retardant le train, dysfonctionnement du matériel du TGV lors de son redémarrage, etc.

L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, chaque arrêt supplémentaire effectuée engendre une hausse du taux de retard de 0,4 points. Or, d'après les estimations calculées dans cette étude, le nombre moyen d'arrêts effectués par liaison en 2019 était de 1,6 et 88 liaisons effectuaient moins de 3 arrêts, contre 16 liaisons effectuant 3 arrêts ou plus. L'impact du nombre d'arrêts sur la ponctualité n'est donc pas très fort d'après le modèle.

La longueur estimée de ligne parcourue sur le réseau classique par liaison TGV :

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LongueurRéseauClassique, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.2433  -4.0637  -0.6713   3.2438  21.6097

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.869312    0.235987  58.772 < 2e-16 ***
LongueurRéseauClassique -0.012811    0.001746  -7.338 5.28e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.376 on 810 degrees of freedom
Multiple R-squared:  0.06233,    Adjusted R-squared:  0.06118
F-statistic: 53.85 on 1 and 810 DF,  p-value: 5.276e-13

```

D'après le modèle, plus la distance parcourue sur le réseau classique par un TGV est grande, moins il est susceptible de connaître un retard, ce qui est contraire à l'intuition relativement à l'effet de cette variable. L'effet sur le taux de retard est significatif au seuil de 0,1%.

Il n'y a pas d'explication évidente à la corrélation observée ici.

- **Le niveau du trafic de l'ensemble des circulations de trains (y compris fret et « haut-le-pied ») sur la section la plus intensément utilisée du réseau emprunté par la liaison TGV en 2000 :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ TMJA_max_Total_Liaison_2000, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.5047  -4.2852  -0.8723   3.3143  21.6223

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.718784   0.587370   16.55 < 2e-16 ***
TMJA_max_Total_Liaison_2000 0.011377   0.002032    5.60 2.94e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.447 on 810 degrees of freedom
Multiple R-squared:  0.03727, Adjusted R-squared:  0.03608
F-statistic: 31.36 on 1 and 810 DF, p-value: 2.936e-08
```

D'après le modèle, plus la section dimensionnante des lignes ferroviaires empruntées par la liaison TGV connaît un trafic intense (en nombre moyen annuel de circulations de trains par jour, y compris trains de fret et trafic « haut-le-pied »), moins la liaison est ponctuelle, ce qui est conforme à l'intuition. L'effet sur le taux de retard est significatif au seuil de 1%.

La valeur du coefficient indique qu'en moyenne, une liaison qui traverse un tronçon dimensionnant du réseau ferré emprunté par 100 trains de plus présentera un taux de retard plus élevé de 1,1 points.

En effet, lorsqu'une circulation TGV emprunte des lignes très fréquentées, elle est davantage exposée au risque de subir un retard qui se propage sur le réseau d'une circulation de train à l'autre, et cela sans que ce retard ne puisse nécessairement être « rattrapé » au cours du trajet, puisque le débit de circulations de trains sur la ligne est déjà très élevé et que la gestion du trafic en situation dégradée doit assurer au mieux la sécurité des voyageurs.

D'après la carte SNCF « TRAFIC TOTAL 2000 » publiée en 2000 par la Direction du Contrôle de gestion de la SNCF, le débit journalier des différents tronçons du réseau ferré emprunté par les TGV varie de 162,8 circulations de trains/jour - pour les TGV au départ ou à destination de Paris Gare de l'Est (axe Est) - à 363,7 – pour les TGV au départ ou à destination de Paris Gare de Lyon (axe Sud-Est). D'après le modèle, cette différence de débit entre les sections dimensionnantes des axes Est et Sud-Est du réseau LGV entraînerait une différence de 2,3 points dans le taux de retard.

Ponctualité des trains Intercités par liaisons :

- **Le nombre de circulations ferroviaires à l'année sur la liaison :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbCirc_Liaison_an, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.666  -4.250  -1.424   3.063  27.490

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.5299061  0.5149896  26.272  <2e-16 ***
NbCirc_Liaison_an -0.0001585  0.0002899  -0.547   0.585
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.487 on 429 degrees of freedom
Multiple R-squared:  0.0006963, Adjusted R-squared:  -0.001633
F-statistic: 0.2989 on 1 and 429 DF,  p-value: 0.5849
```

Le nombre de circulations de trains à l'année sur la liaison ne présente pas d'effet significatif sur le taux de retard au seuil de 10%.

- **La durée moyenne du trajet (en minutes) sur la liaison au cours de l'année considérée :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ TpsTrajet, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-12.654  -4.020  -1.116   2.958  24.799

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.201538   0.524471  21.358 < 2e-16 ***
TpsTrajet    0.007483   0.001519   4.925 1.21e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.313 on 429 degrees of freedom
Multiple R-squared:  0.05351, Adjusted R-squared:  0.05131
F-statistic: 24.26 on 1 and 429 DF, p-value: 1.206e-06

```

D'après le modèle, plus le temps de trajet sur la liaison (en minutes) est long, plus le taux de retard de cette liaison est élevé, ce qui est conforme à l'intuition. L'effet de la durée du trajet sur le taux de retard est significatif au seuil de 1%.

La valeur du coefficient indique qu'en moyenne, une heure de trajet supplémentaire engendre une hausse du taux de retard de 0,45 points. L'impact sur le taux de retard n'est donc pas très prononcé.

- **La distance moyenne du trajet (en km) sur la liaison :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ DistanceMoy, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-12.448  -3.905  -1.096   2.950  24.419

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.212697   0.570844  17.89 < 2e-16 ***
DistanceMoy  0.007364   0.001158   6.36 5.18e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.203 on 429 degrees of freedom
Multiple R-squared:  0.08616, Adjusted R-squared:  0.08403
F-statistic: 40.45 on 1 and 429 DF, p-value: 5.182e-10

```

D'après le modèle, plus la distance kilométrique ferroviaire entre les 2 gares est grande, plus le taux de retard de cette liaison est élevé, ce qui est conforme à l'intuition, et avec le précédent modèle portant sur la durée du trajet. L'effet de la distance kilométrique ferroviaire sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, un trajet en train Intercités plus long de 100 km présente un taux de retard plus élevé de 0,44 points.

- **La vitesse moyenne du trajet (en km/h) des circulations de trains sur la liaison :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ VitesseMoyenne, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.297  -4.312  -1.482   3.019  27.669

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.619819   1.351799   9.336  <2e-16 ***
VitesseMoyenne  0.007008   0.013430   0.522   0.602
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.487 on 429 degrees of freedom
Multiple R-squared:  0.0006343, Adjusted R-squared:  -0.001695
F-statistic: 0.2723 on 1 and 429 DF, p-value: 0.6021
```

La vitesse moyenne des circulations de trains sur la liaison Intercités ne présente pas d'impact significatif au seuil de 10% sur le taux de retard.

- **La fait que la liaison Intercités soit ou non une liaison de nuit :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LiaisonNuit, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-12.557  -4.073  -1.346   2.638  25.590

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.8746     0.3397  37.896 < 2e-16 ***
LiaisonNuit   2.5131     0.8199   3.065  0.00231 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.419 on 429 degrees of freedom
Multiple R-squared:  0.02143, Adjusted R-squared:  0.01915
F-statistic: 9.395 on 1 and 429 DF, p-value: 0.002313

```

D'après le modèle, les liaisons Intercités de nuit sont moins ponctuelles que les autres liaisons Intercités. L'effet sur le taux de retard est significatif au seuil de 1%.

La valeur du coefficient indique qu'en moyenne, un trajet en train Intercités de nuit présente un taux de retard plus élevé de 2,5 points. Il n'y a pas d'explication évidente à cette moins bonne ponctualité des liaisons de nuit.

- **Le fait que le train Intercités parte ou non de Paris intra-muros :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ DépartParis, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-11.505  -4.325  -1.302   2.830  26.643

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.6553     0.3962  31.944 < 2e-16 ***
DépartParis   1.6796     0.6364   2.639  0.00862 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.437 on 429 degrees of freedom
Multiple R-squared:  0.01597, Adjusted R-squared:  0.01368
F-statistic: 6.964 on 1 and 429 DF, p-value: 0.008618

```

D'après le modèle, les trains Intercités partant de Paris sont moins ponctuels que les trains Intercités qui partent d'ailleurs en France. L'effet sur le taux de retard est significatif au seuil de 1%.

La valeur du coefficient indique qu'en moyenne, si l'on compare deux échantillons de 100 circulations de Intercités chacun, le premier constitué uniquement de trains partant de Paris, et le second de trains ne partant pas de Paris, entre 1 et 2 trains de plus seront en retard dans le premier échantillon.

- **Le fait que le train Intercités arrive ou non à Paris intra-muros :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ ArrivéeParis, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-11.164  -4.161  -1.339   3.089  26.983

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.9944     0.3919  35.708 < 2e-16 ***
ArrivéeParis -1.8313     0.6392  -2.865  0.00438 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.428 on 429 degrees of freedom
Multiple R-squared:  0.01877,    Adjusted R-squared:  0.01648
F-statistic: 8.207 on 1 and 429 DF,  p-value: 0.004378
```

D'après le modèle, les trains Intercités arrivant à Paris sont moins ponctuels que les trains Intercités ayant pour terminus une gare ailleurs en France. L'effet sur le taux de retard est significatif au seuil de 1%.

La valeur du coefficient indique qu'en moyenne, si l'on compare deux échantillons de 100 circulations de Intercités chacun, le premier constitué uniquement de trains arrivant à Paris, et le second de trains n'arrivant pas à Paris, entre 1 et 2 trains de plus seront en retard dans le second échantillon.

- **Le fait que la liaison Intercités parte de Paris ou arrive à Paris intra-muros (liaison radiale) :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ Liaison_radiale, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.596  -4.262  -1.438   2.985  27.712

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   13.4370    0.6425  20.915 <2e-16 ***
Liaison_radiale -0.1715    0.7353  -0.233  0.816
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.489 on 429 degrees of freedom
Multiple R-squared:  0.0001268, Adjusted R-squared:  -0.002204
F-statistic: 0.0544 on 1 and 429 DF,  p-value: 0.8157

```

Le fait que la liaison Intercités soit une liaison radiale ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

- **Le fait que le train Intercités parte ou non de Marseille-Saint-Charles :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ DépartMarseille, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.295  -4.122  -1.343   2.428  27.853

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   13.1248    0.3089  42.489 < 2e-16 ***
DépartMarseille  9.7682    2.2673   4.308 2.04e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.353 on 429 degrees of freedom
Multiple R-squared:  0.04147, Adjusted R-squared:  0.03924
F-statistic: 18.56 on 1 and 429 DF,  p-value: 2.041e-05

```

D'après le modèle, les trains Intercités partant de Marseille Saint-Charles sont moins ponctuels que les TGV qui partent d'ailleurs en France. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des trains Intercités partant de Marseille Saint-Charles est supérieur de plus de 9,8 points à celui des trains Intercités partant d'ailleurs en France. Une explication possible à ce constat pourrait être qu'il y a une saturation du réseau ferroviaire au niveau du nœud ferroviaire de Marseille Saint-Charles. Une autre explication

possible pourrait être que la gare de Marseille Saint-Charles soit davantage affectée par des problèmes de gestion du trafic en gare.

- **Le fait que le train Intercités arrive ou non à Marseille-Saint-Charles :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ ArrivéeMarseille, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-11.3559  -4.3441  -0.7653   3.4037  22.8047

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.6743     0.1959   64.70 < 2e-16 ***
ArrivéeMarseille  4.6516     1.0743    4.33 1.68e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.488 on 810 degrees of freedom
Multiple R-squared:  0.02262, Adjusted R-squared:  0.02142
F-statistic: 18.75 on 1 and 810 DF, p-value: 1.678e-05
```

D'après le modèle, les trains Intercités arrivant en gare de Marseille Saint-Charles sont moins ponctuels que les trains Intercités qui arrivent à une gare située ailleurs en France. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des trains Intercités arrivant à Marseille est supérieur de plus de 4,7 points à celui des trains Intercités à destination d'une autre gare en France. Une explication possible à ce constat pourrait être qu'il y a une saturation du réseau ferroviaire au niveau du nœud ferroviaire de Marseille Saint-Charles. Une autre explication possible pourrait être que la gare de Marseille Saint-Charles soit davantage affectée par des problèmes de gestion du trafic en gare.

- **Le fait que le train Intercités parte ou non de Lyon (Lyon-Part-Dieu et Lyon-Perrache) :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ NbCirc_Liaison_an, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.666  -4.250  -1.424   3.063  27.490

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.529061  0.5149896  26.272  <2e-16 ***
NbCirc_Liaison_an -0.0001585  0.0002899  -0.547   0.585
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.487 on 429 degrees of freedom
Multiple R-squared:  0.0006963, Adjusted R-squared:  -0.001633
F-statistic: 0.2989 on 1 and 429 DF,  p-value: 0.5849

```

Le fait que la liaison Intercités parte de Lyon Part-Dieu ou Lyon-Perrache ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

On note que le résultat est différent de celui observé pour les TGV partant de Lyon-Part-Dieu, dont le taux de retard était significativement plus élevé que celui des autres TGV. Il n'y a pas d'explication évidente à ce stade.

- **Le fait que le train Intercités arrive ou non à Lyon (Lyon-Part-Dieu et Lyon-Perrache) :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ ArrivéeLyon, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.510  -4.172  -1.482   2.946  27.637

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.340     0.314  42.481  <2e-16 ***
ArrivéeLyon  -2.925     2.916  -1.003   0.316
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.481 on 429 degrees of freedom
Multiple R-squared:  0.00234, Adjusted R-squared:  1.465e-05
F-statistic: 1.006 on 1 and 429 DF,  p-value: 0.3164

```

Le fait que la liaison Intercités arrive à Lyon Part-Dieu ou Lyon-Perrache ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

On note que le résultat est différent de celui observé pour les TGV arrivant à Lyon-Part-Dieu, dont le taux de retard était significativement plus élevé que celui des autres TGV. Il n'y a pas d'explication évidente à ce stade.

- **Le nombre de nœuds ferroviaires majeurs (Paris, Lyon, Marseille) traversés par la liaison Intercités :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbNoeudsMajeursTraversés, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-13.989  -4.203  -1.137   2.949  24.377

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.3897    0.5994  17.334 < 2e-16 ***
NbNoeudsMajeursTraversés  3.2147    0.5709   5.631 3.25e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.262 on 429 degrees of freedom
Multiple R-squared:  0.06882,    Adjusted R-squared:  0.06665
F-statistic: 31.7 on 1 and 429 DF,  p-value: 3.251e-08
```

D'après le modèle, plus une circulation Intercités traverse les nœuds ferroviaires majeurs (Paris, Lyon, Marseille) et plus la probabilité d'occurrence d'un retard est élevée, ce qui est conforme à l'intuition relativement au niveau de saturation de ces nœuds ferroviaires majeurs sur le réseau ferré français. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, chaque nœud majeur supplémentaire traversé par une liaison se traduit par une hausse du taux de retard de 3,2 points. La traversée d'un nœud ferroviaire majeur augmente donc significativement la probabilité d'occurrence d'un retard pour une circulation Intercités.

- **Le fait que la liaison Intercités traverse ou non le nœud ferroviaire bordelais :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LiaisonTraverseBordeaux, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-14.713  -3.826  -1.019   2.648  28.430

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.5474     0.3146  39.888 < 2e-16 ***
LiaisonTraverseBordeaux  6.5401     0.9236   7.081 5.88e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.14 on 429 degrees of freedom
Multiple R-squared:  0.1047,    Adjusted R-squared:  0.1026
F-statistic: 50.15 on 1 and 429 DF,  p-value: 5.876e-12

```

D'après le modèle, les trains Intercités traversant le nœud ferroviaire bordelais sont moins ponctuelles que les autres trains Intercités. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des liaisons concernées est plus élevé de 6,5 points. L'impact de cette variable est donc très important.

Au-delà de l'effet d'engorgement qui peut se produire au niveau du nœud ferroviaire à proprement parler, le moins bon niveau de ponctualité des liaisons passant par Bordeaux est influencé par les mauvaises performances de ponctualité de la liaison Bordeaux – Marseille, pour laquelle il n'y a pas d'explication évidente *a priori*.

- **Le fait que la liaison Intercités traverse ou non le nœud ferroviaire nantais :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LiaisonTraverseNantes, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.274  -4.196  -1.467   2.529  27.873

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.104     0.320  40.952 <2e-16 ***
LiaisonTraverseNantes  3.351     1.303   2.572  0.0105 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.44 on 429 degrees of freedom
Multiple R-squared:  0.01518,    Adjusted R-squared:  0.01289
F-statistic: 6.614 on 1 and 429 DF,  p-value: 0.01045

```

D'après le modèle, les trains Intercités traversant le nœud ferroviaire nantais sont moins ponctuelles que les autres trains Intercités. L'effet sur le taux de retard est significatif au seuil de 5%, mais pas au seuil 1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des liaisons concernées est plus élevé de 3,4 points. L'impact de cette variable est donc important.

Ce résultat semble *a priori* en accord avec la situation d'engorgement constatée par Réseau Ferré de France en 2014 au niveau de l'étoile ferroviaire de Nantes et des infrastructures de la gare de Nantes dans l'étude technique consacrée au nœud ferroviaire nantais dans le cadre du débat public consacré aux LGV BPL et SEA [4].

- **Le fait que la liaison Intercités traverse ou non le nœud ferroviaire d'Orléans :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ LiaisonTraverseOrléans, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.053  -4.288  -1.373   2.578   28.094

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.8830     0.3582  35.970 <2e-16 ***
LiaisonTraverseOrléans  1.7043     0.7188   2.371  0.0182 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.447 on 429 degrees of freedom
Multiple R-squared:  0.01293, Adjusted R-squared:  0.01063
F-statistic: 5.621 on 1 and 429 DF, p-value: 0.01819
```

D'après le modèle, les trains Intercités traversant le nœud ferroviaire d'Orléans sont légèrement moins ponctuelles que les autres trains Intercités. L'effet sur le taux de retard est significatif au seuil de 5%, mais pas au seuil de 1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des liaisons concernées est plus élevé de 1,7 points.

Au-delà de l'effet d'engorgement qui peut se produire au niveau du nœud ferroviaire à proprement parler, il n'y a pas d'explication évidente *a priori* à cette moins bonne ponctualité relative pour les liaisons concernées.

- **Le fait que la liaison Intercités traverse ou non le nœud ferroviaire de Tours / Saint-Pierre-des-Corps :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LiaisonTraverseTours, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.519  -4.314  -1.491   2.937  27.628

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.3489    0.3273  40.791  <2e-16 ***
LiaisonTraverseTours -0.4860    1.1021  -0.441   0.659
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.488 on 429 degrees of freedom
Multiple R-squared:  0.0004531, Adjusted R-squared:  -0.001877
F-statistic: 0.1945 on 1 and 429 DF, p-value: 0.6594

```

Le fait que la liaison Intercités traverse Tours ne présente pas d'impact significatif sur le taux de retard au seuil de 10%. Le résultat est comparable à celui obtenu pour les TGV, et semble donc indiquer que le nœud ferroviaire de Tours – Saint-Pierre-des-Corps ne génère pas de dégradation majeure de la ponctualité des trains.

- **Le fait que la liaison Intercités traverse ou non le nœud ferroviaire Brive-la-Gaillarde :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LiaisonTraverseBrive, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.419  -4.234  -1.424   2.969  27.728

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.2491    0.3337  39.708  <2e-16 ***
LiaisonTraverseBrive 0.4635    0.9515   0.487   0.626
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.487 on 429 degrees of freedom
Multiple R-squared:  0.0005529, Adjusted R-squared:  -0.001777
F-statistic: 0.2373 on 1 and 429 DF, p-value: 0.6264

```

Le fait que la liaison Intercités traverse Brive-la-Gaillarde ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

- **Le fait que la liaison Intercités traverse ou non le nœud ferroviaire de Toulouse Matabiau :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ LiaisonTraverseToulouse, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-11.014  -4.213  -1.285   2.654  28.170

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.8077     0.3368  38.025 < 2e-16 ***
LiaisonTraverseToulouse  3.0257     0.8299   3.646 0.000299 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.391 on 429 degrees of freedom
Multiple R-squared:  0.03006, Adjusted R-squared:  0.0278
F-statistic: 13.29 on 1 and 429 DF, p-value: 0.000299
```

D'après le modèle, les trains Intercités traversant le nœud ferroviaire de Toulouse Matabiau sont moins ponctuelles que les autres trains Intercités. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des liaisons concernées est plus élevé de 3,0 points. L'impact de cette variable est donc important.

Ce résultat semble *a priori* en accord avec la nécessité d'augmenter la capacité d'accueil des trains en gare de Toulouse- Matabiau soulevée par SNCF Réseau dans son rapport de concertation préalable au projet de modernisation du nœud ferroviaire de Toulouse [\[5\]](#).

- **Le fait que la liaison Intercités traverse ou non le nœud ferroviaire de Narbonne :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LiaisonTraverseNarbonne, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-14.108  -3.910  -1.218   2.461  28.231

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.7468    0.3152  40.439 < 2e-16 ***
LiaisonTraverseNarbonne  6.1807    1.0479   5.898 7.45e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.241 on 429 degrees of freedom
Multiple R-squared:  0.07501, Adjusted R-squared:  0.07286
F-statistic: 34.79 on 1 and 429 DF, p-value: 7.453e-09

```

D'après le modèle, les trains Intercités traversant le nœud ferroviaire de Narbonne sont moins ponctuelles que les autres trains Intercités. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des liaisons concernées est plus élevé de 6,2 points. L'impact de cette variable est donc très important.

Au-delà de l'effet d'engorgement qui peut se produire au niveau du nœud ferroviaire à proprement parler, le moins bon niveau de ponctualité des liaisons passant par Narbonne est influencé par les mauvaises performances de ponctualité de la liaison Bordeaux – Marseille, pour laquelle il n'y a pas d'explication évidente *a priori*.

- **Le fait que la liaison Intercités traverse ou non le nœud ferroviaire de Nevers :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LiaisonTraverseNevers, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.626  -4.193  -1.450   3.085  27.521

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.4561    0.3264  41.224 <2e-16 ***
LiaisonTraverseNevers -1.7015    1.0993  -1.548  0.122
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.471 on 429 degrees of freedom
Multiple R-squared:  0.005553, Adjusted R-squared:  0.003235
F-statistic: 2.396 on 1 and 429 DF, p-value: 0.1224

```

Le fait que la liaison Intercités traverse Nevers ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

- **Le fait que la liaison Intercités traverse ou non le nœud ferroviaire de Dijon :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ LiaisonTraverseDijon, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-15.076  -4.008  -1.377   3.168  23.128

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.9619    0.3179  40.778 < 2e-16 ***
LiaisonTraverseDijon  4.9447    1.2048   4.104 4.86e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.365 on 429 degrees of freedom
Multiple R-squared:  0.03778,    Adjusted R-squared:  0.03554
F-statistic: 16.84 on 1 and 429 DF,  p-value: 4.86e-05
```

D’après le modèle, les trains Intercités traversant le nœud ferroviaire de Dijon sont moins ponctuelles que les autres trains Intercités. L’effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu’en moyenne, le taux de retard des liaisons concernées est plus élevé de 4,9 points. L’impact de cette variable est donc très important.

Au-delà de l’effet d’engorgement qui peut se produire au niveau du nœud ferroviaire à proprement parler, on peut noter que l’ensemble des liaisons concernées sont des liaisons de nuit : Paris – Briançon, Paris - Bourg-Saint-Maurice/ Saint-Gervais-les-Bains, Paris – Nice (ex Train Bleu). Pour autant, il n’y a pas d’explication évidente à la moins bonne ponctualité relative de ces liaisons *a priori*.

- **Le fait que la liaison Intercités traverse ou non le nœud ferroviaire de Caen :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LiaisonTraverseCaen, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.668  -4.177  -1.342   3.096  27.480

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.4978    0.3304  40.851  <2e-16 ***
LiaisonTraverseCaen -1.7210    0.9901  -1.738  0.0829 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.466 on 429 degrees of freedom
Multiple R-squared:  0.006993, Adjusted R-squared:  0.004679
F-statistic: 3.021 on 1 and 429 DF, p-value: 0.08289

```

Le fait que la liaison Intercités traverse Caen ne présente pas d'impact significatif sur le taux de retard au seuil de 5%.

- **Le fait que la liaison Intercités traverse ou non le nœud ferroviaire de Rouen :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LiaisonTraverseRouen, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.742  -4.237  -1.209   2.932  27.405

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.5720    0.3214  42.225  < 2e-16 ***
LiaisonTraverseRouen -3.5816    1.1796  -3.036  0.00254 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.42 on 429 degrees of freedom
Multiple R-squared:  0.02104, Adjusted R-squared:  0.01876
F-statistic: 9.219 on 1 and 429 DF, p-value: 0.002541

```

D'après le modèle, les trains Intercités traversant Rouen sont plus ponctuels que les autres trains Intercités. L'effet sur le taux de retard est significatif au seuil de 1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des trains Intercités traversant Rouen est inférieur de plus de 3,6 points à celui des autres trains Intercités. La différence est donc

marquée, et le passage par le nœud ferroviaire rouennais n'a pas d'impact négatif sur la ponctualité des trains Intercités.

Le fait que les liaisons concernées sont nettement plus courtes (184 km) que la moyenne des liaisons Intercités du panel (420 km) fournit peut-être un élément d'explication possible, parmi d'autres, à la meilleure ponctualité relative de ces liaisons.

- **Le fait que la liaison Intercités traverse ou non Amiens :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ LiaisonTraverseAmiens, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.316  -4.192  -1.509   2.828   27.832

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.1458     0.3218  40.849  <2e-16 ***
LiaisonTraverseAmiens  2.4675     1.2626   1.954  0.0513 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.46 on 429 degrees of freedom
Multiple R-squared:  0.008824, Adjusted R-squared:  0.006514
F-statistic: 3.819 on 1 and 429 DF, p-value: 0.05131
```

Le fait que la liaison Intercités traverse Amiens ne présente pas d'impact significatif sur le taux de retard au seuil de 5%.

- **Le fait que la liaison Intercités traverse ou non Saint-Quentin :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ LiaisonTraverseSaintQuentin, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.539  -4.180  -1.444   3.046  27.608

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      13.3689    0.3181  42.022 <2e-16 ***
LiaisonTraverseSaintQuentin -1.6932    1.6512  -1.025  0.306
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.481 on 429 degrees of freedom
Multiple R-squared:  0.002445, Adjusted R-squared:  0.0001197
F-statistic: 1.051 on 1 and 429 DF, p-value: 0.3057

```

Le fait que la liaison Intercités traverse Saint-Quentin (Aisne) ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

- **Le fait que le train Intercités parte ou bien arrive ou non au niveau d'un des nœuds ferroviaires secondaires retenus :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ DépartOuArr_NoeudSecondaire, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.213  -4.144  -1.172   2.890  28.721

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      12.2565    0.4147  29.556 < 2e-16 ***
DépartOuArr_NoeudSecondaire  2.3318    0.6181   3.772 0.000184 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.384 on 429 degrees of freedom
Multiple R-squared:  0.03211, Adjusted R-squared:  0.02985
F-statistic: 14.23 on 1 and 429 DF, p-value: 0.0001843

```

D'après le modèle, les trains Intercités dont le départ ou l'arrivée s'effectue au niveau de l'un des nœuds ferroviaires secondaires retenus (Bordeaux, Nantes, Toulouse, Orléans, Tours, Brive-la-Gaillarde, Narbonne, Nevers, Dijon, Caen, Amiens, Saint-Quentin) sont moins ponctuels que les autres trains Intercités. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard de ces trains Intercités est supérieur de près de 2,3 points à celui des autres trains Intercités.

Néanmoins, les résultats des modèles de régression testés précédemment pour chacun de ces nœuds ferroviaires secondaires montrent des disparités importantes quant à l'effet du passage par l'un de ces nœuds sur le taux de retard. Ainsi, les trains Intercités traversant les nœuds ferroviaires

de Bordeaux, Nantes, Toulouse, Orléans, Dijon ou Narbonne sont moins ponctuels que les autres trains Intercités, tandis que ceux qui traversent Rouen sont plus ponctuels. L'absence de significativité statistique de la régression ne permet pas de conclure quant à un éventuel impact du passage d'un train Intercités par les nœuds ferroviaires de Tours/Saint-Pierre-des-Corps, Brive-la-Gaillarde, Nevers, Caen, Amiens ou Saint-Quentin sur la probabilité qu'il subisse un retard.

- **Le nombre de nœuds ferroviaires secondaires traversés par la liaison Intercités :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbNoeudsSecondaires, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-13.100  -4.054  -1.113   2.522   28.300

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.9304    0.4826  22.651 < 2e-16 ***
NbNoeudsSecondaires  1.7473    0.2785   6.274 8.62e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.21 on 429 degrees of freedom
Multiple R-squared:  0.08404,    Adjusted R-squared:  0.08191
F-statistic: 39.36 on 1 and 429 DF,  p-value: 8.616e-10
```

D'après le modèle, plus une circulation Intercités traverse de nœuds ferroviaires secondaires retenus (Bordeaux, Nantes, Toulouse, Orléans, Tours, Brive-la-Gaillarde, Narbonne, Nevers, Dijon, Caen, Amiens, Saint-Quentin), plus la probabilité d'occurrence d'un retard est élevée, ce qui est conforme à l'intuition relativement à la situation d'engorgement que connaissent les circulations ferroviaires au niveau de ces nœuds ferroviaires secondaires du réseau ferré français. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, chaque nœud secondaire supplémentaire traversé par une liaison se traduit par une hausse du taux de retard de 1,7 points. On peut observer que la hausse du taux de retard est légèrement moins forte que celle provoquée par le passage par un nœud majeur (Paris, Lyon, Marseille), qui était de 3,2 points.

On peut également noter que l'impact du passage par un nœud ferroviaire secondaire du panel retenu sur le taux de retard annuel moyen d'une liaison Intercités et quantitativement très proche (1,7 points) de celui obtenu dans la modélisation de l'effet équivalent du passage par un nœud ferroviaire secondaire du réseau LGV sur les liaisons TGV (1,6 points).

- **Le fait que la liaison appartienne ou non à l'axe Intercités « Nord-Normandie » :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ Axe_Nord_Normandie, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-11.160  -4.095  -1.087   3.100  26.987

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.9903    0.3874  36.110 < 2e-16 ***
Axe_Nord_Normandie -1.8902    0.6440  -2.935  0.00351 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.425 on 429 degrees of freedom
Multiple R-squared:  0.01969,    Adjusted R-squared:  0.0174
F-statistic: 8.615 on 1 and 429 DF,  p-value: 0.003512
```

D'après le modèle, les trains Intercités de l'axe « Nord-Normandie » du réseau Intercités de 2016 (présenté en Annexe 12) sont plus ponctuels que les autres trains Intercités. L'effet sur le taux de retard est significatif au seuil de 1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des liaisons de l'axe « Nord-Normandie » est moins élevé de 1,9 points.

Le fait que les liaisons concernées sont nettement plus courtes (240 km) que la moyenne des liaisons Intercités du panel (420 km) fournit peut-être un élément d'explication possible à la meilleure ponctualité relative de cet axe Intercités.

- **Le fait que la liaison appartienne ou non à l'axe Intercités « Méditerranée-Atlantique » :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ Axe_Med_Atlantique, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-13.033  -4.124  -1.190   2.716  28.259

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.7186    0.3245  39.200 < 2e-16 ***
Axe_Med_Atlantique  4.6888    0.9166   5.115 4.73e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.3 on 429 degrees of freedom
Multiple R-squared:  0.05749, Adjusted R-squared:  0.05529
F-statistic: 26.17 on 1 and 429 DF, p-value: 4.732e-07

```

D'après le modèle, les trains Intercités de l'axe « Méditerranée-Atlantique » du réseau Intercités de 2016 (présenté en Annexe 12) sont moins ponctuels que les autres trains Intercités. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des liaisons de l'axe « Méditerranée-Atlantique » est plus élevé de 4,7 points. La différence est donc prononcée.

Il n'y a pas d'explication évidente *a priori* à cette moins bonne ponctualité relative de l'axe « Méditerranée-Atlantique ».

- **Le fait que la liaison appartienne ou non à l'axe Intercités « Paris – Centre - Sud-Ouest » :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ Axe_Paris_Centre_SudOuest, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.053  -4.288  -1.373   2.578  28.094

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.8830    0.3582  35.970 <2e-16 ***
Axe_Paris_Centre_SudOuest  1.7043    0.7188   2.371  0.0182 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.447 on 429 degrees of freedom
Multiple R-squared:  0.01293, Adjusted R-squared:  0.01063
F-statistic: 5.621 on 1 and 429 DF, p-value: 0.01819

```

D'après le modèle, les trains Intercités de l'axe « Paris – Centre - Sud-Ouest » du réseau Intercités de 2016 (présenté en Annexe 12) sont moins ponctuels que les autres trains Intercités. L'effet sur le taux de retard est significatif au seuil de 5%, mais pas au seuil de 1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des liaisons de l'axe « Paris – Centre - Sud-Ouest » est plus élevé de 1,7 points.

Le fait que les liaisons concernées sont plus longues (580 km) que la moyenne des liaisons Intercités du panel (420 km) fournit peut-être un élément d'explication possible à la moins bonne ponctualité relative de cet axe Intercités.

On peut également noter que la part des liaisons de nuit parmi les liaisons de cet axe Intercités est plus de 2 fois plus grande (42%) que celle des liaisons de nuit du panel à l'échelle nationale (19%).

- **Le fait que la liaison appartienne ou non à l'axe Intercités « Est – Centre » :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ Axe_Est_Centre, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-9.469 -4.482 -1.472  2.878 29.166

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.8437     0.3609  38.354 < 2e-16 ***
Axe_Est_Centre -2.0324     0.7018  -2.896  0.00397 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.427 on 429 degrees of freedom
Multiple R-squared:  0.01917, Adjusted R-squared:  0.01689
F-statistic: 8.386 on 1 and 429 DF, p-value: 0.003974
```

D'après le modèle, les trains Intercités de l'axe « Est – Centre » du réseau Intercités de 2016 (présenté en Annexe 12) sont plus ponctuels que les autres trains Intercités. L'effet sur le taux de retard est significatif au seuil de 1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des liaisons de l'axe « Est - Centre » est moins élevé de 2,0 points.

Il n'y a pas d'explication évidente *a priori* à cette meilleure ponctualité relative de l'axe « Est - Centre ».

- **Le fait que le trajet soit effectué en sens Est -> Ouest :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :


```

Call:
lm(formula = TxRetard ~ Est_Ouest, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.576  -4.239  -1.459   2.879   27.571

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.4065     0.3544  37.831  <2e-16 ***
Est_Ouest    -0.4510     0.7509  -0.601   0.548
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.486 on 429 degrees of freedom
Multiple R-squared:  0.0008401, Adjusted R-squared:  -0.001489
F-statistic: 0.3607 on 1 and 429 DF,  p-value: 0.5484

```

Le fait que le trajet soit effectué en sens Est -> Ouest ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

- **Le fait que le trajet soit effectué en sens Ouest -> Est :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ Ouest_Est, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.878  -4.195  -1.427   2.922  28.372

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.7086     0.3521  38.930  <2e-16 ***
Ouest_Est    -1.8070     0.7461  -2.422   0.0159 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.445 on 429 degrees of freedom
Multiple R-squared:  0.01349, Adjusted R-squared:  0.01119
F-statistic: 5.866 on 1 and 429 DF,  p-value: 0.01585

```

D'après le modèle, les trains Intercités empruntant un trajet dans la direction Ouest -> Est sont plus ponctuels que les autres trains Intercités. L'effet sur le taux de retard est significatif au seuil de 5%, mais pas au seuil 1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard de ces liaisons est moins élevé de 1,8 points.

Il n'y a pas d'explication évidente *a priori* à cette meilleure ponctualité relative.

- **Le fait que le trajet soit effectué en sens Nord -> Sud :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ Nord_Sud, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-12.460  -4.153  -1.300   2.664  25.687

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.4864    0.3643  34.277 < 2e-16 ***
Nord_Sud     2.8037    0.6737   4.161 3.82e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.362 on 429 degrees of freedom
Multiple R-squared:  0.0388,    Adjusted R-squared:  0.03656
F-statistic: 17.32 on 1 and 429 DF,  p-value: 3.822e-05
```

D'après le modèle, les trains Intercités empruntant un trajet dans la direction Nord -> Sud sont moins ponctuels que les autres trains Intercités. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard de ces liaisons est plus élevé de 2,8 points.

Le fait que les liaisons concernées sont légèrement plus longues (490 km) que la moyenne des liaisons Intercités du panel (420 km) fournit peut-être un élément d'explication possible à la moins bonne ponctualité relative des liaisons concernées.

On peut également noter que la part des liaisons partant de Paris parmi les liaisons effectuant un trajet en sens Nord -> Sud est près de 2 fois plus grande (72%) que celle des liaisons partant de Paris dans l'ensemble du panel à l'échelle nationale (38%).

- **Le fait que le trajet soit effectué en sens Sud -> Nord :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ Sud_Nord, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.483  -4.268  -1.485   2.985  27.664

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.31326    0.36855  36.123  <2e-16 ***
Sud_Nord     -0.02552    0.69557  -0.037   0.971
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.489 on 429 degrees of freedom
Multiple R-squared:  3.138e-06, Adjusted R-squared:  -0.002328
F-statistic: 0.001346 on 1 and 429 DF,  p-value: 0.9707

```

Le fait que le trajet soit effectué en sens Sud -> Nord ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

- **Le nombre estimé d'arrêts moyen par liaison Intercités au cours d'un Jour Ouvrable de Base :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ NbMoyenArrêts_2020, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.697  -4.390  -1.347   2.644  27.451

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.29281    0.54915  26.03  <2e-16 ***
NbMoyenArrêts_2020 -0.15317    0.07027  -2.18   0.0298 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.453 on 429 degrees of freedom
Multiple R-squared:  0.01095, Adjusted R-squared:  0.008647
F-statistic: 4.751 on 1 and 429 DF,  p-value: 0.02983

```

D'après le modèle, plus une circulation Intercités effectue d'arrêts, moins elle est susceptible de connaître des retards. L'effet sur le taux de retard est significatif au seuil de 5%, mais pas au seuil de 1%.

Il est possible que la durée de l'arrêt au niveau d'un « petit arrêt » - pas au niveau d'un nœud ferroviaire – offre une marge de détente supérieure à la durée des « petits incidents », permettant ainsi le rattrapage des « petits retards » éventuellement occasionnés. Ce serait une explication éventuellement possible malgré les incidents qui peuvent survenir lors d'un arrêt (par exemple les

blocages de porte (volontaire ou non)), mais qui ne sont peut-être pas si fréquents que cela aux « petits arrêts » que l'on peut rencontrer sur les liaisons Intercités.

- **Le niveau du trafic de l'ensemble des circulations de trains (y compris fret et « haut-le-pied ») sur la section la plus intensément utilisée du réseau emprunté par la liaison Intercités en 2000 :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ TMJA_max_Total_Liaison_2000, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.528  -4.266  -1.433   2.986  27.619

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.3586096  0.5542480  24.102  <2e-16 ***
TMJA_max_Total_Liaison_2000 -0.0003322  0.0028951  -0.115    0.909
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.489 on 429 degrees of freedom
Multiple R-squared:  3.069e-05, Adjusted R-squared:  -0.0023
F-statistic: 0.01316 on 1 and 429 DF,  p-value: 0.9087
```

D'après le modèle, le niveau du trafic pour l'année 2000 (ensemble des circulations de trains) sur la section la plus intensément utilisée des lignes ferroviaires empruntées par la liaison Intercités ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

- **Le niveau du trafic de l'ensemble des circulations de trains (y compris fret et « haut-le-pied ») sur la section la plus intensément utilisée du réseau emprunté par la liaison Intercités en 2000 en dehors de Paris intra-muros :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ TMJA_max_Total_Liaison_2000_horsParis,
    data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.355  -4.284  -1.474   3.047  27.889

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.085203   0.578796  22.608  <2e-16 ***
TMJA_max_Total_Liaison_2000_horsParis  0.001751   0.003863   0.453    0.65
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.487 on 429 degrees of freedom
Multiple R-squared:  0.000479, Adjusted R-squared:  -0.001851
F-statistic: 0.2056 on 1 and 429 DF,  p-value: 0.6505

```

D'après le modèle, le niveau du trafic pour l'année 2000 (ensemble des circulations de trains) sur la section la plus intensément utilisée des lignes ferroviaires, hors Paris intra-muros, empruntées par la liaison Intercités ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

Ponctualité des TER par ancienne région :

- **Le nombre de circulations TER à l'année d'après la base de l'AQST :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbCirc_an_AQST, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.8822 -1.7486 -0.3656  1.5693  9.4247

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.663257   0.461008  16.623  <2e-16 ***
NbCirc_an_AQST 0.007783   0.003575   2.177  0.0312 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.885 on 134 degrees of freedom
Multiple R-squared:  0.03416, Adjusted R-squared:  0.02696
F-statistic:  4.74 on 1 and 134 DF, p-value: 0.03122
```

Le nombre de circulations de trains à l'année sur la liaison a un effet significatif au seuil de 5%, mais pas de 1%.

La valeur du coefficient montre qu'un ajout de 1000 circulations TER à l'année entraîne une baisse du taux de retard de 0,008 points environ. Sur l'année 2016, l'étendue entre les régions du nombre de circulations TER à l'année vaut environ 275 000 circulations TER, ce qui se traduit d'après le modèle par un impact d'environ 2,1 points de retard des TER entre la région où il circule le plus de TER (Rhône-Alpes) et celle où il en circule le moins (Basse-Normandie). Au sein de la série, l'écart-type est d'environ 71 000 circulations TER, se traduisant par une différence de taux de retard de 0,55 points en moyenne d'après le modèle.

- **Le nombre moyen de circulations quotidiennes à l'heure de pointe du matin :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ NbCirc_HPM, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.8845 -1.8174 -0.2999  1.5474  9.2858

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.46862    0.46798  15.959 < 2e-16 ***
NbCirc_HPM   0.03397    0.01299   2.614  0.00996 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.863 on 134 degrees of freedom
Multiple R-squared:  0.04853,    Adjusted R-squared:  0.04143
F-statistic: 6.835 on 1 and 134 DF,  p-value: 0.009961

```

Le nombre de circulations de trains à l'heure de pointe du matin (HPM) sur la liaison a un effet significatif au seuil de 1%.

La valeur du coefficient montre qu'un ajout de 100 circulations TER à l'heure de pointe du matin entraîne une baisse du taux de retard de 3,4 points environ. Sur l'année 2016, l'étendue entre les régions est de 73 circulations TER à l'HPM, ce qui se traduit d'après le modèle par une différence d'environ 2,5 points de retard des TER entre la région où il circule le plus de TER à l'HPM (Rhône-Alpes) et celle où il en circule le moins (Basse-Normandie). Au sein de la série, l'écart-type est d'environ 19 circulations TER, se traduisant par une différence de taux de retard de 0,66 points en moyenne d'après le modèle.

- **Le nombre moyen de circulations quotidiennes à l'heure de pointe du soir :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ NbCirc_HPS, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.6865 -1.8949 -0.3525  1.6075  9.5616

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.62144    0.49339   15.447  <2e-16 ***
NbCirc_HPS   0.02556    0.01227    2.083  0.0392 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.889 on 134 degrees of freedom
Multiple R-squared:  0.03136,    Adjusted R-squared:  0.02413
F-statistic: 4.339 on 1 and 134 DF,  p-value: 0.03916

```

Le nombre de circulations de trains à l'heure de pointe du soir (HPS) sur la liaison a un effet significatif au seuil de 5%, mais pas au seuil de 1%.

La valeur du coefficient montre qu'un ajout de 100 circulations TER à l'heure de pointe du soir (HPS) entraîne une baisse du taux de retard de 2,6 points environ. Pour l'année 2016, l'étendue entre les régions est de 82 circulations TER à l'HPS, ce qui se traduit d'après le modèle par une différence d'environ 2,1 points de retard des TER entre la région où il circule le plus de TER à l'HPS (Rhône-Alpes) et celle où il en circule le moins (Basse-Normandie). Au sein de la série, l'écart-type est d'environ 21 circulations TER, se traduisant par une différence de taux de retard de 0,53 points en moyenne d'après le modèle.

- **La distance moyenne (en km) parcourue par les TER :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ DistanceMoy, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-4.3715 -1.9818 -0.5574  1.4119 10.2254

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.08426    1.10108    3.709 0.000303 ***
DistanceMoy  0.05281    0.01283    4.116 6.69e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.766 on 134 degrees of freedom
Multiple R-squared:  0.1123,    Adjusted R-squared:  0.1056
F-statistic: 16.94 on 1 and 134 DF,  p-value: 6.692e-05

```


D'après le modèle, plus la distance kilométrique ferroviaire parcourue par les TER est grande, plus le taux de retard des TER pour la région concernée est élevé, ce qui est conforme à l'intuition, et avec le précédent modèle portant sur la durée du trajet. L'effet de la distance kilométrique ferroviaire sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, un trajet en TER plus long de 100 km présente un taux de retard plus élevé de 5,3 points. Sur l'année 2016, l'étendue entre les régions de la distance moyenne parcourue par les TER est de 73 km, ce qui se traduit d'après le modèle par un impact d'environ 3,8 points de retard des TER entre la région où les TER parcourent le plus de distance (Languedoc-Roussillon) et celle où il en circule le moins (Alsace). Au sein de la série, l'écart-type est d'environ 19 km, se traduisant par une différence de taux de retard d'1,0 point en moyenne.

- **Le nombre de passagers-km transportés (en millions) dans l'année :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ VoyKm_an, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.6415 -1.6228 -0.4361  1.2976  8.7113

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.0713806   0.3630820   19.476 < 2e-16 ***
VoyKm_an     0.0021451   0.0004187    5.124 1.02e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.684 on 134 degrees of freedom
Multiple R-squared:  0.1638,    Adjusted R-squared:  0.1576
F-statistic: 26.25 on 1 and 134 DF,  p-value: 1.022e-06
```

D'après le modèle, plus le nombre de passagers-km transportés par les TER est grand, plus le taux de retard des TER pour la région concernée est élevé. L'effet de la distance kilométrique ferroviaire sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient montre qu'une hausse de 1 milliard de voyageurs-km TER entraîne une hausse du taux de retard de 2,1 points environ. Sur l'année 2016, l'étendue entre les régions est de 2,46 milliards de voy-km transportés, ce qui se traduit d'après le modèle par une différence d'environ 5,3 points de retard des TER entre la région qui en comptabilise le plus (Rhône-Alpes) et celle qui en comptabilise le moins (Limousin). Au sein de la série, l'écart-type est de 568 millions de voy-km transportés, se traduisant par une différence de taux de retard de 1,2 points en moyenne d'après le modèle.

- **La capacité d'emport moyenne des TER en nombre de sièges :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbPlacesMoy, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.1557 -1.9850 -0.4458  1.6517  9.4290

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.764511   0.849407   7.964 6.34e-13 ***
NbPlacesMoy  0.006170   0.002872   2.148  0.0335  *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.886 on 134 degrees of freedom
Multiple R-squared:  0.0333,    Adjusted R-squared:  0.02608
F-statistic: 4.616 on 1 and 134 DF,  p-value: 0.03348
```

D'après le modèle, plus la capacité d'emport du matériel roulant TER est importante, plus le taux de retard des TER est élevé. L'effet de la distance kilométrique ferroviaire sur le taux de retard est significatif au seuil de 5%, mais pas au seuil de 1%.

La valeur du coefficient montre qu'une hausse de 100 sièges de la capacité d'emport moyenne du matériel roulant TER s'accompagne d'une hausse du taux de retard de 0,6 points environ. L'impact de cette variable selon le modèle est donc modéré.

S'il n'y a pas d'explication évidente à la corrélation entre la capacité d'emport du matériel roulant et le taux de retard, on peut supposer qu'un matériel roulant plus capacitairé réponde à une plus grande demande en nombre de voyageurs. Les rames transportant davantage de voyageurs, il est possible que le risque de retards liés aux voyageurs soit plus important - par exemple, malaise voyageur, blocage (volontaire ou non) des portes par un voyageur retardant le train, etc...

- **Le taux d'occupation moyen des TER :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ TauxOccupationMoy, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-6.057 -1.832 -0.272  1.556  9.413

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.10895    1.40666   2.921  0.00409 **
TauxOccupationMoy 0.17752    0.05589   3.177  0.00185 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.831 on 134 degrees of freedom
Multiple R-squared:  0.07003,    Adjusted R-squared:  0.06309
F-statistic: 10.09 on 1 and 134 DF,  p-value: 0.00185

```

D'après le modèle, plus le taux d'occupation moyen des TER est élevé, plus le taux de retard des TER est important. L'effet de la distance kilométrique ferroviaire sur le taux de retard est significatif au seuil de 1%.

La valeur du coefficient montre qu'une hausse de 10 points de pourcentage du taux d'occupation moyen des TER entraîne une hausse du taux de retard de 1,8 points environ. L'impact de cette variable selon le modèle n'est donc pas négligeable.

Pour donner un élément d'explication à la corrélation observée, on peut supposer que lorsque les rames TER transportent plus de voyageurs, elles sont davantage susceptibles de connaître des retards liés aux voyageurs - par exemple, malaise voyageur, blocage (volontaire ou non) des portes par un voyageur retardant le train, etc...

- **Le nombre de sièges-km TER offerts par habitant :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ SiegesKmParHab, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.0427 -2.0785 -0.4049  1.6716  9.7691

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.513e+00  4.969e-01  17.132  <2e-16 ***
SiegesKmParHab -2.302e-06  4.167e-04  -0.006    0.996
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.935 on 134 degrees of freedom
Multiple R-squared:  2.278e-07, Adjusted R-squared:  -0.007462
F-statistic: 3.053e-05 on 1 and 134 DF,  p-value: 0.9956

```

Le nombre de sièges-km TER offerts par habitant ne présente pas d'impact significatif au seuil de 10% sur le taux de retard.

- **Le nombre de voyageurs-km TER par habitant :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ VoyKmParHab, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.2230 -2.0759 -0.3976  1.6722  9.7812

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.3359609  0.5129504  16.25  <2e-16 ***
VoyKmParHab  0.0006991  0.0017937   0.39   0.697
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.934 on 134 degrees of freedom
Multiple R-squared:  0.001132, Adjusted R-squared:  -0.006322
F-statistic: 0.1519 on 1 and 134 DF,  p-value: 0.6973

```

Le niveau de la demande en voyageurs-km TER par habitant ne présente pas d'impact significatif au seuil de 10% sur le taux de retard.

- **L'intensité d'utilisation du réseau ferroviaire hors LGV par les services TER :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ IntensitéUsageTER, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.7342 -1.9619 -0.3249  1.7372  9.5076

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    7.96731    0.66805  11.926 <2e-16 ***
IntensitéUsageTER 0.03083    0.03516   0.877  0.382
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.927 on 134 degrees of freedom
Multiple R-squared:  0.005706, Adjusted R-squared:  -0.001714
F-statistic: 0.7689 on 1 and 134 DF,  p-value: 0.3821
```

L'intensité d'utilisation du réseau ferroviaire hors LGV par les services TER – en train-km TER par km de ligne par jour - ne présente pas d'impact significatif au seuil de 10% sur le taux de retard.

- **La longueur (en km de ligne) du réseau ferré hors LGV :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ LongueurRéseau_HorsLGV, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-4.2255 -1.8189 -0.2313  1.2336  9.7356

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.7771911    0.7458388   6.405 2.35e-09 ***
LongueurRéseau_HorsLGV 0.0030471    0.0005793   5.260 5.57e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.672 on 134 degrees of freedom
Multiple R-squared:  0.1711, Adjusted R-squared:  0.1649
F-statistic: 27.66 on 1 and 134 DF,  p-value: 5.569e-07
```

D'après le modèle, plus le réseau ferré hors LGV est long, plus le taux de retard des TER est élevé. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient montre qu'une hausse de 1000 km de la longueur du réseau ferroviaire hors LGV s'accompagne d'une hausse du taux de retard de 3,0 points environ. Sur l'année 2016, l'étendue entre les régions est de 1610 km de ligne, ce qui se traduit d'après le modèle par une différence d'environ 4,9 points de retard des TER entre la région dont le réseau ferré hors LGV est le plus long (Rhône-Alpes) et celle où il est le plus court (Basse-Normandie). Au sein de la série, l'écart-type est de 385 km de ligne, se traduisant par une différence de taux de retard de 1,2 points en moyenne d'après le modèle.

Il n'y a pas d'explication évidente *a priori* à la corrélation observée. Il est possible que les trajets TER soient plus longs dans les plus grandes régions, augmentant le risque d'occurrence d'un retard pour les TER de ces régions.

- **La densité (en km de ligne pour 100 km²) du réseau ferré hors LGV :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ DensitéRéseau_HorsLGV, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-4.944 -2.120 -0.664  1.811  9.042

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.8288    0.8297  14.257 < 2e-16 ***
DensitéRéseau_HorsLGV -0.6578    0.1576  -4.173 5.36e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.761 on 134 degrees of freedom
Multiple R-squared:  0.115,    Adjusted R-squared:  0.1084
F-statistic: 17.42 on 1 and 134 DF,  p-value: 5.364e-05
```

D'après le modèle, moins le réseau ferré hors LGV est dense, plus le taux de retard des TER est élevé. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient montre qu'une hausse de 1 km de ligne hors LGV/100 km² de superficie s'accompagne d'une baisse du taux de retard de 0,66 points environ. Sur l'année 2016, au sein de la série, l'écart-type est de 1,6 km de ligne/100 km², se traduisant par une différence de taux de retard de 1,1 points en moyenne d'après le modèle.

Il n'est pas impossible que le signe de la corrélation observée soit lié au fait qu'un réseau ferroviaire plus dense permette d'éviter pour partie la saturation des lignes ferroviaires fréquentées par les TER, par exemple en répartissant de manière plus homogène le trafic sur les différentes lignes du réseau par effet de maillage.

- **La part des voies des catégories UIC 5 à 6 dans le réseau ferroviaire régional hors LGV :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ PartCatUIC_5_6, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.9883 -1.7684 -0.1225  1.5870  9.1355

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.46154    0.56504   18.515 < 2e-16 ***
PartCatUIC_5_6 -0.06600    0.01732   -3.811  0.00021 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.788 on 134 degrees of freedom
Multiple R-squared:  0.09781,    Adjusted R-squared:  0.09107
F-statistic: 14.53 on 1 and 134 DF,  p-value: 0.0002098
```

Les lignes du réseau ferroviaire peuvent être classées selon la classification de l'Union Internationale des Chemins de fer (UIC). La classification et son application au Réseau Ferré National (RFN) sont présentées par SNCF Réseau dans un article web relatif au débat public sur la Voie Ferrée Centre Europe Atlantique et disponible en ligne⁹. SNCF Réseau explique notamment que :

« Le groupe UIC 1 correspond à des lignes très chargées et, à l'opposé, le groupe UIC 9 correspond à des lignes très faiblement chargées. Les lignes à grande vitesse ainsi que les grandes lignes du réseau ferré national appartiennent en principe aux groupes UIC 1 à 4, à l'exception de certains axes, notamment transversaux, qui relèvent des groupes UIC 5 à 6. Les lignes faiblement chargées des groupes UIC 7 à 9, avec voyageurs (AV) ou sans voyageurs (SV), correspondent en général au réseau capillaire, généralement régional. »

D'après le modèle, plus la part des voies des catégories UIC 5 à 6 dans le réseau ferré régional hors LGV – correspondant pour partie aux lignes TER interrégionales - est grande, moins le taux de retard des TER est élevé. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient montre qu'une hausse de 10 points pourcentage de cette part s'accompagne d'une diminution du taux de retard de 0,66 points environ.

Nous n'avons pas à ce stade identifié d'explication évidente au signe de la corrélation observée.

⁹ Lien vers l'article : <https://cpdp.debatpublic.fr/cdpd-vfcea/classification-uic-union-internationale-chemins-fer-ligne-nevers-chagny.html>

- **La part des voies des catégories UIC 7 à 9 avec voyageurs (AV) dans le réseau ferroviaire régional hors LGV :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ PartCatUIC_7_9_AV, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-4.7603 -2.1213 -0.4676  1.7239  9.5057

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.47376    0.59422  15.943  <2e-16 ***
PartCatUIC_7_9_AV -0.02916    0.01633  -1.786   0.0764 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.901 on 134 degrees of freedom
Multiple R-squared:  0.02324,    Adjusted R-squared:  0.01595
F-statistic: 3.188 on 1 and 134 DF,  p-value: 0.07643
```

La part des voies des catégories UIC 7 à 9 avec voyageurs (AV) dans le réseau ferré régional hors LGV – correspondant aux « petites lignes », soit les lignes TER infrarégionales - ne présente pas d'impact significatif au seuil de 5% sur le taux de retard des TER.

- **L'âge moyen du réseau ferroviaire régional hors LGV :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :


```

Call:
lm(formula = TxRetard ~ AgeRéseau_HorsLGV, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-4.8498 -2.1030 -0.1139  1.8060  9.4987

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.02933    1.85963   5.931 2.42e-08 ***
AgeRéseau_HorsLGV -0.07753    0.05671  -1.367   0.174
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.915 on 134 degrees of freedom
Multiple R-squared:  0.01375, Adjusted R-squared:  0.006394
F-statistic: 1.869 on 1 and 134 DF, p-value: 0.1739

```

L'âge moyen du réseau ferroviaire régional hors LGV ne présente pas d'impact significatif au seuil de 10% sur le taux de retard des TER.

- **L'âge moyen des lignes des catégories UIC 5 à 6 du réseau ferroviaire régional :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ AgeRéseauCatUIC_5_6, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-4.354 -2.068 -0.513  1.764 10.023

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.97277    1.04388   6.680 5.84e-10 ***
AgeRéseauCatUIC_5_6 0.05502    0.03627   1.517   0.132
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.91 on 134 degrees of freedom
Multiple R-squared:  0.01688, Adjusted R-squared:  0.009542
F-statistic: 2.301 on 1 and 134 DF, p-value: 0.1317

```

L'âge moyen des lignes UIC 5 à 6 – correspondant pour partie aux lignes TER interrégionales - du réseau ferroviaire régional ne présente pas d'impact significatif au seuil de 10% sur le taux de retard.

- **L'âge moyen des lignes des catégories UIC 7 à 9 AV du réseau ferroviaire régional :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ AgeRéseauCatUIC_7_9_AV, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.1916 -1.9657 -0.2627  1.7840  9.5077

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.13898    1.18864   8.530 2.75e-14 ***
AgeRéseauCatUIC_7_9_AV -0.04125    0.02943  -1.402   0.163
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.914 on 134 degrees of freedom
Multiple R-squared:  0.01445,    Adjusted R-squared:  0.007094
F-statistic: 1.964 on 1 and 134 DF,  p-value: 0.1633
```

L'âge moyen des lignes UIC 7 à 9 AV – correspondant aux lignes TER au niveau de trafic le plus faible en général - du réseau ferroviaire régional ne présente pas d'impact significatif au seuil de 10% sur le taux de retard.

- **La densité des lignes UIC 5-6 (en km de voie pour 100 km²) :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ DensitéRéseau_CatUIC_5_6, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.1067 -1.9763 -0.4131  1.8546  8.8661

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.3830    0.4843  21.438 <2e-16 ***
DensitéRéseau_CatUIC_5_6 -0.7574    0.1713  -4.423  2e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.742 on 134 degrees of freedom
Multiple R-squared:  0.1274,    Adjusted R-squared:  0.1209
F-statistic: 19.56 on 1 and 134 DF,  p-value: 1.997e-05
```

D'après le modèle, moins le réseau des lignes UIC 5-6 est dense, plus le taux de retard des TER est élevé. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient montre qu'une hausse de 1 km de voie UIC 5-6/100 km² de superficie s'accompagne d'une baisse du taux de retard de 0,76 points environ. Sur l'année 2016, au sein de la série, l'écart-type est de 1,4 km de voie UIC 5-6/100 km², se traduisant par une différence de taux de retard de 1,1 points en moyenne d'après le modèle.

Il n'est pas impossible que le signe de la corrélation observée soit lié au fait qu'un réseau ferroviaire plus dense permette d'éviter pour partie la saturation des lignes ferroviaires fréquentées par les TER, notamment en répartissant de manière plus homogène le trafic sur les différentes lignes du réseau.

- **La densité des lignes UIC 7-9 AV (en km de voie pour 100 km²) :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ DensitéRéseau_CatUIC_7_9_AV, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.3381 -1.6838 -0.0969  1.5706  8.7340

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.1001     0.5224  21.247 < 2e-16 ***
DensitéRéseau_CatUIC_7_9_AV -1.0094     0.1833  -5.506 1.81e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.651 on 134 degrees of freedom
Multiple R-squared:  0.1845,    Adjusted R-squared:  0.1784
F-statistic: 30.31 on 1 and 134 DF,  p-value: 1.808e-07
```

D'après le modèle, moins le réseau des lignes UIC 7-9 AV est dense, plus le taux de retard des TER est élevé. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient montre qu'une hausse de 1 km de voie UIC 7-9 (AV)/100 km² de superficie s'accompagne d'une baisse du taux de retard de 1 point environ. Sur l'année 2016, au sein de la série, l'écart-type est de 1,2 km de voie UIC 7-9 (AV)/100 km², se traduisant par une différence de taux de retard de 1,3 points en moyenne d'après le modèle.

Il n'est pas impossible que le signe de la corrélation observée soit lié au fait qu'un réseau ferroviaire des « petites lignes » plus dense permette d'éviter pour partie la saturation des lignes ferroviaires fréquentées par les TER, notamment en répartissant de manière plus homogène le trafic sur les différentes lignes du réseau.

- **Le nombre de liaisons infrarégionales de l'offre TER :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbLiaisonsInfrarégionales, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.1769 -1.6011 -0.1654  1.3094  9.5533

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.5299815   0.4758098   13.724 < 2e-16 ***
NbLiaisonsInfrarégionales 0.0021386   0.0004482    4.771 4.72e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.714 on 134 degrees of freedom
Multiple R-squared:  0.1452,    Adjusted R-squared:  0.1388
F-statistic: 22.77 on 1 and 134 DF,  p-value: 4.716e-06
```

D'après le modèle, plus l'offre de TER de la région compte de liaisons infrarégionales, plus le taux de retard des TER est élevé. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient montre qu'un ajout de 1000 liaisons infrarégionales à l'offre TER s'accompagne d'une hausse du taux de retard de 2,1 points environ. Sur l'année 2016, au sein de la série, l'écart-type est de 544 liaisons infrarégionales, se traduisant par une différence de taux de retard de 1,2 points en moyenne d'après le modèle.

Un élément d'explication envisageable à la corrélation observée pourrait être qu'un nombre de liaisons infrarégionales plus élevé s'accompagnerait peut-être d'une hausse du nombre de retards liés à la gestion du trafic, et notamment des retards liés aux effets de réseau (propagation d'un retard d'une liaison TER infrarégionale à une autre).

Le nombre de liaisons interrégionales de l'offre TER :

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbLiaisonsInterrégionales, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-3.9192 -1.9512 -0.3788  1.3503  9.8070

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.8209325   0.5612560   10.371 < 2e-16 ***
NbLiaisonsInterrégionales 0.0047523   0.0009053    5.249 5.83e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.673 on 134 degrees of freedom
Multiple R-squared:  0.1706,    Adjusted R-squared:  0.1644
F-statistic: 27.56 on 1 and 134 DF,  p-value: 5.832e-07
```

D'après le modèle, plus l'offre de TER de la région compte de liaisons interrégionales, plus le taux de retard des TER est élevé. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient montre qu'un ajout de 1000 liaisons interrégionales à l'offre TER s'accompagne d'une hausse du taux de retard de 4,7 points environ. Sur l'année 2016, au sein de la série, l'écart-type est de 283 liaisons interrégionales, se traduisant par une différence de taux de retard de 1,3 points en moyenne d'après le modèle.

Un élément d'explication envisageable à la corrélation observée pourrait être qu'un nombre de liaisons interrégionales plus élevé s'accompagnerait peut-être d'une hausse du nombre de retards liés à la gestion du trafic, et notamment des retards liés aux effets de réseau (propagation d'un retard d'une liaison TER interrégionale à une autre).

Le nombre de communes desservies dans l'offre TER :

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbCommunesDesservies, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.5034 -1.5821 -0.1349  1.2430  9.8836

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.992252   0.654699   9.153 8.13e-16 ***
NbCommunesDesservies 0.022053   0.005345   4.126 6.44e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.765 on 134 degrees of freedom
Multiple R-squared:  0.1127,    Adjusted R-squared:  0.1061
F-statistic: 17.02 on 1 and 134 DF,  p-value: 6.445e-05
```

D'après le modèle, plus l'offre de TER de la région dessert de communes, plus le taux de retard des TER est élevé. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient montre qu'un ajout de 100 communes desservies à l'offre TER s'accompagne d'une hausse du taux de retard de 2,2 points environ. Sur l'année 2016, au sein de la série, l'écart-type est de 46 communes desservies, se traduisant par une différence de taux de retard de 1,0 point en moyenne d'après le modèle.

Pour tenter de fournir un élément d'explication à la corrélation observée, on peut envisager qu'un nombre plus important de communes desservies signifie soit davantage de liaisons infrarégionales ou interrégionales dans l'offre régionale de TER – ce qui pourrait peut-être entraîner davantage de retards liés aux effets de réseaux, comme expliqué au paragraphe précédent - soit un nombre moyen d'arrêts plus important sur les liaisons TER de la région. Un nombre moyen d'arrêts plus élevé peut potentiellement entraîner une hausse du nombre de retard. En effet, certains événements peuvent survenir lors d'un arrêt en gare et retarder le départ du train : par exemple, malaise voyageur, blocage (volontaire ou non) des portes par un voyageur retardant le train, dysfonctionnement du matériel roulant lors de son redémarrage, ou parfois des infrastructures en sortie de gare, etc. Nous étudierons plus en détail la corrélation entre le nombre moyen d'arrêts effectués par les circulations de TER et le taux de retard des TER par ancienne région dans un autre paragraphe.

- **La part de la population de chaque région résidant dans une commune située à moins de 10 km d'une gare régionale :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ PartPopInf10kmGareTER, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.920 -1.970 -0.407  1.592  9.225

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.89530     3.07501   0.942  0.3481
PartPopInf10kmGareTER 0.06623     0.03615   1.832  0.0692 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.899 on 134 degrees of freedom
Multiple R-squared:  0.02443, Adjusted R-squared:  0.01715
F-statistic: 3.356 on 1 and 134 DF, p-value: 0.06918
```

La part de la population de chaque région résidant dans une commune située à moins de 10 km d'une gare régionale ne présente pas d'impact significatif au seuil de 5% sur le taux de retard.

- **Le ratio des effectifs TER pour 100 000 trains-km réalisés à fin 2015 :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ EffectifsPour100000trkm_2015, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-4.9836 -2.1183 -0.3544  1.6958  9.7256

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.20292     1.13873   7.204 3.82e-11 ***
EffectifsPour100000trkm_2015 0.01898     0.06861   0.277  0.782
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.934 on 134 degrees of freedom
Multiple R-squared:  0.000571, Adjusted R-squared: -0.006887
F-statistic: 0.07656 on 1 and 134 DF, p-value: 0.7824
```

Le ratio des effectifs TER pour 100 000 trains-km réalisés dans la région à fin 2015 ne présente pas d'impact significatif au seuil de 10% sur le taux de retard.

Le nombre moyen d'arrêts effectués par les circulations TER en 2018 (données de l'ART disponibles uniquement pour les nouvelles régions) :

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbMoyArrêts_Nv11eReg_2018, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-4.5194 -1.4508  0.2561  1.3814  5.3901

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -5.0657     1.1585  -4.373 2.44e-05 ***
NbMoyArrêts_Nv11eReg_2018  1.5346     0.1294  11.856 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.051 on 134 degrees of freedom
Multiple R-squared:  0.512,    Adjusted R-squared:  0.5083
F-statistic: 140.6 on 1 and 134 DF,  p-value: < 2.2e-16
```

D'après le modèle, plus le nombre moyen d'arrêts effectués en 2018 par les TER dans la nouvelle région est important, plus le taux de retard des TER est élevé. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient montre qu'un arrêt supplémentaire effectué par une circulation TER se traduit en moyenne par un taux de retard plus élevé de 1,5 points. L'effet de cette variable est donc très prononcé.

Pour expliquer cette corrélation, on peut penser que le risque d'occurrence de retards est important lors d'un arrêt. En effet, certains événements peuvent survenir lors d'un arrêt en gare et retarder le départ du train : par exemple, malaise voyageur, blocage (volontaire ou non) des portes par un voyageur retardant le train, dysfonctionnement du matériel roulant lors de son redémarrage, ou parfois des infrastructures en sortie de gare, etc...

Le nombre moyen d'arrêts effectués par les circulations TER en 2020 sur le panel des liaisons les plus fréquentées par ancienne région (avant la réforme territoriale du 1^{er} janvier 2016) :

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :


```

Call:
lm(formula = TxRetard ~ NbMoyArrêts_2020_PpalesLiaisons, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-4.8033 -1.7662 -0.0257  1.2759  8.9265

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)         4.9248    0.5874   8.384 6.21e-14 ***
NbMoyArrêts_2020_PpalesLiaisons  0.7029    0.1069   6.577 9.86e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.552 on 134 degrees of freedom
Multiple R-squared:  0.244,    Adjusted R-squared:  0.2384
F-statistic: 43.26 on 1 and 134 DF,  p-value: 9.856e-10

```

D'après le modèle, plus le nombre moyen d'arrêts effectués en 2020 par les TER sur les principales liaisons en nombre de circulations par an (panel retenu par l'ART, présenté en annexe 16) par ancienne région est important, plus le taux de retard des TER est élevé. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient montre qu'un arrêt supplémentaire effectué par une circulation TER se traduit en moyenne par un taux de retard plus élevé de 0,7 points. L'effet de cette variable est donc très prononcé.

Les éléments d'explication fournis dans le paragraphe précédent pour les nouvelles régions administratives s'appliquent à cette variable également.

On peut remarquer que la p-value du test de Fisher est plus élevée que celle du modèle précédent portant sur le nombre moyen d'arrêts pour l'intégralité des liaisons TER par nouvelle région. Le modèle précédent semble donc légèrement plus fiable.

L'impact d'un arrêt supplémentaire sur le taux de retard est également de l'ordre de 2 fois plus faible avec cette variable qu'avec la précédente. Cela peut peut-être s'expliquer par des arrêts plus nombreux sur les liaisons à faible trafic qui ne sont pas prises en compte dans le panel des liaisons TER les plus fréquentées selon l'ART. Un autre élément d'explication pourrait être une potentielle moins bonne correspondance entre le nombre moyen d'arrêts effectués par les TER en 2020 (cette variable) et le nombre moyen d'arrêts sur la période d'étude 2013-2019, par rapport à la correspondance obtenue avec le nombre moyen d'arrêts effectués par les TER en 2018 (variable précédente), du fait de changements survenus dans les politiques d'arrêts entre 2018 et 2020.

- **Le niveau du trafic journalier moyen annuel (TJMA) des trains TER sur la section de ligne la plus intensément utilisée dans la région en 2018 :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ TMJA_max_TER_2018, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.502 -1.988 -0.202  1.565  9.373

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    7.139881   0.589963  12.102  <2e-16 ***
TMJA_max_TER_2018 0.020538   0.008038   2.555   0.0117 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.866 on 134 degrees of freedom
Multiple R-squared:  0.04645,    Adjusted R-squared:  0.03934
F-statistic: 6.528 on 1 and 134 DF,  p-value: 0.01174

```

D'après le modèle, plus le trafic des TER au niveau de la section dimensionnante du réseau régional est intense, plus le taux de retard des TER de cette région est élevé, ce qui est conforme à l'intuition. L'effet sur le taux de retard est significatif au seuil 5%, mais pas au seuil 1%.

La valeur du coefficient indique qu'en moyenne, une région dont la section dimensionnant du réseau ferré en nombre de circulations journalières est empruntée par 100 TER de plus présentera un taux de retard plus élevé de 2,1 points.

En effet, lorsqu'une circulation TER emprunte des lignes plus fréquentées, elle est davantage exposée au risque de subir un retard qui se propage sur le réseau d'une circulation de train à l'autre, et cela sans que ce retard ne puisse nécessairement être « rattrapé » au cours du trajet, puisque les distances entre 2 arrêts peuvent être courtes dans le cas des TER et que la gestion du trafic en situation dégradée doit assurer au mieux la sécurité des voyageurs.

D'après la carte SNCF « LE TRAFIC VOYAGEURS RÉGIONAUX » - de l'atlas du réseau publié (édition 2020) par SNCF Réseau, le débit journalier des différents tronçons du réseau TER par ancienne région varie de 160 circulations TER/jour pour la région Rhône-Alpes, à 20 circulations TER/jour pour la région Poitou-Charentes. D'après le modèle, cette différence de débit entre les sections dimensionnantes des réseaux TER de ces deux régions entraînerait une différence de 2,5 points dans le taux de retard.

- **L'estimation du taux de circulations de TER traversant le nœud ferroviaire régional retenu sur le panel des liaisons les plus fréquentées par ancienne région (panel retenu par l'ART) en 2018 :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ TxCircPpalesLiaisonsTraversantNoeudRégional,
    data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-4.9918 -2.0822 -0.4222  1.6624  9.8087

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.721320   1.041978   8.370 6.71e-14 ***
TxCircPpalesLiaisonsTraversantNoeudRégional -0.002635   0.012619  -0.209  0.835
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.935 on 134 degrees of freedom
Multiple R-squared:  0.0003252, Adjusted R-squared:  -0.007135
F-statistic: 0.04359 on 1 and 134 DF,  p-value: 0.8349

```

L'estimation du taux de circulations de TER traversant le nœud ferroviaire régional retenu (liste présentée en annexe 16) sur le panel des liaisons les plus fréquentées par ancienne région (panel retenu par l'ART, présenté en annexe 16), en 2018, ne présente pas d'impact significatif au seuil de 10% sur le taux de retard.

- **L'estimation de l'âge moyen du parc de matériel roulant TER par ancienne région :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ AgeMoyen_MaterielRoulant, data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-5.7581 -2.0890 -0.5029  1.8424  9.3367

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      6.60827   1.07890   6.125 9.43e-09 ***
AgeMoyen_MaterielRoulant 0.12224   0.06747   1.812  0.0723 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.9 on 134 degrees of freedom
Multiple R-squared:  0.02391, Adjusted R-squared:  0.01662
F-statistic: 3.282 on 1 and 134 DF,  p-value: 0.07228

```

D'après le modèle, l'âge moyen estimé du parc de matériel roulant TER par ancienne région (tableau de synthèse et méthodologie de l'estimation présentées en annexe 17) ne présente pas d'impact significatif sur le taux de retard des TER au seuil 5%.

Ponctualité des RER et Transilien d'Île-de-France par ligne :

- **La longueur totale de la ligne :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ LongueurLigne, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-6.0247 -2.6535 -0.6253  2.1096  9.8467

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.986169   0.756325  11.881 < 2e-16 ***
LongueurLigne 0.013395   0.005013   2.672  0.00896 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.372 on 89 degrees of freedom
Multiple R-squared:  0.07427, Adjusted R-squared:  0.06387
F-statistic: 7.141 on 1 and 89 DF, p-value: 0.008962
```

La longueur totale de la ligne a un effet significatif au seuil de 1%.

La valeur du coefficient montre qu'une ligne plus longue de 100 km de ligne présente en moyenne un taux de retard plus élevé de 1,3 points. Sur l'année 2019, l'étendue entre lignes de la longueur totale de ligne vaut environ 225 km, ce qui se traduit d'après le modèle par un impact d'environ 3,0 points de retard des RER et Transilien entre la ligne la plus longue (J) et la ligne la plus courte (U). Au sein de la série, l'écart-type est d'environ 73 km, se traduisant par une différence de taux de retard de 1,0 point en moyenne d'après le modèle.

- **Le nombre total de gares de la ligne :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ NbArrêts, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-6.5167 -2.3608 -0.4299  2.3109  9.9549

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.34045    0.81713  11.431  <2e-16 ***
NbArrêts     0.03547    0.01817   1.952   0.054  .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.432 on 89 degrees of freedom
Multiple R-squared:  0.04107, Adjusted R-squared:  0.0303
F-statistic: 3.812 on 1 and 89 DF, p-value: 0.05403

```

Le nombre total de gares de la ligne ne présente pas d'impact significatif sur le taux de retard au seuil de 5%.

- **La distance moyenne entre 2 gares de la ligne :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ DistanceMoyEntreArrêts, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-5.6297 -2.7351 -0.3583  2.3653  7.9810

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.4573    0.8465  11.173  <2e-16 ***
DistanceMoyEntreArrêts  0.3626    0.2109   1.719   0.0891 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.448 on 89 degrees of freedom
Multiple R-squared:  0.03213, Adjusted R-squared:  0.02125
F-statistic: 2.954 on 1 and 89 DF, p-value: 0.08913

```

La distance moyenne entre deux gares de la ligne ne présente pas d'impact significatif sur le taux de retard au seuil de 5%.

- **Le nombre de circulations quotidiennes sur la ligne en 2019 :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbCircJour_2019, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-6.2175 -2.5267 -0.0256  2.5136 10.1058

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.430558   0.792187  11.904 <2e-16 ***
NbCircJour_2019 0.003341   0.001756   1.902  0.0604 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.436 on 89 degrees of freedom
Multiple R-squared:  0.03908, Adjusted R-squared:  0.02828
F-statistic: 3.619 on 1 and 89 DF, p-value: 0.06035
```

Le nombre de circulations quotidiennes sur la ligne ne présente pas d'impact significatif sur le taux de retard au seuil de 5%.

- **Le nombre de voyageurs quotidiens sur la ligne :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbVoyageursJour, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-7.1552 -2.3178 -0.2769  2.4151  9.9114

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.709383   0.518071  18.741 < 2e-16 ***
NbVoyageursJour 0.002918   0.001043   2.799  0.00628 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.36 on 89 degrees of freedom
Multiple R-squared:  0.08092, Adjusted R-squared:  0.0706
F-statistic: 7.836 on 1 and 89 DF, p-value: 0.006279
```

Le nombre de voyageurs quotidiens sur la ligne un effet significatif au seuil de 1%.

La valeur du coefficient montre qu'une ligne empruntée par 1 million de voyageurs quotidiens de plus présente en moyenne un taux de retard plus élevé de 2,9 points. Sur l'année 2019, l'étendue entre lignes du nombre de voyageurs quotidiens transportés vaut environ 1,4 millions, ce qui se traduit d'après le modèle par un impact d'environ 4,0 points de retard des RER et Transilien entre la ligne qui transporte le plus de voyageurs (A) et la ligne qui en transporte le moins (K). Au sein de la série, l'écart-type est d'environ 411 000 voyageurs, se traduisant par une différence de taux de retard de 1,2 points en moyenne d'après le modèle.

- **Le nombre de branches de la ligne :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ NbBranches, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-6.3525 -2.5295 -0.5148  2.3925  9.8382

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.4741     0.9654   9.813 7.78e-16 ***
NbBranches   0.3377     0.2326   1.452  0.15
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.464 on 89 degrees of freedom
Multiple R-squared:  0.02314,    Adjusted R-squared:  0.01216
F-statistic: 2.108 on 1 and 89 DF,  p-value: 0.1501
```

Le nombre de circulations quotidiennes sur la ligne ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

- **Le volume de l'offre réalisée à l'année en trains-km :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ TrKm, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-6.1748 -2.2938 -0.5123  2.1442 10.9023

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.5387     0.6735  12.677 < 2e-16 ***
TrKm          0.4181     0.1088   3.844 0.000227 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.246 on 89 degrees of freedom
Multiple R-squared:  0.1424,    Adjusted R-squared:  0.1327
F-statistic: 14.77 on 1 and 89 DF,  p-value: 0.0002271

```

Le volume de l'offre annuelle sur la ligne en trains-km réalisés présente un effet significatif au seuil de 0,1%.

La valeur du coefficient montre qu'une ligne produisant une offre annuelle plus élevée de 1 million de trains-km réalisés affiche en moyenne un taux de retard plus élevé de 0,4 point. Sur l'année 2019, l'étendue entre lignes du nombre de trains-km réalisés vaut environ 9,7 millions de trains-km, ce qui se traduit d'après le modèle par un impact d'environ 4,0 points de retard des RER et Transilien entre la ligne dont l'offre annuelle est la plus importante (A) et la ligne qui produit le moins de trains-km à l'année (K). Au sein de la série, l'écart-type est d'environ 3,2 millions de trains-km réalisés, se traduisant par une différence de taux de retard de 1,3 points en moyenne d'après le modèle.

- **Le nombre de voyageurs-km transportés :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ VoyKm, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-7.6974 -2.2911 -0.3418  2.2726  9.9175

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.7037     0.4702  20.636 < 2e-16 ***
VoyKm         0.5756     0.1714   3.359 0.00115 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.302 on 89 degrees of freedom
Multiple R-squared:  0.1125,    Adjusted R-squared:  0.1025
F-statistic: 11.28 on 1 and 89 DF,  p-value: 0.001154

```

Le nombre de voyageurs-km transportés sur la ligne présente un effet significatif au seuil de 1%.

La valeur du coefficient montre qu'une ligne transportant 1 milliard de voyageurs-km supplémentaires affiche en moyenne un taux de retard plus élevé de 0,58 point. Sur l'année 2019, l'étendue entre lignes du nombre de voyageurs-km transportés vaut environ 7,95 milliards de voyageurs-km, ce qui se traduit d'après le modèle par un impact d'environ 4,6 points de retard des RER et Transilien entre la ligne dont le trafic annuel est le plus important (A) et celle dont le trafic annuel est le plus faible (K). Au sein de la série, l'écart-type est d'environ 2,4 milliards de voyageurs-km transportés, se traduisant par une différence de taux de retard de 1,4 points en moyenne d'après le modèle.

- **La durée de l'intervalle minimal entre 2 trains à l'hyperpointe dans la section dimensionnante de la ligne :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ IntervalleMin_2019, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-5.8643 -2.8843 -0.2202  2.2727  8.0016

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.349653   0.663928  15.589  <2e-16 ***
IntervalleMin_2019  0.001804   0.002360   0.764   0.447
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.493 on 89 degrees of freedom
Multiple R-squared:  0.006519, Adjusted R-squared:  -0.004644
F-statistic: 0.584 on 1 and 89 DF, p-value: 0.4468
```

La durée de l'intervalle minimal entre 2 trains à l'hyperpointe dans la section dimensionnante de la ligne ne présente pas d'impact significatif sur le taux de retard au seuil de 10%. L'hyperpointe désigne la période durant laquelle la demande de transport est la plus forte. Elle peut être infra-horaire, par exemple 7h45-8h15, à la différence de l'heure de pointe, qui représente une heure complète.

- **Le nombre de trains par sens à l'heure de pointe du matin :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ NbTrainsHPM_2019, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-5.6930 -2.6230 -0.1404  2.1857 10.3916

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.69262    0.91175   9.534 2.94e-15 ***
NbTrainsHPM_2019 0.11315    0.04567   2.478  0.0151 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.39 on 89 degrees of freedom
Multiple R-squared:  0.06452, Adjusted R-squared:  0.05401
F-statistic: 6.138 on 1 and 89 DF, p-value: 0.01512

```

Le nombre de trains par sens à l'heure de pointe du matin présente un effet significatif au seuil de 5%, mais pas au seuil de 1%.

La valeur du coefficient indique qu'une production à l'heure de pointe du matin (HPM) plus élevée de 10 trains/sens se traduit en moyenne par un taux de retard plus élevé de 1,1 point. Sur l'année 2019, l'étendue entre les lignes de la production à l'HPM est de 27 trains HPM/sens, ce qui se traduit d'après le modèle par un impact d'environ 3,0 points de retard des RER et Transilien entre les lignes dont l'offre à l'HPM est la plus importante (B et D avec 32 trains/sens dans le tunnel entre Châtelet-Les Halles et la Gare du Nord) et la ligne dont l'offre à l'HPM est la plus faible (K). Au sein de la série, l'écart-type est d'environ 8 trains/sens, se traduisant par une différence de taux de retard de 0,9 points en moyenne d'après le modèle.

- **Le fait qu'une ou plusieurs section(s) des voies de la ligne soi(en)t partagée(s) entre plusieurs lignes :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ VoiesPartagées, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-5.9459 -2.9659 -0.2359  2.6191  8.8941

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.8746    0.9720  11.188 <2e-16 ***
VoiesPartagées -0.1187    1.0499  -0.113   0.91
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.505 on 89 degrees of freedom
Multiple R-squared:  0.0001436, Adjusted R-squared: -0.01109
F-statistic: 0.01279 on 1 and 89 DF, p-value: 0.9102

```

Le fait qu'une ou plusieurs section(s) des voies de la ligne soi(en)t partagée(s) entre plusieurs lignes ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

- **L'âge moyen du matériel roulant de la ligne :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ AgeMoyen_MaterielRoulant_1erJanvier,
    data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-7.3907 -2.4099 -0.0351  2.0651  7.5483

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)         7.59617    0.84168   9.025 3.33e-14 ***
AgeMoyen_MaterielRoulant_1erJanvier  0.14514    0.03524   4.118 8.53e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.212 on 89 degrees of freedom
Multiple R-squared:  0.1601,    Adjusted R-squared:  0.1506
F-statistic: 16.96 on 1 and 89 DF,  p-value: 8.527e-05
```

L'âge moyen du matériel roulant présente un effet sur le taux de retard significatif au seuil de à 0,1%.

La valeur du coefficient indique qu'en moyenne, un parc de matériel roulant sur la ligne plus âgé de 10 ans s'accompagne d'un taux de retard annuel plus élevé de 1,5 points, ce qui est conforme à l'intuition. En effet, un matériel roulant âgé présente un risque de panne ou d'incident matériel plus important. Il est donc susceptible de connaître davantage de retard en moyenne.

- **Le nombre moyen d'arrêts effectués par les circulations de trains sur la ligne à l'heure de pointe du matin :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ NbMoyArrêtsHPM_2019, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-5.8888 -2.5083 -0.0061  2.2690  9.7000

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.4345     0.9067   9.302 8.89e-15 ***
NbMoyArrêtsHPM_2019  0.1783     0.0637   2.799 0.00629 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.36 on 89 degrees of freedom
Multiple R-squared:  0.08089,    Adjusted R-squared:  0.07056
F-statistic: 7.833 on 1 and 89 DF,  p-value: 0.00629

```

Le nombre d'arrêts effectués par les trains de la ligne à l'HPM présente un effet significatif au seuil de 1%.

La valeur du coefficient indique que chaque arrêt supplémentaire à l'heure de pointe du matin (HPM) se traduit en moyenne par un taux de retard plus élevé de 0,18 points. Pour expliquer cette corrélation, on peut supposer que le risque d'occurrence de retards est important lors d'un arrêt. En effet, certains événements peuvent survenir lors d'un arrêt en gare et retarder le départ du train : par exemple, malaise voyageur, blocage (volontaire ou non) des portes par un voyageur retardant le train, dysfonctionnement du matériel roulant lors de son redémarrage, ou parfois des infrastructures ferroviaires en sortie de gare (par exemple la signalisation), etc.

- **Le nombre moyen d'arrêts effectués dans Paris intra-muros par les circulations de trains sur la ligne à l'heure de pointe du matin :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ NbMoyArrêtsHPMdansParis_2019, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-5.5724 -2.7650 -0.4518  2.5899  9.2263

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     10.2173     0.4827  21.168 <2e-16 ***
NbMoyArrêtsHPMdansParis_2019  0.2063     0.1189   1.736  0.086 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.447 on 89 degrees of freedom
Multiple R-squared:  0.03275,    Adjusted R-squared:  0.02188
F-statistic: 3.014 on 1 and 89 DF,  p-value: 0.08603

```

Le nombre d'arrêts effectués dans Paris intra-muros par les circulations de trains sur la ligne à l'HPM ne présente pas d'impact significatif sur le taux de retard au seuil de 5%.

- **Le temps de parcours total moyen des circulations sur la ligne à l'heure de pointe du matin :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ TpsParcoursMoyHPM_2019, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-7.1427 -1.5004 -0.4461  1.1516  9.2620

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.63326    1.21288    1.347   0.182
TpsParcoursMoyHPM_2019 0.17867    0.02305    7.751 1.41e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.708 on 89 degrees of freedom
Multiple R-squared:  0.403,    Adjusted R-squared:  0.3963
F-statistic: 60.07 on 1 and 89 DF,  p-value: 1.406e-11
```

Le temps de parcours total moyen des circulations sur la ligne à l'HPM présente un effet significatif au seuil de 0,1%.

La valeur du coefficient indique qu'un temps moyen de parcours total à l'HPM plus élevé de 10 minutes se traduit en moyenne par un taux de retard plus élevé de 1,8 point. La corrélation observée s'explique probablement par le fait que plus un trajet en RER ou Transilien dure longtemps, plus il est soumis longtemps au risque d'occurrence d'un incident entraînant un retard.

Sur l'année 2019, l'étendue entre les lignes du temps de parcours total à l'HPM est de 41 minutes, ce qui se traduit d'après le modèle par un impact d'environ 7,3 points de retard des RER et Transilien entre la ligne dont le temps moyen de parcours total à l'HPM est le plus long (D) et celle dont le temps moyen de parcours à l'HPM est le plus court (E). Au sein de la série, l'écart-type est d'environ 13 minutes, se traduisant par une différence de taux de retard de 2,3 points en moyenne d'après le modèle.

- **La durée moyenne d'un arrêt (temps d'échange voyageurs) à l'heure de pointe du matin :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ DuréeMoyInterstation_2019, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-5.7201 -2.8411 -0.3014  2.5027  8.3015

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.8065     1.1343   7.764 1.32e-11 ***
DuréeMoyInterstation_2019  0.4926     0.2694   1.828  0.0708 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.441 on 89 degrees of freedom
Multiple R-squared:  0.0362,    Adjusted R-squared:  0.02537
F-statistic: 3.343 on 1 and 89 DF,  p-value: 0.07084

```

La durée moyenne d'un arrêt (temps d'échange voyageurs) à l'HPM ne présente pas d'impact significatif sur le taux de retard au seuil de 5%.

- **Le type de contrôle de vitesse utilisé depuis 2015 :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ TypeContrôleVitesse, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-6.8731 -2.1557 -0.5833  2.4067  9.8867

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      9.7633     0.4147  23.545 < 2e-16 ***
TypeContrôleVitesse  1.8749     0.4508   4.159 7.34e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.207 on 89 degrees of freedom
Multiple R-squared:  0.1627,    Adjusted R-squared:  0.1533
F-statistic: 17.3 on 1 and 89 DF,  p-value: 7.339e-05

```

D'après le modèle, plus le type de contrôle de vitesse utilisé depuis 2015 est « favorable à la fluidification du trafic » d'après l'audit EPFL [6], moins la ligne est ponctuelle, ce qui est contraire à l'intuition relativement à l'effet de cette variable.

L'effet sur le taux de retard est significatif au seuil de 0,1%.

- **Le fait que la Gare du Nord appartienne ou non à la ligne RER ou Transilien :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ GareDuNord, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-6.5443 -2.7244 -0.1044  2.2007  8.2957

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.5144     0.4388  23.961  <2e-16 ***
GareDuNord    0.8398     0.7911   1.062   0.291
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.483 on 89 degrees of freedom
Multiple R-squared:  0.01251,    Adjusted R-squared:  0.00141
F-statistic: 1.127 on 1 and 89 DF,  p-value: 0.2913
```

Le fait que la Gare du Nord appartienne ou non à la ligne ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

- **Le fait que la Gare de Lyon appartienne ou non à la ligne RER ou Transilien :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ GareDeLyon, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-7.3129 -2.1889  0.0911  2.0041  9.8311

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.8189     0.3622  27.110  < 2e-16 ***
GareDeLyon    4.1340     0.7540   5.483  3.87e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.03 on 89 degrees of freedom
Multiple R-squared:  0.2525,    Adjusted R-squared:  0.2441
F-statistic: 30.06 on 1 and 89 DF,  p-value: 3.866e-07
```

D'après le modèle, les lignes de RER ou Transilien traversant la Gare de Lyon sont moins ponctuelles que les autres lignes RER ou Transilien. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des lignes RER ou Transilien empruntant la Gare de Lyon est supérieur de 4,1 points à celui des autres lignes. La moins bonne ponctualité de ces lignes semble en accord avec la qualification de « nœud ferroviaire majeur » donnée à la gare de Lyon par le CGEDD [7], le phénomène de saturation constaté par le CGEDD au niveau du secteur de la gare de Lyon [8], et les contraintes d'exploitation des lignes RER D et Transilien R entraînées par le partage des voies avec des trains TER Bourgogne-Franche-Comté, Intercités de la ligne Paris - Clermont-Ferrand, ainsi que des trains de fret, évoquées par RFF et la SNCF dans le document de préparation au schéma directeur de la ligne Transilien R [9].

- **Le fait que la Gare Saint-Lazare appartienne ou non à la ligne RER ou Transilien :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```
Call:
lm(formula = TxRetard ~ GareStLazare, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-5.7496 -2.7777 -0.1557  2.4393  9.0904

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.5596     0.3952  26.717  <2e-16 ***
GareStLazare   1.3861     1.0077   1.376   0.172
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.468 on 89 degrees of freedom
Multiple R-squared:  0.02082,    Adjusted R-squared:  0.009815
F-statistic: 1.892 on 1 and 89 DF,  p-value: 0.1724
```

Le fait que la Gare Saint-Lazare appartienne ou non à la ligne ne présente pas d'impact significatif sur le taux de retard au seuil de 10%.

- **Le fait que la gare RER de Châtelet – Les Halles appartienne ou non à la ligne RER ou Transilien :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :


```

Call:
lm(formula = TxRetard ~ ChâteletLesHalles, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-6.9519 -2.2971 -0.0219  2.0429  9.7229

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.9271     0.3751  26.468 < 2e-16 ***
ChâteletLesHalles  3.6648     0.7807   4.694 9.66e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.138 on 89 degrees of freedom
Multiple R-squared:  0.1984,    Adjusted R-squared:  0.1894
F-statistic: 22.03 on 1 and 89 DF,  p-value: 9.659e-06

```

D'après le modèle, les lignes de RER ou Transilien traversant la gare RER de Châtelet – Les Halles sont moins ponctuelles que les autres lignes RER ou Transilien. L'effet sur le taux de retard est significatif au seuil de 0,1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des lignes RER ou Transilien empruntant la gare RER de Châtelet – Les Halles est supérieur de 3,7 points à celui des autres lignes. La moins bonne ponctualité de ces lignes semble en accord le phénomène de saturation à l'heure de pointe constatée par l'EPFL au niveau de cette gare ainsi qu'au niveau du tunnel reliant cette gare avec la Gare du Nord [\[6\]](#).

- **Le fait que la gare de l'Est appartienne ou non à la ligne RER ou Transilien :**

Les résultats des tests statistiques effectués pour ce modèle sont synthétisés dans le tableau suivant :

```

Call:
lm(formula = TxRetard ~ GareParisEst, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-6.3314 -2.5736  0.1886  2.4336  8.5086

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.1414     0.3868  28.804 <2e-16 ***
GareParisEst -2.3957     0.9862  -2.429  0.0171 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.394 on 89 degrees of freedom
Multiple R-squared:  0.06219,    Adjusted R-squared:  0.05165
F-statistic: 5.902 on 1 and 89 DF,  p-value: 0.01714

```

D'après le modèle, les lignes de RER ou Transilien traversant la gare de l'Est sont davantage ponctuelles que les autres lignes RER ou Transilien. L'effet sur le taux de retard est significatif au seuil de 5%, mais pas au seuil de 1%.

La valeur du coefficient indique qu'en moyenne, le taux de retard des lignes RER ou Transilien empruntant la gare de l'Est est inférieur de 2,4 points à celui des autres lignes.

Annexe 8 : Tableau de colinéarité entre les variables explicatives du taux de retard des liaisons TGV

	NbCirc_Liaison_an	DuréeTrajet	DistanceMoyenne	VitesseMoyenne	AgeLGV_axe	DépartParis	ArrivéeParis	Liaison_radiale	DépartMarseille	ArrivéeMarseille	DépartLyon	ArrivéeLyon	DépartOuArr_Lyon_Marseille	NbNoeudsMajeursTraversés	LiaisonPasseParAvignon	LiaisonPasseParBordeaux	LiaisonPasseParDijon	LiaisonPasseParLille	LiaisonPasseParRennes	LiaisonPasseParLeMans	LiaisonPasseParTours	DépartOuArr_NoedSecondaire	NbNoeudsSecondaires	NbBifurcationsLGV	NbBifurcationsLGV_horsIDF	NbBifurcationsRéseauClassique	TMJA_max_Liaison_2019	NbMoyenArrêts_2019	LongueurRéseauClassique	
NbCirc_Liaison_an	X																													
DuréeTrajet		X	X	X																										
DistanceMoyenne		X	X	X																										
VitesseMoyenne		X	X	X																										
AgeLGV_axe					X																									
DépartParis						X		X																						
ArrivéeParis							X	X																						
Liaison_radiale						X	X	X																						
DépartMarseille									X					X	X															
ArrivéeMarseille										X				X	X															
DépartLyon											X			X	X															
ArrivéeLyon												X		X	X															
DépartOuArr_Lyon_Marseille									X	X	X	X	X	X	X															
NbNoeudsMajeursTraversés					X	X	X	X	X	X	X	X	X	X	X															
LiaisonPasseParAvignon															X							X	X	X	X					
LiaisonPasseParBordeaux																X						X	X	X	X					
LiaisonPasseParDijon																	X					X	X	X	X					
LiaisonPasseParLille																		X				X	X	X	X					
LiaisonPasseParRennes																			X			X	X	X	X					
LiaisonPasseParLeMans																				X		X	X	X	X					
LiaisonPasseParTours																					X	X	X	X	X					
DépartOuArr_NoedSecondaire															X	X	X	X	X	X	X	X	X	X	X					
NbNoeudsSecondaires															X	X	X	X	X	X	X	X	X	X	X					
NbBifurcationsLGV															X						X	X	X	X						
NbBifurcationsLGV_horsIDF															X						X	X	X	X						
NbBifurcationsRéseauClassique																										X				
TMJA_max_Liaison_2019																											X			
NbMoyenArrêts_2019																												X		
LongueurRéseauClassique																													X	

Code couleur : **en jaune**, variables présentant le plus grand nombre de relations de colinéarité avec d'autres variables.

Annexe 9 : Première analyse des coefficients associés aux variables explicatives du modèle de régression linéaire multiple retenu pour les services TGV

```

Call:
lm(formula = TxRetard ~ DépartParis + DépartLyon + LiaisonPasseParAvignon +
    LiaisonPasseParRennes + LiaisonPasseParDijon + DépartOuArr_NoeudSecondaire +
    NbMoyenArrêts_2019 + TMJA_max_Total_Liaison_2000, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-13.1148  -3.2723  -0.1854   2.8244  18.7606

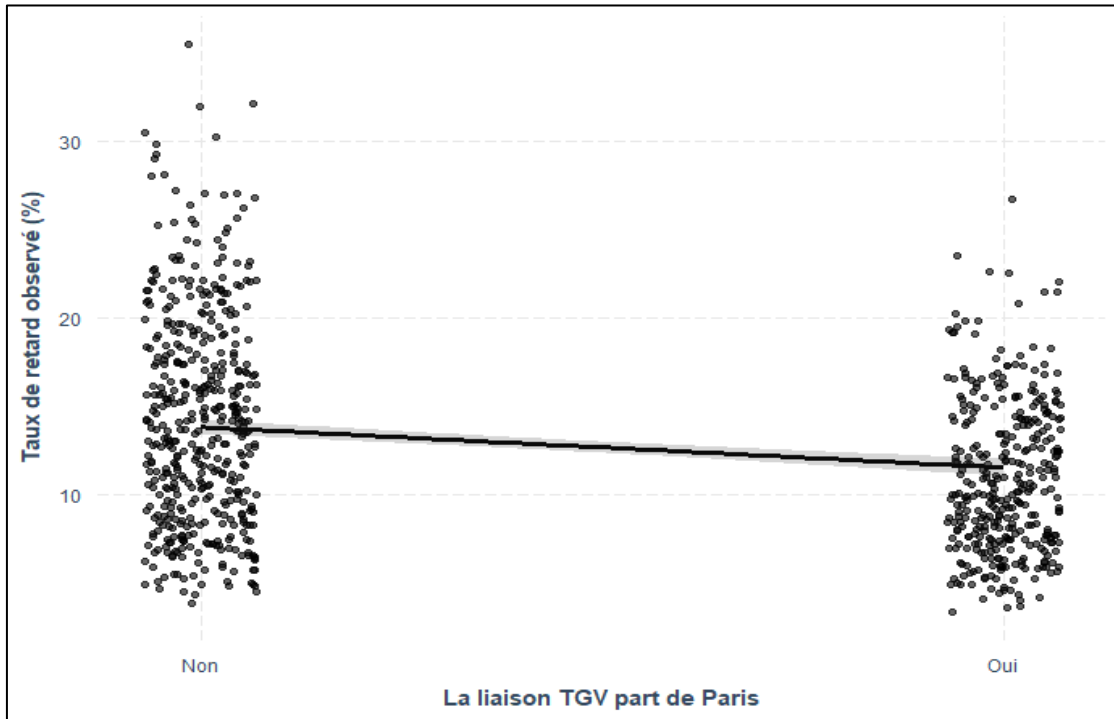
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      10.986504   0.669113   16.420 < 2e-16 ***
DépartParis      -2.291971   0.335944   -6.822 1.76e-11 ***
DépartLyon       4.890913   0.754562    6.482 1.58e-10 ***
LiaisonPasseParAvignon
2.460692   0.489295    5.029 6.08e-07 ***
LiaisonPasseParRennes
-2.849098   0.566522   -5.029 6.08e-07 ***
LiaisonPasseParDijon
-4.838904   0.743421   -6.509 1.33e-10 ***
DépartOuArr_NoeudSecondaire
3.113007   0.402450    7.735 3.10e-14 ***
NbMoyenArrêts_2019
0.314262   0.104626    3.004 0.00275 **
TMJA_max_Total_Liaison_2000
0.005308   0.002261    2.348 0.01912 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.583 on 803 degrees of freedom
Multiple R-squared:  0.3244,    Adjusted R-squared:  0.3177
F-statistic: 48.2 on 8 and 803 DF,  p-value: < 2.2e-16

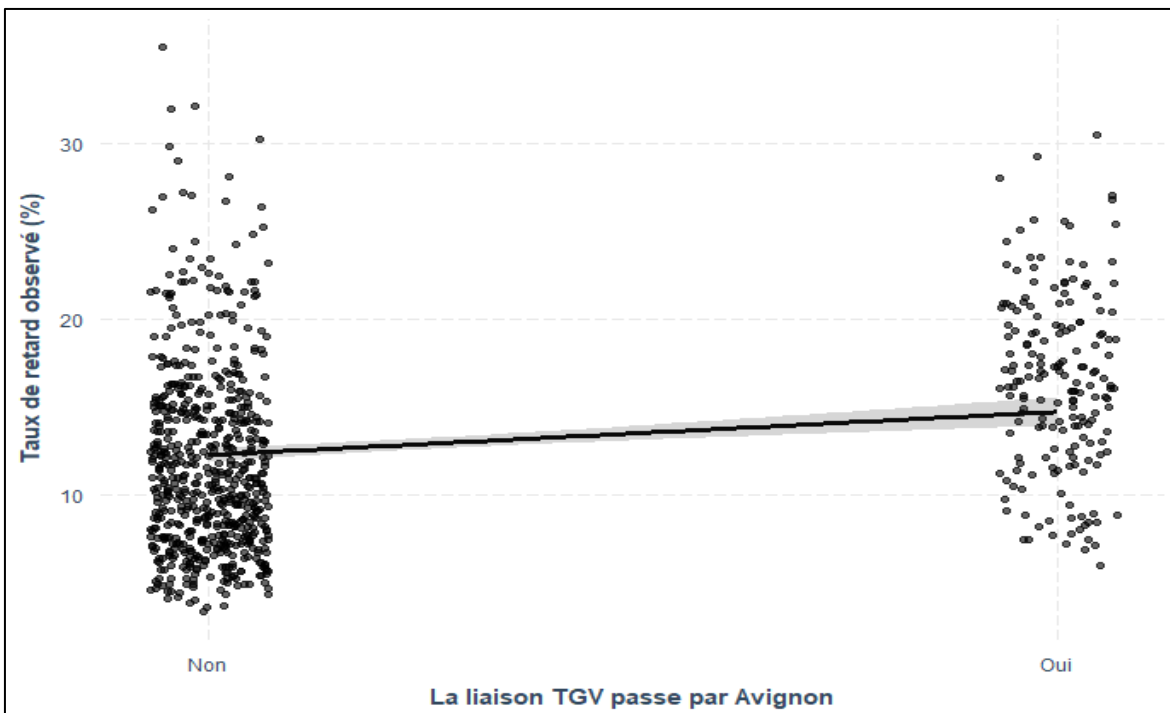
```

D'après le modèle retenu, une liaison TGV ayant pour départ Paris présente en moyenne un taux de retard plus faible d'environ 2,3 points, alors qu'une liaison TGV au départ de Lyon Part-Dieu présente en moyenne un taux de retard plus élevé de 4,9 points. Une explication possible à ce dernier point pourrait être qu'il y a une saturation du réseau ferroviaire au niveau du nœud ferroviaire lyonnais, ce qui rejoint le constat effectué par plusieurs experts ferroviaires dans la littérature spécialisée. [\[1\]](#)[\[2\]](#)[\[3\]](#)

Le graphique ci-dessous illustre la meilleure ponctualité des liaisons partant de Paris :



Le modèle indique également que les liaisons TGV qui passent par Avignon sur la LGV Méditerranée et desservent ou bien partent du littoral méditerranéen sont moins ponctuelles. Au-delà de l'impact du nœud ferroviaire majeur de Marseille Saint-Charles, une explication partielle de la moins bonne ponctualité de ces liaisons pourraient être l'impact des événements météorologiques qui surviennent dans ces régions et détériorent les infrastructures ferroviaires (fortes chaleurs en été, épisodes méditerranéens, ou des problèmes de congestion, ou autres). Le graphique ci-dessous illustre cette relation :



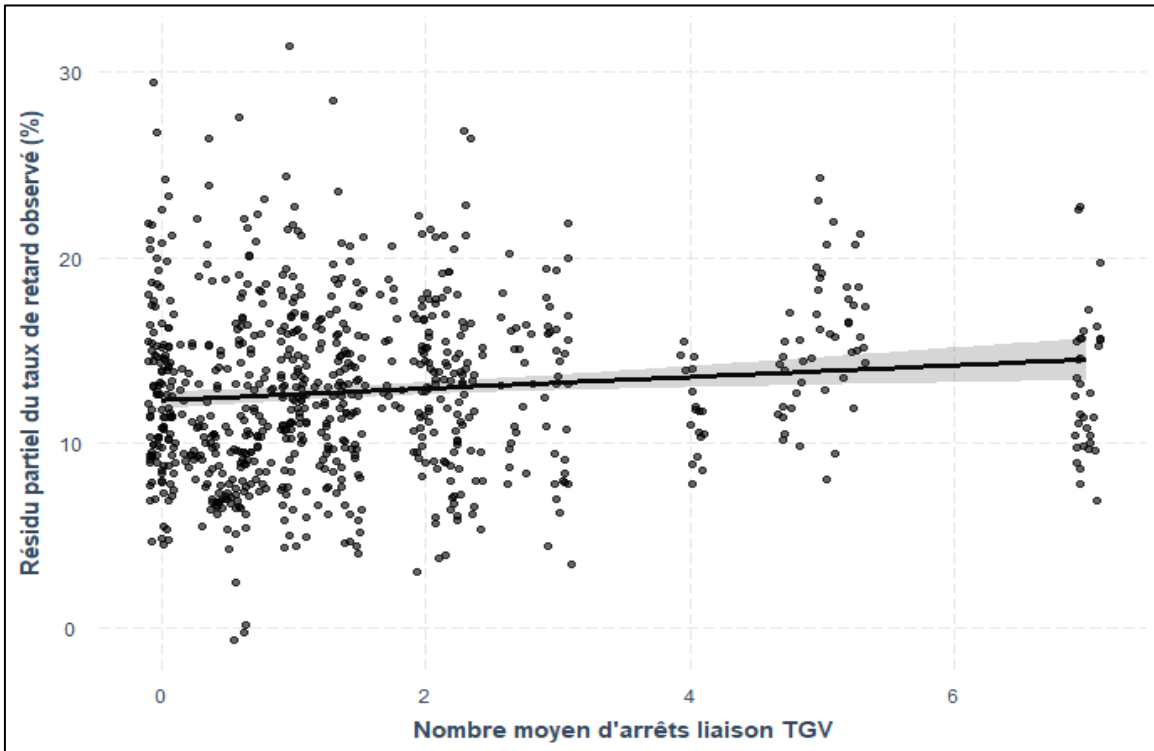
La valeur du coefficient associé au passage de la liaison TGV par Rennes indique qu'en moyenne, le taux de retard des liaisons concernées est moins élevé de 2,8 points. Les liaisons TGV qui traversent Rennes empruntent toutes la LGV Bretagne-Pays de la Loire. Cette LGV est très récente (mise en service en 2017) et était relativement peu intensément utilisée en 2018 d'après la carte « Le trafic des trains aptes à la Grande Vitesse » de l'atlas du réseau ferré en France – édition 2020, publié par SNCF Réseau. Les voies ferroviaires et les autres infrastructures ferroviaires de la LGV Bretagne-Pays de la Loire sont donc *a priori* en très bon état pour le moment, ce qui pourrait expliquer pour partie la bonne ponctualité des trains sur les liaisons passant par Rennes. En outre, le nœud ferroviaire de Rennes a fait l'objet d'investissements importants de capacité lors de la construction de la LGV Bretagne-Pays de la Loire, ce qui a probablement contribué à la bonne ponctualité des liaisons TGV qui le traversent.

Un raisonnement similaire pourrait être invoqué pour expliquer en partie la nettement meilleure ponctualité des liaisons TGV traversant l'agglomération de Dijon (-4,8 points de retard en moyenne). En effet, ces liaisons TGV empruntent la LGV Rhin-Rhône. Cette LGV est en effet récente (mise en service en 2011) et était relativement peu intensément utilisée en 2018 d'après l'atlas du réseau ferré en France – édition 2020, publié par SNCF Réseau.

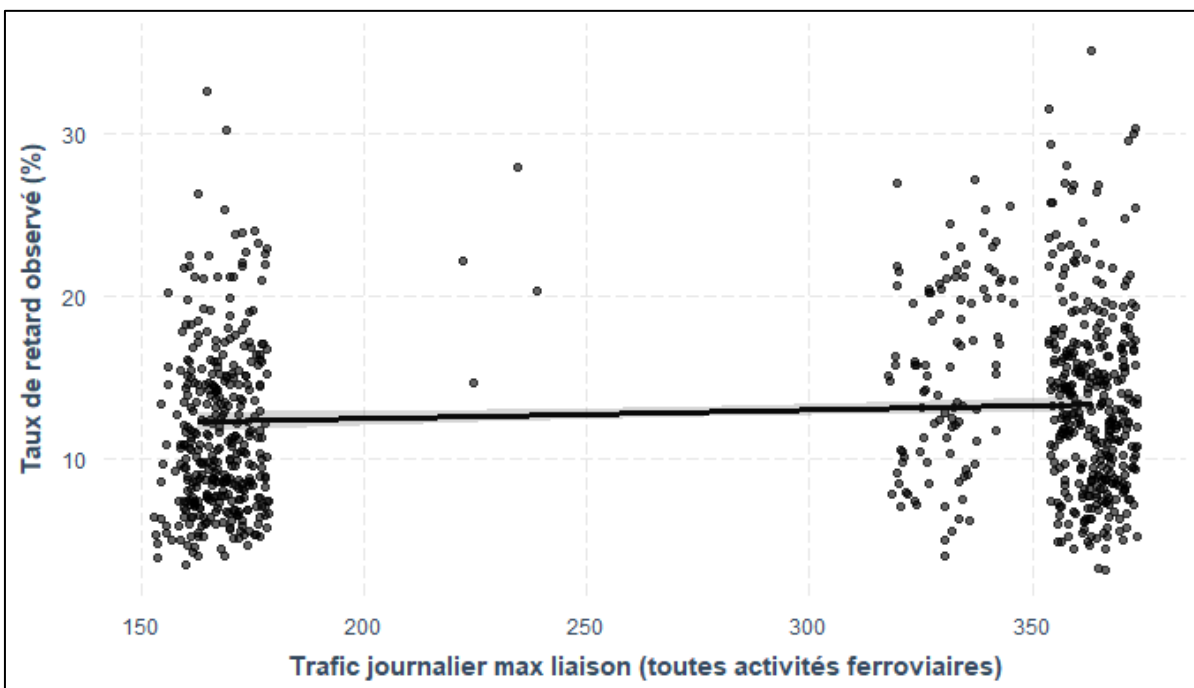
Le modèle indique également que le taux de retard des TGV partant ou arrivant au niveau d'un des nœuds ferroviaires secondaires du réseau ferré retenus (Avignon, Bordeaux, Le Mans, Lille, Tours) est supérieur de près de 3,1 points à celui des autres TGV. Du fait de la saturation du réseau ferré observée au niveau de ces nœuds, il est possible qu'un plus grand nombre de TGV soient impactés si un retard s'y produit. Les difficultés de gestion du trafic au niveau d'un nœud ferroviaire pourraient également accroître le risque de retard.

Selon le modèle retenu, plus une circulation TGV effectue d'arrêts, plus elle est susceptible de connaître des retards, ce qui est conforme à l'intuition relativement à l'effet de cette variable. En revanche, on peut noter que l'impact du nombre d'arrêts sur la ponctualité n'est pas très prononcé d'après le modèle (+0,3 points par arrêt supplémentaire).

Le graphique ci-dessous illustre l'effet du nombre d'arrêts sur la ponctualité des liaisons TGV. En axe des ordonnées, le résidu partiel du taux de retard observé est représenté, et non plus le taux de retard observé. Il s'agit en fait du taux de retard observé auquel on retranche l'effet des autres variables de la régression, afin d'identifier plus facilement la relation de corrélation graphiquement. On note que pour un niveau donné du nombre d'arrêts moyen, le résidu partiel du taux de retard observé présente une forte dispersion, ce qui parasite la lecture de l'effet de cette variable explicative.



La valeur du coefficient associé au niveau du trafic moyen journalier pour l'année 2000 de l'ensemble des circulations de trains (y compris fret et « haut-le-pied ») sur le tronçon dimensionnant du réseau ferré emprunté par la liaison TGV indique qu'en moyenne, si ce tronçon est emprunté par 100 trains de plus par jour, il présentera un taux de retard plus élevé de 0,5 points. Le graphique ci-dessous illustre la corrélation identifiée par le modèle retenu :



Un élément partiel d'explication pourrait être que lorsqu'une circulation TGV emprunte des lignes très fréquentées, elle soit davantage exposée au risque de subir un retard qui se propagerait sur le réseau d'une circulation de train à la suivante, et cela sans que ce retard ne puisse nécessairement être « rattrapé » au cours du trajet, puisque le débit de circulations de trains sur la ligne est déjà très élevé.

Annexe 10 : Distances ferroviaires par liaison Intercités

Les données sont issues du Mémento statistique 2018 de la SNCF pour les 5 premières relations, et de la base de données Rome2Rio pour les 35 suivantes :

Relation	Distance (km)	Liaison AQST sens_1	Liaison AQST sens_2
Source	Mémento statistique de la SNCF - 2018		
Paris - Caen	239	CAEN => PARIS ST LAZARE	PARIS ST LAZARE => CAEN
Paris - Le Havre	228	LE HAVRE => PARIS ST LAZARE	PARIS ST LAZARE => LE HAVRE
Paris - Limoges	400	LIMOGES => PARIS AUSTERLITZ	PARIS AUSTERLITZ => LIMOGES
Lyon-Part-Dieu - Nantes (via Massy)	807	LYON PART DIEU => NANTES	NANTES => LYON PART DIEU
Bordeaux - Marseille	680	BORDEAUX ST JEAN => MARSEILLE ST CHARLES	MARSEILLE ST CHARLES => BORDEAUX ST JEAN
Source	Rome2Rio	Liaison AQST sens_1	Liaison AQST sens_2
Albi/Rodez - Paris	734	ALBI/RODEZ => PARIS AUSTERLITZ	PARIS AUSTERLITZ => ALBI/RODEZ
Amiens - Paris	131	AMIENS => PARIS NORD	PARIS NORD => AMIENS
Bayonne - Toulouse	321	BAYONNE => TOULOUSE MATABIAU	TOULOUSE MATABIAU => BAYONNE
Belfort - Paris	441	BELFORT => PARIS EST	PARIS EST => BELFORT
Béziers - Clermont	385	BEZIERS => CLERMONT-FERRAND	CLERMONT-FERRAND => BEZIERS
Bordeaux - Limoges	226	BORDEAUX ST JEAN => LIMOGES	LIMOGES => BORDEAUX ST JEAN
Bordeaux - Nantes	372	BORDEAUX ST JEAN => NANTES	NANTES => BORDEAUX ST JEAN
Bordeaux - Nice	901	BORDEAUX ST JEAN => NICE VILLE	NICE VILLE => BORDEAUX ST JEAN
Boulogne Ville - Paris	253	BOULOGNE VILLE => PARIS NORD	PARIS NORD => BOULOGNE VILLE

Bourg St Maurice/St Gervais - Paris	620	BOURG ST MAURICE/ST GERVAIS => PARIS AUSTERLITZ	PARIS AUSTERLITZ => BOURG ST MAURICE/ST GERVAIS
Bourges - Paris	231	BOURGES => PARIS AUSTERLITZ	PARIS AUSTERLITZ => BOURGES
Briançon - Paris	766	BRIANCON => PARIS AUSTERLITZ	PARIS AUSTERLITZ => BRIANCON
Caen - Tours	275	CAEN => TOURS	TOURS => CAEN
Cahors/Brive - Paris	593	CAHORS/BRIVE => PARIS AUSTERLITZ	PARIS AUSTERLITZ => CAHORS/BRIVE
Cambrai/Maubeuge - Paris	206	CAMBRAI/MAUBEUGE => PARIS NORD	PARIS NORD => MAUBEUGE/CAMBRAI
Cerbere - Paris	958	CERBERE => PARIS AUSTERLITZ	PARIS AUSTERLITZ => CERBERE
Cerbere/Toulouse - Paris	958	CERBERE/TOULOUSE => PARIS AUSTERLITZ	PARIS AUSTERLITZ => CERBERE/TOULOUSE
Cherbourg - Paris	369	CHERBOURG => PARIS ST LAZARE	PARIS ST LAZARE => CHERBOURG
Clermont-Ferrand - Nîmes	294	CLERMONT-FERRAND => NIMES	NIMES => CLERMONT-FERRAND
Clermont-Ferrand - Paris	418	CLERMONT-FERRAND => PARIS BERCY (VOYAGEURS)	PARIS BERCY (VOYAGEURS) => CLERMONT-FERRAND
Granville - Paris	325	GRANVILLE => MONTPARNASSE-VAUGIRARD	MONTPARNASSE-VAUGIRARD => GRANVILLE
Hendaye - Paris	813	HENDAYE => PARIS AUSTERLITZ	PARIS AUSTERLITZ => HENDAYE
Hendaye - Toulouse	354	HENDAYE => TOULOUSE MATABIAU	TOULOUSE MATABIAU => HENDAYE
Latour De Carol - Paris	870	LATOUR DE CAROL => PARIS AUSTERLITZ	PARIS AUSTERLITZ => LATOUR DE CAROL
Maubeuge - Paris	227	MAUBEUGE => PARIS NORD	PARIS NORD => MAUBEUGE
Nevers - Paris	252	NEVERS => PARIS BERCY (VOYAGEURS)	PARIS BERCY (VOYAGEURS) => NEVERS
Orléans - Paris	121	ORLEANS => PARIS AUSTERLITZ	PARIS AUSTERLITZ => ORLEANS
Rouen - Paris	139	ROUEN R.D. => PARIS ST LAZARE	PARIS ST LAZARE => ROUEN R.D.
St Quentin - Paris	153	ST QUENTIN => PARIS NORD	PARIS NORD => ST QUENTIN
Toulouse - Paris	707	TOULOUSE MATABIAU => PARIS AUSTERLITZ	PARIS AUSTERLITZ => TOULOUSE MATABIAU
Toulouse/Rodez - Paris	655	TOULOUSE MATABIAU/RODEZ => PARIS AUSTERLITZ	PARIS AUSTERLITZ => TOULOUSE MATABIAU/RODEZ
Tours - Paris	233	TOURS => PARIS AUSTERLITZ	PARIS AUSTERLITZ => TOURS

Trouville-Deauville - Paris	218	TROUVILLE-DEAUVILLE => PARIS ST LAZARE	PARIS ST LAZARE => TROUVILLE-DEAUVILLE
Troyes - Paris	166	TROYES => PARIS EST	PARIS EST => TROYES
Paris - Nice	1088	PARIS AUSTERLITZ => NICE VILLE	NICE VILLE => PARIS AUSTERLITZ

Annexe 11 : Estimation du nombre d'arrêts moyen par liaison Intercités au cours d'un Jour Ouvrable de Base en 2020

Une estimation du nombre d'arrêts moyen des trains circulant sur chaque liaison au cours d'un Jour Ouvrable de Base (JOB) - le 3 novembre 2020 – a été calculé pour l'amplitude horaire retenue de 6h à 22h. Les données utilisées ont été recueillies sur le site <https://www.horaires-de-trains.fr/> et proviennent de l'OpenData de la SNCF. Pour simplifier l'étude, nous avons considéré que le nombre d'arrêts moyen ne dépend pas en première approche du sens de circulation sur la liaison Intercités. Si cette hypothèse semble raisonnable pour un grand nombre de liaisons, elle n'est pas nécessairement exacte sur la totalité des liaisons.

Liaison Sens 1	Liaison Sens 2	NbArretsMoyen
ALBI/RODEZ => PARIS AUSTERLITZ	PARIS AUSTERLITZ => ALBI/RODEZ	10,0
AMIENS => PARIS NORD	PARIS NORD => AMIENS	2,9
BAYONNE => TOULOUSE MATABIAU	TOULOUSE MATABIAU => BAYONNE	5,0
BELFORT => PARIS EST	PARIS EST => BELFORT	8,0
BEZIERS => CLERMONT-FERRAND	CLERMONT-FERRAND => BEZIERS	22,0
BORDEAUX ST JEAN => LIMOGES	LIMOGES => BORDEAUX ST JEAN	5,0
BORDEAUX ST JEAN => NANTES	NANTES => BORDEAUX ST JEAN	3,0
BORDEAUX ST JEAN => NICE VILLE	NICE VILLE => BORDEAUX ST JEAN	12,8
BOULOGNE VILLE => PARIS NORD	PARIS NORD => BOULOGNE VILLE	7,6
BOURG ST MAURICE/ST GERVAIS => PARIS AUSTERLITZ	PARIS AUSTERLITZ => BOURG ST MAURICE/ST GERVAIS	5,0
BOURGES => PARIS AUSTERLITZ	PARIS AUSTERLITZ => BOURGES	1,3
BRIANCON => PARIS AUSTERLITZ	PARIS AUSTERLITZ => BRIANCON	9,0
CAEN => TOURS	TOURS => CAEN	9,0
CAHORS/BRIVE => PARIS AUSTERLITZ	PARIS AUSTERLITZ => CAHORS/BRIVE	7,4
CAMBRAI/MAUBEUGE => PARIS NORD	PARIS NORD => MAUBEUGE/CAMBRAI	9,0
CERBERE => PARIS AUSTERLITZ	PARIS AUSTERLITZ => CERBERE	12,0
CERBERE/TOULOUSE => PARIS AUSTERLITZ	PARIS AUSTERLITZ => CERBERE/TOULOUSE	9,0
CHERBOURG => PARIS ST LAZARE	PARIS ST LAZARE => CHERBOURG	5,0
CLERMONT-FERRAND => NIMES	NIMES => CLERMONT-FERRAND	14,0
CLERMONT-FERRAND => PARIS BERCY (VOYAGEURS)	PARIS BERCY (VOYAGEURS) => CLERMONT-FERRAND	4,0
GRANVILLE => MONTPARNASSE-VAUGIRARD	MONTPARNASSE-VAUGIRARD => GRANVILLE	8,0
HENDAYE => PARIS AUSTERLITZ	PARIS AUSTERLITZ => HENDAYE	5,0
HENDAYE => TOULOUSE MATABIAU	TOULOUSE MATABIAU => HENDAYE	8,0

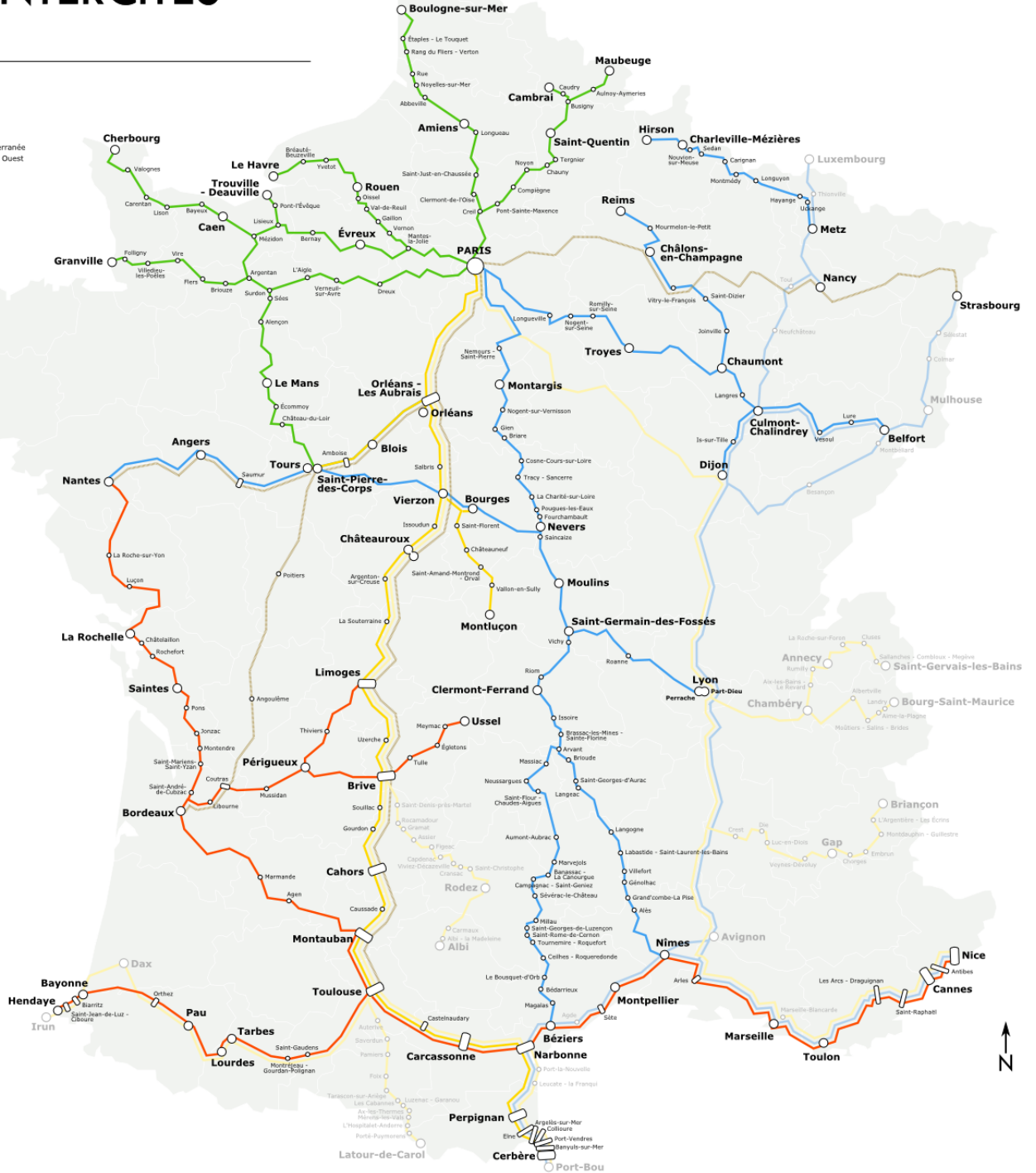
LATOUR DE CAROL => PARIS AUSTERLITZ	PARIS AUSTERLITZ => LATOUR DE CAROL	12,0
MAUBEUGE => PARIS NORD	PARIS NORD => MAUBEUGE	4,0
NEVERS => PARIS BERCY (VOYAGEURS)	PARIS BERCY (VOYAGEURS) => NEVERS	3,3
ORLEANS => PARIS AUSTERLITZ	PARIS AUSTERLITZ => ORLEANS	1,0
ROUEN R.D. => PARIS ST LAZARE	PARIS ST LAZARE => ROUEN R.D.	2,6
ST QUENTIN => PARIS NORD	PARIS NORD => ST QUENTIN	3,5
TOULOUSE MATABIAU => PARIS AUSTERLITZ	PARIS AUSTERLITZ => TOULOUSE MATABIAU	9,7
TOULOUSE MATABIAU/RODEZ => PARIS AUSTERLITZ	PARIS AUSTERLITZ => TOULOUSE MATABIAU/RODEZ	10,0
TOURS => PARIS AUSTERLITZ	PARIS AUSTERLITZ => TOURS	7,2
TROUVILLE-DEAUVILLE => PARIS ST LAZARE	PARIS ST LAZARE => TROUVILLE-DEAUVILLE	4,0
TROYES => PARIS EST	PARIS EST => TROYES	2,0
PARIS AUSTERLITZ => NICE VILLE	NICE VILLE => PARIS AUSTERLITZ	5,0
CAEN => PARIS ST LAZARE	PARIS ST LAZARE => CAEN	1,5
LE HAVRE => PARIS ST LAZARE	PARIS ST LAZARE => LE HAVRE	2,7
LIMOGES => PARIS AUSTERLITZ	PARIS AUSTERLITZ => LIMOGES	3,3
LYON PART DIEU => NANTES	NANTES => LYON PART DIEU	9,0
BORDEAUX ST JEAN => MARSEILLE ST CHARLES	MARSEILLE ST CHARLES => BORDEAUX ST JEAN	8,3

Annexe 12 : Réseau Intercités en 2016

Réseau INTERCITES 2016

Rouen ○ destination principale
Banyuls-sur-Mer ○ destination secondaire

- réseau Nord-Normandie
- réseau Atlantique-Méditerranée
- réseau Paris-Centre-Sud Ouest
- réseau Est-Centre
- Intercités 100% Éco
- Intercités de nuit



Crédits : Lucas Destrem, Wikimedia Commons, 2016, licence Creative Commons Attribution-Share Alike 4.0 International

Annexe 13 : Tableau de colinéarité entre les variables explicatives du taux de retard des liaisons Intercités

	NbCirc_Li aison_an	DuréeTraj et Moy	Distance Moy	VitesseM oyenne	LiaisonN uit	DépartPa ris	ArrivéePa ris	Liaison_r adiale	DépartM arseille	ArrivéeM arseille	DépartLy on	ArrivéeLy on	NbNoeud sMajeurs Traversés	LiaisonTr averseBo rdeaux	LiaisonTr averseNa ntes	LiaisonTr averseTo ulouse	LiaisonTr averseTo urs	LiaisonTr averseOrl éans	LiaisonTr averseBri ve	LiaisonTr averseNa rbonne	LiaisonTr averseNe vers	LiaisonTr averseDij on	LiaisonTr averseCa en
NbCirc_Liaison_an	X																						
DuréeTrajet		X	X	X																			
DistanceMoy		X	X	X									X										
VitesseMoyenne		X	X	X																			
LiaisonNuit					X																		
DépartParis						X		X					X										
ArrivéeParis							X	X					X										
Liaison_radiale						X	X	X					X										
DépartMarseille									X				X										
ArrivéeMarseille										X			X										
DépartLyon											X		X										
ArrivéeLyon												X	X										
NbNoeudsMajeursTraversés			X			X	X	X	X	X	X	X	X										
LiaisonTraverseBordeaux													X										
LiaisonTraverseNantes														X									
LiaisonTraverseToulouse															X								
LiaisonTraverseTours																X							
LiaisonTraverseOrléans																	X						
LiaisonTraverseBrive																		X					
LiaisonTraverseNarbonne																			X				
LiaisonTraverseNevers																				X			
LiaisonTraverseDijon																					X		
LiaisonTraverseCaen																						X	
LiaisonTraverseRouen																							
LiaisonTraverseAmiens																							
LiaisonTraverseSaint-Quentin																							
DépartOuArr_NoeadSecondaire														X	X	X	X	X	X	X	X	X	X
NbNoeudsSecondaires			X											X	X	X	X	X	X	X	X	X	X
Axe_Nord_Normandie																							X
Axe_Med_Atlantique														X		X				X			
Axe_Paris_Centre_SudOuest																	X						
Axe_Est_Centre																							
Est_Ouest																							
Ouest_Est																							
Nord_Sud																							
Sud_Nord																							
NbMoyenArrêts_2020			X																				
TMJA_max_Total_Liaison_2000																							
TMJA_max_Total_Liaison_2000_hors Paris																							

Code couleur : **en jaune**, variables présentant le plus grand nombre de relations de colinéarité avec d'autres variables.

	LiaisonTraverseNarbonne	LiaisonTraverseNevers	LiaisonTraverseDijon	LiaisonTraverseCaen	LiaisonTraverseRouen	LiaisonTraverseAmiens	LiaisonTraverseSaint-Quentin	DépartOuArr_NoeudSecondaire	NbNoeudsSecondaires	Axe_Nord_Normandie	Axe_Med_Atlantique	Axe_Paris_Centre_SudOuest	Axe_Est_Centre	Sens_Est_Ouest	Sens_Ouest_Est	Sens_Nord_Sud	Sens_Sud_Nord	NbMoyenArrêts_2020	TMJA_max_Total_Liaison_2000	TMJA_max_Total_Liaison_2000_horsParis
NbCirc_Liaison_an																				
DuréeTrajet																				
DistanceMoy								X										X		
VitesseMoyenne																				
LiaisonNuit																				
DépartParis																				
ArrivéeParis																				
Liaison_radiale																				
DépartMarseille																				
ArrivéeMarseille																				
DépartLyon																				
ArrivéeLyon																				
NbNoeudsMajeursTraversés																				
LiaisonTraverseBordeaux											X									
LiaisonTraverseNantes								X	X											
LiaisonTraverseToulouse								X	X		X									
LiaisonTraverseTours								X	X											
LiaisonTraverseOrléans								X	X			X								
LiaisonTraverseBrive								X	X											
LiaisonTraverseNarbonne	X							X	X			X								
LiaisonTraverseNevers		X						X	X											
LiaisonTraverseDijon			X					X	X											
LiaisonTraverseCaen				X				X	X	X										
LiaisonTraverseRouen					X			X	X	X										
LiaisonTraverseAmiens						X		X	X	X										
LiaisonTraverseSaint-Quentin							X	X	X	X										
DépartOuArr_NoeudSecondaire	X	X	X	X	X	X	X	X	X											
NbNoeudsSecondaires	X	X	X	X	X	X	X	X	X											
Axe_Nord_Normandie				X	X	X	X													
Axe_Med_Atlantique	X																			
Axe_Paris_Centre_SudOuest																				
Axe_Est_Centre																				
Sens_Est_Ouest																				
Sens_Ouest_Est																				
Sens_Nord_Sud																				
Sens_Sud_Nord																				
NbMoyenArrêts_2020																				
TMJA_max_Total_Liaison_2000																				
TMJA_max_Total_Liaison_2000_horsParis																				

Code couleur : **en jaune**, variables présentant le plus grand nombre de relations de colinéarité avec d'autres variables.

Annexe 14 : Première analyse des coefficients associés aux variables explicatives du modèle de régression linéaire multiple retenu pour les services Intercités

```
Call:
lm(formula = TxRetard ~ LiaisonNuit + NbNoeudsMajeursTraversés +
    LiaisonTraverseBordeaux + LiaisonTraverseToulouse + LiaisonTraverseDijon +
    LiaisonTraverseRouen + DépartOuArr_NoeudSecondaire + Axe_Est_Centre +
    Nord_Sud, data = data_PonctuIC_liaisons)

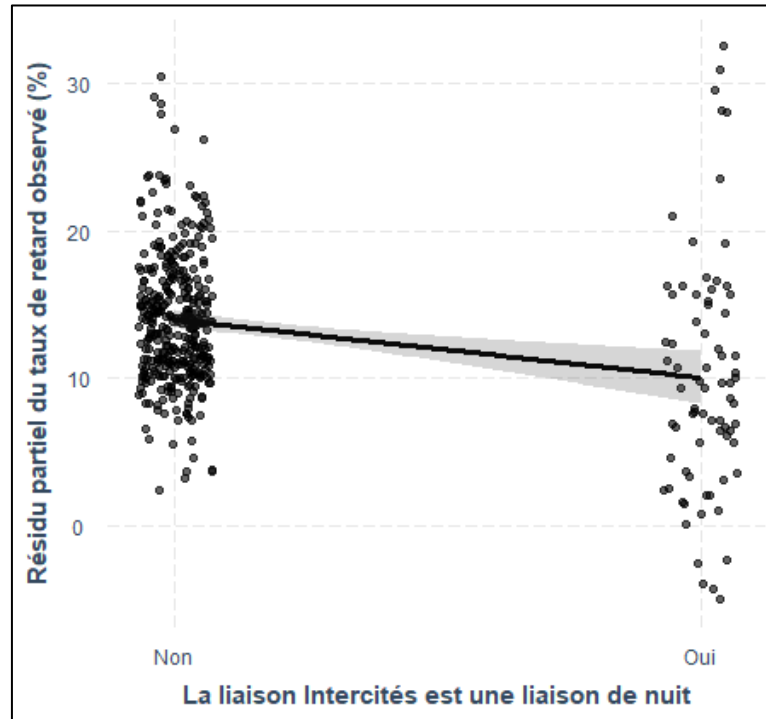
Residuals:
    Min       1Q   Median       3Q      Max
-15.0183  -3.2128  -0.3527   2.6899  22.4246

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)         6.9484     0.8298   8.374 8.36e-16 ***
LiaisonNuit         -3.9108     1.0465  -3.737 0.000212 ***
NbNoeudsMajeursTraversés
                    5.2014     0.7097   7.329 1.19e-12 ***
LiaisonTraverseBordeaux
                    7.2998     0.8684   8.406 6.59e-16 ***
LiaisonTraverseToulouse
                    2.8881     0.8259   3.497 0.000521 ***
LiaisonTraverseDijon
                    4.2044     1.7489   2.404 0.016645 *
LiaisonTraverseRouen
                   -2.9840     1.0266  -2.907 0.003848 **
DépartOuArr_NoeudSecondaire
                    1.6491     0.6238   2.644 0.008505 **
Axe_Est_Centre     -2.2518     0.7699  -2.925 0.003633 **
Nord_Sud            2.4554     0.5921   4.147 4.08e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.288 on 421 degrees of freedom
Multiple R-squared:  0.3482,    Adjusted R-squared:  0.3343
F-statistic: 24.99 on 9 and 421 DF,  p-value: < 2.2e-16
```

D'après le modèle retenu, une liaison Intercités de nuit présente en moyenne un taux de retard plus faible d'environ 3,9 points. Le plus faible niveau de trafic des liaisons de nuit relativement aux autres liaisons Intercités pourrait expliquer pour partie ce meilleur niveau de ponctualité. Les retards liés à la gestion du trafic et aux effets de réseau – propagation d'un retard d'une circulation de train à la suivante – ont en effet moins de chance de survenir lorsque la fréquence de circulation de trains est faible.

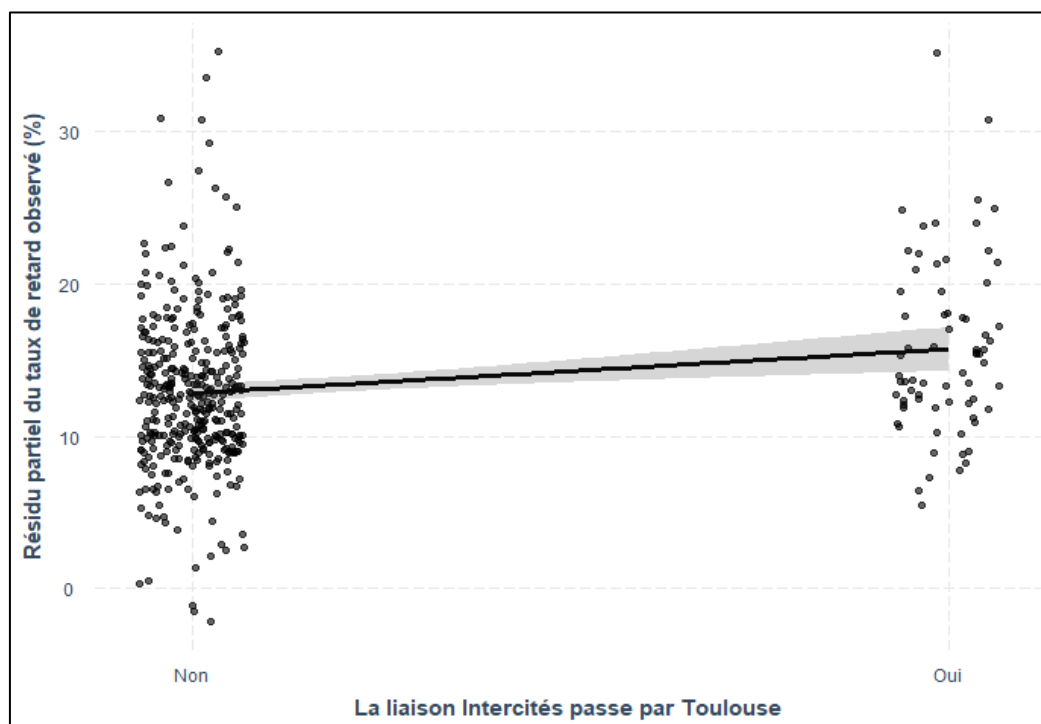
Le graphique ci-dessous illustre la meilleure ponctualité des liaisons de nuit. La zone ombrée autour de la droite de régression correspond à la plage d'incertitude statistique sur cette droite. En axe des ordonnées, le résidu partiel du taux de retard observé est représenté. Il s'agit en fait du taux de retard observé auquel on retranche l'effet des autres variables de la régression, afin d'identifier plus facilement graphiquement la relation de corrélation pour cette variable spécifiquement :



Le modèle suggère qu'en moyenne, chaque nœud majeur supplémentaire traversé par une liaison Intercités se traduit par une forte hausse du taux de retard (+5,2 points/ nœud majeur traversé).

Le modèle indique également que les liaisons Intercités traversant les nœuds ferroviaires de Bordeaux, Toulouse et Dijon sont moins ponctuelles. Au-delà de l'effet d'engorgement qui peut se produire au niveau du nœud ferroviaire à proprement parler, on peut noter que le moins bon niveau de ponctualité des liaisons passant par Bordeaux est influencé par les mauvaises performances de ponctualité de la liaison Bordeaux – Marseille, pour laquelle il n'y a pas d'explication évidente *a priori*. La moins bonne ponctualité des liaisons Intercités passant par Toulouse-Matabiau semble en accord avec la nécessité d'augmenter la capacité d'accueil des trains en gare de Toulouse-Matabiau soulevée par SNCF Réseau dans son rapport de concertation préalable au projet de modernisation du nœud ferroviaire de Toulouse [5].

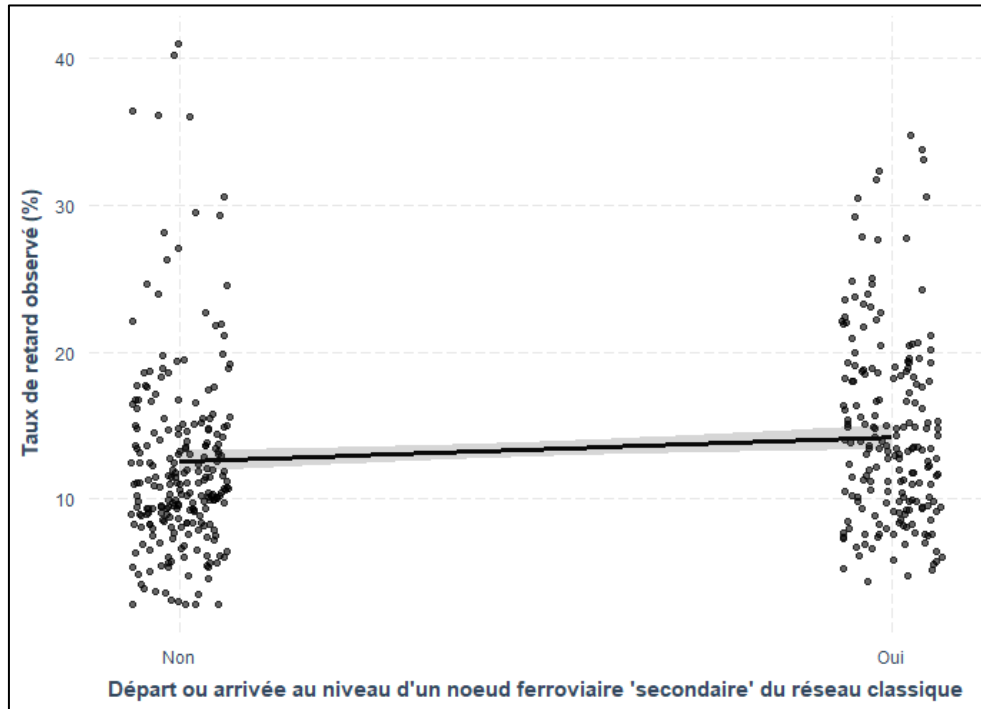
Le graphique ci-dessous montre les moins bonnes performances de ponctualité en moyenne des liaisons passant par Toulouse-Matabiau :



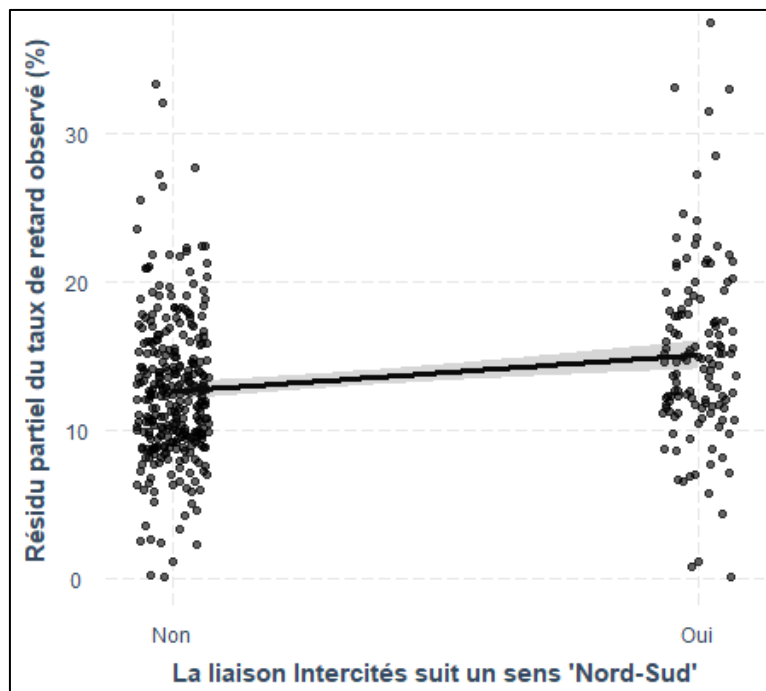
La valeur du coefficient associé au passage de la liaison par Rouen indique qu'en moyenne, le taux de retard des trains Intercités traversant Rouen est inférieur de près de 3,0 points à celui des autres trains Intercités. Le fait que les liaisons concernées sont nettement plus courtes (184 km) que la moyenne des liaisons Intercités du panel (420 km) fournit peut-être un élément d'explication possible à la meilleure ponctualité relative de ces liaisons.

Selon le modèle retenu, les trains Intercités dont le départ ou l'arrivée s'effectue au niveau de l'un des nœuds ferroviaires secondaires retenus (Bordeaux, Nantes, Toulouse, Orléans, Tours, Brive-la-Gaillarde, Narbonne, Nevers, Dijon, Caen, Amiens, Saint-Quentin) sont également moins ponctuels que les autres trains Intercités (+1,6 points de retard). Du fait de la saturation du réseau ferré observée au niveau de certains de ces nœuds, il est possible qu'un plus grand nombre de trains Intercités soient impactés si un retard s'y produisait. Les difficultés de gestion du trafic au niveau d'un nœud ferroviaire pourraient également accroître le risque de retard.

Le graphique suivant montre que la distribution des taux de retard présente une moyenne plus élevée lorsque la liaison part de ou arrive au niveau d'un des nœuds ferroviaires secondaires retenus :



Enfin, le modèle retenu suggère que les liaisons Intercités de l'axe « Est – Centre » du réseau Intercités de 2016 (présenté en Annexe 12) sont plus ponctuelles que les autres (-2,3 points de retard), et que les trains Intercités empruntant un trajet dans la direction Nord -> Sud sont moins ponctuels que les autres (+2,5 points de retard). Nous n'avons pas identifié à ce stade d'explication évidente à l'effet de ces variables sur la ponctualité des trains Intercités. Sur le graphique ci-après, on peut observer une distribution du taux de retard décalée vers le haut lorsqu'une liaison Intercités est parcourue en sens Nord -> Sud :



Annexe 15 : Nombre moyen de circulations TER quotidiennes aux heures de pointe par ancienne région en 2016

Des données de l'ART portant sur le nombre moyen de circulations TER quotidiennes par tranche horaire ont été utilisées. Ces données sont issues d'une base de l'ART accessible en ligne¹⁰.

Pour chaque région (anciennes régions administratives), ont été retenues comme heures de pointe du matin et du soir les tranches horaires durant lesquelles le nombre moyen de circulations quotidiennes est le plus élevé.

Le tableau ci-dessous synthétise les données utilisées.

Région	Heure de Pointe du Matin (HPM)	Nb moyen de circulations quotidiennes à l'HPM en 2016	Heure de Pointe du Soir (HPS)	Nb moyen de circulations quotidiennes à l'HPS en 2016
Rhône-Alpes	[7-8h[82,0	[17-18h[93,5
Nord-Pas-de-Calais	[7-8h[66,5	[17-18h[69,1
Alsace	[7-8h[55,4	[17-18h[59,9
PACA	[7-8h[44,9	[17-18h[42,9
Pays de la Loire	[7-8h[40,5	[17-18h[50,4
Lorraine	[7-8h[38,9	[18-19h[39,7
Picardie	[7-8h[35,6	[17-18h[40,3
Aquitaine	[7-8h[34,7	[17-18h[33,7
Bretagne	[7-8h[31,3	[17-18h[38,3
Centre	[7-8h[25,5	[18-19h[33,1
Midi-Pyrénées	[7-8h[25,4	[17-18h[31,5
Bourgogne	[7-8h[21,9	[17-18h[26,8
Haute-Normandie	[7-8h[19,9	[18-19h[22,9
Auvergne	[7-8h[18,5	[17-18h[22,7
Franche-Comté	[7-8h[17,7	[17-18h[18,9
Limousin	[7-8h[14,5	[17-18h[15,8
Languedoc-Roussillon	[7-8h[14,4	[17-18h[17,3
Champagne-Ardenne	[7-8h[14,1	[17-18h[19,1
Poitou-Charentes	[7-8h[10,8	[17-18h[14,8
Basse-Normandie	[7-8h[9,4	[18-19h[11,7

¹⁰ Lien vers la base de données de l'ART : <https://www.autorite-transport.fr/observatoire-des-transport/jeux-de-donnees-en-open-data/>

Annexe 16 : Nombre moyen d'arrêts effectués par les trains TER sur les liaisons les plus fréquentées par ancienne région

Des données de l'ART portant sur le nombre moyen de circulations TER quotidiennes sur les liaisons TER les plus fréquentées ont été utilisées. Ces données sont issues du Bilan du marché français du transport ferroviaire de voyageurs et de marchandises de l'ART (volume 2), disponible en ligne¹¹.

Une estimation du nombre d'arrêts moyen des trains circulant sur chacune des liaisons TER du panel des liaisons les plus fréquentées, au cours d'un Jour Ouvrable de Base (JOB) - le mardi 27 octobre 2020 – a été calculé pour l'amplitude horaire retenue de 6h à 20h.

Les données utilisées ont été recueillies sur le site <https://www.horaires-de-trains.fr/> et proviennent de l'OpenData de la SNCF. Pour simplifier l'étude, nous avons considéré que le nombre d'arrêts moyen ne dépend pas en première approche du sens de circulation sur la liaison TER. Si cette hypothèse semble raisonnable pour un grand nombre de liaisons, elle n'est pas nécessairement exacte sur la totalité des liaisons TER étudiées.

Le départ et le terminus ne sont pas pris en compte dans le calcul du nombre moyen d'arrêts.

Pour chaque région (anciennes régions administratives), une moyenne du nombre d'arrêts pondérée du nombre de circulations quotidiennes par liaison a été calculée.

Le panel des liaisons TER les plus fréquentées retenu par l'ART se compose pour 2/3 de liaisons infrarégionales, et pour 1/3 de liaisons interrégionales. Il ne contenait initialement aucune liaison TER d'Auvergne. Les principales liaisons en volume de trafic (voy-km) pour cette région - Clermont-Ferrand - Moulins-sur-Allier, Clermont-Ferrand - Lyon Part Dieu, Clermont-Ferrand – Langeac, ont donc été ajoutées au panel.

A ensuite été calculé pour chaque région le taux des principales liaisons TER empruntant le principal nœud ferroviaire régional principal retenu. La sélection de ces nœuds principaux se base sur les niveaux de trafic journalier indiqué par les cartes l'atlas de SNCF Réseau 2020, « le trafic voyageurs régionaux » (pages 44-45), et « le trafic total voyageurs et fret » (pages 50-51).

Le tableau ci-dessous présente les résultats pour ces variables :

Région	Principal nœud ferroviaire régional retenu	Nombre moyen d'arrêts pour les principales liaisons TER	Taux des principales liaisons TER traversant le principal nœud ferroviaire régional retenu
Alsace	Strasbourg	3,8	100
Aquitaine	Bordeaux	7,8	93
Auvergne	Clermont-Ferrand	4,2	100
Basse-Normandie	Caen	1,7	75

¹¹ Lien vers la publication de l'ART : <https://www.autorite-transport.fr/observatoire-des-transport/marche-du-transport-ferroviaire/>

Bourgogne	Dijon	5,2	92
Bretagne	Rennes	5,0	100
Centre	Tours	5,7	60
Champagne-Ardenne	Reims	2,1	63
Franche-Comté	Besançon	3,1	78
Haute-Normandie	Rouen	4,6	80
Languedoc-Roussillon	Nîmes	10,0	68
Limousin	Limoges	8,8	74
Lorraine	Metz	4,9	91
Midi-Pyrénées	Toulouse	6,6	100
Nord-Pas-de-Calais	Lille	4,7	100
PACA	Marseille ou Nice	6,3	95
Pays de la Loire	Nantes	3,0	77
Picardie	Amiens	4,6	23
Poitou-Charentes	Poitiers	5,5	50
Rhône-Alpes	Lyon	4,4	91

Le tableau de la page suivante détaille les résultats par liaison TER du panel des liaisons les plus fréquentées retenu par l'ART dans son bilan du marché ferroviaire de voyageurs 2018 (volume 1) :

Liaison TER	Nombre circ./jour	Type liaison TER	Région Origine	Région Destination	Nombre moyen d'arrêts pour les principales liaisons TER
Sélestat - Strasbourg	98	infrarégionale	Alsace	Alsace	1,4
Cannes - Nice	84	infrarégionale	PACA	PACA	7,1
Antibes - Nice	84	infrarégionale	PACA	PACA	4,7
Metz Ville - Nancy	81	infrarégionale	Lorraine	Lorraine	4,8
Aubagne - Marseille Saint-Charles	80	infrarégionale	PACA	PACA	2,0
Colmar - Mulhouse	78	infrarégionale	Alsace	Alsace	2,4
Avignon Centre - Avignon TGV	70	infrarégionale	PACA	PACA	0,0
Lille Flandres - Valenciennes	69	infrarégionale	Nord-Pas-de-Calais	Nord-Pas-de-Calais	4,3
Bordeaux Saint-Jean - Libourne	67	infrarégionale	Aquitaine	Aquitaine	2,0
Marseille Saint-Charles - Toulon	66	infrarégionale	PACA	PACA	5,2
Saint-Pierre-des-Corps - Tours	66	infrarégionale	Centre	Centre	0,0
Beaune - Dijon	65	infrarégionale	Bourgogne	Bourgogne	1,3
Douai - Lille Flandres	64	infrarégionale	Nord-Pas-de-Calais	Nord-Pas-de-Calais	4,4
Hazebrouck - Lille Flandres	61	infrarégionale	Nord-Pas-de-Calais	Nord-Pas-de-Calais	4,5
Biganos Factice - Bordeaux Saint-Jean	60	infrarégionale	Aquitaine	Aquitaine	3,0
Haguenau - Strasbourg	60	infrarégionale	Alsace	Alsace	5,1
Creil - Paris Gare du Nord	60	inter-régionale	Picardie	Île-de-France	1,2
Bourgoin-Jallieu - Lyon Part Dieu	59	infrarégionale	Rhône-Alpes	Rhône-Alpes	0,8
Colmar - Strasbourg	58	infrarégionale	Alsace	Alsace	1,2
Mulhouse - Strasbourg	55	infrarégionale	Alsace	Alsace	2,0
Lyon Part Dieu - Saint-Étienne Châteaureux	53	infrarégionale	Rhône-Alpes	Rhône-Alpes	2,6
Chalon-sur-Saône - Dijon	50	infrarégionale	Bourgogne	Bourgogne	4,2
Chartres - Paris Montparnasse	50	inter-régionale	Centre	Île-de-France	6,0
Lens - Lille Flandres	48	infrarégionale	Nord-Pas-de-Calais	Nord-Pas-de-Calais	6,7
Chantilly - Gouvieux - Paris Gare du Nord	45	inter-régionale	Picardie	Île-de-France	1,0
Montpellier Saint-Roch - Nîmes	44	infrarégionale	Languedoc-Roussillon	Languedoc-Roussillon	3,0
Béthune - Lille Flandres	44	infrarégionale	Nord-Pas-de-Calais	Nord-Pas-de-Calais	5,8
Les Aubrais - Paris Austerlitz	43	inter-régionale	Centre	Île-de-France	0,8
Besançon Viotte - Dijon	43	inter-régionale	Franche-Comté	Bourgogne	6,0
Béziers - Montpellier Saint-Roch	42	infrarégionale	Languedoc-Roussillon	Languedoc-Roussillon	4,0

Agde - Montpellier Saint-Roch	42	infrarégionale	Languedoc-Roussillon	Languedoc-Roussillon	2,7
Montpellier Saint-Roch - Sète	42	infrarégionale	Languedoc-Roussillon	Languedoc-Roussillon	1,7
Grenoble - Lyon Part Dieu	41	infrarégionale	Rhône-Alpes	Rhône-Alpes	4,0
Montpellier Saint-Roch - Narbonne	41	infrarégionale	Languedoc-Roussillon	Languedoc-Roussillon	4,8
Arcachon - Bordeaux Saint-Jean	38	infrarégionale	Aquitaine	Aquitaine	9,0
Champagne-Ardenne TGV - Reims	37	infrarégionale	Champagne-Ardenne	Champagne-Ardenne	2,0
Angers Saint-Laud - Nantes	35	infrarégionale	Pays de la Loire	Pays de la Loire	2,1
Bayeux - Caen	35	infrarégionale	Basse-Normandie	Basse-Normandie	0,8
Blois - Chambord - Orléans	34	infrarégionale	Centre	Centre	3,4
Bourg-en-Bresse - Lyon Part Dieu	34	infrarégionale	Rhône-Alpes	Rhône-Alpes	6,6
Lyon Part Dieu - Villefranche-sur-Saône	29	infrarégionale	Rhône-Alpes	Rhône-Alpes	1,0
Lille Flandres - Saint-Omer	29	infrarégionale	Nord-Pas-de-Calais	Nord-Pas-de-Calais	3,0
Besançon Franche-Comté TGV - Besançon Viotte	29	infrarégionale	Franche-Comté	Franche-Comté	1,0
Lyon Part Dieu - Mâcon	29	inter-régionale	Rhône-Alpes	Franche-Comté	3,0
Orléans - Paris Austerlitz	29	inter-régionale	Centre	Île-de-France	3,1
Beauvais - Paris Gare du Nord	29	inter-régionale	Picardie	Île-de-France	6,5
Montauban Ville Bourbon - Toulouse Matabiau	28	infrarégionale	Midi-Pyrénées	Midi-Pyrénées	3,4
Dunkerque - Lille Flandres	28	infrarégionale	Nord-Pas-de-Calais	Nord-Pas-de-Calais	3,5
Caen - Lisieux	28	infrarégionale	Basse-Normandie	Basse-Normandie	1,4
Grenoble - Valence TGV Rhône-Alpes Sud	28	infrarégionale	Rhône-Alpes	Rhône-Alpes	5,1
Compiègne - Paris Gare du Nord	28	inter-régionale	Picardie	Île-de-France	3,6
Orléans - Tours	27	infrarégionale	Centre	Centre	7,0
Paris Est - Troyes	27	inter-régionale	Île-de-France	Champagne-Ardenne	2,0
Lyon Part Dieu - Roanne	26	infrarégionale	Rhône-Alpes	Rhône-Alpes	6,0
Alès - Nîmes	26	infrarégionale	Languedoc-Roussillon	Languedoc-Roussillon	2,0
Nantes - Saint-Nazaire	26	infrarégionale	Pays de la Loire	Pays de la Loire	3,9
Paris Bercy Bourgogne - Pays d'Auvergne - Sens	26	inter-régionale	Île-de-France	Bourgogne	0,0
Lyon Part Dieu - Valence	25	infrarégionale	Rhône-Alpes	Rhône-Alpes	4,5
Rennes - Saint-Malo	25	infrarégionale	Bretagne	Bretagne	5,2
Dieppe - Rouen Rive Droite	25	infrarégionale	Haute-Normandie	Haute-Normandie	5,3
Chambéry - Challes-les-Eaux - Lyon Part Dieu	24	infrarégionale	Rhône-Alpes	Rhône-Alpes	2,5
Montpellier Saint-Roch - Perpignan	24	infrarégionale	Languedoc-Roussillon	Languedoc-Roussillon	9,0

Albi - Toulouse Matabiau	24	infrarégionale	Midi-Pyrénées	Midi-Pyrénées	5,6
Dijon - Lyon Part Dieu	24	inter-régionale	Bourgogne	Rhône-Alpes	8,0
Avignon Centre - Nîmes	24	inter-régionale	PACA	Languedoc-Roussillon	3,0
Arras - Lille Flandres	23	infrarégionale	Nord-Pas-de-Calais	Nord-Pas-de-Calais	1,0
Bordeaux Saint-Jean - Périgueux	23	infrarégionale	Aquitaine	Aquitaine	7,5
La Roche-sur-Yon - Nantes	23	infrarégionale	Pays de la Loire	Pays de la Loire	3,4
Rennes - Saint-Brieuc	23	infrarégionale	Bretagne	Bretagne	5,5
Lamballe - Rennes	23	infrarégionale	Bretagne	Bretagne	3,7
Avignon Centre - Montpellier Saint-Roch	23	inter-régionale	PACA	Languedoc-Roussillon	8,3
Limoges Bénédictins - Périgueux	21	inter-régionale	Limousin	Aquitaine	4,6
Angers Saint-Laud - Cholet	19	infrarégionale	Pays de la Loire	Pays de la Loire	3,6
Rennes - Vannes	19	infrarégionale	Bretagne	Bretagne	5,5
Amiens - Lille Flandres	18	inter-régionale	Picardie	Nord-Pas-de-Calais	3,3
Carcassonne - Toulouse Matabiau	18	inter-régionale	Languedoc-Roussillon	Midi-Pyrénées	8,3
Alençon - Le Mans	18	inter-régionale	Basse-Normandie	Pays de la Loire	4,2
Agen - Bordeaux Saint-Jean	17	infrarégionale	Aquitaine	Aquitaine	7,8
Nantes - Rennes	17	inter-régionale	Pays de la Loire	Bretagne	0,0
Nancy - Strasbourg	17	inter-régionale	Lorraine	Alsace	5,0
Annecy - Lyon Part Dieu	16	infrarégionale	Rhône-Alpes	Rhône-Alpes	4,0
Châlons-en-Champagne - Paris Est	16	inter-régionale	Champagne-Ardenne	Île-de-France	3,0
Épernay - Paris Est	16	inter-régionale	Champagne-Ardenne	Île-de-France	2,0
Angoulême - Saintes	15	infrarégionale	Poitou-Charentes	Poitou-Charentes	2,8
Angoulême - Cognac	15	infrarégionale	Poitou-Charentes	Poitou-Charentes	2,0
Lisieux - Trouville - Deauville	15	infrarégionale	Basse-Normandie	Basse-Normandie	1,8
Caen - Rouen Rive Droite	15	inter-régionale	Basse-Normandie	Haute-Normandie	4,5
Marseille Saint-Charles - Nice	14	infrarégionale	PACA	PACA	5,0
La Rochelle - Poitiers	14	infrarégionale	Poitou-Charentes	Poitou-Charentes	8,6
Niort - Poitiers	14	infrarégionale	Poitou-Charentes	Poitou-Charentes	4,3
Dijon - Paris Bercy	10	inter-régionale	Bourgogne	Île-de-France	8,0
Metz Ville - Strasbourg	10	inter-régionale	Lorraine	Alsace	6,4
Brest - Rennes	9	infrarégionale	Bretagne	Bretagne	6,6
Paris Saint-Lazare - Vernon - Giverny	8	inter-régionale	Île-de-France	Haute-Normandie	1,4

Évreux Normandie - Paris Saint-Lazare	8	inter-régionale	Haute-Normandie	Île-de-France	1,2
Limoges Bénédictins - Poitiers	8	inter-régionale	Limousin	Poitou-Charentes	7,8
Paris Austerlitz - Tours	7	inter-régionale	Île-de-France	Centre	4,3
Paris Austerlitz - Saint-Pierre-des-Corps	7	inter-régionale	Île-de-France	Centre	3,3
Bordeaux Saint-Jean - Limoges Bénédictins	7	inter-régionale	Aquitaine	Limousin	8,0
Brive-la-Gaillarde - Toulouse Matabiau	6	inter-régionale	Limousin	Midi-Pyrénées	9,0
Bordeaux Saint-Jean - La Rochelle	6	inter-régionale	Aquitaine	Poitou-Charentes	2,0
Paris Saint-Lazare - Rouen Rive Droite	4	inter-régionale	Île-de-France	Haute-Normandie	3,4
Bordeaux Saint-Jean - Brive-la-Gaillarde	4	inter-régionale	Aquitaine	Limousin	12,0
Clermont-Ferrand - Moulins-sur-Allier	34	infrarégionale	Auvergne		4,1
Clermont-Ferrand - Lyon Part Dieu	8	inter-régionale	Auvergne	Rhône-Alpes	3,5
Clermont-Ferrand - Langeac	6	infrarégionale	Auvergne		5,7

Annexe 17 : Estimation de l'âge moyen du matériel roulant TER par ancienne région

Des données sur l'âge moyen en fin d'année du parc de matériel roulant des TER par ancienne région ont été recueillies dans les rapports d'observations définitives des Enquêtes Transports Express Régionaux - Exercices 2012 et suivants – des Chambres Régionales des Comptes.

Afin de compléter ces données et d'obtenir une estimation de l'âge moyen en fin d'année du parc de matériel roulant des TER par ancienne région sur la période 2013-2019, des modèles de parc ont été construits. Ils reposent sur des données détaillées de la composition du parc de matériel roulant, et sur les dates de mises en services et de radiations fournies par les sources de données suivantes :

- Chambres Régionales des Comptes, rapports d'observations définitives, Enquêtes Transports Express Régionaux, Exercices 2012 et suivants ;
- Fiches Wikipédia par ancienne région sur les TER, et sur le parc de matériel roulant TER.

Les chiffres fournis par les Chambres régionales des Comptes (indiqués en **vert**), et les estimations obtenues à l'aide des modèles de parc du matériel roulant TER sont regroupés dans le tableau ci-dessous :

Ancienne Région	Estimation AQST de l'âge moyen du matériel roulant en fin d'année*						
	2013	2014	2015	2016	2017	2018	2019
Alsace	21,5	21,4	21,7	20,4	18,3	19,1	20,1
Aquitaine	18	17,7	15,4	12,8	10,3	10,6	11,2
Auvergne	15,7	16,7	16,4	17,5	17,5	17,5	17,4
Basse-Normandie	12,9	11,8	10	8,3	9,8	10,8	11,8
Bourgogne	19,1	19,7	20,4	18,9	19,5	19,2	20,2
Bretagne	18,1	18	17	16,9	15,2	13,3	13,7
Centre	19,9	20,9	19,8	18,1	19,1	19,4	19,8
Champagne-Ardenne	10,7	11,7	12,7	13,2	14,2	14	12,2
Franche-Comté	21,5	22,5	22,1	21,3	19,7	17,2	16,6
Haute-Normandie	11,1	12,1	12,4	12,7	13,7	14,7	15,7
Languedoc-Roussillon**	14,2	14,8	15,9	16,8	17,7	15	15,4
Limousin	12,7	13,7	13	14	13,6	13,1	14,1
Lorraine	10,2	10,9	11,9	14,5	13,9	14,8	15,8
Midi-Pyrénées	18,8	18,4	18,4	16,8	16,3	13,7	12,1
Nord-Pas-de-Calais	12,4	14,2	14,5	14,3	14,1	15,2	16,2
PACA	18,7	19,1	19,1	18,6	19,3	18,6	20,1
Pays de la Loire	10,5	11,3	10,6	9,9	9,9	10,6	11,1
Picardie	18,4	18,2	18,5	18	16,9	14,4	16,6
Poitou-Charentes	7,9	8,4	6,3	7,3	8,3	9,3	10,3
Rhône-Alpes	16,1	17	17,5	17,6	18	19	20

* En **vert**, chiffres communiqués par les Chambres Régionales des Comptes (rapports d'observations définitives, Enquêtes Transports Express Régionaux, Exercices 2012 et suivants)

** Compte tenu de l'âge du matériel roulant pour le « Train Jaune », l'âge moyen du parc est ici présenté hors Train Jaune (Z100 mis en service entre 1908 et 1912, et Z200 mis en service en 1909).

Le détail de la composition du parc de matériel roulant TER modélisé par ancienne région est fourni ci-après :

Les cellules vides correspondent à des effectifs nuls.

Alsace : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
X 73500	3	3	3	3	3	3	5
X 73900	12	12	12	12	12	12	12
X 76500 (AGC*)	22	22	22	22	28	28	28
Z 27500 (AGC)	6	6	6	6	6	6	6
B 82500 (AGC)	7	7	7	7	7	7	7
Z 11500	22	22	22	0			
RRR (rames réversibles régionales)	22	22	22	22	22	22	22
Corail (rames de 6 voitures)	22	22	22	22	22	22	22
BB 67400	6	6	6	6	6	6	6
BB 25500	10	10	10	10	9	8	8
BB 26000	14	14	14	14	14	17	17
Porteur Polyvalent Grande capacité (Coradia Liner)	0	2	7	16	37	39	39
Porteur Polyvalent Moyenne capacité (Régiolis et Coradia Liner)	0	6	6	6	7	7	7
TOTAL	146	154	159	146	173	177	179

* AGC = Autorail grande capacité

Lorraine : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
X 73500	6	6	6	6	6	11	11
X 73900	7	7	7	7	7	7	7
X 76500 (AGC*)	9	9	9	8	8	8	8
Z 27500 (AGC)	32	32	32	32	32	32	32

Z 11500	0	0	0	22	17	15	15
Z 24500	25	25	25	25	25	25	25
RRR (rames réversibles régionales)	8	8	8	8	2	2	2
BB 25500	6	6	6	0	0	0	0
Porteur Polyvalent Moyenne capacité (Régiolis et Coradia Liner)	5	8	8	8	9	9	9
TOTAL	98	101	101	116	106	109	109

* AGC = Autorail grande capacité

Champagne-Ardenne : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
X 73500	8	8	8	8	8	1	0
X 76500 (AGC*)	23	23	23	23	23	23	23
Z 27500 (AGC)	13	13	13	13	13	13	13
B 82500 (AGC)	8	8	8	8	8	8	8
BB 15000	7	7	7	6	6	4	0
TOTAL	59	59	59	58	58	49	44

* AGC = Autorail grande capacité

Aquitaine : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
B 81500	27	27	27	27	27	27	27
B 84500					0	3	5
X 72500	17	17	17	17	17	18	18
X 73500	18	18	20	20	20	20	20

Z 51500		10	14	22	22	22	22
Z 56300		2	9	16	24	24	24
X 2200	23	23	14	8	0		
Z 7300	22	22	15	8	0		
BB 7200	6	6	6	5	5	3	2
Z 21500	6	0					
TOTAL	119	125	122	123	115	117	118

Limousin : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
B 81500	12	12	12	12	12	12	12
B 84500						5	5
X 72500	6	6	6	6	6	6	6
X 73500	24	24	24	24	24	24	23
X 2200	10	10	4	4	0		
TOTAL	52	52	46	46	42	47	46

Poitou-Charentes : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
B 81500	12	12	12	12	12	12	12
B 84500						5	5
X 72500	6	6	6	6	6	6	6
X 73500	24	24	24	24	24	24	23
X 2200	10	10	4	4	0		

TOTAL	34	36	41	41	41	41	41
--------------	-----------	-----------	-----------	-----------	-----------	-----------	-----------

Rhône-Alpes : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
BB 22200	29	29	29	29	29	29	29
BB 67300	2	2	2	2	0		
BB 67400							
B 81500	41	41	41	41	41	41	41
B 82500	28	28	28	28	28	28	28
B 84500							
X 72500	18	18	18	18	16	16	16
X 73500	43	43	43	43	43	43	43
X 76500					0		
Z 850	3	3	3	3	3	3	3
Z 850	6	6	6	6	6	6	6
U 52500	24	24	24	24	24	24	24
Z 9500	14	14	14	14	13	13	13
Z 9600	9	9	9	9	9	9	9
Z 23500	16	16	16	16	16	16	16
Z 24500	60	60	60	60	60	60	60
Z 27500	15	15	15	15	15	15	15
Z 55500		2	15	38	40	40	40
RRR (rames réversibles régionales)							
Corail (rames de 7 voitures)	31	31	31	31	29	29	29
TOTAL	339	310	323	346	343	343	343

Auvergne : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
BB 67300	2	2	0				
BB 67400	4	4	4	4	4	2	0
B 84500			0	6	12	12	12
X 73500	36	36	43	43	43	43	43
X 76500	26	26	26	26	26	26	26
RRR (rames réversibles régionales)	8	8	8	8	8	8	8
Corail (rames de 7 voitures)	6	6	4	4	4	2	0
TOTAL	76	76	81	87	93	91	89

Haute-Normandie : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
B 82500	9	9	9	9	9	9	9
B 85900	0	0	5	10	10	10	10
X 73500	14	14	14	14	14	14	14
Z 26500	16	16	16	16	16	16	16
Z 27500	22	22	22	22	22	22	22
VO2N (rames de 7 voitures)	6	6	6	6	6	6	6
V2N (rames de 7 voitures)	5	5	5	5	5	5	5
RRR (rames réversibles régionales)	6	6	6	6	6	6	6
TOTAL	78	78	83	88	88	88	88

Basse-Normandie : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
B 84500	0	6	6	14	15	15	15
X 73500			0	6	11	11	11
X 76500	14	14	14	14	14	14	14
Z 27500	9	9	9	9	9	9	9
X 4750	7	7	3	0			
TOTAL	30	36	32	43	49	49	49

Languedoc-Roussillon : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
B 81500	10	10	10	10	10	10	10
B 84500							3
X 73500			11	11	11	11	11
Z 100	13	13	13	13	13	13	13
Z 150	2	2	2	2	2	2	2
Z 7300	11	10	10	10	10	3	3
Z 7500	3	3	3	3	3	3	3
Z 27500	38	38	38	38	38	38	38
BB 7200	4	4	4	4	4	0	
Rames Corail (7 voitures)	3	3	3	3	3	0	
TOTAL	84	80	91	91	91	80	83

Midi-Pyrénées : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
B 81500	18	18	18	18	18	18	18
B 83500		6	13	15	15	18	18
X 73500	26	26	26	26	26	26	26
Z 7300	15	15	15	15	15	15	11
Z 27500	19	19	19	19	19	19	19
Z 54900				9	10	15	15
Z 56300					5	15	18
BB 7200	21	21	21	18	17	8	0
BB 9300	2	0					
BB 8500	7	7	7	3	0	0	
X 72500	8	8	8	8	8	1	0
TOTAL	116	120	127	131	133	135	125

Centre : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
BB 7200	17	17	17	17	17	17	17
BB 26000	2	2	2	2	2	2	2
B 81500	1	1	1	1	1	1	1
B 84500						1	3
X 72500	15	15	15	15	15	15	15
X 73500	20	20	20	20	20	20	20
X 74500	5	5	5	5	5	5	5
Z 7300	15	15	15	15	15	13	11
Z 21500	17	17	17	17	17	17	17

Z 26500	12	12	12	12	12	12	12
Z 27500	7	7	7	7	7	7	7
Z 55500			9	14	14	14	14
Z 56500						2	2
VO2N (rames de 7 voitures)	6	6	3	0			
BB 8500	10	10	8	0			
X 4750	2	2	1	0			
TOTAL	129	129	132	125	125	126	126

Bretagne : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
B 82500	8	8	8	8	8	8	8
X 73500	17	17	17	17	17	17	17
Z 21500	19	19	19	19	19	19	19
Z 27500	14	14	14	14	14	14	14
Z 55500		5	6	10	19	22	26
RRR (rames réversibles régionales)	11	11	11	11	11	11	11
Z 9600	5	5	4	1	0		
BB 25500	16	15	10	10	4	0	
X 2100	21	21	16	16	5	0	
TOTAL	111	115	105	106	97	91	95

Pays de la Loire : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
B 84500		3	10	10	10	10	10
U 53500	24	24	24	24	24	24	24
X 72500	10	10	10	10	10	10	10
X 73500	12	12	12	12	12	12	12
X 76500	17	17	17	17	17	17	17
Z 21500	21	21	21	21	21	21	21
Z 24500	13	13	13	13	13	13	13
Z 27500	18	18	18	18	18	18	18
Z 51500				5	10	10	10
Z 56500						5	11
BB 25500	3	3	1	1	0		
Z 9600	16	16	11	2	0		
TOTAL	134	137	137	133	135	140	146

Provence-Alpes-Côte d'Azur : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
BB 22200	27	27	27	27	26	18	18
BB 67400	8	8	8	6	5	3	3
B 81500	24	24	24	24	24	16	8
B 84500		1	6	10	10	10	10
X 72500	19	19	19	17	17	17	17
X 76500	8	8	8	8	8	8	8
Z 23500	30	30	30	30	30	30	30
Z 26500	15	15	15	15	15	15	15

Z 55500		4	8	16	16	16	16
RRR (rames réversibles régionales)	12	12	12	12	12	6	6
rames Corail (6 voitures)	14	14	14	14	14	8	8
TOTAL	157	162	171	179	177	147	139

Pour les régions Bourgogne et Franche-Comté, la composition du parc de matériel roulant TER n'est présentée que sur la période 2016-2019. Les données sur l'âge moyen du matériel roulant TER fournies par la Chambre régionale de la région Bourgogne-Franche-Comté couvraient en effet déjà la période 2013-2015.

Bourgogne : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2016	31/12/2017	31/12/2018	31/12/2019
BB 7200	19	19	19	19
BB 25500	0			
Y 8000	2	2	2	2
B 81500	38	38	37	37
Rames Corail (9 voitures)	14	14	14	14
V2N (rames de 8 voitures)	3	3	3	3
X 73500	15	15	20	20
BB 25500	0			
Z 27500	10	10	10	10
Z 51500		2	9	9
TOTAL	101	103	114	114

Franche-Comté : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2016	31/12/2017	31/12/2018	31/12/2019
BB 22200	7	7	1	1
BB 25500	4	1		
RRR (rames réversibles régionales)	3	0		
Rames Corail (9 voitures)	4	4	4	4
X 73500	24	24	28	28
X 76500	5	5	5	5
Z 27500	4	4	4	4
Z 51500		7	7	9
Z 9600		4	4	2
Z 9500	4	4	0	
TOTAL	55	60	53	53

Nord-Pas-de-Calais : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
BB 22200		12	15	15	15	15	15
B 82500	30	30	30	30	30	30	30
X 76500	15	15	15	5	5		
Z 23500	33	33	33	33	33	33	33
Z 24500	47	47	47	47	47	47	47
Z 55500			1	6	18	18	18
VO2N (rames de 5 voitures)	10	10	10	10	10	10	10
BB 67400	3	0					
BB 25600	15	12	6				
TOTAL	153	159	157	146	158	153	153

Picardie : Estimation des effectifs du matériel roulant TER en fin d'année :

Séries	31/12/2013	31/12/2014	31/12/2015	31/12/2016	31/12/2017	31/12/2018	31/12/2019
BB 15000	14	14	16	15	15	15	11
BB 22200	2	2	2	2	2	2	17
B 82500	16	16	16	16	16	16	16
B 84500		7	7	15	17	17	27
X 72500	6	6	6	6	6	6	4
X 76500	24	24	24	24	29	34	34
Z 26500	23	23	23	23	23	23	23
Z 55500					0	7	7
BB 25600	15	12	6				
V2N (rames de 9 voitures)	9	9	9	9	9	9	9
BB 67400	14	14	11	8	8	0	0
X 73500	12	12	12	12	12	0	0
BB17000	2	2					
RRR (rames réversibles régionales)	33	24	21	21			
Rames Corail (rames de 7 voitures)	6	6	6	6	5	3	0
TOTAL	161	159	153	157	142	132	148

Annexe 18 : Tableau de colinéarité entre les variables explicatives du taux de retard des TER

	NbCirc_an_AQST	NbCirc_H_PM	NbCirc_H_PS	DistanceMoy	VoyKm_an	NbPlacesMoy	TauxOccupationMoy	SiegesKmParHab	VoyageurKmParHab	IntensitéUsageTER	LongueurRéseau_HorsLGV	DensitéRéseau_HorsLGV	PartCatUIC_5_6	PartCatUIC_7_9_AV	AgeRéseau_HorsLGV	AgeRéseauCatUIC_5_6	AgeRéseauCatUIC_7_9_AV	DensitéRéseau_CatUIC_5_6	DensitéRéseau_CatUIC_7_9_AV	NbLiaisonsInfrarégionales	NbLiaisonsInterrégionales	NbCommunesDesservies	PartPopInf10kmGareTER
NbCirc_an_AQST	X	X	X		X															X			
NbCirc_HPM	X	X	X		X															X			
NbCirc_HPS	X	X	X		X															X			
DistanceMoy				X																			
VoyKm_an	X	X	X		X															X			
NbPlacesMoy						X																	
TauxOccupationMoy							X																
SiegesKmParHab								X	X														
VoyageurKmParHab								X	X														
IntensitéUsageTER										X													
LongueurRéseau_HorsLGV											X												
DensitéRéseau_HorsLGV												X											
PartCatUIC_5_6													X										
PartCatUIC_7_9_AV														X									
AgeRéseau_HorsLGV															X								
AgeRéseauCatUIC_5_6																X							
AgeRéseauCatUIC_7_9_AV																	X						
DensitéRéseau_CatUIC_5_6																		X					
DensitéRéseau_CatUIC_7_9_AV																			X				
NbLiaisonsInfrarégionales	X	X	X		X															X		X	
NbLiaisonsInterrégionales																					X		
NbCommunesDesservies																						X	
PartPopInf10kmGareTER																							X
EffectifsPour100000trkm_2015																							
NbMoyArrêts_NvleReg_2018																							
NbMoyArrêts_2020-PpalesLiaisons																							
TMJA_max_TER_2018	X	X	X		X															X			
TxCircPpalesLiaisonsTraversant-NoeudRégional																							
AgeMoyen_MaterielRoulant																							

Code couleur : **en jaune**, variables présentant des relations de colinéarité avec d'autres variables.

	SiegesKmParHab	VoyageurKmParHab	IntensitéUsageTER	LongueurRéseau_HorsLGV	DensitéRéseau_HorsLGV	PartCatUIC_5_6	PartCatUIC_7_9_AV	AgeRéseau_HorsLGV	AgeRéseauCatUIC_5_6	AgeRéseauCatUIC_7_9_AV	DensitéRéseau_CatUIC_5_6	DensitéRéseau_CatUIC_7_9_AV	NbLiaisonsInfrarégionales	NbLiaisonsInterrégionales	NbCommunesDesservies	PartPopInf10kmGareTER
NbCirc_an_AQST													X			
NbCirc_HPM													X			
NbCirc_HPS													X			
DistanceMoy																
VoyKm_an													X			
NbPlacesMoy																
TauxOccupationMoy																
SiegesKmParHab	X	X														
VoyageurKmParHab	X	X														
IntensitéUsageTER			X													
LongueurRéseau_HorsLGV				X												
DensitéRéseau_HorsLGV					X											
PartCatUIC_5_6						X										
PartCatUIC_7_9_AV							X									
AgeRéseau_HorsLGV								X								
AgeRéseauCatUIC_5_6									X							
AgeRéseauCatUIC_7_9_AV										X						
DensitéRéseau_CatUIC_5_6											X					
DensitéRéseau_CatUIC_7_9_AV												X				
NbLiaisonsInfrarégionales													X		X	
NbLiaisonsInterrégionales														X		
NbCommunesDesservies													X		X	
PartPopInf10kmGareTER																X
EffectifsPour100000trkm_2015																
NbMoyArrêts_NvleReg_2018																
NbMoyArrêts_2020-PpalesLiaisons																
TMJA_max_TER_2018													X			
TxCircPpalesLiaisonsTraversant-NoeudRégional																
AgeMoyen_MaterielRoulant																

Code couleur : **en jaune**, variables présentant des relations de colinéarité avec d'autres variables.

Annexe 19 : Estimation de la fréquentation quotidienne par ligne de RER et Transilien sur la période 2013-2019

Des données - provenant de l'OpenData de la SNCF, des comités de ligne, ou des sites Internet officiels des lignes de RER et Transilien - sur le nombre de voyageurs quotidiens par ligne ont été exploitées.

Les données sur les années manquantes ont été estimées rapidement par continuité sur la base de simplification d'un taux de croissance annuel fixe (calculé avec les données recueillies) sur la période de 2013-2019.

Le tableau ci-dessous synthétise les données utilisées. Les données fournies par la documentation apparaissent **en gras**, les autres valeurs du tableau étant des estimations.

Ligne/Année	Estimation AQST du nombre de voyageurs quotidiens par ligne [en milliers]						
	2013	2014	2015	2016	2017	2018	2019
A	936	958	980	1 090	1 200	1 300	1 400
B	847	870	893	916	939	962	983
C	516	528	541	553	565	577	590
D	550	583	615	627	639	651	662
E	356	361	367	373	373	373	373
H	203	214	226	238	250	255	260
J	226	240	254	269	272	275	278
K	10	11	12	13	15	16	17
L	291	301	306	311	316	321	328
N	133	133	133	133	133	133	133
P	93	97	101	105	109	113	116
R	50	50	50	51	51	51	51
U	51	52	52	52	53	53	53

Annexe 20 : Estimation du nombre moyen d'arrêts par circulation par ligne RER et Transilien à l'heure de pointe du matin

Pour chaque ligne, les données provenant des fiches horaires les plus récentes identifiées ont été exploitées pour calculer, le nombre moyen d'arrêts par circulation à l'heure de pointe du matin (HPM) durant un Jour Ouvrable de Base (JOB), le nombre moyen d'arrêts dans Paris intra-muros, le temps de parcours moyen des trains à l'HPM en minutes et la durée moyenne entre 2 arrêts (temps d'échange voyageurs compris).

Le tableau ci-dessous présente les résultats obtenus :

Ligne	NbMoyArrêtsHPM_2019	NbMoyArrêtsHPMdansParis_2019	TpsParcours MoyHPM [en minutes]	DuréeMoyInterstation [en minutes]
A	22,3	5	68	2,92
B	24,8	7	64	2,48
C	18,8	10,9	54	2,73
D	16,3	2,2	73	4,22
E	10,5	3	32	2,78
H	12,7	0,8	36	2,63
J	7	1	49	6,13
K	8,5	1	49	5,16
L	12,6	1,8	38	2,79
N	8,9	1	48	4,85
P	7,1	0,7	47	5,8
R	10	0,6	64	5,82
U	11	0	43	3,58

Annexe 21 : Estimation de l'âge moyen du matériel roulant RER et Transilien par ligne

Afin d'estimer l'âge moyen en début d'année du parc de matériel roulant des RER et Transilien par ligne sur la période 2013-2019, des modèles de parc ont été construits. Ils reposent sur des données détaillées de la composition du parc de matériel roulant, et sur les dates de mises en services et de radiations fournies par les sources de données suivantes :

- Fiches Wikipédia sur les lignes de Transilien et sur les différentes séries du matériel roulant composant leurs parcs ;
- Articles de la presse spécialisée portant sur les mouvements de parc de matériel roulant (mises en service et radiations).

Les estimations obtenues à l'aide des modèles de parc du matériel roulant RER et Transilien sont regroupées dans le tableau ci-dessous :

Ligne	Estimation de l'âge moyen du matériel roulant en début d'année						
	2013	2014	2015	2016	2017	2018	2019
A	25	18	11	5,6	6,4	7,4	8,4
B	28,8	29,8	30,8	31,8	32,8	33,7	34,7
C	20,1	21,1	22,1	23,1	24,1	25,1	26,1
D	24,9	25,4	26,1	26,9	27,8	28,3	22
E	15	16	17	15,9	16,9	17,5	18,5
H	9,6	10,6	11,6	11,1	4,8	5,7	6,7
J	28,3	24,8	23,5	16	16,8	15,8	13,9
K	44	45,4	46,4	47,4	1	2	3
L	32,6	29,5	25,5	22,7	23,4	19,8	14,8
N	24,4	24	24,4	25,1	26,3	27,3	28,3
P	19	14,5	14	14	14,2	15,2	16,2
R	24,9	25,4	26,1	26,9	27,8	28,3	22
U	26	27	28	29	30	31	32

Le détail de la composition du parc de matériel roulant des RER et Transilien modélisé par ligne est fourni ci-après :

Les cellules vides correspondent à des effectifs nuls.

RER A : Estimation des effectifs du matériel roulant en début d'année :

Séries	01/01/2013	01/01/2014	01/01/2015	01/01/2016	01/01/2017	01/01/2018	01/01/2019
MS 61	105	51	12	0			
MI 84	27	27	27	0			
MI 2N RATP	43	43	43	43	43	43	43
MI 09	58	81	106	132	140	140	140
TOTAL	233	202	188	175	183	183	183

RER B : Estimation des effectifs du matériel roulant en début d'année :

Séries	01/01/2013	01/01/2014	01/01/2015	01/01/2016	01/01/2017	01/01/2018	01/01/2019
MI 79 - Z8100	117	117	117	117	117	116	116
MI 84	37	37	37	37	37	42	42
TOTAL	154	154	154	154	154	158	158

RER C : Estimation des effectifs du matériel roulant en début d'année :

Séries	01/01/2013	01/01/2014	01/01/2015	01/01/2016	01/01/2017	01/01/2018	01/01/2019
Z 5600	36	36	36	36	36	36	36
Z 8800	35	35	35	35	35	35	35
Z 20500			0	46	46	46	45
Z 20900	54	54	54	54	54	54	54
TOTAL	125	125	125	171	171	171	170

RER D : Estimation des effectifs du matériel roulant en début d'année :

Séries	01/01/2013	01/01/2014	01/01/2015	01/01/2016	01/01/2017	01/01/2018	01/01/2019
Z 5300	34	29	27	25	23	22	0
Z 5600	9	9	9	9	9	9	9
Z 20500	138	138	138	138	134	134	134
Z 57000					0	2	30
TOTAL	181	176	174	172	166	167	173

RER E : Estimation des effectifs du matériel roulant en début d'année :

Séries	01/01/2013	01/01/2014	01/01/2015	01/01/2016	01/01/2017	01/01/2018	01/01/2019
MI 2N SNCF (Z22500)	53	53	53	53	53	53	53
Z 50000			0	8	8	10	10
TOTAL	53	53	53	61	61	63	63

Transilien H : Estimation des effectifs du matériel roulant en début d'année :

Séries	01/01/2013	01/01/2014	01/01/2015	01/01/2016	01/01/2017	01/01/2018	01/01/2019
BB 17000	12	12	12	8	0		
RIB-RIO*	8	8	8	8	0		
Z 50000	82	82	82	82	82	82	82
TOTAL	102	102	102	98	82	82	82

* Rames réversibles : rames inox de banlieue (RIB) ou rames inox omnibus (RIO)

Transilien J : Estimation des effectifs du matériel roulant en début d'année :

Séries	01/01/2013	01/01/2014	01/01/2015	01/01/2016	01/01/2017	01/01/2018	01/01/2019
BB 27300	42	42	42	42	42	42	40
BB 17000	18	13	18	3	3	3	0
RIB-RIO	18	13	18	0			
VB2N	44	42	43	42	42	42	40
Z 50000	0	13	36	55	57	76	97
TOTAL	122	123	157	142	144	163	177

Transilien K : Estimation des effectifs du matériel roulant en début d'année :

Séries	01/01/2013	01/01/2014	01/01/2015	01/01/2016	01/01/2017	01/01/2018	01/01/2019
BB 17000	17	17	17	15	0		
RIB-RIO	16	16	16	14	0		
VB2N	2						
Z 50000				0	16	16	16
TOTAL	35	33	33	29	16	16	16

Transilien L : Estimation des effectifs du matériel roulant en début d'année :

Séries	01/01/2013	01/01/2014	01/01/2015	01/01/2016	01/01/2017	01/01/2018	01/01/2019
Z 6400	75	75	74	74	72	63	43
Z 20500	17	17	0				
Z 50000	0	13	36	55	55	76	97
TOTAL	92	105	110	129	127	139	140

Transilien N : Estimation des effectifs du matériel roulant en début d'année :

Séries	01/01/2013	01/01/2014	01/01/2015	01/01/2016	01/01/2017	01/01/2018	01/01/2019
BB 27300	25	25	25	25	25	25	25
BB 7600	14	14	14	14	14	14	14
Z 5300	13	7	5	0			
Z 8800				18	22	22	22
VB2N	39	39	39	39	39	39	39
TOTAL	91	85	83	96	100	100	100

Transilien P : Estimation des effectifs du matériel roulant en début d'année :

Séries	01/01/2013	01/01/2014	01/01/2015	01/01/2016	01/01/2017	01/01/2018	01/01/2019
BB 67400	9	9	9	9	9	9	9
BB 17000			0	1	1	1	1
RIB-RIO	8	8	8	10	10	10	10
B 82500	24	24	24	24	24	24	24
Z 20500	15	15	15	21	21	21	21
U 25500	1	1	1	1	1	1	1
Z 50000	0	22	31	39	45	45	45
TOTAL	57	79	88	105	111	111	111

Transilien R : Estimation des effectifs du matériel roulant en début d'année :

Séries	01/01/2013	01/01/2014	01/01/2015	01/01/2016	01/01/2017	01/01/2018	01/01/2019
Z 5300	34	29	27	25	23	22	0
Z 5600	9	9	9	9	9	9	9
Z 20500	138	138	138	138	134	134	134
Z 57000					0	2	30
TOTAL	181	176	174	172	166	167	173

Transilien U : Estimation des effectifs du matériel roulant en début d'année :

Séries	01/01/2013	01/01/2014	01/01/2015	01/01/2016	01/01/2017	01/01/2018	01/01/2019
Z 8800	22	22	22	22	22	22	22
TOTAL	22	22	22	22	22	22	22

Annexe 22 : Estimation de la moyenne du Trafic Moyen Journalier Annuel par voie et de son niveau sur la section la plus chargée

Afin d'estimer la moyenne du Trafic Moyen Journalier Annuel (TMJA) par voie, ainsi que son niveau sur la section la plus chargée par ligne de RER et Transilien, nous avons estimé le niveau du TMJA par voie sur les différents tronçons de chaque ligne à l'aide des sources de données suivantes :

- Le nombre de voies par tronçon : la carte du réseau physique issue de l'atlas SNCF Réseau – édition 2020, ainsi que les schémas fonctionnels et des plans des voies issues des schémas directeurs et des études techniques de SNCF Réseau ;
- La longueur du tronçon en km : données fournies par Rome2Rio ;
- Le niveau du TMJA Fret + TER + Grandes Lignes : données issues de la Carte SNCF - TRAFIC TOTAL 2000 (SNCF / Direction du Contrôle de gestion (2000)). Des données assez anciennes (2000) ont été utilisées car les cartes plus récentes à disposition n'offrent pas un aussi bon niveau de précision : le découpage du réseau par tronçons n'est pas aussi fin ;
- Le niveau du TMJA pour les Transilien fourni par l'atlas SNCF Réseau 2020, carte "Trafic Transilien" (page 46-47)
- La répartition du trafic des différentes activités ferroviaires et lignes de RER et Transilien par voie, estimée à partir des schémas de desserte par ligne et des schémas fonctionnels des grandes gares parisiennes fournis par les schémas directeurs et des études techniques de SNCF Réseau.

Pour chaque ligne RER ou Transilien, la moyenne du TMJA par voie est calculée en prenant la moyenne pondérée de la longueur de ligne par tronçon. Le niveau du TMJA sur la section la plus chargée du réseau francilien empruntée par la ligne est la valeur maximale de cette variable parmi les différents tronçons.

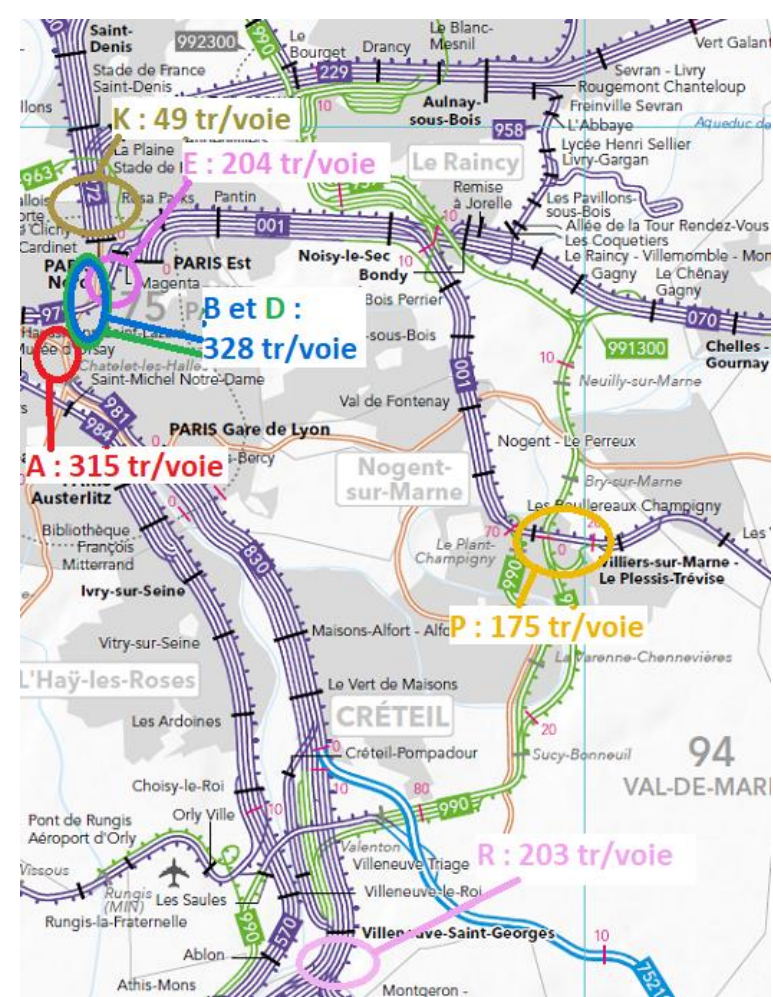
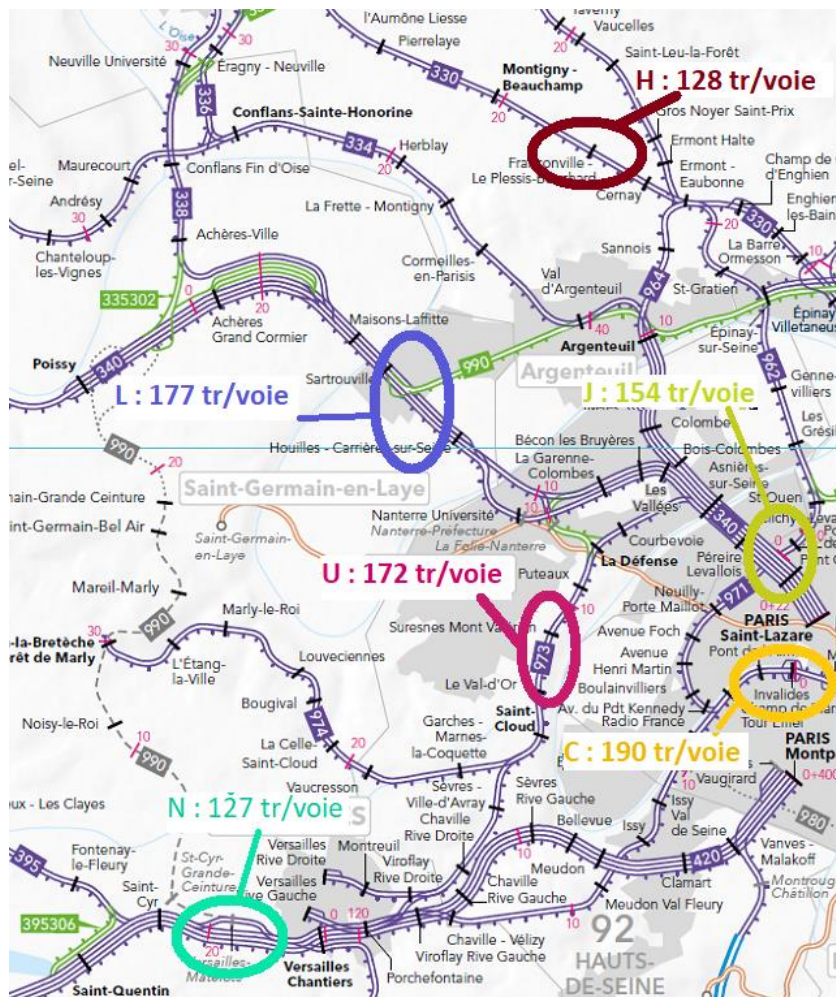
Les estimations obtenues et la section la plus chargée par ligne sont présentées dans le tableau ci-dessous :

Ligne	Moyenne du TMJA tous services ferroviaires par voie	TMJA tous services ferroviaires par voie section la plus chargée	Section la plus chargée
A	106	315	Tronçon central Gare de Lyon - Auber
B	164	328	Tunnel Châtelet-Les Halles - Gare du Nord
C	90	190	Champ de Mars-Tour Eiffel - Gare d'Austerlitz
D	80	328	Tunnel Châtelet-Les Halles - Gare du Nord
E	124	204	Tunnel Haussmann-Saint-Lazare - Magenta
H	64	128	Montigny-Beauchamp - Ermont-Eaubonne
J	75	154	Asnières-sur-Seine - Gare Saint-Lazare (groupes IV, V et VI)
K	37	49	Saint-Denis - Gare du Nord (RER B sur voies dédiées depuis RER B Nord+ en 2013)
L	112	177	Maisons-Laffitte - Nanterre-Université
N	60	127	Versailles - St-Cyr
P	69	175	Villiers-sur-Marne - Les Boullereaux
R	75	203	Villeneuve-Saint-Georges - Combs la Ville
U	107	172	Saint-Cloud - La Défense

La carte ci-dessous – adaptée de la carte du réseau physique francilien de l’atlas SNCF Réseau édition 2020 – présente la localisation géographique des sections les plus chargées par ligne de RER et Transilien.

Estimation du nombre de circulations de trains quotidiennes (toutes activités ferroviaires) par voie utilisée dans la section présentant le plus haut niveau de trafic pour chaque ligne de RER et Transilien

Crédits fond de carte : SNCF Réseau, atlas du réseau ferré français, édition 2020



Note de lecture : le nombre de trains par voie indiqué s'applique uniquement aux voies utilisées par la ligne de RER ou Transilien considérée, et pas aux éventuelles autres voies parallèles utilisées par d'autres lignes RER ou Transilien ou d'autres activités ferroviaires. Pour les voies retenues, l'ensemble du trafic (voyageurs Transilien, TER, Intercités, TGV + fret et « haut-le-pied ») est comptabilité.

Le détail des trafics par voie et par tronçon pour chaque ligne de RER et Transilien est présenté dans les tableaux ci-dessous. Le nombre de voies support retenu tient compte des éventuelles informations relatives à des parcours des trains sur voies dédiées (doublet de voies dit « doublet banlieue » ou doublet de voies dit « doublet grandes lignes » en cas de 2x2 voies) lorsque ces informations étaient communiquées dans des documents publics comme les schémas directeurs de ligne et autres études techniques publiques.

RER A :

tronçon	Nb voies support	Longueur du tronçon (km de ligne)	TMJA Transilien	TMJA Fret+GL+TER	TMJA total	TMJA total/voie
Cergy-Le-Haut - Conflans fin d'Oise	2	11,5	182	0	182	91
Conflans fin d'Oise - Maisons-Laffitte	2	8,9	202	0	202	101
Poissy- Maisons-Laffitte	2	9	117	0	117	59
Maisons-Laffitte - Sartrouville	2	1,4	354	0	354	177
Sartrouville - Nanterre-Préfecture	2	7	354	0	354	177
Marne-La-Vallée - Vincennes	2	29	100	0	100	50
Boissy-Saint-Léger - Vincennes	2	16,1	50	0	50	25
Saint-Germain - Nanterre-Préfecture	2	11,3	54	0	54	27
Tronçon central	2	18,3	630	0	630	315

RER B :

tronçon	Nb voies support	Longueur du tronçon (km de ligne)	TMJA Transilien	TMJA Fret+GL+TER	TMJA total	TMJA total/voie
Aéroport CDG - Aulnay-sous-Bois	2	14,2	256	0	256	128
Mitry-Claye - Aulnay-sous-Bois	2	12,2	195	0	195	98
Aulnay-sous-Bois - La Plaine Stade de France	2	9,9	447	0	447	224
La Plaine Stade de France - Gare du Nord	2	4,6	447	0	447	224
Gare du Nord - Châtelet	2	2,3	656	0	656	328
Châtelet - Bourg-la-Reine	2	10,7	570	0	570	285
Bourg-la-Reine - Robinson	2	3,2	125	0	125	63
Bourg-la-Reine - Massy-Palaiseau	2	7,9	445	0	445	223
Massy-Palaiseau - Orsay-ville	2	7,1	187	0	187	94
Orsay-ville - Saint-Rémy-les-Chevreuses	2	8,5	125	0	125	63

RER C :

tronçon	Nb voies support	Longueur du tronçon (km de ligne)	TMJA Transilien	TMJA Fret+GL+TER	TMJA total	TMJA total/voie
---------	------------------	-----------------------------------	-----------------	------------------	------------	-----------------

Pontoise - Montigny-Beauchamp	2	9	167	5,4	172,4	86,2
Montigny-Beauchamp - Ermont-Eaubonne	2	6,1	251	5,1	256,1	128
Ermont-Eaubonne - Champ de Mars - Tour Eiffel	2	19	180	0	180	90
Champ de Mars-Tour Eiffel - Gare d'Austerlitz	2	6,8	380	0	380	190
Gare d'Austerlitz - Choisy-le-Roi	4*	9,6	425	135	560	140
Choisy-le-Roi - Juvisy	4*	9,6	370	180	550	138
Juvisy - Savigny	4*	2,8	370	180	550	138
Savigny - Massy-Palaiseau	2	11,6	120	23	143	72
Choisy-le-Roi - Massy-Palaiseau	2	24,1	110	51,5	161,5	81
Savigny - Brétigny	4*	9,6	178	207	385	96
Brétigny - Saint-Martin-d'Étampes	2	26,1	95	64	159	80
Brétigny - Dourdan-La Forêt	2	25,2	94	10	104	52
Massy-Palaiseau - Versailles-Chantiers	2	14,6	74	33	107	54
Champ de Mars-Tour Eiffel - Viroflay-Rive-Gauche	2	12,1	230	0	230	115
Viroflay-Rive-Gauche - Versailles-Chantier	2	2,7	177,5	0	177,5	89
Versailles - St-Cyr	2	4,9	367,5	0	367,5	184
St-Cyr - Saint-Quentin-en-Yvelines	2	2,5	243	0	243	122

* Le doublet de voies sur lequel circule les trains varie selon les missions RER C sur ce tronçon.

RER D :

tronçon	Nb voies support	Longueur du tronçon (km de ligne)	TMJA Transilien	TMJA Fret+GL+TER	TMJA total	TMJA total/voie
Creil - Orry-la-Ville - Coye	2	15,2	98	0	98	49
Orry-la-Ville - Coye - Goussainville	2	15,6	98	0	98	49
Goussainville - Garges	2	7,3	215	0	215	108
Garges - Saint-Denis	2	6	278	0	278	139
Saint-Denis - Gare du Nord	2	6,2	340	0	340	170
Gare du Nord - Châtelet	2	2,3	656	0	656	328
Châtelet - Gare de Lyon	2	2,7	274	0	274	137
Gare de Lyon - Créteil-Pompadour (voies 1bis et 2bis)	2	9,8	274	0	274	137
Créteil-Pompadour - Villeneuve	2	4,8	274	138	412	206
Villeneuve-Saint-Georges - Combs la Ville	2	11,5	37	227	264	132
Combs la Ville - Melun	2	18,2	27	197	224	112
Villeneuve-Saint-Georges - Juvisy	4*	6,7	273	101	374	94
Juvisy - Corbeil (Ouest)	2	13,3	155	15	170	85
Juvisy - Corbeil (Est)	2	11,2	116	0	116	58
Corbeil - Melun	2	24,4	65	11	76	38
Corbeil - Malesherbes	2	44,1	61	5	66	33

* Le doublet de voies utilisé varie selon les missions RER D pour certaines sections de ce tronçon.

RER E :

tronçon	Nb voies support	Longueur du tronçon (km de ligne)	TMJA Transilien	TMJA Fret+GL+TER	TMJA total	TMJA total/voie
Tournan - Villiers-sur-Marne	2	20	159	51	210	105
Villers-sur-Marne - Nogent-Le Perreux	2	2,2	299	50	349	175
Nogent-Le Perreux - Rosny Bois Perrier	2	7,4	231,5	0	231,5	116
Rosny Bois Perrier - Noisy-le-Sec	2	2,3	231,5	0	231,5	116
Chelles-Gournay - Noisy-le-Sec	2	9,4	231,5	4	235,5	118
Noisy-le-Sec - Magenta	2	8,5	407	0	407	204
Magenta - Hausmann - Saint-Lazare	2	2,5	407	0	407	204

Transilien H :

tronçon	Nb voies support	Longueur du tronçon (km de ligne)	TMJA Transilien	TMJA Fret+GL+TER	TMJA total	TMJA total/voie
Creil - Persan-Beaumont	2	21,2	70	37	107	54
Persan-Beaumont - Valmondois	2	9,2	54	21	75	38
Valmondois - Pontoise	2	10,4	55	21	76	38
Pontoise - Montigny-Beauchamp	2	9	167	5,4	172,4	86
Montigny-Beauchamp - Ermont-Eaubonne	2	6,1	251	5,1	256,1	128
Valmondois - Ermont-Eaubonne	2	15	100	0	100	50
Ermont-Eaubonne - Épinay-Villetaneuse	2	5,2	250	2	252	126
Persan-Beaumont - Montsoul	2	12,1	67	36	103	52
Luzarches - Montsoul	1	11	52	0	52	52
Montsoul - Sarcelles	2	9,6	119	0	119	60
Sarcelles - Épinay-Villetaneuse	2	5,6	155	0	155	78
Épinay-Villetaneuse- Saint-Denis	4	3	360	40	400	100
Saint-Denis - Gare du Nord (groupe de voies DV/RV (départ Valmondois / retour Valmondois) et DM/RM (départ Montsoul – retour Montsoul))	4	6,2	360	40	400	100

Transilien J :

tronçon	Nb voies support	Longueur du tronçon (km de ligne)	TMJA Transilien	TMJA Fret+GL+TER	TMJA total	TMJA total/voie
Gisors - Pontoise	2	39,6	26	0	26	13
Pontoise - Conflans-Sainte-Honorine	2	7,1	90	2	92	46
Conflans-Sainte-Honorine - La Frette-Montigny	2	6,7	181	31	212	106

La Frette-Montigny - Argenteuil	2	7,6	220	31	251	126
Ermont-Eaubonne - Argenteuil	2	4,3	180	66	246	123
Argenteuil - Asnières-sur-Seine	4	9,6	330	30	360	90
Vernon - Mantes	2	23,7	115	65	180	90
Mantes - Conflans-Sainte-Honorine	2	32,4	72	30	102	51
Mantes - Épône-Mézières	3	7,4	115	132	247	82
Épône-Mézières - Poissy	2	22,8	94	133	227	114
Poissy- Houilles	2	13	94	145	239	119,5
Houilles - Asnières-sur-Seine	2	9,4	94	105	199	99,5
Asnières-sur-Seine - Gare Saint-Lazare (groupes de voies IV, V et VI)	6	4,4	509	412	921	154

Transilien K :

tronçon	Nb voies support	Longueur du tronçon (km de ligne)	TMJA Transilien	TMJA Fret+GL+TER	TMJA total	TMJA total/voie
Crépy-en-Valois - Mitry-Claye	2	34,1	34	23	57	28,5
Mitry-Claye - Aulnay-sous-Bois	2	12,2	34	57	91	46
Aulnay-sous-Bois - La Plaine Stade de France	2	9,9	34	64	98	49
La Plaine Stade de France - Gare du Nord	2	4,6	34	64	98	49

Transilien L :

tronçon	Nb voies support	Longueur du tronçon (km de ligne)	TMJA Transilien	TMJA Fret+GL+TER	TMJA total	TMJA total/voie
Cergy-Le-Haut - Conflans fin d'Oise	2	11,5	182	0	182	91
Conflans fin d'Oise - Maisons-Laffitte	2	8,9	202	0	202	101
Maisons-Laffitte - Sartrouville	2	1,4	354	0	354	177
Satrouville - Nanterre-Université	2	6,2	354	0	354	177
Nanterre-Université - Asnières-sur-Seine	2	5,8	260	0	260	130
Versailles-Rive-Droite - Viroflay-Rive-Droite	2	2,5	110	1,7	111,7	56
Viroflay-Rive-Droite - Saint-Cloud	2	6,2	225	1,7	226,7	113
Saint-Cloud - La Défense	2	5,7	343	1,7	344,7	172
La Défense - Asnières-sur-Seine	2	3,8	260	1,6	261,6	131
Saint-Nom-la-Bretèche - Saint-Cloud	2	15,6	120	0	120	60
Asnières-sur-Seine - Gare Saint-Lazare (groupes de voies II, III et V)	6	4,4	657	387	1044	174

Transilien N :

tronçon	Nb voies support	Longueur du tronçon (km de ligne)	TMJA Transilien	TMJA Fret+GL+TER	TMJA total	TMJA total/voie
Rambouillet - La Verrière	2	15,4	103	76	179	90
La Verrière - Saint-Quentin-en-Yvelines	4	8,2	196	105	301	75
Dreux - Plaisir-Grignon	2	48,8	47	18	65	33
Mantes-la-Jolie - Épône-Mézières	3	8,6	115	132	247	82
Épône-Mézières - Plaisir-Grignon	2	20	41	31	72	36
Plaisir-Grignon - Saint-Cyr	2	10,8	129	50	179	90
St-Cyr - Saint-Quentin-en-Yvelines	4	2,5	260	106	366	92
Versailles - St-Cyr	4	4,9	389	117	506	127
Versailles-Chantiers - Viroflay-Rive-Gauche	4	2,7	389	0	389	97
Viroflay-Rive-Gauche - Paris Montparnasse (doublets de voies 1-2 et 1bis-2bis)	4	13,5	239	83	322	81

Transilien P :

tronçon	Nb voies support	Longueur du tronçon (km de ligne)	TMJA Transilien	TMJA Fret+GL+TER	TMJA total	TMJA total/voie
La Ferté-Milon - Trilport	2	29,6	30	7	37	19
Château-Thierry - Trilport	2	44,1	40	131	171	86
Trilport - Meaux	2	6	69	138	207	104
Meaux - Esbly	2	8	130	138	268	134
Crécy-la-Chapelle - Esbly	2	9,8	41	0	41	21
Esbly - Lagny-Thorigny	2	8,8	141	136	277	139
Lagny - Vaires-Torcy	4	4,8	242	136	378	95
Vaires-Torcy - Chelles-Gournay	4	4,2	242	177	419	105
Chelles-Gournay - Noisy-le-Sec	2	9,4	108,5	103	211,5	106
Coulommiers - Tournan	1	30,8	35	0	35	35
Provins - Tournan	2	56,6	35	48	83	42
Tournan - Villiers-sur-Marne	2	20	159	51	210	105
Villers-sur-Marne - Nogent-Le Perreux	2	2,2	299	50	349	175
Nogent-Le Perreux - Rosny Bois Perrier	2	7,4	67,5	81	148,5	74
Rosny Bois Perrier - Noisy-le-Sec	2	2,3	67,5	81	148,5	74
Noisy-le-Sec - Paris-Est	4	8,5	304	160	464	116

Transilien R :

tronçon	Nb voies support	Longueur du tronçon (km de ligne)	TMJA Transilien	TMJA Fret+GL+TER	TMJA total	TMJA total/voie
Gare de Lyon - Créteil-Pompadour (voies 1 et 2)	2	9,8	178	144	322	161
Créteil-Pompadour - Villeneuve	2	4,8	178	138	316	158
Villeneuve-Saint-Georges - Combs la Ville	2	11,5	178	227	405	203
Combs la Ville - Melun	2	18,2	178	197	375	188
Melun - Moret	2	22,5	83	34	117	59
Melun - Montereau	2	36,1	37	24	61	31
Montereau - Moret	2	11,8	37	18	55	28
Moret - Montargis	2	50,8	46	19	65	33

Transilien U :

tronçon	Nb voies support	Longueur du tronçon (km de ligne)	TMJA Transilien	TMJA Fret+GL+TER	TMJA total	TMJA total/voie
La Verrière - Saint-Quentin-en-Yvelines	2	8,2	179	0	179	90
Saint-Quentin-en-Yvelines - St-Cyr	2	2,5	243	0	243	122
St-Cyr - Versailles	2	4,9	367,5	0	367,5	184
Versailles-Chantiers - Viroflay-Rive-Droite	2	3,1	78	1,7	79,7	40
Viroflay-Rive-Droite - Saint-Cloud	2	6,2	225	1,7	226,7	113
Saint-Cloud - La Défense	2	5,7	343	1,7	344,7	172

Annexe 23 : Tableau de colinéarité entre les variables explicatives du taux de retard des RER et Transilien

	LongueurLigne	NbArrêts	DistanceMoyEntreArrêts	NbCircJours	NbVoyageursJour	NbBranches	TrKm	IntervalleMin_2019	NbTrainsHPM_2019	VoiesPartagées	AgeMoyen_MatérielRoulant_1erJanvier	NbMoyArrêtsHPM_2019	NbMoyArrêtsHPMdansParis_2019	TpsParcoursMoyHPM_2019	DuréeMoyInterstation_2019	TypeContrôleVitesse	GareDuNord	GareDeLyon	GareStLazare	ChâteletLesHalles
LongueurLigne																				
NbArrêts						X														
DistanceMoyEntreArrêts															X					
NbCircJours																				
NbVoyageursJour							X		X			X								
NbBranches		X																		
TrKm					X				X											
IntervalleMin_2019																				
NbTrainsHPM_2019					X		X													
VoiesPartagées																				
AgeMoyen_MatérielRoulant_1erJanvier																				
NbMoyArrêtsHPM_2019					X															
NbMoyArrêtsHPMdansParis_2019																				
TpsParcoursMoyHPM_2019																				
DuréeMoyInterstation_2019			X																	
TypeContrôleVitesse																				
GareDuNord																				
GareDeLyon																				
GareStLazare																				
ChâteletLesHalles																				

Code couleur : **en jaune**, variables présentant des relations de colinéarité avec d'autres variables

Annexe 24 : Effet de l'âge moyen du matériel roulant sur le taux de retard des RER et Transilien par ligne hors ligne K

L'annexe 21 présente les données utilisées pour la variable de l'âge moyen du matériel roulant en début d'année par ligne RER ou Transilien. On peut remarquer une nette rupture entre 2016 et 2017 pour la ligne K. En effet, l'intégralité du matériel roulant de cette ligne a été remplacé le 2 septembre 2016 (remplacement de rames réversibles type RIB par des rames Francilien Z 50000). On pourrait alors imaginer que l'effet de l'âge du matériel roulant sur la ponctualité par ligne soit biaisé par cette rupture, et qu'il soit non significatif en retirant la ligne K de l'échantillon étudié. Nous avons testé cette éventualité. Les résultats sont présentés ci-dessous.

```
Call:
lm(formula = TxRetard ~ +TMJA_Tot_ParVoie_Max + DistanceMoyTrain_2019 +
    AgeMoyen_MaterielRoulant_1erJanvier + GareDeLyon + GareStLazare +
    GareDuNord + GareParisEst, data = data_PonctuTransilien_lignes_HorsLigneK)

Residuals:
    Min       1Q   Median       3Q      Max
-5.4637 -1.3546  0.0187  1.4126  4.9031

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.20290    1.33313   -0.902  0.36975
TMJA_Tot_ParVoie_Max  0.01672    0.00520    3.216  0.00191 **
DistanceMoyTrain_2019  0.10534    0.02116    4.977 3.92e-06 ***
AgeMoyen_MaterielRoulant_1erJanvier  0.09707    0.03574    2.716  0.00817 **
GareDeLyon         2.03852    0.86064    2.369  0.02040 *
GareStLazare       5.54951    0.77162    7.192 3.81e-10 ***
GareDuNord         1.14413    0.64481    1.774  0.08001 .
GareParisEst       1.22045    0.74329    1.642  0.10473
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.033 on 76 degrees of freedom
Multiple R-squared:  0.6791,    Adjusted R-squared:  0.6496
F-statistic: 22.98 on 7 and 76 DF,  p-value: < 2.2e-16
```

Le test statistique montre que l'effet de l'âge du matériel roulant reste significatif au seuil 1% et de même signe. Il est en revanche légèrement atténué en retirant la ligne K (+1,0 point de retard/vieillessement de 10 ans du parc de matériel roulant, contre +1,2 points en intégrant la ligne K).

Annexe 25 : Cartes SNCF Réseau - mode d'espacement des trains et lignes équipées de contrôle de vitesse en 2018

MODES D'ESPACEMENT DES TRAINS

SITUATION AVRIL 2018

(fermetures de l'année 2018 prises en compte)



- Légende**
- Signalisation en cabine (LGV)
 - Block automatique lumineux
 - Block automatique autre
 - Block manuel
 - Cantonnement téléphonique
 - Autre (signalisation simplifiée)
 - En France, LGV avec signalisation latérale (BAL + KVB opérationnel, ETCS 1)



- NANTES** Préfecture de région
QUIMPER Préfecture de département
Brest Ville, gare importante
Lannion Autre ville, bifurcation importante

LIGNES ÉQUIPÉES DE CONTRÔLE DE VITESSE Hors zones de gares

SITUATION AVRIL 2018

(fermetures de l'année 2018 prises en compte)



- Légende**
- ETCS 2
 - ETCS 1
 - TVM 430
 - TVM 300
 - KVB
 - Lignes non équipées
 - LGV



- NANTES** Préfecture de région
QUIMPER Préfecture de département
Brest Ville, gare importante
Lannion Autre ville, bifurcation importante

Sources : Données Réseau RGI, avril 2018
Document non contractuel

Sources : Données Réseau RGI, avril 2018
Document non contractuel



Annexe 26 : Modèles testant l'effet du trafic par branche pour l'ensemble des lignes RER et Transilien

Afin de tester l'influence du niveau de trafic par branche sur le taux de retard des RER et Transilien par branche sur la période 2013-2018¹², des modèles de régressions linéaires multiples ont été implémentés. Les données de l'OMNIL utilisées correspondent au taux de ponctualité voyageurs Île-de-France Mobilités annuel pour les voyageurs qui descendent à l'une des gares de la branche.

4 modèles ont été retenus dans le but de vérifier la robustesse statistique de l'impact identifié du niveau de trafic de la branche sur son taux de retard annuel. Les modèles retenus sont ceux qui optimisent un ensemble de critères statistiques de sélection et pour lesquels aucun signe des coefficients associés aux variables explicatives – hormis celui de la variable portant sur le niveau de trafic de la branche - ne semble contre-intuitif ou inexplicable.

Les 2 premiers modèles intègrent la variable sur le niveau maximal du Trafic Moyen Journalier Annuel (TMJA) par voie au sein de la branche : le premier prend également en compte le niveau du TMJA par voie au niveau de la section la plus chargée empruntée par les trains de la branche sur l'ensemble de leur parcours, qui ne se limite pas à la branche et traverse généralement la section la plus chargée de la ligne ; le second modèle ne prend pas en compte cette variable.

Les 2 derniers modèles se décomposent de la même manière mais porte sur le niveau moyen du TMJA par voie au sein de la branche, ce qui pourrait *a priori* fournir des résultats différents par rapport au niveau maximal.

Les résultats suivants ont été obtenus :

¹² Pour l'année 2019, les données publiées par l'OMNIL ne couvrent que la période de janvier à juillet 2019. En cas de saisonnalité du taux de ponctualité, il est possible que les disparités entre branches de l'effet de saisonnalité influencent la ponctualité, puis la validité des résultats obtenus avec la modélisation statistique. Nous n'avons donc pas intégré les données par branches de 2019 dans la modélisation.

```
Call:
lm(formula = TxRetard ~ TMJA_ParVoie_Max_Branche + TMJA_ParVoie_Max_ParcoursCompleet +
  GareStLazare + VoiesPartagées_Transilien + AgeMoyen_MaterielRoulant_1erJanvier,
  data = data_PonctuTransilien_branches)

Residuals:
    Min       1Q   Median       3Q      Max
-11.2117  -2.8445  -0.6353   2.3642  14.6112

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.563081   0.980604   1.594 0.112081
TMJA_ParVoie_Max_Branche
-0.021408     0.005302  -4.037 7.01e-05 ***
TMJA_ParVoie_Max_ParcoursCompleet
 0.056963     0.004002  14.235 < 2e-16 ***
GareStLazare   3.542354     0.950947   3.725 0.000237 ***
VoiesPartagées_Transilien
 1.124743     0.571961   1.966 0.050246 .
AgeMoyen_MaterielRoulant_1erJanvier
 0.036296     0.040434   0.898 0.370148

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.553 on 276 degrees of freedom
(12 observations deleted due to missingness)
Multiple R-squared:  0.4632,    Adjusted R-squared:  0.4534
F-statistic: 47.62 on 5 and 276 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = TxRetard ~ NbArrêts + AgeMoyen_MaterielRoulant_1erJanvier +
  GareStLazare + GareParisEst + TMJA_ParVoie_Max_Branche +
  ChâteletLesHalles + VoiesPartagées_Transilien, data = data_PonctuTransilien_branches)

Residuals:
    Min       1Q   Median       3Q      Max
-10.4801  -2.9770  -0.6341   2.3087  13.9939

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.257769   1.279853   2.545  0.0115 *
NbArrêts       0.298182   0.065982   4.519 9.24e-06 ***
AgeMoyen_MaterielRoulant_1erJanvier
 0.203134     0.038978   5.211 3.69e-07 ***
GareStLazare   5.468496     1.021409   5.354 1.82e-07 ***
GareParisEst   1.769573     0.973834   1.817  0.0703 .
TMJA_ParVoie_Max_Branche
-0.034266     0.005661  -6.053 4.66e-09 ***
ChâteletLesHalles
 8.173377     0.792686  10.311 < 2e-16 ***
VoiesPartagées_Transilien
 2.944772     0.632975   4.652 5.12e-06 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.33 on 274 degrees of freedom
(12 observations deleted due to missingness)
Multiple R-squared:  0.518,    Adjusted R-squared:  0.5056
F-statistic: 42.06 on 7 and 274 DF,  p-value: < 2.2e-16
```

```

Call:
lm(formula = TxRetard ~ TMJA_ParVoie_Moy_Branche + TMJA_ParVoie_Max_ParcoursComple +
  GareStLazare + VoiesPartagées_Transilien + AgeMoyen_MaterielRoulant_1erJanvier,
  data = data_PonctuTransilien_branches)

Residuals:
    Min       1Q   Median       3Q      Max
-10.895  -2.815  -0.548   2.319  13.090

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.863340   0.935631   4.129 4.77e-05 ***
TMJA_ParVoie_Moy_Branche
-0.058769    0.008127  -7.232 4.32e-12 ***
TMJA_ParVoie_Max_ParcoursComple
0.054069    0.003415  15.831 < 2e-16 ***
GareStLazare  2.814621    0.869564   3.237 0.00135 **
VoiesPartagées_Transilien
1.882661    0.574481   3.277 0.00118 **
AgeMoyen_MaterielRoulant_1erJanvier
0.112349    0.034534   3.253 0.00128 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.366 on 288 degrees of freedom
Multiple R-squared:  0.4991,    Adjusted R-squared:  0.4904
F-statistic: 57.38 on 5 and 288 DF,  p-value: < 2.2e-16

```

```

Call:
lm(formula = TxRetard ~ NbArrêts + AgeMoyen_MaterielRoulant_1erJanvier +
  GareDeLyon + GareStLazare + GareParisEst + TMJA_ParVoie_Moy_Branche +
  ChâteletLesHalles + VoiesPartagées_Transilien, data = data_PonctuTransilien_branches)

Residuals:
    Min       1Q   Median       3Q      Max
-11.2561  -3.1278  -0.6586   2.2611  14.9968

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.132062   1.280306   4.008 7.81e-05 ***
NbArrêts      0.252995   0.063377   3.992 8.34e-05 ***
AgeMoyen_MaterielRoulant_1erJanvier
0.207558    0.035974   5.770 2.07e-08 ***
GareDeLyon    1.249908    0.847568   1.475 0.141397
GareStLazare  4.371053    0.961178   4.548 8.03e-06 ***
GareParisEst  2.709299    1.017124   2.664 0.008169 **
TMJA_ParVoie_Moy_Branche
-0.046043    0.008227  -5.597 5.13e-08 ***
ChâteletLesHalles
6.593027    1.037110   6.357 8.14e-10 ***
VoiesPartagées_Transilien
2.413971    0.670979   3.598 0.000378 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.326 on 285 degrees of freedom
Multiple R-squared:  0.5134,    Adjusted R-squared:  0.4997
F-statistic: 37.58 on 8 and 285 DF,  p-value: < 2.2e-16

```

D'après la p-value associée au test de Fisher sur la régression, les modèles retenus sont significatifs au seuil de 0,1%. Les coefficients de détermination ajustés ou R^2 ajustés des modèles retenus varient entre 0,45 et 0,506, ce qui signifie que 45% à 50,6% de la variance du taux de retard moyen annuel des RER et Transilien par branche sur la période 2013-2018 sont expliqués par les modèles.

Globalement, nous retrouvons une majorité de variables dont l'effet sur le taux de retard des RER et Transilien a déjà été identifié avec le modèle par ligne retenu dans la section 3.2.4.1, comme notamment le passage par les grandes gares parisiennes, le temps de parcours des trains (équivalent en première approximation à la longueur du parcours dans le modèle par ligne), l'âge du matériel roulant et le niveau du trafic au niveau de la section la plus chargée empruntée par les trains sur l'ensemble de leur parcours.

D'après les modèles retenus, les branches dont certaines sections sont partagées entre plusieurs lignes de RER et Transilien sont moins ponctuelles que les autres branches. La significativité de ce résultat passe le seuil de 1% pour 3 des 4 modèles retenus, et celui de 0,1% pour 2 de ces modèles. En moyenne, le taux de retard de ces branches est plus élevé de 1,1 à 2,9 points selon le modèle choisi parmi les 4 modèles retenus. Ce résultat semble en accord avec l'intuition selon laquelle lorsqu'une branche est partiellement partagée entre plusieurs lignes de RER et Transilien, son exploitation est rendue plus difficile puisque les plans de transport des lignes doivent être coordonnés, y compris en situation perturbée suite à un incident.

En revanche, le fait que la branche comporte des sections partagées avec d'autres activités ferroviaires (TER, Grandes Lignes ou fret) a également été testé et ne présente pas d'impact significatif sur le taux de retard par branche à l'issue de la modélisation.

D'après les 4 modèles retenus, le niveau du trafic par voie au sein la branche présente un impact sur son taux de retard contraire à l'intuition : en moyenne, les branches qui connaissent le trafic le plus faible ne sont pas les plus ponctuelles. Le résultat est significatif au seuil de 0,1%, aussi bien en prenant en compte le niveau maximal ou le niveau moyen du trafic au sein de la branche.

Les 4 modèles précédents semblent indiquer que l'intensité du trafic par voie au sein de la branche n'est pas une cause de non-ponctualité lorsqu'on considère l'ensemble des lignes des réseaux RER et Transilien.

Toutefois, on observe qu'un trafic par voie plus élevé est corrélé avec une moins bonne ponctualité de la branche en se restreignant à certaines lignes de RER et Transilien. 2 autres modèles sur la ponctualité des RER et Transilien par branche pour des groupes de lignes ont ainsi été construits : le premier portant sur le niveau moyen du trafic par voie de la branche, et le second sur le niveau maximal du trafic par voie de la branche. Ces deux modèles sont présentés dans le corps du document en section 3.2.4.4.

Annexe 27 : Modèle testant l'effet du nombre de voies de service en terminus de branche pour l'ensemble des lignes RER et Transilien

On peut penser qu'un nombre de voies de garage en bout de branche plus élevé s'accompagne d'une meilleure ponctualité. En effet, si le terminus/départ de la branche dispose de nombreuses voies de service (voies de garage notamment), il sera plus aisé pour l'exploitant d'adapter le plan de circulation en cas d'incident par l'ajout ou la suppression de trains puisqu'il y aura davantage de rames en réserve dans le sens du départ et de voies disponibles pour accueillir des rames dans le sens de l'arrivée. On s'attend alors à ce qu'en moyenne sur l'année, cette souplesse améliore la ponctualité de la branche concernée d'une ligne de RER et Transilien donnée.

Afin de vérifier statistiquement l'effet de cette variable sur le taux de retard des RER et Transilien par branche sur la période 2013-2018, le modèle de régression linéaire multiple ci-dessous a été implémenté. Les données de l'OMNIL utilisées correspondent au taux de ponctualité voyageurs Île-de-France Mobilités annuel pour les voyageurs qui descendent à l'une des gares de la branche.

```
Call:
lm(formula = TxRetard ~ +NbArrêts + AgeMoyen_MaterielRoulant_1erJanvier +
    GareStLazare + GareParisEst + GareDeLyon + ChâteletLesHalles +
    VoiesPartagées_Transilien + NbVoiesGarageTerminus, data = data_PonctuTransilien_branches)

Residuals:
    Min       1Q   Median       3Q      Max
-12.757  -2.879  -0.488   2.486  16.527

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.42542    1.34995   3.278 0.001177 **
NbArrêts          0.14197    0.06510   2.181 0.030032 *
AgeMoyen_MaterielRoulant_1erJanvier 0.13800    0.03477   3.969 9.17e-05 ***
GareStLazare     3.58516    1.00352   3.573 0.000416 ***
GareParisEst     2.30307    1.07392   2.145 0.032854 *
GareDeLyon       1.99930    0.87613   2.282 0.023244 *
ChâteletLesHalles 7.06217    1.10565   6.387 7.04e-10 ***
VoiesPartagées_Transilien 1.08914    0.64832   1.680 0.094090 .
NbVoiesGarageTerminus -0.02241    0.05016  -0.447 0.655385
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.497 on 279 degrees of freedom
(6 observations deleted due to missingness)
Multiple R-squared:  0.4844,    Adjusted R-squared:  0.4696
F-statistic: 32.77 on 8 and 279 DF,  p-value: < 2.2e-16
```

On peut constater que l'effet du nombre de voies de service en bout de branche ne passe pas le test de significativité à 10%, même si le signe négatif du coefficient associé à cette variable est conforme à l'intuition : davantage de voies de service en bout de ligne participent d'une meilleure ponctualité à long-terme, pour les raisons exposées deux paragraphes plus haut.

En revanche, l'effet passe le seuil de significativité lorsqu'on se restreint à certains groupes de lignes, comme on peut le constater dans le modèle retenu pour les lignes A, H et P présenté pages 79-80.

Annexe 28 : Comparaison des résultats pour des différents services ferroviaires

Comparaison des modèles TGV et Intercités

Les modèles retenus pour les TGV et les services Intercités sont rappelés ci-dessous :

Pour les TGV :

```
Call:
lm(formula = TxRetard ~ DépartParis + DépartLyon + LiaisonPasseParAvignon +
  LiaisonPasseParRennes + LiaisonPasseParDijon + DépartOuArr_NoedSecondaire +
  NbMoyenArrêts_2019 + TMJA_max_Total_Liaison_2000, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-13.1148  -3.2723  -0.1854   2.8244  18.7606

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.986504   0.669113  16.420 < 2e-16 ***
DépartParis    -2.291971   0.335944  -6.822 1.76e-11 ***
DépartLyon     4.890913   0.754562   6.482 1.58e-10 ***
LiaisonPasseParAvignon
 2.460692     0.489295   5.029 6.08e-07 ***
LiaisonPasseParRennes
 -2.849098     0.566522  -5.029 6.08e-07 ***
LiaisonPasseParDijon
 -4.838904     0.743421  -6.509 1.33e-10 ***
DépartOuArr_NoedSecondaire
 3.113007     0.402450   7.735 3.10e-14 ***
NbMoyenArrêts_2019
 0.314262     0.104626   3.004 0.00275 **
TMJA_max_Total_Liaison_2000
 0.005308     0.002261   2.348 0.01912 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.583 on 803 degrees of freedom
Multiple R-squared:  0.3244,    Adjusted R-squared:  0.3177
F-statistic: 48.2 on 8 and 803 DF,  p-value: < 2.2e-16
```

31,8% de la variance du taux de retard moyen annuel par liaisons TGV sur la période 2012-2019 est expliquée par le modèle TGV, contre 33,4% pour le modèle Intercités. Les parts de la variance expliqués sont donc très proches. On peut supposer que cela suggère des similitudes dans l'impact des événements causes de retard qui ne sont pas pris en compte par les variables explicatives retenues, en particulier :

- Les évènements météorologiques majeurs (inondations et pluies intenses, tempêtes, épisodes neigeux et de froid, canicule et fortes chaleurs, etc.) ;
- Les mouvements sociaux ;
- Les autres évènements externes au transport, dont notamment les actes de malveillance, les accidents de personne, les collisions des trains avec des véhicules, ou avec des animaux, ...);

- Les incidents liés au dysfonctionnements des infrastructures ferroviaires, par exemple les ruptures de caténaire ou les pannes de signalisation ;
- Les dysfonctionnements du matériel roulant, avec notamment les pannes du matériel moteur ;
- Les incidents voyageurs (malaise d'un voyageur, blocage involontaire des portes par un voyageur, etc.)

Pour les Intercités :

```
Call:
lm(formula = TxRetard ~ LiaisonNuit + NbNoeudsMajeursTraversés +
    LiaisonTraverseBordeaux + LiaisonTraverseToulouse + LiaisonTraverseDijon +
    LiaisonTraverseRouen + DépartOuArr_NoeudSecondaire + Axe_Est_Centre +
    Nord_Sud, data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-15.0183  -3.2128  -0.3527   2.6899  22.4246

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      6.9484    0.8298   8.374 8.36e-16 ***
LiaisonNuit     -3.9108    1.0465  -3.737 0.000212 ***
NbNoeudsMajeursTraversés
LiaisonTraverseBordeaux
                  7.2998    0.8684   8.406 6.59e-16 ***
LiaisonTraverseToulouse
                  2.8881    0.8259   3.497 0.000521 ***
LiaisonTraverseDijon
                  4.2044    1.7489   2.404 0.016645 *
LiaisonTraverseRouen
                 -2.9840    1.0266  -2.907 0.003848 **
DépartOuArr_NoeudSecondaire
                  1.6491    0.6238   2.644 0.008505 **
Axe_Est_Centre  -2.2518    0.7699  -2.925 0.003633 **
Nord_Sud         2.4554    0.5921   4.147 4.08e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.288 on 421 degrees of freedom
Multiple R-squared:  0.3482,    Adjusted R-squared:  0.3343
F-statistic: 24.99 on 9 and 421 DF,  p-value: < 2.2e-16
```

Afin de comparer l'impact des variables explicatives des modèles retenus sur le taux de retard des TGV et des trains Intercités, nous avons tenté de croiser ces modèles, en appliquant le modèle « TGV » aux données des trains Intercités, et le modèle « trains Intercités » aux données portant sur les TGV. Les réseaux TGV et Intercités n'étant pas parfaitement similaires, certaines variables ne sont pas applicables lors du croisement entre les services. Par exemple, aucune liaison Intercités de la base de l'AQST ne traverse Rennes ou Avignon, comme le montre le plan du réseau Intercités fourni en annexe 12. De même, aucune liaison TGV de la base de l'AQST ne traverse Rouen, et elle ne contient pas de liaison « TGV de nuit ».

L'application du modèle « Intercités » aux données sur les TGV retourne les résultats suivants :

```
Call:
lm(formula = TxRetard ~ NbNoeudsMajeursTraversés + LiaisonPasseParBordeaux +
    LiaisonPasseParToulouse + LiaisonPasseParDijon + Nord_Sud +
    DépartOuArr_NoeudSecondaire, data = data_PonctuTGV_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-10.0662  -3.9048  -0.7357   3.0942  20.3282

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)         9.4927    0.4661  20.367 < 2e-16 ***
NbNoeudsMajeursTraversés  2.1711    0.2998   7.241 1.04e-12 ***
LiaisonPasseParBordeaux  -1.8681    1.2888  -1.449  0.14760
LiaisonPasseParToulouse   5.5271    1.8188   3.039  0.00245 **
LiaisonPasseParDijon     -4.2161    0.7783  -5.417  8.00e-08 ***
Nord_Sud                -0.4167    0.3961  -1.052  0.29311
DépartOuArr_NoeudSecondaire  3.0080    0.4578   6.571  8.96e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.123 on 805 degrees of freedom
Multiple R-squared:  0.1537,    Adjusted R-squared:  0.1473
F-statistic: 24.36 on 6 and 805 DF,  p-value: < 2.2e-16
```

La part de la variance expliquée par le modèle est de 14,7%, soit nettement moins qu'avec le modèle « TGV » retenu. Cela suggère que le mode de fonctionnement des activités TGV et Intercités est assez différent, ce que nous pourrions le cas échéant confirmer après application du modèle « TGV » aux données de ponctualité des trains Intercités.

Le passage par le nœud ferroviaire de Bordeaux, et le fait que le trajet soit effectué du Nord vers le Sud n'ont pas d'effet statistiquement significatif au seuil de 10%.

En comparant les coefficients associés aux variables explicatives lorsqu'on applique le modèle « Intercités » aux données Intercités et TGV, on peut remarquer que les liaisons TGV passant par Dijon sont davantage ponctuelles que les autres liaisons TGV, tandis que les liaisons Intercités passant par Dijon sont moins ponctuelles que les autres liaisons Intercités. Un élément d'explication possible concernant les TGV pourrait être que les liaisons TGV concernées empruntent la LGV Rhin-Rhône, récente (mise en service en 2011) et relativement peu intensément utilisée en 2018 en comparaison aux autres LGV d'après l'atlas du réseau ferré en France – édition 2020 (SNCF Réseau). En revanche, au-delà de l'effet d'engorgement qui peut se produire au niveau du nœud ferroviaire, il n'y a pas d'explication évidente à ce stade à la moins bonne ponctualité des liaisons Intercités traversant Dijon.

Les coefficients associés à l'effet sur le taux de retard du passage par un nœud ferroviaire majeur (Paris, Lyon, Marseille) sont de même signe lorsqu'on applique le modèle « Intercités » aux données Intercités et TGV. En revanche, la dégradation de la ponctualité liée au passage par un nœud ferroviaire majeure est près de 2,4 fois plus importante pour les services Intercités.

On note une situation inverse concernant l'effet sur le taux de retard du passage par le nœud ferroviaire de Toulouse, qui engendre une dégradation de la ponctualité près de 2 fois plus importante sur les services Intercités que sur les TGV d'après ce modèle.

Les coefficients associés à l'impact sur le taux de retard d'un départ ou d'une arrivée située au niveau d'un nœud secondaire ne sont pas totalement comparables, puisque les nœuds secondaires retenus dépendent du service : seuls 3 nœuds ferroviaires (Bordeaux, Dijon, Tours et Toulouse) sont communs aux 2 listes, sur 8 nœuds ferroviaires retenus pour les TGV, et 13 pour les trains Intercités.

Néanmoins, on peut observer que le coefficient de cette variable est du même signe dans les 2 modèles (les liaisons ayant pour origine ou terminus un nœud ferroviaire secondaire parmi les nœuds retenus sont moins ponctuelles que les autres), et que l'impact sur le taux de retard est près de 2 fois plus important dans le modèle TGV que dans le modèle « Intercités ».

L'application du modèle « TGV » aux données sur les trains Intercités retourne les résultats suivants :

```
Call:
lm(formula = TxRetard ~ DépartParis + DépartLyon + LiaisonTraverseDijon +
    DépartOuArr_NoedSecondaire + NbMoyenArrêts_2020 + TMJA_max_Total_Liaison_2000,
    data = data_PonctuIC_liaisons)

Residuals:
    Min       1Q   Median       3Q      Max
-15.834  -3.760  -1.101   3.188  23.918

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.675633   1.073184   9.948 < 2e-16 ***
DépartParis     1.698127   0.641925   2.645  0.00846 **
DépartLyon     -2.291854   2.794113  -0.820  0.41254
LiaisonTraverseDijon
 6.496354     1.333152   4.873 1.56e-06 ***
DépartOuArr_NoedSecondaire
 3.394063     0.636557   5.332 1.58e-07 ***
NbMoyenArrêts_2020
 -0.041957    0.077101  -0.544  0.58660
TMJA_max_Total_Liaison_2000
 0.001831     0.003376   0.542  0.58782
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.135 on 424 degrees of freedom
Multiple R-squared:  0.1164,    Adjusted R-squared:  0.1039
F-statistic: 9.312 on 6 and 424 DF,  p-value: 1.329e-09
```

La part de la variance expliquée par le modèle est de 10,4%, soit nettement moins qu'avec le modèle « Intercités » retenu. Cela confirme l'hypothèse de la page précédente, selon laquelle l'utilisation d'un modèle spécifique par activité est plus appropriée parce que les déterminants de la ponctualité des activités TGV et Intercités sont assez différents car le mode de fonctionnement des activités TGV et Intercités est assez différent. A première vue, ce résultat n'est pas invraisemblable puisque les TGV roulent pour 2/3 de leurs parcours - voire plus depuis 2017 et la mise en service des LGV Bretagne-Pays de la Loire et Sud Europe Atlantique - sur les LGV. Or certains aléas pouvant entraîner d'importants retards sont plus rares sur les LGV que sur le réseau classique car les infrastructures sont globalement récentes, bien entretenues et aux normes. Les TGV sont également moins soumis aux aléas liés à la mixité de trafic que les trains Intercités, et s'arrêtent également moins fréquemment que les trains Intercités.

Le fait que la liaison Intercités parte de Lyon (gares de Lyon-Perrache ou Lyon-Part-Dieu), l'estimation du nombre moyen d'arrêts effectués (gares de départ et de terminus n'ont compris), et le niveau du trafic de l'ensemble des circulations de trains (y compris fret et « haut-le-pied ») sur la

section la plus intensément utilisée du réseau emprunté par la liaison Intercités en 2000, n'ont pas d'effet statistiquement significatif au seuil de 10%.

En comparant les coefficients associés aux variables explicatives lorsqu'on applique le modèle « TGV » aux données Intercités et TGV, on peut remarquer que les liaisons TGV partant de Paris sont davantage ponctuelles que les autres liaisons TGV, tandis que les liaisons Intercités partant de Paris sont moins ponctuelles que les autres liaisons Intercités. Il n'y a pas d'explication évidente à cette divergence à ce stade.

On note également que les coefficients associés au passage par le nœud ferroviaire de Dijon sont opposés lorsqu'on applique les modèles aux services TGV et Intercités. Un élément d'explication possible a été présenté précédemment dans l'analyse de l'application du modèle « Intercités » aux données TGV.

En revanche, l'impact d'un départ ou d'une arrivée au niveau d'un nœud ferroviaire secondaire sur le taux de retard est de même signe et de proportions similaires quand on applique ce modèle aux données TGV et Intercités.

Comparaison des modèles TER et RER et Transilien d'Île-de-France

Afin de comparer l'impact des variables explicatives des modèles retenus sur le taux de retard des TER et des trains régionaux d'Île-de-France (RER et Transilien), nous avons tenté de croiser ces modèles, en appliquant le modèle « TER » aux données des trains régionaux d'Île-de-France, et le modèle « RER et Transilien » aux données portant sur les TER. Les réseaux TER et Transilien n'étant pas parfaitement similaires, certaines variables ne sont pas applicables lors du croisement entre les services. Par exemple, la gare de Lyon se situe dans Paris et non en province. De même, il n'y a pas de variable évidente permettant de transposer la densité des lignes des catégories UIC 7 à 9 (AV) par ancienne région au réseau Transilien d'Île-de-France par ligne.

L'application du modèle « TER » aux données sur les RER et Transilien retourne les résultats suivants :

```
Call:
lm(formula = TxRetard ~ DistanceMoyTrain_2019 + NbMoyArrêtsHPM_2019 +
    VoyKm, data = data_PonctuTransilien_lignes)

Residuals:
    Min       1Q   Median       3Q      Max
-7.0595 -2.2302 -0.5956  1.8326  8.8305

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.64671    1.65341    3.415 0.000970 ***
DistanceMoyTrain_2019 0.09132    0.02263    4.035 0.000117 ***
NbMoyArrêtsHPM_2019  0.06121    0.13561    0.451 0.652825
VoyKm           0.27267    0.37872    0.720 0.473456
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.062 on 87 degrees of freedom
Multiple R-squared:  0.2537,    Adjusted R-squared:  0.228
F-statistic: 9.858 on 3 and 87 DF,  p-value: 1.156e-05
```


La part de la variance expliquée par le modèle est de 22,8%, soit nettement moins qu'avec le modèle « Transilien » retenu. Cela suggère que le mode de fonctionnement des activités TER en province et RER et Transilien en Île-de-France, ainsi que les caractéristiques de leurs réseaux, sont assez différents, ce que nous pourrions le cas échéant confirmer et commenter après application du modèle « Transilien » aux données de ponctualité des TER de province.

Le nombre de voyageurs-km transportés et le nombre moyen d'arrêts effectués à l'heure de pointe n'ont pas d'effet statistiquement sur le taux de retard par ligne significatif au seuil de 10%.

En comparant les coefficients associés à la distance moyenne parcourue par les trains régionaux pour les 2 activités, on peut remarquer que l'impact sur le taux de retard d'un ajout de 10 km à ces variables est plus important pour les trains régionaux d'Île-de-France (+0,91 points) que pour les TER (+0,25 points). Cependant, la comparaison de l'impact d'un allongement absolu n'est pas complètement pertinente puisque la moyenne sur le panel des régions de la distance moyenne parcourue par les TER (86 km) est de l'ordre du double de la moyenne sur le panel des lignes RER et Transilien de la distance moyenne parcourue par les trains (42 km). Toutefois, un allongement relatif de 20% de la distance parcourue par les trains régionaux entraînerait selon le modèle « TER » une dégradation plus forte de la ponctualité pour les RER et Transilien (+0,76 points de retard) que pour les TER (+0,4 points).

L'application du modèle « Transilien » aux données sur les TER retourne les résultats suivants :

```
Call:
lm(formula = TxRetard ~ NbCirc_an_AQST + DistanceMoy + AgeMoyen_MaterielRoulant_1er
  Janvier,
  data = data_PonctuTER_régions)

Residuals:
    Min       1Q   Median       3Q      Max
-4.2463 -1.5702 -0.1638  1.1764  9.6066

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -0.66821    1.48374  -0.450   0.653
NbCirc_an_AQST    0.01667    0.00382   4.363 2.57e-05 ***
DistanceMoy       0.08070    0.01343   6.009 1.71e-08 ***
AgeMoyen_MaterielRoulant_1erJanvier  0.03860    0.06263   0.616   0.539
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.555 on 132 degrees of freedom
Multiple R-squared:  0.2537,    Adjusted R-squared:  0.2368
F-statistic: 14.96 on 3 and 132 DF,  p-value: 1.936e-08
```

La part de la variance expliquée par le modèle est de 23,7%, soit nettement moins qu'avec le modèle « TER » retenu. Cela confirme l'hypothèse de la page précédente, selon laquelle l'utilisation d'un modèle spécifique par activité est plus appropriée parce que les déterminants de la ponctualité des activités TER et RER et Transilien sont pour partie différents.

A première vue, divers éléments laissent penser que ce résultat n'est pas invraisemblable. Des contraintes d'exploitation ferroviaire spécifiques semblent peser en Île-de-France du fait de l'ampleur

de l'offre et de la demande de déplacements en trains régionaux : d'après le CGDD¹³, les RER et Transilien ont transporté 19,5 milliards de voyageurs-km en 2017 pour la seule région Île-de-France, quand l'ensemble des TER de province (toutes régions confondues) en ont transporté 13,7 milliards. D'après les estimations de l'AQST à partir des données des mémentos statistiques de la SNCF et les données de l'ART sur les longueurs de lignes ferroviaires exploitées par ancienne région et par activité, l'intensité d'utilisation des lignes ferroviaires d'Île-de-France par les services Transilien était plus de 6,4 fois plus élevée en 2017 que l'intensité d'utilisation moyenne des lignes du RFN (hors LGV) par les services TER. Or les causes prédominantes de retards ne sont pas nécessairement les mêmes lorsque la ligne est exploitée à un niveau proche de sa limite de capacité.

De surcroît, le nombre moyen¹⁴ d'arrêts effectués à l'heure de pointe du matin par les RER et Transilien en 2019 était de 14,8 arrêts d'après les calculs de l'AQST, quand le nombre moyen d'arrêts effectués par les TER (sur l'ensemble de la journée) atteignait 11,7 arrêts pour la région Sud-PACA dans laquelle les TER s'arrêtent le plus (données 2018, ART). Les lignes franciliennes présentent également la particularité de toutes traverser Paris intra-muros, et cela principalement en souterrain, une autre spécificité de l'exploitation du réseau ferroviaire francilien.

Lorsqu'on applique le modèle « Transilien » aux données TER, l'âge moyen du matériel roulant en début d'année n'a pas d'effet sur le taux de retard statistiquement significatif au seuil de 10%.

En comparant les coefficients associés au nombre de circulations journalières lorsqu'on applique le modèle « Transilien » aux données TER et Transilien, on remarque que l'ajout de 100 circulations quotidiennes de plus a un impact relativement similaire pour les TER (+0,55 points de retard) et les RER et Transilien (+0,39 points).

En revanche, un allongement relatif de 20% de la distance parcourue par les trains régionaux entraînerait selon le modèle « Transilien » une dégradation plus forte de la ponctualité pour TER (+1,3 points de retard) que pour les RER et Transilien (+0,73 points). Ce résultat est l'opposé de celui obtenu pour cette même variable en appliquant le modèle « TER » successivement aux TER et aux RER et Transilien. Ce signe de non-robustesse confirme que l'utilisation d'un modèle spécifique par activité semble plus appropriée.

¹³ CGDD/SDES, Chiffres clés du transport – Édition 2019

¹⁴ Moyenne pondérée du nombre de circulations par ligne de RER ou Transilien