

# Prédire et expliquer les retards au décollage : Une étude de cas à l'aéroport international de Paris - Charles de Gaulle

Thibault Falque\*,\*\*, Bertrand Mazure\*\*, Karim Tabia\*\*

\*Exakis Nelite, Paris, France

\*\*CRIL, Université d'Artois & CNRS

[falque,mazure,tabia@cril.fr](mailto:falque,mazure,tabia@cril.fr)

<http://www.cril.univ-artois.fr>

**Résumé.** Dans le contexte très concurrentiel où évoluent les acteurs du transport aérien tels que les aéroports, l'optimisation des ressources est indispensable pour améliorer ses services et maîtriser ses coûts. Outre les problématiques générales telles que la maintenance prédictive et la prévision des flux de passagers, de nombreuses problématiques spécifiques peuvent bénéficier des avancées récentes de l'apprentissage automatique et d'une plus grande disponibilité des données. Cet article traite d'un problème réel et difficile: la prédiction des retards au décollage des vols. Nous étudions le cas de l'aéroport international Paris Charles de Gaulle (CDG) en partant des spécificités de ce problème à CDG jusqu'à la proposition d'une modélisation puis d'une solution et l'analyse des résultats sur des données réelles de l'aéroport. Ces travaux sont en cours d'intégration dans le système d'information d'Aéroports de Paris.

## 1 Introduction

Dans le contexte d'un aéroport, il existe plusieurs problèmes typiques de l'intelligence artificielle, comme la planification, l'optimisation, la simulation et la prédiction. En effet, il existe de nombreux problèmes de transport aérien où les systèmes d'aide à la décision sont utilisés tout en intégrant des composants d'intelligence artificielle.

Avant la crise sanitaire du COVID-19, les prévisions de l'Association du Transport Aérien International (IATA) montrent que le nombre de passagers va doubler d'ici 2036, pour atteindre les 7,8 milliards de passagers. Certes, la pandémie de COVID-19 a considérablement ralenti le trafic aérien, surtout en 2020 et début 2021, mais la pression concurrentielle est toujours d'actualité même en période d'activité réduite. Ces derniers mois, le trafic aérien a repris dans plusieurs régions du monde. En France, l'aéroport Paris-Charles de Gaulle (Paris-CDG), principal hub d'Air France, verra son nombre de passagers augmenter de 35 à 40 millions pour atteindre 100 millions en 2036. Cela va nécessiter environ 400 mouvements supplémentaires d'appareils (décollages ou atterrissages) par jour. Face au contexte très concurrentiel du domaine aérien, le projet de développement de l'aéroport Paris-CDG ne prévoit pas d'extension ou de nouvelle piste. Il devient donc primordial d'améliorer :

- les flux des passagers en proposant un parcours simplifié, fluide et personnalisé ;

- la ponctualité des vols en anticipant dans la mesure du possible les retards ;
- les mouvements d'avions avec une planification optimisée et adaptative des ressources comme les parkings d'avions, les banques d'enregistrement, les jetées à bagages, etc.

Le problème des retards des vols par exemple (retards à l'atterrissage et au décollage) n'a pas seulement des conséquences financières immédiates. Ces retards peuvent aussi provoquer, par effet de chaîne, d'autres retards et problèmes qui influent sur les retards et les classements des compagnies aériennes et des aéroports.

Dans (Xu et al., 2008), les auteurs soulignent qu'environ 84% des retards sont générés par les aéroports. Le problème de la prédiction des retards est étudié dans quelques travaux, mais il concerne principalement les retards à l'atterrissage ou au roulage. Le retard au départ du parking (délai entre l'heure prévue et l'heure réelle à laquelle un avion quitte son poste de stationnement ou sa porte) dépend de plusieurs facteurs, dont beaucoup sont spécifiques à chaque aéroport (tels que l'importance du trafic et des ressources de l'aéroport, les processus passagers, les conditions météorologiques, le contrôle du trafic aérien, etc.) À Paris-CDG, il y a un décollage presque toutes les minutes, et le moindre retard peut avoir des conséquences en cascade sur le programme de décollage, qui peuvent prendre plusieurs heures pour être résorbées et revenir à une situation normale. Prévoir et prédire les retards en temps réel permettra d'anticiper leur gestion en prévoyant des plans de gestion des retards et des ajustements aux horaires établis tels que l'horaire des passerelles, des comptoirs d'enregistrement et des jetées à bagages.

L'une des spécificités de notre travail est que nous nous sommes concentrés sur des données très fluctuantes dues à la pandémie COVID-19. Les principales contributions de l'article sont l'étude de la situation existante à Paris-CDG, des besoins et l'identification ensuite de trois tâches principales : la prédiction en temps réel des retards parking, la prévision et l'explicitabilité des prédictions. Concernant la modélisation, nous avons identifié cinq catégories de données pour notre problème : données de vol, données sur l'avancement des processus passagers (sécurité, embarquement, etc.), données météorologiques, données sur la situation des retards actuels, etc. Enfin, nous avons réalisé une étude empirique pour effectuer la sélection des données, des caractéristiques et des modèles et nous fournissons un aperçu des résultats obtenus.

## 2 Retard avions : état de l'art

Le problème des retards avions est un problème assez connu. Différents modèles parmi les forêts aléatoires, les machines à vecteurs de support et la régression logistique ont été étudiés (Natarajan et al., 2018) pour prédire si un vol sera retardé ou non. Un modèle de régression logistique pour prédire une classe de vols départ est proposé dans (Nigam et Govinda, 2017). Dans (Venkatesh et al., 2017), les auteurs étudient les retards à l'arrivée et proposent différentes approches pour prédire si un vol spécifique sera retardé ou non. Dans (Ibrahim et al., 2021), les auteurs ont comparé différentes approches d'apprentissage automatique (forêt aléatoire, régression logistique, classifieur naïf bayésien et arbres de décision) pour la prédiction des retards, à l'arrivée. Dans (Tang, 2021), une comparaison de sept modèles binaires est réalisée. Dans (Yi et al., 2021), les auteurs ont proposé plusieurs approches empilées pour l'ensemble de données concernant les vols de l'aéroport international Boston Logan de janvier à décembre 2019. Nous trouvons également différentes études sur l'impact de différents facteurs

sur les retards des vols. Par exemple, les auteurs de (Wang et al., 2003) ont étudié l'impact des connexions de vol sur le retard. Dans (Markovic et al., 2008), une étude statistique sur l'impact de la météo à l'aéroport de Francfort est proposée. Dans (Yogita Borse et al., 2020), les auteurs se sont concentrés sur les données météorologiques comme caractéristique principale pour prédire la classe de retard. Dans (Esmailzadeh et Mokhtarimousavi, 2020), un modèle de machine à vecteur de support (SVM) est utilisé. Basée sur 20 jours, cette dernière étude examine certaines causes des retards du trafic aérien dans les trois principaux aéroports de la ville de New York. Dans l'étude (Cai et al., 2021), une approche d'apprentissage profond pour la prédiction des retards de vol en tenant compte d'un scénario multi-aéroports est proposée. En ce qui concerne les approches basées sur la régression, les auteurs (Rebollo et Balakrishnan, 2014) ont proposé des approches basées sur la classification et la régression avec des forêts aléatoires pour les aéroports américains. À notre connaissance, il n'existe qu'un seul travail visant à prédire le retard au décollage, mais il ne tente de prédire qu'une heure avant l'heure de décollage estimée, et il est destiné au Centre de contrôle de l'espace aérien supérieur de Maastricht (Dalmau-Codina et al., 2019).

### 3 Aéroport Paris-CDG

Nous donnons dans cette section des informations factuelles sur l'aéroport de Paris-CDG. L'aéroport de Paris-CDG est le plus important aéroport de France. Il a été ouvert en 1974 pour faire face à la saturation de l'aéroport de Paris Orly (le principal aéroport parisien avant l'ouverture de Paris-CDG). Il est situé au nord de Paris, et c'est le hub d'Air France. Cette compagnie représente 50% du trafic de Paris-CDG. Trois terminaux principaux numérotés de 1 à 3 composent Paris-CDG.

En ce qui concerne le trafic à l'aéroport de Paris-CDG, il y a plus d'un départ de vol par minute. Il y a environ 720000 vols par an, soit environ 2000 vols par jour. En moyenne, il y a 145 passagers par vol. À Paris-CDG, les ressources sont actuellement planifiées à l'aide de solutions alimentées par des solveurs de contraintes. Parmi les ressources critiques, on trouve les parkings (ou « stands ») affectés aux vols. Lorsqu'un « stand » n'est pas libéré à temps, cela peut entraîner des changements de planification complexes et provoquer des retards en cascade. Il est donc important d'anticiper et de prévoir ces retards le plus précisément possible, de les expliquer et de proposer des actions pour les limiter. Avant de présenter le problème, nous allons introduire d'abord quelques termes. Nous appelons *rotation* l'ensemble composé d'un vol arrivée et d'un vol départ. En général, cet ensemble est composé de deux vols, mais il peut n'y en avoir qu'un seul, auquel cas le vol commence une nouvelle rotation. Par conséquent, le temps de la rotation est le temps entre l'arrivée d'un avion (atterrissage) et son départ (décollage). Un vol a une heure de départ prévue, appelée SOBT (Scheduled Off-Block Time), à laquelle il est censé quitter son parking. Le moment où le vol quitte effectivement son poste de stationnement est appelé AOBT. Le retard est alors la durée entre l'AOBT et la SOBT.

#### 3.1 Jalons avant le départ parking de l'avion

Nous présentons ici les principaux jalons précédant le « pushback » d'un avion. Ces jalons sont présentés dans la Fig. 1.

## Prédire et expliquer les retards au décollage: Une étude de cas à l'aéroport Paris - CDG

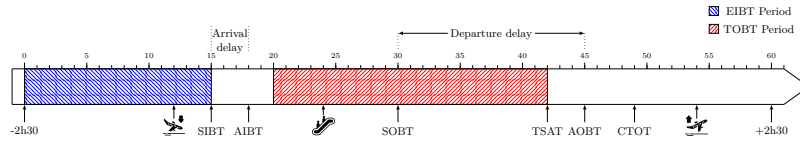


FIG. 1 – Jalons d'un vol à Paris-CDG

Avant l'arrivée du vol à l'aéroport, le vol estime son heure d'arrivée au parking en envoyant des EIBT (Estimated In-Block Time). L'AIBT (Actual In-Block Time) correspond à l'heure réelle où le vol arrive à son stand. Nous appelons « retard à l'arrivée » la différence entre l'AIBT et la SIBT. L'embarquement commence entre 1 heure et 30 minutes avant la SOBT. Pendant la période de rotation, la compagnie aérienne envoie une estimation de l'heure à laquelle l'avion sera prêt au système de gestion : TOBT (Target Off-Block Time). En cas de trafic dense ou saturé dans le ciel ou d'encombrement sur la piste, le contrôle aérien peut « créneauter » un vol, c'est-à-dire forcer son décollage entre une heure calculée appelée CTOT (Calculated Take-Off Time) et CTOT plus 15 minutes. Si l'avion ne décolle pas pendant cette période, il peut être à nouveau « créneauté ». Le dernier jalon est la TSAT qui est l'heure fournie par le contrôle aérien qui tient compte de la TOBT, du CTOT et/ou de la situation du trafic et est l'heure à laquelle un avion peut espérer obtenir l'approbation de son démarrage ou de son « pushback ».

### 3.2 Les retards avion à Paris-CDG

Paris-CDG est classé<sup>1</sup> à la 10e place en 2018 en termes de ponctualité. Environ la moitié des vols arrivent à l'heure, mais seulement 20% décollent à l'heure. Une étude des retards à Paris-CDG a mis en évidence différentes causes (ex. congestion, pannes, incidents à l'aéroport, parcours passagers, etc.) de ces retards à différentes phases (stationnement/poussée, taxis, etc.). La Figure. 2 donne une vue d'ensemble des retards sur l'année considérée dans notre étude (août 2021 - septembre 2022). Il convient de noter qu'au cours de cette période, certains terminaux ont été fermés, tandis que d'autres ont réouvert. Sur cette période, la proportion de vols présentant des retards est de 82%. La Figure. 2a montre le nombre de vols retardés par terminal. On peut observer que les deux terminaux correspondant à Air France (2E et 2F) ont le plus de vols retardés. La Figure. 2b montre les retards cumulés (en minutes) pour chaque journée de la période considérée. On peut observer une augmentation durant l'été 2022, qui correspond à une reprise du trafic et à la réouverture de certains terminaux mais également aux jours de panne des trieurs bagages ayant provoqué de nombreux retards. La Figure. 2c représente la somme et la moyenne des retards pour chaque tranche horaire de la journée, allant de  $p1$  (6h-8h) à  $p6$  (20 h - 23 h). Les périodes du matin, en particulier  $p2$  (9h - 11h) et  $p3$  (12h - 14h), cumulent la plupart des retards. Comme la majorité des retards se produisent le matin et qu'il existe un effet de retards en cascade, il est crucial de prévoir et de gérer avec précision ces retards dans ces tranches horaires.

1. selon OAG Flightview

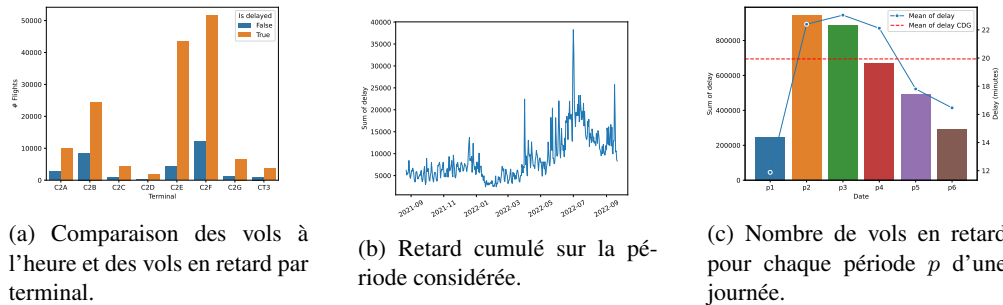


FIG. 2 – Aperçu des retards « off-block » à Paris-CDG de Août 2021 à Septembre 2022

## 4 Définition du problème et des objectifs

Rappelons que nous appelons retard au décollage, la durée séparant l'heure où l'avion quitte son poste de stationnement de l'heure de départ parking programmé. Soit  $Y$  la variable cible à prédire pour un échantillon d'entrée décrivant le vol étudié. Nous distinguons deux tâches de régression : la prédiction en temps réel et la prévision.

### 4.1 Prédiction en temps réel du retard parking

Le problème considéré ici est celui de la prédiction, à tout moment  $t$ , du retard parking  $Y$  (exprimé en minutes) que le vol aura effectivement. Ces prédictions sont mises à jour toutes les 5 minutes jusqu'à ce que le vol quitte son parking. En effet, pour chaque slot (chaque slot dure 5 minutes), nous pouvons acquérir de nouvelles données pertinentes à partir du système d'information opérationnel de Paris-CDG qui peuvent être utilisées pour mettre à jour les prédictions (ceci est valable pour les variables dynamiques telles que les conditions météorologiques, l'avancement des processus passagers, etc.) Les prédictions de retard parking en temps réel sont destinées principalement à : attirer l'attention des managers en temps réel sur les vols susceptibles d'avoir des retards importants et pouvant avoir des conséquences en cascade ; expliquer et identifier les causes et mettre en place des actions si nécessaire pour résoudre la situation.

### 4.2 Prévision du retard

On appelle prévisions la prédiction des retards avant l'ouverture des vols. Cela peut être quelques heures ou plusieurs jours avant le vol. Évidemment, dans notre cas, la prévision ne peut pas s'appuyer sur des informations dynamiques sur l'évolution des processus des passagers, les conditions météorologiques, etc. De même, de telles prévisions n'ont pas besoin d'être mises à jour puisque les données sont disponibles à l'avance. De telles prévisions peuvent servir à : établir à l'avance plusieurs plans et mesures d'atténuation en fonction des retards prévus. identifier les causes et anticiper l'effet d'enchaînement et ses conséquences ; utiliser les prévisions pour faire des simulations plausibles sur la congestion, les files d'attente en fonction des prévisions de retard, puis envisager des solutions et des plans de gestion.

## 5 Définition du modèle

Dans cette section, nous présentons et motivons les informations essentielles actuellement disponibles dans le système d'information opérationnel de Paris-CDG et qui sont susceptibles d'être pertinentes pour la prédiction des retards parkings. Les variables proposées sont issues de l'analyse d'un rapport récent sur les retards au décollage à Paris-CDG et de l'analyse d'une année entière de données réelles. Nous avons finalement établi 5 catégories de variables, présentées dans les sections suivantes.

### 5.1 Caractéristiques de base d'un vol (Basic Flight Features)

Ce sont les caractéristiques de base d'un vol et elles ne changent pas dans le temps (ces caractéristiques sont énumérées dans le Tableau 1). Par exemple, le nom de la compagnie aérienne qui effectue le vol, le type d'avion, le code IATA de la destination du vol, le terminal, le type douanier (nationale, Schengen ou international) et la saison IATA (été ou hiver).

Variable	Description	Type	Exemple
Compagnie	Code unique de la compagnie	Catégorique	AF
Type avion	Code du type avion	Catégorique	77W
Destination	Code IATA de l'aéroport de destination	Catégorique	JFK
Terminal	Code du terminal de CDG	Catégorique	C2E
Douane	Critère douanier	Catégorique	Schengen
Saison	Le code de la saison IATA	Catégorique	W
Semaine	Un index de semaine depuis une date de référence	Numérique	1000
Jour	Jour de la semaine (1-7)	Numérique	1
Bus	Vrai si l'avion utilise un bus pour son accès.	Booléen	True
Parking	Vrai si le parking d'arrivée et le parking de départ sont les mêmes.	Booléen	False
SOBT	La SOBT exprimée en minutes depuis minuit.	Timestamp	360
Rotation	Durée de la rotation (différence entre SOBT et SIBT)	Numérique	300
Pax Count	Nombre de passagers (estimé)	Numérique	140
Total Pif Passenger	Nombre de passagers devant passer par le PIF	Numérique	75
Service Type	Catégorie du transport	Catégorique	J

TAB. 1 – Variables basiques du vol (BCF)

### 5.2 Caractéristiques sur les jalons du vol (Off-block Milestone Features)

Ces jalons sont établis par les différents acteurs tels que le contrôle aérien, la compagnie aérienne et l'aéroport. Pour cette étude, nous nous sommes concentrés sur les jalons autour de la SOBT (-2h30 à +2h30, voir Figure. 1). Cette période est divisée en 60 créneaux de 5 minutes. Chaque nouveau jalon (parmi TOBT, CTOT, TSAT) possède un horodatage. Ces caractéristiques sont listées dans le Tableau 2.

### 5.3 Variables concernant les retards précédents (PCFDF)

Pour chaque créneau de chaque vol, nous calculons la proportion de vols en retard et la durée moyenne de ces retards. À un instant  $t$ , ces caractéristiques sont calculées pendant une fenêtre temporelle  $w$  qui peut aller de quelques minutes à quelques heures. L'ensemble de ces nouvelles variables est présenté dans le Tableau 3. Cet ensemble est nommé dans la suite *PCFDF* (Previous and current flights delay features).

Variable	Description	Type
Retard à l'arrivée	Temps en minutes entre la dernière EIBT et la SIBT.	Numérique
$TOBT_{diff}$	Temps en minutes entre la dernière TOBT et la SOBT.	Numérique
$TOBT_{count}$	Le nombre de TOBT	Numérique
$CTOT_{diff}$	Temps en minutes entre la dernière CTOT et la SOBT.	Numérique
$TSAT_{diff}$	Temps en minutes entre la dernière TSAT et la SOBT.	Numérique

TAB. 2 – Jalons du vol (OMF)

Variable	Description	Type
Retard aéroport	Retard moyen à l'aéroport pendant les $w$ dernières minutes (indépendamment du terminal)	Numérique
Retard terminal	Retard moyen dans le terminal du vol considéré pendant les $w$ dernières minutes	Numérique
Retard compagnie	Retard moyen de la compagnie du vol considéré pendant les $w$ dernières minutes	Numérique
Pourcentage des vols en retard sur l'aéroport	Pourcentage des vols en retard à l'aéroport pendant les $w$ dernières minutes (indépendamment du terminal)	Numérique
Pourcentage des vols en retard sur le terminal	Pourcentage des vols en retard dans le terminal du vol considéré pendant les $w$ dernières minutes	Numérique
Pourcentage des vols en retard de la même compagnie	Pourcentage des vols en retard de la même compagnie du vol considéré pendant les $w$ dernières minutes	Numérique

TAB. 3 – Variables des retards des vols précédents (PCFDF)

#### 5.4 Caractéristiques Météo (Weather Condition Features)

Certaines conditions météorologiques telles qu'une faible visibilité et des vents forts sont connues pour être des facteurs susceptibles de provoquer des retards au décollage, et donc de retarder le départ du vol de son parking. Les variables météo considérées dans cette étude sont résumées dans le Tableau 4.

Variable	Description	Type
Procédures de visibilité faible	Vrai si la procédure de visibilité faible est enclenchée, faux sinon	Booléen
Taux d'humidité (en pourcent)		Numérique
Vitesse du vent (en m/sec)		Numérique
Pression de l'air (en hectoPascal)		Numérique
Température (en degrés Celsius)		Numérique

TAB. 4 – Variables conditions météo (WCF)

## 5.5 Caractéristiques des flux passagers (Passenger Flow Features)

Ces variables permettent d'obtenir à tout moment des informations sur l'avancement de certains processus passagers, qui peuvent causer un retard au départ du parking. En particulier, l'information pertinente est le pourcentage au créneau  $t$  de passagers qui ont embarqué ou qui ont déjà passé les contrôles de sécurité. Ces caractéristiques sont utilisées uniquement pour prédire et mettre à jour en temps réel (pour chaque créneau d'un vol) la prédiction du retard au départ du parking.

# 6 Extraction, prétraitement et sélection de données

## 6.1 Extraction et prétraitement des données

Le système d'information opérationnel de Paris-CDG (appelé AOP pour Airport Operation Plan) collecte une grande quantité d'informations pour chaque vol et sa progression. Pour nos tâches de prédiction, à chaque créneau horaire, une nouvelle entrée de vol est créée avec un horodatage associé et des données mises à jour. Il est donc possible de retracer le statut d'un vol jusqu'à son départ. Par conséquent, pour les caractéristiques statiques, nous les extrayons une seule fois. Pour les caractéristiques dynamiques, telles que les retards des autres vols, il s'agit de variables calculées que nous effectuons avec des requêtes sur les vols passés. Par exemple, pour calculer la proportion de vols qui ont été retardés au cours des dernières  $w$  minutes, il faut passer en revue tous les vols impliqués dans la fenêtre de temps  $w$ . Les caractéristiques de retard des vols précédents et actuels (PCFDF) sont calculées après extraction avec différentes fenêtres  $w$ . Dans notre étude, nous avons considéré les données d'une année (août 2021-septembre 2022) et construit un ensemble de données comprenant 10633920 lignes et 31 colonnes (chaque vol est répété 60 fois avec des valeurs dynamiques pour chaque slot).

## 6.2 Sélection de variables

Une fois notre ensemble de données extrait et prétraité, nous avons procédé à la sélection des variables afin de confirmer nos intuitions et d'éliminer les attributs qui s'avèreraient non pertinents pour nos tâches de prédiction. Nous avons d'abord effectué une analyse de corrélation simple entre chaque caractéristique et la variable cible (le retard au départ du parking). La Figure. 3 présente les résultats du coefficient de corrélation de *Pearson*.

Nous pouvons voir sur la Figure. 3 que les variables les plus pertinentes au slot 0 sont la différence entre la SOBT et la TSAT avec un score *Pearson* de 0,4, la différence entre la TOBT et la SOBT avec un score *Pearson* de 0,33. En revanche, les variables représentant la durée de la rotation (*Rotation*) et le retard d'un vol à l'arrivée (*Arrival delay*) semblent peu pertinentes. Ces variables ont un score *Pearson* négatif de  $-0,66$  et  $-0,29$  respectivement. Pour le créneau 30, l'ordre des variables essentielles se confirme. Les durées entre la SOBT et la TSAT ou la TOBT sont les variables essentielles avec un score *Pearson* de 0,74 et 0,68.

L'importance de la variable TSAT s'explique par le fait qu'il s'agit de l'une des dernières informations obtenues pour un vol avant qu'il ne quitte son parking et que le départ a lieu le plus souvent à l'heure indiquée par la TSAT. La variable *Percent Flight Airport*, représentant la proportion de vols retardés sur l'ensemble de l'aéroport (pour le calcul des valeurs nous avons utilisé la fenêtre de temps de 10 minutes), devient plus importante avec une progression de



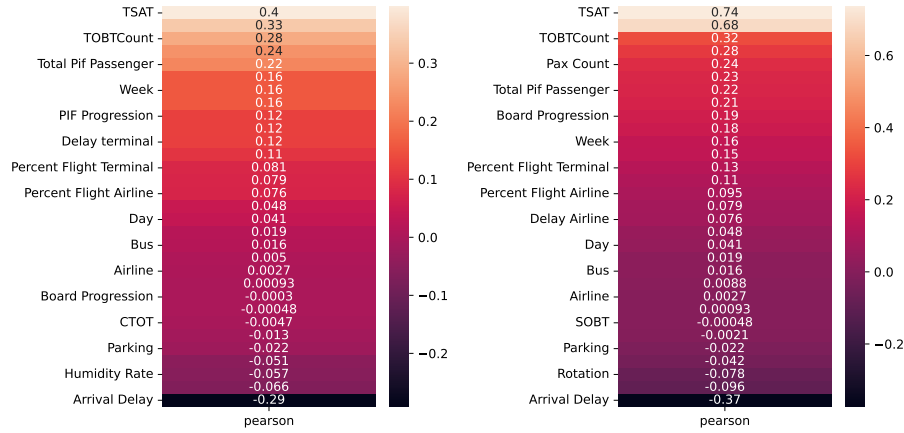


FIG. 3 – Résultats du coefficient de corrélation de *Pearson* aux slots 0 (figure de gauche) et aux slots 30 (figure de droite)

son score entre le slot 0 et le slot 30 passant de 0,16 à 0,23. Enfin, nous pouvons noter que les variables dynamiques (en particulier, la variable concernant la progression de l'embarquement) ont un score en progression. Cette progression montre l'importance de l'utilisation des données de vols dynamiques pour les prédictions en temps réel.

Afin de valider les conclusions sur les scores de corrélation obtenus, nous avons effectué une autre analyse empirique en faisant varier l'ensemble des variables utilisées pour la prédiction des retards. Nous avons également fait varier la fenêtre  $w$  utilisée pour calculer les variables dynamiques à partir de l'historique proche et la quantité d'historique à utiliser (douze mois, neuf mois, six mois ou trois mois). Nous avons utilisé un modèle d'ensemble d'arbres de régression boosté appelé *LightGBM*<sup>2</sup> (Ke et al., 2017). Le principe de ce type de modèle est de construire une séquence d'arbres où chaque arbre corrige le précédent. Le résultat final est la somme des résultats de chaque arbre de la séquence. Dans la suite, nous désignons par  $\mathcal{D}_w^m$  le jeu de données avec  $m$  le nombre de mois utilisés ( $m \in \{3, 6, 9, 12\}$ ) et  $w$  la durée de la fenêtre utilisée pour le calcul des variables dynamiques ( $w \in \{10, 30, 60\}$ ).  $\mathcal{V}$  désigne l'ensemble des variables utilisées pour le jeu de données ( $\mathcal{V} \in \{\{BFF\}, \{BFF, WCF, OMF, PFF\}, \{BFF, PCFDF, WCF, OMF, PFF\}\}$ ).

Pour notre étude, nous avons testé notre configuration sur 40 jours (du 12 août au 21 septembre 2022). Pour chaque journée, nous avons entraîné le modèle jusqu'à la veille du jour du test et l'avons évalué sur le jour du test. Le Tableau 5 montre les configurations optimales. Les erreurs et le score R2 présentés dans ce tableau sont pour les 40 jours testés. Les hyperparamètres sont notés *#Tree/#Leaves/LearningRate*.

Le Tableau 5<sup>3</sup> présente les résultats pour les différentes configurations. L'utilisation des variables dynamiques apporte un véritable gain permettant de réduire de moitié l'erreur et d'améliorer sensiblement le score R2. Ainsi la meilleure configuration consiste à utiliser l'intégralité de l'historique (12 mois), avec une fenêtre de temps de 60 minutes.

2. <https://lightgbm.readthedocs.io/>

3. L'intégralité des résultats est disponible en annexe dans un fichier CSV.

Dataset	Variables ( $\mathcal{V}$ )	Hyperparamètres	MAE	RMSE	R2
$\mathcal{D}_{60}^{12}$	$BFF, PCFDF, WCF, OMF, PFF$	75/256/0.05	9.645	13.858	0.731
$\mathcal{D}_{10}^{12}$	$BFF, WCF, OMF, PFF$	75/256/0.05	9.709	14.009	0.725
$\mathcal{D}_{60}^{12}$	$BFF, PCFDF, WCF, OMF, PFF$	32/64/0.05	10.376	14.500	0.705
$\mathcal{D}_{12}^{12}$	$BFF, PCFDF, WCF, OMF, PFF$	32/128/0.05	10.473	14.584	0.702
$\mathcal{D}_{60}^3$	$BFF, PCFDF, WCF, OMF, PFF$	32/128/0.05	11.637	15.414	0.667
$\mathcal{D}_{10}^{12}$	$BFF, WCF, OMF, PFF$	32/64/0.05	13.922	20.812	0.393
$\mathcal{D}_{60}^9$	$BFF$	75/256/0.05	17.443	26.576	0.0108

TAB. 5 – Résultats de la sélection des caractéristiques et des données historiques.

## 7 Prédiction en temps réel

Date	MAE	Baseline model			
2022-08-13	8.196	18.814	2022-09-10	7.429709	16.339208
2022-08-14	8.490	18.176	2022-09-11	8.311289	19.677116
2022-08-15	8.026	16.877	2022-09-12	7.088036	16.854864
2022-08-16	9.911	20.869	2022-09-13	7.252672	15.125720
2022-08-17	8.288	18.728	2022-09-14	8.613775	17.317014
2022-09-05	9.340	19.727	2022-09-15	7.553985	17.969956
			2022-09-16	31.250517	65.622123

TAB. 6 – MAE pour la prédiction en temps réel.

Le Tableau 6 présente l'erreur moyenne absolue (MAE) pour un sous-ensemble des journées (pour des raisons d'espace) testées avec le meilleur modèle sélectionné dans la section précédente. Nous remarquons que le modèle proposé est plus précis et plus stable en terme d'erreur que le modèle de référence (e.g un modèle prédisant tout le temps la moyenne du retard à CDG). Cependant, nous pouvons également remarquer que le 16 septembre l'erreur est très importante.<sup>4</sup> Les retards pouvaient aller jusqu'à 4h30 (la moyenne du retard sur la journée était de 1h20). De plus comme la période couverte pour chaque vol est comprise entre -2h30 et +2h30, nous ne collectons pas toutes les données sur ces « retards longs ».

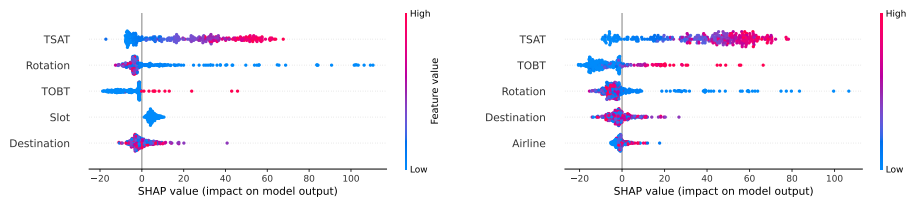


FIG. 4 – Valeurs de SHAP pour la journée du 16 septembre 2022 aux slots 0 (figure de gauche) et slots 30 (figure de droite)

La Figure. 4 donne un aperçu des caractéristiques les plus importantes pour notre modèle (pour des raisons d'espace, uniquement les 5 plus importantes<sup>5</sup>). Les caractéristiques sont

4. Le jeudi 16 septembre, les retards sont dus à une grève des contrôleurs aériens.
5. L'intégralité des figures sont disponibles en annexes.

triées par la somme des valeurs SHAP sur tous les échantillons au slot 0 et au slot 30. La couleur représente la valeur de la caractéristique (*rouge* correspond à *high*, *bleu* à *low*). Cela révèle par exemple que lorsque le slot est à 30, une valeur élevée pour la TSAT augmente le retard prédit. Enfin, les valeurs de SHAP confirment largement les coefficients de *Pearson* de la section 6.2 et l'importance de l'impact des variables TSAT et TOBT.

## 8 Conclusions

La ponctualité est une question sensible dans les grands aéroports et les plates-formes de correspondance pour l'expérience passagers. Dans cet article, nous avons abordé le problème de la prédiction des retards au départ des parkings à l'aéroport de Paris-CDG, l'un des plus importants aéroports du monde et la plaque tournante de la compagnie aérienne Air France. Notre étude a commencé par l'analyse du problème (son ampleur, sa forme, ses causes, etc.) et des besoins (prévisions et prédiction en temps réel) à Paris-CDG. À partir de cette analyse, nous avons proposé deux types de catégories de caractéristiques susceptibles d'être utiles pour la prédiction des retards : les caractéristiques statiques utilisées pour la tâche de prévision et les caractéristiques dynamiques (qui peuvent être mises à jour en temps réel) utilisées pour la prédiction en temps réel des retards. L'étape suivante a consisté à construire un pipeline pour extraire les données brutes dont nous avons besoin du système d'information opérationnel de Paris-CDG. Ceci nous a permis de construire un jeu de données représentant une année d'activité réelle. Nous avons réalisé une étude empirique pour effectuer la sélection des caractéristiques et des données, puis la sélection des modèles. Une des spécificités de notre travail est que nous avons travaillé avec des données très fluctuantes en raison de la pandémie de COVID-19 et de ses conséquences en termes de restrictions des voyages aériens à plusieurs reprises, ainsi que d'autres aléas du trafic aérien. Les résultats obtenus montrent clairement qu'il est possible de prévoir certains retards bien mieux que le modèle de référence. Ce résultat peut être grandement amélioré en explorant systématiquement d'autres modèles et leurs meilleurs hyperparamètres. L'un des éléments cruciaux pour notre application, en plus de l'amélioration de la précision, sera l'explicabilité des prédictions, en particulier l'identification d'explications qui peuvent aider à la gestion des retards.

## Références

- Cai, K., Y. Li, Y.-P. Fang, et Y. Zhu (2021). A Deep Learning Approach for Flight Delay Prediction Through Time-Evolving Graphs. *IEEE Trans. Intell. Transport. Syst.*, 1–11.
- Dalmau-Codina, R., F. Ballerini, H. Naessens, S. Belkoura, et S. Wangnick (2019). Improving the predictability of take-off times with machine learning a case study for the maastricht upper area control centre area of responsibility.
- Esmailzadeh, E. et S. Mokhtarimousavi (2020). Machine Learning Approach for Flight Departure Delay Prediction and Analysis.
- Ibrahem, A., H. Elbeh, et H. M. Mousa (2021). A Comparative Analysis of Models for Predicting Airline Arrival Delays. pp. 5.

- Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, et T.-Y. Liu (2017). LightGBM : a highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, Long Beach, California, USA, pp. 3149–3157. Curran Associates Inc.
- Markovic, D., T. Hauf, P. Röhner, et U. Spehr (2008). A statistical study of the weather impact on punctuality at Frankfurt Airport. *Meteorological Applications* 15(2), 293–303.
- Natarajan, V., S. Meenakshisundaram, G. Balasubramanian, et S. Sinha (2018). A Novel Approach : Airline Delay Prediction Using Machine Learning. In *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 1081–1086.
- Nigam, R. et K. Govinda (2017). Cloud based flight delay prediction using logistic regression. In *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, pp. 662–667.
- Rebollo, J. J. et H. Balakrishnan (2014). Characterization and prediction of air traffic delays. *Transportation Research Part C : Emerging Technologies* 44, 231–241.
- Tang, Y. (2021). Airline Flight Delay Prediction Using Machine Learning Models. In *2021 5th International Conference on E-Business and Internet*, Singapore Singapore, pp. 151–154. ACM.
- Venkatesh, V., A. Arya, P. Agarwal, S. Lakshmi, et S. Balana (2017). Iterative machine and deep learning approach for aviation delay prediction. *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)*.
- Wang, Schaefer, et Wojcik (2003). Flight connections and their impacts on delay propagation. In *22nd Digital Avionics Systems Conference Proceedings (Cat No 03CH37449) DASC-03*, Indianapolis, IN, USA, pp. 5.B.4–5.1. IEEE.
- Xu, N., L. Sherry, et K. B. Laskey (2008). Multifactor Model for Predicting Delays at U.S. Airports. *Transportation Research Record* 2052(1), 62–71.
- Yi, J., H. Zhang, H. Liu, G. Zhong, et G. Li (2021). Flight Delay Classification Prediction Based on Stacking Algorithm. *Journal of Advanced Transportation* 2021, 1–10.
- Yogita Borse, Dhruvin Jain, Shreyash Sharma, Viral Vora, Aakash Zaveri, et K J Somaiya College of Engineering (2020). Flight Delay Prediction System. *IJERT V9(03)*, IJERTV9IS030148.

## Summary

Punctuality is a sensitive issue in large airports and hubs for passenger experience and for controlling operational costs. This paper presents a real and challenging problem of predicting and explaining flight off-block delays. We study the case of the international airport Paris Charles de Gaulle (Paris-CDG) starting from the specificities of this problem at Paris-CDG until the proposal of modelings then solutions and the analysis of the results on real data covering an entire year of activity. The proof of concept provided in this paper allows us to believe that the proposed approach could help improving the management of delays and reduce the impact of the resulting consequences.