

THALES



THALES AIR SYSTEMS SA
Parc Tertiaire SILIC
3 avenue Charles Lindbergh
94628 Rungis CEDEX

Université de Technologie de
Compiègne
Rue du Dr Schweitzer
60200 Compiègne

Stage Ingénieur - TN10 - Génie Informatique

Étude de la prédictibilité des points caractéristiques d'un vol

04 février 2013 - 31 juillet 2013

GAUMONT Noé
noe.gaumont@gmail.com

Maître de stage : M. Areski HADJAZ
Suiveur UTC : M. Aziz MOUKRIM

Remerciements

Je tiens à remercier tout particulièrement les personnes suivantes pour tout ce qu'elles m'ont apporté durant mes 25 semaines de stage à Thales Air System à Rungis

Tout d'abord, je souhaite remercier mon tuteur, Areski Hadjaz, pour m'avoir offert la possibilité de découvrir un laboratoire d'innovation et le domaine de la gestion de trafic aérien, m'avoir suivi et conseillé durant tout mon stage. Grâce à lui, j'ai eu la chance d'évoluer dans une petite équipe tournée vers l'innovation et le développement de solutions innovantes.

Mes remerciements vont également à Gaétan Marceau-Caron pour toute l'aide qu'il m'a apportée au long de ce stage. Il m'a notamment permis de découvrir le domaine du *machine learning*. Plus généralement, je remercie également l'ensemble de l'équipe de l'Innovation Lab de Thales et toutes les personnes qui m'ont accueilli durant mon stage.

Table des matières

Acronyme	4
1 Introduction	5
2 Présentation de Thales air systems	6
2.1 L'innovation Lab	7
3 Présentation du sujet et du contexte	8
3.1 Données utilisées	8
3.2 Évaluation d'une méthode	9
4 Études de méthodes de prédiction	10
4.1 Méthodes existantes	10
4.1.1 Prédictions de Ground TP	10
4.1.2 Méthodes utilisant le modèle BADA	12
4.2 Méthodes développées	14
4.2.1 Prédiction par extrapolation	14
4.2.2 Modèle semblable à la BADA	16
4.3 Autres pistes explorées	22
4.3.1 Corrélation de l'erreur avec la position géographique	22
4.3.2 Correction des prédictions existantes	22
4.4 Conclusion technique	23
5 Organisation de mon stage	25
5.1 Déroulement chronologique	25
5.2 Temps forts du stage	26
6 Conclusion	28
Bibliographie	29
Table des figures	30
A Annexes	31

Acronyme

ADEP Aéroport de Départ.

ADES Aéroport d'Arrivée.

ADS-B Automatic Dependent Surveillance-Broadcast.

ATC Air Traffic Contrôle.

ATM Air Traffic Management.

BADA Base of Aircraft Data. Fichier contenant des caractéristiques techniques d'un avion (vitesse, taux de montée...) en fonction d'un niveau de vol.

ETA Estimated Time of Arrival. Temps d'arrivée de l'avion sur la piste d'atterrissage.

FL Flight Level. Unité de mesure utilisée pour le niveau de vol d'un avion.
410 FL = 41000 pieds = 12496.8 mètres.

RFL Requested Flight Level.

ROC Rate Of Climb. Taux de montée de l'avion. Il est données en feet/s.
Si il est négatif, c'est équivalent à un Rate of Descent.

SESAR Single European Sky ATM Research.

TAS True Air Speed. Vitesse réelle de l'avion par rapport à l'air.

TOC Top Of Climb. Point d'inflexion de la trajectoire en fin de phase de montée.

TOD Top Of Descent. Point d'inflexion de la trajectoire avant la phase de descente.

TP Trajectory Prediction.

UTC Université de Technologie de Compiègne.

Introduction

La gestion de trafic aérien est une tâche critique qui regroupe beaucoup d'acteurs à différents niveaux. Les centres de contrôle *en route* sont responsables de grandes zones de trafic, généralement à haute altitude, comportant des appareils se déplaçant d'un aéroport à un autre. Les centres de contrôle d'*approche* sont dédiés à un aéroport ou à quelques aéroports proches les uns des autres. Ils contrôlent les séquences de décollage et d'atterrissage. Les centres de contrôle *locaux* sont situés le plus souvent dans des tours de contrôle, et autorisent les appareils à décoller ou atterrir en s'assurant de la sécurité des voies dans l'aérodrome. Les échanges entre ces différents acteurs sont nombreux afin d'assurer le suivi et la sécurité de chaque vol. En effet, chacun de ces acteurs est responsable d'un secteur spécifique. Lors d'un changement de secteur d'un avion, les contrôleurs initiaux doivent s'assurer que l'avion est toujours suivi par un autre contrôleur.

Pour assurer une gestion du trafic optimale, l'ensemble de ces acteurs se base en partie sur les prédictions de plan de vol disponibles pour chaque avion. Pour les contrôleurs, les prédictions sont utiles à deux horizons différents : temps réel et court terme. En temps réel, elles permettent d'assurer la sécurité de chaque vol. À court terme (entre 20 et 60 min), elles permettent d'anticiper les changements de secteur et ainsi faciliter le suivi d'un vol, mais elles sont également utiles pour prévoir la congestion d'un secteur. Les prédictions à court terme ont été moins étudiées que celle en temps réel car elles sont moins critiques. Cependant, avec l'augmentation du trafic, il devient très intéressant, d'une part, de mesurer les performances de ces prédictions à court terme et d'autre part de proposer de nouvelles approches pour établir des prédictions plus fiables.

C'est dans ce contexte de recherche que s'est déroulé mon stage avec pour objectif d'expérimenter des techniques de prédiction innovantes afin d'améliorer les prédictions à court terme. Tous les travaux entrepris relèvent de la validation de principe et se limitent à la prédiction temporelle pour une mise en place facilitée mais peuvent être étendus à la prédiction spatiale. Pour valider une méthode, les prédictions obtenues par cette méthode sont comparées à celles faites par les systèmes actuels. L'ensemble a été réalisé en C++ et les résultats obtenus ont été visualisés avec R.

Dans ce rapport, seront présentés dans un premier temps, Thales Air System ainsi que la cellule d'innovation qui m'ont accueilli. Dans un second temps, les différentes approches mises en place sont expliquées. Enfin l'organisation du travail est détaillée.

Présentation de Thales air systems

Thales Air Systems est une société française du groupe Thales. Ses domaines d'activité sont le guidage électronique de missiles, les systèmes de surveillance et de défense, les radars surface, et la gestion du trafic aérien, ou en anglais Air Traffic Management (ATM). Par sa *Business Line* ATM, Thales Air Systems est présent en France, Australie, Royaume Uni, Allemagne, Italie et aux Etats-Unis. La *Business Line* est principalement dédiée à la marque de produits TopSky - Air Traffic Contrôle (ATC), qui consiste en des systèmes logiciels complets pour le contrôle du trafic aérien. TopSky fournit aux contrôleurs des outils fiables et des interfaces homme-machine intuitives pour la surveillance du trafic aérien.

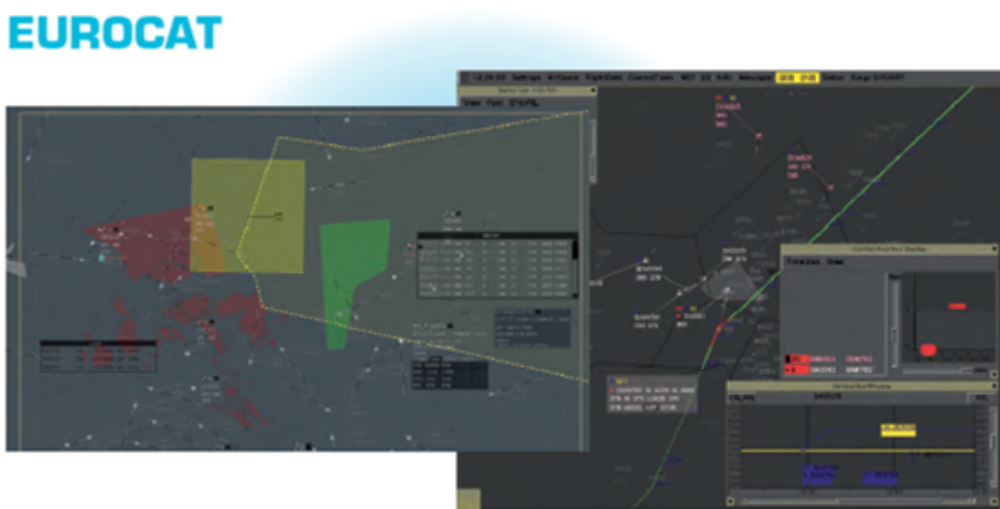


FIGURE 2.1 – Capture d'écran de TopSky

Thales Air Systems développe également d'autres produits liés à l'ATM, notamment un système de gestion de flux, qui permet de prévoir, d'optimiser et de planifier le trafic aérien. Pour finir, Thales Air Systems est le principal acteur industriel du programme européen Single European Sky ATM Research (SESAR), dirigé par l'organisation européenne pour la sécurité de la navigation aérienne¹ dans laquelle de nombreux industriels travaillent ensemble pour redéfinir l'espace aérien de l'Europe, et concevoir des systèmes d'ATM de nouvelle génération interoperables.

1. EuroControl

2.1 L'innovation Lab

Créée en 2009 au sein de la direction technique de la *business line* ATM, l'innovation Lab est une petite cellule d'innovation regroupant 3 ingénieurs, un doctorant ainsi que des stagiaires. Son fondateur, Areski HADJAZ, avait pour but le développement de projets innovants à l'intérieur de la *business line*. L'innovation Lab effectue une veille permanente autour des nouvelles technologies, et réfléchit à la possibilité d'adapter ces technologies au domaine du contrôle aérien, puis de les mettre rapidement en œuvre au travers de démonstrateurs. Ces démonstrateurs sont développés à travers des projets à court terme, d'environ 6 mois, dont l'objectif est la réalisation d'une maquette simulant une situation opérationnelle.

La conception des maquettes s'effectue en collaboration avec des opérateurs, des contrôleurs, des experts en facteurs humains, des ergonomes, des designers, et toutes personnes susceptibles d'utiliser ou d'améliorer le concept. Le résultat de ce travail collaboratif permet de déterminer la faisabilité du projet et s'il est envisageable de le soumettre à la communauté des contrôleurs aériens.

Un travail de communication est ensuite accompli, en interne comme en externe, pour à la fois informer les employés de Thales sur les réalisations du Lab et pour montrer aux clients que Thales est une entreprise innovante, capable de s'adapter aux nouvelles technologies et de proposer des solutions en rapport avec les problématiques émergentes liées au domaine du contrôle aérien. Tous les projets du Lab sont développés suivant au moins l'un de ces deux axes :

Interactions Ou comment combiner les nouvelles méthodes d'interaction homme-machine avec les besoins opérationnels des contrôleurs. Les sujets ont pour but d'améliorer la position de travail d'un contrôleur avec des technologies comme l'eye-tracking, les surfaces multi-touch, la reconnaissance spatiale et vocale ou la réalité augmentée.

Optimisation Ou comment optimiser les trajets aériens, la planification des vols, réaliser des systèmes les plus automatisés possibles en proposant des solutions viables : résolution de conflit, apprentissage, optimisation globale, systèmes autonomes, efficacité du vol.

Les travaux réalisés durant ce stage s'inscrivent dans le cadre de ce dernier axe.

Présentation du sujet et du contexte

L'objectif du stage est d'étudier les prédictions des trajectoires afin d'améliorer les prévisions de congestion des secteurs aériens et ainsi permettre l'émergence d'outils d'optimisation. L'étude ne porte pas sur la prédictibilité de l'ensemble de la trajectoire mais seulement sur trois points caractéristiques d'une trajectoire d'avion (voir la figure 3.1) :

- Le Top Of Climb (TOC), point d'inflexion de la trajectoire en fin de phase de montée.
- Le Top Of Descent (TOD), point d'inflexion de la trajectoire avant la phase de descente.
- L'Estimated Time of Arrival (ETA), heure d'arrivée de l'avion sur la piste d'atterrissage

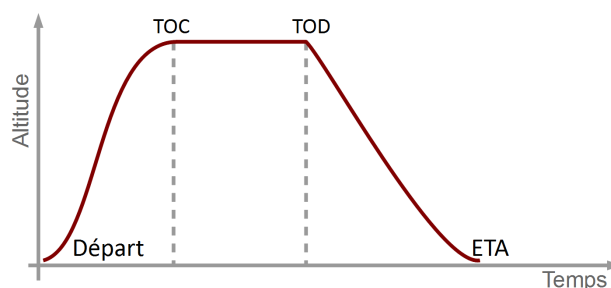


FIGURE 3.1 – Exemple de profil de plan de vol avec les points caractéristiques.

L'étude se limite à ces trois points pour plusieurs raisons. Tout d'abord pour une raison logique, le vol d'un avion se découpe réellement en trois phases¹, pour lesquelles on observe un comportement de l'avion différent. La prédiction du moment de changement de phase est alors intéressante pour les contrôleurs afin préparer les transitions inter-phases ainsi que pour faciliter la coopération entre contrôleurs. En effet, ces prévisions à court termes permettent d'établir une prédiction de la congestion d'un secteur et d'adapter les ressources en conséquence. Deuxième raison principale, ces points sont clairement identifiables et présents² dans chaque trajectoire, ce qui facilite la comparaison des prédictions sur différentes trajectoires. L'étude des points caractéristiques permet de mettre en place des indicateurs de performance pour les différents prédicteurs, qui seront présentés par la suite.

3.1 Données utilisées

Pour comparer plusieurs méthodes de prédiction, un ensemble de trajectoires réelles est utilisé. Les trajectoires réelles proviennent de radars Automatic Dependent Surveillance-Broadcast (ADS-B) qui enregistrent les données envoyées par les avions.

Plusieurs informations sont envoyées par l'avion toute les 5 secondes environ, entre autres : la latitude, la longitude, l'heure courante, l'altitude en Flight Level (FL), le Rate Of Climb (ROC) et la vitesse. Une trajectoire courte est en moyenne constituée de 200 relevés d'information envoyés

1. Phase de montée, d'en route et de descente.

2. Le TOC et le TOD peuvent être confondus sur certaines trajectoires.

par l'avion. Lors du stage, différents relevés de trajectoire ont été étudiés : un relevé du trafic autour de l'aéroport de Dublin durant le mois de juin 2011 et un autre autour de l'aéroport de Malmö durant le mois de novembre 2012. Les enregistrements de Dublin et Malmö contiennent environ 4000 trajectoires. La différence principale entre ces deux aéroports est la diversité des types avions rencontrés qui est beaucoup plus importante dans celui de Malmö. D'un point de vue plus technique, chaque trajectoire est stockée sous un fichier de format *xml*. Un index regroupe l'ensemble des trajectoires disponibles pour l'enregistrement en indiquant quelques informations en plus du chemin d'accès au xml spécifique, de type méta-données, telles que : l'Aéroport d'Arrivée (ADES), l'Aéroport de Départ (ADEP), la date du vol et un identifiant, appelé callsign.

3.2 Évaluation d'une méthode

Lors du test d'une méthode de prédiction sur des données réelles, principalement deux indicateurs sont étudiés : l'exactitude et la précision de la prédiction au cours du temps. L'exactitude est mesurée par la médiane de l'erreur sur le jeu de données de test. La médiane est préférée à la moyenne car elle est beaucoup moins sensible aux valeurs extrêmes. Pour mesurer la précision, l'écart inter-quartiles³ de l'erreur est utilisé (voir la figure 3.2). Il s'agit d'une mesure de dispersion qui s'obtient en faisant la différence entre le premier quartile Q_1 et le troisième quartile Q_3 .

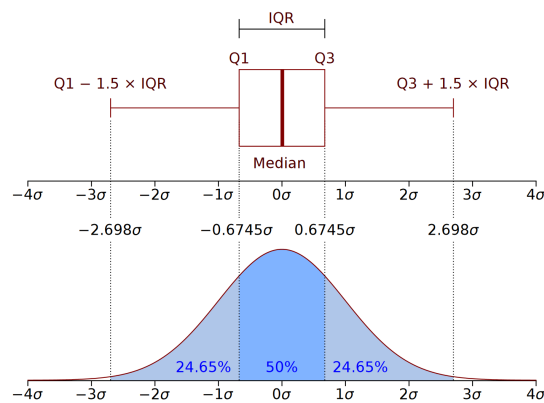


FIGURE 3.2 – Exemple de boîte à moustaches[5].

L'écart inter-quartiles a été préféré à la variance car il est plus simple à calculer et permet d'autres mesures comme l'écart entre le 5^e et le 95^e centile. Plus l'écart inter-quartiles est faible plus la méthode est certaine, c'est à dire que les erreurs de prédiction seront peu dispersées autour de la médiane.

Lors de l'évaluation d'une méthode, ces deux indicateurs d'exactitude et de précision sont visualisés par le biais de boîtes⁴ à moustaches au cours du temps (voir l'exemple 4.3 page 13). Ainsi à chaque pas de temps avant le point à prédire, une boîte à moustache est affichée, ce qui permet l'évaluation. Une bonne méthode se définit alors par une erreur médiane proche de zéro à chaque pas de temps et par un écart inter-quartiles faible ou à défaut qui diminue au cours du temps. Pour faciliter les comparaisons entre les différentes méthodes, seules les prédictions faites sur le TOC sont présentées dans les graphiques qui suivent. Ces graphiques ont été réalisés avec R [7].

3. Inter Quartil Range (IQR) en anglais.

4. Graphique permettant de visualiser la médiane et la dispersion des valeurs (voir la figure 3.2).

Études de méthodes de prédiction

Dans cette partie, différentes méthodes de prédiction sont présentées. Il s'agit tout d'abord des méthodes pré-existantes puis des méthodes développées durant le stage. Pour chacune des méthodes, le principe de fonctionnement est expliqué puis les résultats sont présentés pour enfin conclure sur les avantages et inconvénients de chaque méthode.

4.1 Méthodes existantes

4.1.1 Prédiction de Ground TP

Le composant Ground Trajectory Prediction (TP) est un composant du système ATC Sol. Parmi les différentes missions du système ATC Sol figure le calcul de prédiction de trajectoires en fonction d'un plan de vol. Le plan de vol est une feuille de route constituée de points de passages qui est déposée par le pilote avant le vol. Le système ATC Sol calcule la trajectoire permettant de suivre ce plan de vol en fonction des caractéristiques de l'avion. Au cours du vol, le plan de vol peut être modifié à cause d'ordres provenant de contrôleurs. En temps normal, ce changement est rentré dans le système pour que le système ATC Sol mette à jour la trajectoire prévue. Cependant il peut arriver que le plan de vol ne soit pas modifié et donc que le système ne puisse pas mettre à jour la trajectoire de l'avion. Cela arrive souvent dans les phases de début et de fin de vol car le contrôleur donne beaucoup d'indications à l'avion et elles ne sont pas forcément rentrées dans le système par manque de temps. C'est pourquoi le plan de vol et par conséquent la trajectoire peuvent être erronés.

Dans la trajectoire fournie par le système ATC Sol, le TOC, le TOD et l'ETA sont présents, cela permet d'évaluer les performances des prédictions. Pour évaluer les performances des prédictions sur le TOC, il suffit alors de comparer le temps d'arrivée prédit par les différents plans de vol établis au cours du vol par rapport au TOC réel. En faisant cela on obtient la figure 4.1.

Comme expliqué dans la partie précédente, il y a deux mesures visibles sur ce graphique. La courbe verte représente la médiane de l'erreur sur l'ensemble des 800 trajectoires au cours du temps ce qui permet de visualiser l'exactitude des prédictions. Entre -10 et 0 minutes avant le TOC, la médiane est proche de zéro. Si la prédiction est faite à plus long terme, l'erreur médiane est nettement plus élevée.

Le deuxième indicateur à analyser est la précision de la prédiction. Cet indicateur est représenté par l'écart entre les courbes rouge et bleue qui représente respectivement le premier et troisième quartile. Cet écart est de presque 4 minutes quelque soit l'instant auquel la prédiction est faite. Il diminue peu et uniquement dans les dernières minutes de vols.

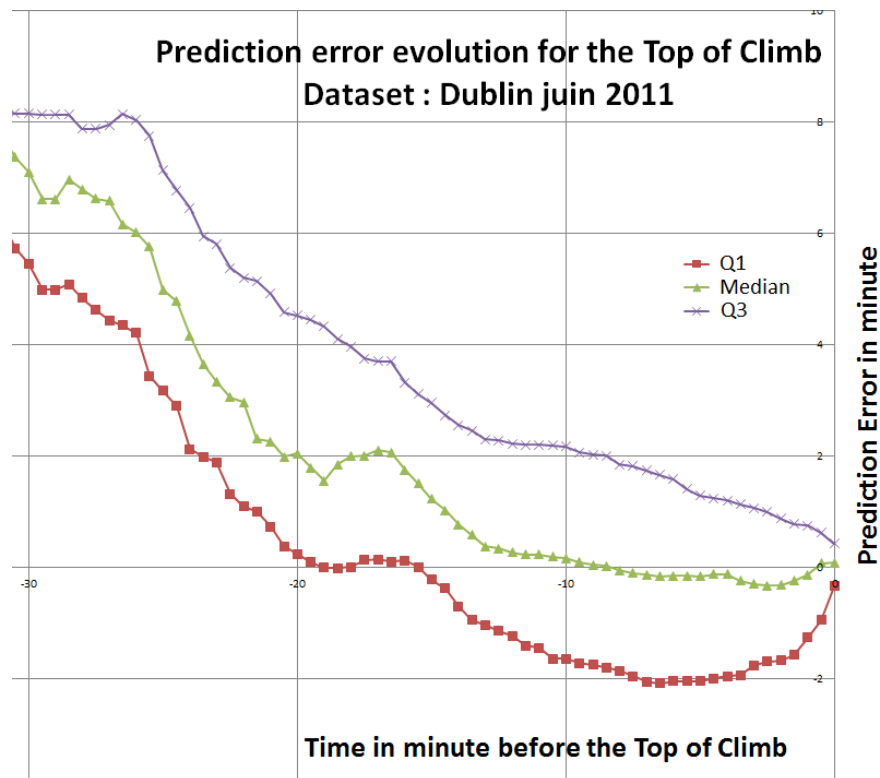


FIGURE 4.1 – Graphique de l’erreur de la prédiction faite par le Ground TP au cours du temps.

Cette première étude permet de mettre en évidence plusieurs choses. Tout d’abord on remarque que l’exactitude est bonne sur le court terme ($t > -10\text{min}$) car la médiane de l’erreur est proche de zéro. La détérioration observée sur le long terme ($t < -10\text{min}$) peut, en partie, être expliquée par un biais de la représentation. En effet, il faut en moyenne 20 minutes¹ à un avion pour décoller et atteindre le TOC. Par conséquent les valeurs visualisées avant -20 minutes sont constituées uniquement par des avions qui montent lentement. Une fois que l’échantillon devient plus représentatif, aux alentours de -15 minutes, l’erreur est acceptable.

Pour l’incertitude en revanche, la situation est beaucoup moins bonne. L’écart inter-quartiles est élevé, entre 3 minutes 30 et 4 minutes, mais en plus il ne se réduit quasiment pas au cours du temps. Il n’y a que dans les dernières minutes que l’incertitude baisse réellement. Si un espace inter-quartiles de 4 minutes peut être acceptable pour une prédiction à long terme, en revanche il n’est pas satisfaisant qu’il ne diminue pas plus avec le temps.

Dans le reste de ce rapport, les méthodes étudiées seront évaluées par rapport à ces résultats obtenus par le Ground TP. L’objectif principal est d’obtenir une incertitude des prédictions plus faible que celle du Ground TP.

1. Pour de plus amples détails sur la répartition TOC voir la figure A.1 page 31.

4.1.2 Méthodes utilisant le modèle BADA

La Base of Aircraft Data (BADA) est un modèle développé et maintenu par EuroControl. Il permet de calculer les performances de l'avion en fonction de l'altitude à laquelle il se trouve. Les performances recouvrent notamment le ROC et la True Air Speed (TAS) qui sont nécessaires pour effectuer une prédiction temporelle. Il existe deux moyens d'utiliser la BADA : le modèle à énergie totale et le modèle tabulaire. Le modèle à énergie totale est un modèle cinétique qui permet, à partir d'un ensemble d'équations polynomiales, de calculer les données voulues. Le modèle tabulaire est un ensemble de tableaux, un par type d'avion (A320, B737 ...), contenant les valeurs minimales, maximales et nominales de chaque caractéristique. Dans le cadre de ce stage, le modèle BADA tabulaire a été étudié.

Le calcul d'une prédiction, avec le modèle BADA, est simple. Les futurs ROC de l'avion sont lus dans le tableau du modèle en fonction des futures altitudes de l'avion. Il suffit alors d'appliquer les différents ROC au cours de la trajectoire à partir de l'altitude courante de l'avion pour connaître le temps de parcours. Cette méthode de prédiction a uniquement besoin de l'altitude courante de l'avion et de l'altitude cible, ce qui facilite grandement sa mise en place. La prédiction issue de cette méthode ne permet pas de calculer un plan de vol mais de calculer les temps de survol des points de passage du plan de vol. Les conditions météorologiques et le trafic aérien environnants ne sont pas pris en compte par ce modèle.

Malgré la simplicité de cette méthode, de bons résultats sont obtenus lors de la prédiction. Pour éviter le biais observé sur la figure 4.1 pour le Ground TP, l'erreur n'est plus visualisée en fonction du temps mais en fonction de l'altitude, voir la figure 4.3. À titre de comparaison, les mêmes données ont été visualisées mais au cours du temps sur la figure 4.2.

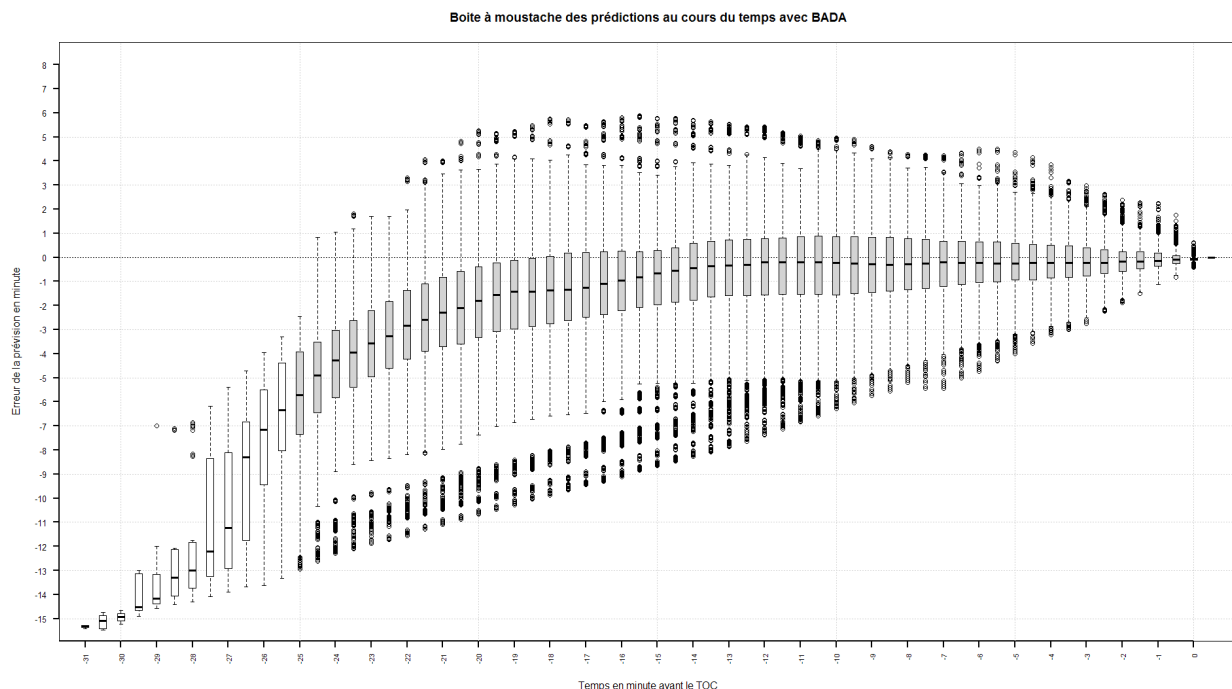


FIGURE 4.2 – Graphique de l'erreur de la prédiction du TOC faite par la BADA au cours du temps.

Tout d'abord sur la figure 4.3, la médiane de l'erreur est proche de zéro. En comparant avec les prédictions du Ground TP, aucune amélioration nette n'est observable. En revanche, l'écart interquartiles est légèrement plus faible avec cette méthode qu'avec le Ground TP et surtout il décroît avec le temps. Bien que l'amélioration soit faible, ces résultats sont intéressants car ils permettent d'obtenir des résultats comparable au Ground TP avec une méthode plus simple. Les calculs faits par le Ground TP pour obtenir une prédiction sont influencés par de nombreux paramètres alors que pour le modèle BADA, les paramètres d'entrés sont au nombre de trois : la position courante de l'avion, la position visée et le type d'avion². Pour les calculs internes, seule une liste de table est utilisée. On gagne donc énormément en terme de lisibilité des paramètres qui ont influé sur les prédictions données. Cependant cette méthode est très peu flexible. En effet, elle se base entièrement sur les tables lues en entrée. Il est donc impossible de s'adapter au comportement d'un avion au cours du vol.

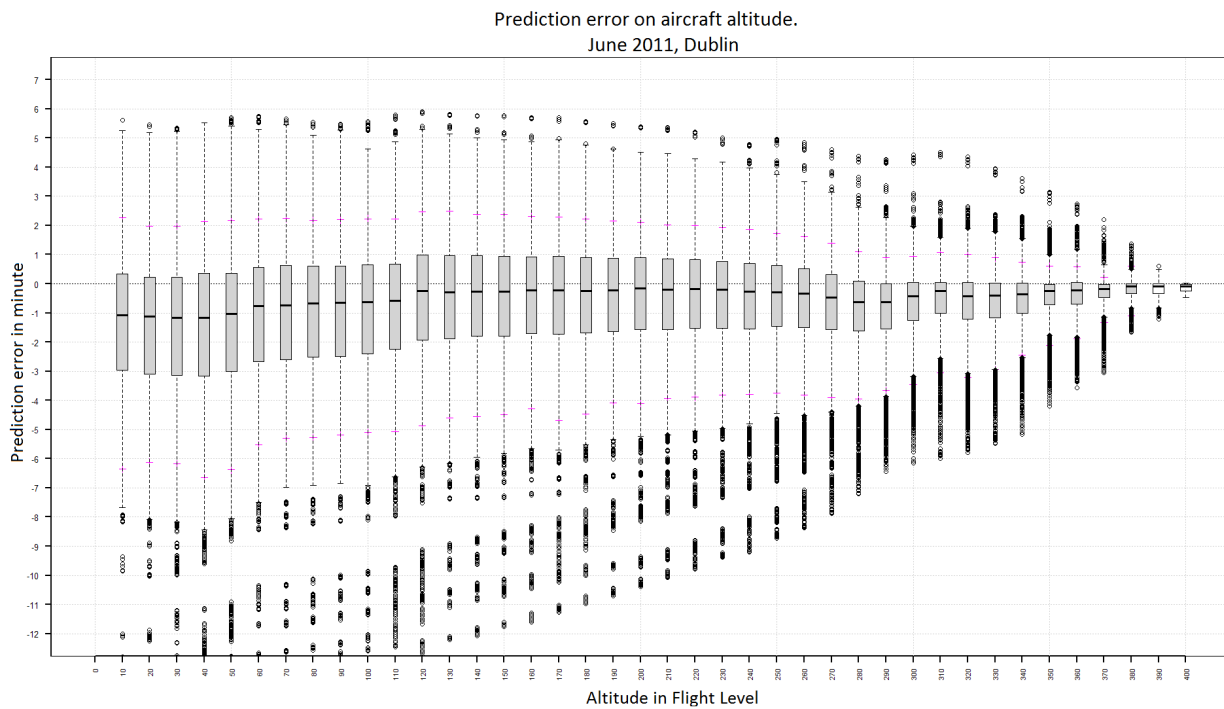


FIGURE 4.3 – Graphique de l'erreur de la prédiction du TOC faite par la BADA en fonction de l'altitude.

2. Si il s'agit d'un nouveau type d'avion, alors il faut créer un nouveau tableau ou trouver une correspondance.

4.2 Méthodes développées

Les nouvelles méthodes élaborées ne se basent pas sur les équations de la dynamique de vol car ces types d'approche ont déjà été explorés. De plus, ces approches pour être précises nécessitent des informations cruciales, telles que le poids l'avion au décollage, qui sont souvent manquantes. Lors de l'élaboration de nouvelles méthodes, l'accent a été mis sur l'étude de méthodes innovantes et sur leur validation par l'utilisation de jeux de données réels.

4.2.1 Prédiction par extrapolation

Les premières approches qui ont été poursuivies sont des techniques d'extrapolation. Le principe est simple : prédire les positions futures à partir des positions passées. Le but est de trouver les coefficients d'un modèle d'équations tel que la distance entre les observations et la courbe résultant du modèle soit minimale ce qui correspond à un problème de moindres carrés. Deux types d'extrapolations ont été testés : les extrapolations linéaires et les extrapolations quadratiques. L'extrapolation linéaire permet de trouver la droite affine correspondant le mieux au profil de l'avion. L'extrapolation quadratique permet elle de trouver une courbe quadratique.

Lors de l'étude des profils de vols verticaux, il a été remarqué qu'un avion a tendance à adopter un profil de vol quadratique lorsque l'on visualise l'altitude en fonction du temps. Cette tendance a bien entendu été vérifiée par le calcul du coefficient de corrélation entre les trajectoires réelles et les différentes extrapolations.

Classiquement dans le problème de moindres carrés, le but est d'expliquer la valeur d'une variable par l'intermédiaire d'un ensemble de variables dites explicatives pondérées par des coefficients qu'il faut calculer. Le problème est défini formellement par :

$$\min_{x \in \mathbf{R}^n} \|Ax - b\|_2^2 \quad o$$

- A est une matrice (m, n) dont chaque ligne est une observation comprenant les variables explicatives.
- x est un vecteur colonne $(n, 1)$ correspondant aux coefficients à trouver.
- b est un vecteur colonne $(n, 1)$ dont chaque ligne est la variable à expliquer.

Dans le cas présent, le but est d'expliquer h_i , l'altitude observée à un instant i , par t_i , le temps depuis le décollage. Si l'on souhaite approcher la trajectoire par une droite affine alors une observation se fera sous la forme $(t_i, 1)$ et x sera de la forme (a, b) . Ainsi on obtient l'équation suivante pour prédire la future altitude de l'avion, $h_{prédit}$, à l'instant t_{futur} :

$$h_{prédit} = a \cdot t_{futur} + b \quad (4.1)$$

Si l'on souhaite approcher la trajectoire par une courbe quadratique alors une observation se fera sous la forme $(t_i^2, t_i, 1)$ et x sera de la forme (a, b, c) . Ainsi on obtient l'équation suivante pour prédire la future altitude de l'avion, $h_{prédit}$, à l'instant t_{futur} :

$$h_{prédit} = a \cdot t_{futur}^2 + b \cdot t_{futur} + c \quad (4.2)$$

Pour réaliser ces calculs, la formalisation de ce problème a été permise par le cours MT09[6] suivi à l'Université de Technologie de Compiègne (UTC). Pour l'application pratique, une librairie d'algèbre linéaire a été utilisée [8] dans un premier temps puis QuadPorg++[2] a été utilisé pour ajouter des contraintes lors de la minimisation.

Une fois les coefficients calculés, effectuer une prédiction est relativement simple. Par exemple pour le TOC, il suffit de calculer à quel instant la courbe, linéaire ou quadratique, intersecte la droite horizontale correspondant à l'altitude du TOC, voir la figure 4.4. Avec une droite linéaire cela ne pose aucun problème, il y a toujours intersection entre l'altitude voulue et la droite car la droite calculée n'est jamais parallèle à l'horizontale. Avec une courbe quadratique, l'existence d'une prédiction n'est pas assurée. En effet le maximum de la parabole peut être inférieur au Requested Flight Level (RFL). Afin d'assurer une certaine stabilité aux prédictions, des contraintes ont été ajoutées à l'extrapolation quadratique. Les contraintes ajoutées sont des intervalles de confiance sur les paramètres. C'est à cause de ces contraintes que j'ai utilisé quadprog[2] qui est un solveur de problème quadratique avec contrainte linéaire.

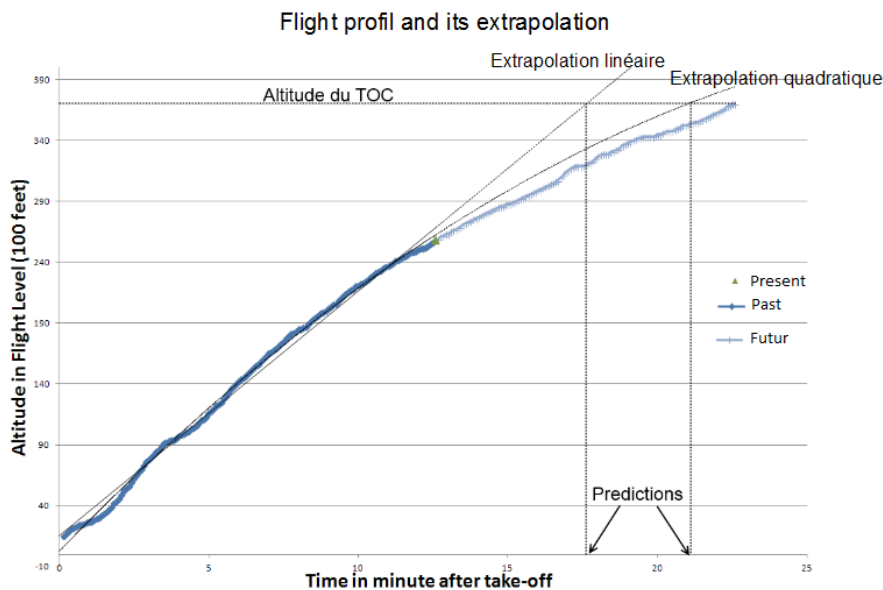


FIGURE 4.4 – Exemple de prédictions faites par extrapolations.

Au cours du vol d'un avion, les positions successives s'ajoutent à l'ensemble des positions observées permettant ainsi d'affiner les prédictions. Ce processus est visible sur les figures A.3 à A.5 en annexe page 32. Sur ces figures ainsi que sur la figure 4.4, on remarque que le modèle quadratique suit mieux le profil que le modèle linéaire. Cependant comme indiqué au dessus, le modèle quadratique ne permet pas d'effectuer une prédiction dans tous les cas de figures. C'est surtout vrai dans les premières minutes de vol car l'extrapolation est encore très variable. L'instabilité du modèle est due à deux choses : d'une part l'instabilité intrinsèque du vol surtout au décollage et d'autre part le fait que la part quadratique du modèle reste relativement faible. En moyenne le coefficient a de l'équation 4.2 est égal à -0.8 et il n'est stabilisé que tardivement dans le vol.

Cette instabilité explique pourquoi le modèle linéaire a été préféré au modèle quadratique. Il faut également ajouter que le but de la méthode est d'obtenir une prédiction exacte et précise du temps de passage sur un point particulier et non de modéliser le comportement de l'avion le long du plan de vol. Or il se trouve que le modèle linéaire bien qu'éloigné de la réalité parvient à prédire un temps de passage cohérent.

Malheureusement malgré toutes les améliorations apportées, ce type de méthode ne donne pas de bons résultats. Ceux-ci sont visibles en annexe page 33. Lorsque l'on visualise l'erreur ainsi que

l'incertitude au cours du temps depuis le décollage, on remarque l'existence d'un biais croissant sur la médiane de l'erreur. Cependant l'incertitude est meilleure avec ce modèle qu'avec le Ground TP. Il faudrait donc penser à d'autres moyens d'informer le système pour essayer corriger ce biais.

4.2.2 Modèle semblable à la BADA

Le fonctionnement de la table BADA, qui a été présenté dans la partie 4.1.2, est un modèle plus souple que le modèle linéaire car au lieu de calculer un taux de montée unique, un taux de montée différent est calculé pour chaque intervalle d'altitudes. Le modèle BADA est donc un modèle linéaire par morceaux défini par des taux de montée successifs. Le modèle BADA est aussi plus robuste que le modèle quadratique car en imposant un taux de montée positif au cours de la montée alors on peut assurer que le profil prédit intersectera l'altitude voulue et donc on peut obtenir une prédiction. C'est pourquoi dans la suite du stage, des alternatives semblables à la BADA ont été développées.

Reconstruction de la BADA

La première étape a été de reconstruire une table BADA à partir des enregistrements de vols réels³ et de comparer les valeurs obtenues avec les valeurs nominales de la BADA. Au début, aucune séparation sur le type d'avion n'était faite car sur le premier enregistrement utilisé, Dublin juin 2011, l'A320 représente environ 85% des avions. Les autres types d'avions étant peu présents, ils n'étaient pas assez représentatifs pour permettre une distinction sur le type d'avion.

Pour obtenir une nouvelle valeur nominale à partir des enregistrements, les ROC observés sont stockés par intervalles d'altitude, les mêmes intervalles que pour le modèle BADA. Une fois les ROC stockés, la médiane de chaque intervalle est calculée. Le résultat est visible sur la figure 4.5.

Les différences entre les médianes apprises et les valeurs nominales sont relativement faibles sur la fin de la montée, pour $FL > 120$, environ 2% de différence. À l'inverse, au début de la phase de montée, $FL < 120$, les écarts sont plus importants. Ces écarts s'expliquent par la grande variabilité des comportements lors des premières minutes de vol mais aussi par les spécificités de l'aéroport qui ne peuvent être pris en compte dans le modèle BADA.

Ce modèle nécessite une phase d'apprentissage avant de pouvoir effectuer une prédiction. Pour éviter tout biais lors de la validation, le jeu de données initial est séparé en deux groupes distincts tirés de manière aléatoire. Un jeu de données d'apprentissage, il représente environ 70% du total et il est utilisé pour l'initialisation du modèle. Un jeu de données de validation, il représente environ 30% du total et c'est sur ce jeu de données que le modèle est évalué.

Une fois la base du modèle mise en place, le but est de le complexifier pour améliorer les prédictions tout en gardant un modèle le plus général possible. Ainsi une attention particulière est portée sur la représentativité des données mais aussi à ce que les modifications apportées ne soient pas spécifiques à un cas particulier. Toute erreur sur l'un ou l'autre de ces aspects entraînerait des prédictions de piètre qualité. C'est pourquoi la séparation selon les différents types d'avions n'a été envisagée que lorsque l'enregistrement de Malmö a été utilisé car les types d'avions y sont beaucoup plus diversifiés. Tout d'abord, les différences de ROC selon les types d'avions ont été visualisées, voir l'exemple entre 2 types d'avions sur la figure A.7 page 34 en annexe. Il apparaît

3. Les données utilisées sont décrites dans la partie 3.1 page 8.

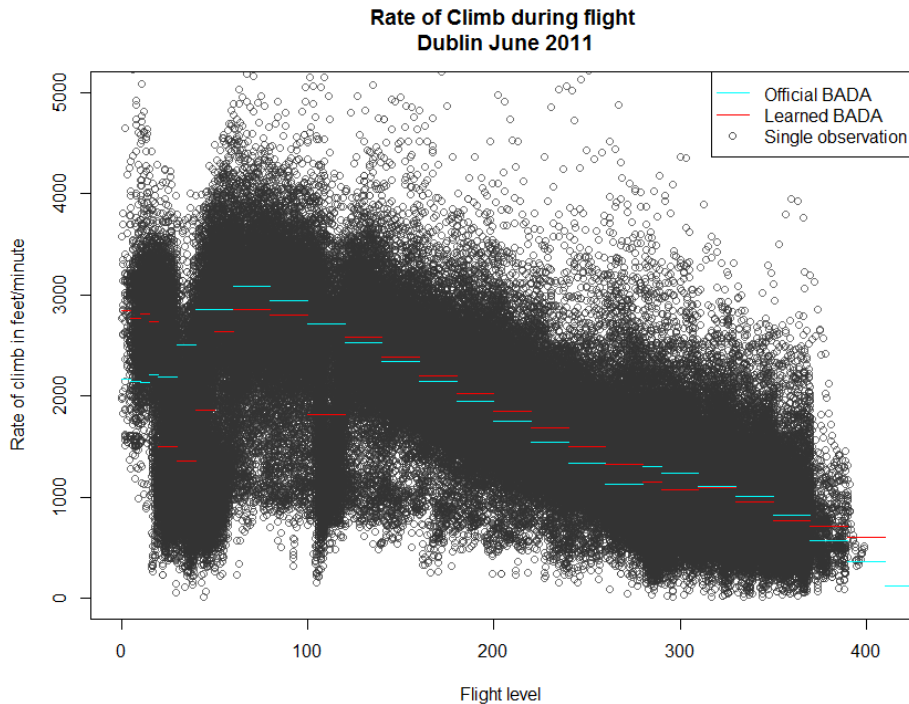


FIGURE 4.5 – ROC au cours de la montée, pour l’ensemble des vols de l’enregistrement de Dublin juin 2011.

qu’il existe des différences qui nécessitent d’être prises en compte dans le modèle mais une séparation systématique n’est pas la solution car elle entrainerait une trop grande fragmentation du jeu de données initial.

Pour palier ce manque d’information du système et éviter la fragmentation des jeux de données, la séparation selon le type se déroule en deux étapes :

1. Fixer un seuil de représentativité pour qu’un type d’avion soit pris en compte. Par exemple, si un type d’avion est représenté par un échantillon de moins de 50 vols alors l’échantillon n’est pas considéré comme représentatif.
2. Pour les échantillons non représentatifs, trouver l’échantillon le plus proche et fusionner les données des deux échantillons ensemble.

Pour comparer deux échantillons, un indice de ressemblance est calculé par la moyenne des indices de ressemblance entre les intervalles des deux échantillons. Cet indice correspond au pourcentage de chevauchement des valeurs stockées dans les deux intervalles. On considère que deux intervalles sont proches si ils se chevauchent beaucoup ce qui se traduit par un indice élevé. De cette manière, deux échantillons sont proches si leurs intervalles respectifs sont proches. Lors des fusions, une table d’indirection est aussi créée pour faire la correspondance entre le type d’avion voulu et les types d’avions sauvegardés.

Une fois toutes ces modifications appliquées, un modèle proche de la BADA est obtenu. Pour comparer les performances entre ce modèle et le modèle BADA, 10 tests sont effectués à la suite au lieu d’un seul car la répartition des vols dans les jeux d’apprentissage et de validation est aléatoire. Cela permet également d’avoir une idée de la stabilité du modèle. Les résultats obtenus sont

très encourageants. Dans la figure 4.6, l'ensemble d'indicateurs présenté permet de comparer les résultats obtenus par le modèle BADA officiel et le modèle appris.

En haut à gauche est présentée la médiane de l'erreur⁴ au cours de la montée pour les deux modèles. On remarque tout de suite que la médiane du modèle appris est meilleure que celle du modèle BADA. D'ailleurs, la médiane de l'erreur pour le modèle appris est toujours en dessous de 30 secondes, ce qui est un très bon résultat.

Le graphique en haut à droite permet de savoir dans quelle mesure les prédictions ont été améliorées. Pour chaque paire de prédictions, BADA officielle et BADA apprise, la différence est mesurée. Une différence négative signifie que la prédiction a été meilleure avec le modèle appris. On remarque qu'en moyenne les prédictions ont été améliorées de 37 secondes.

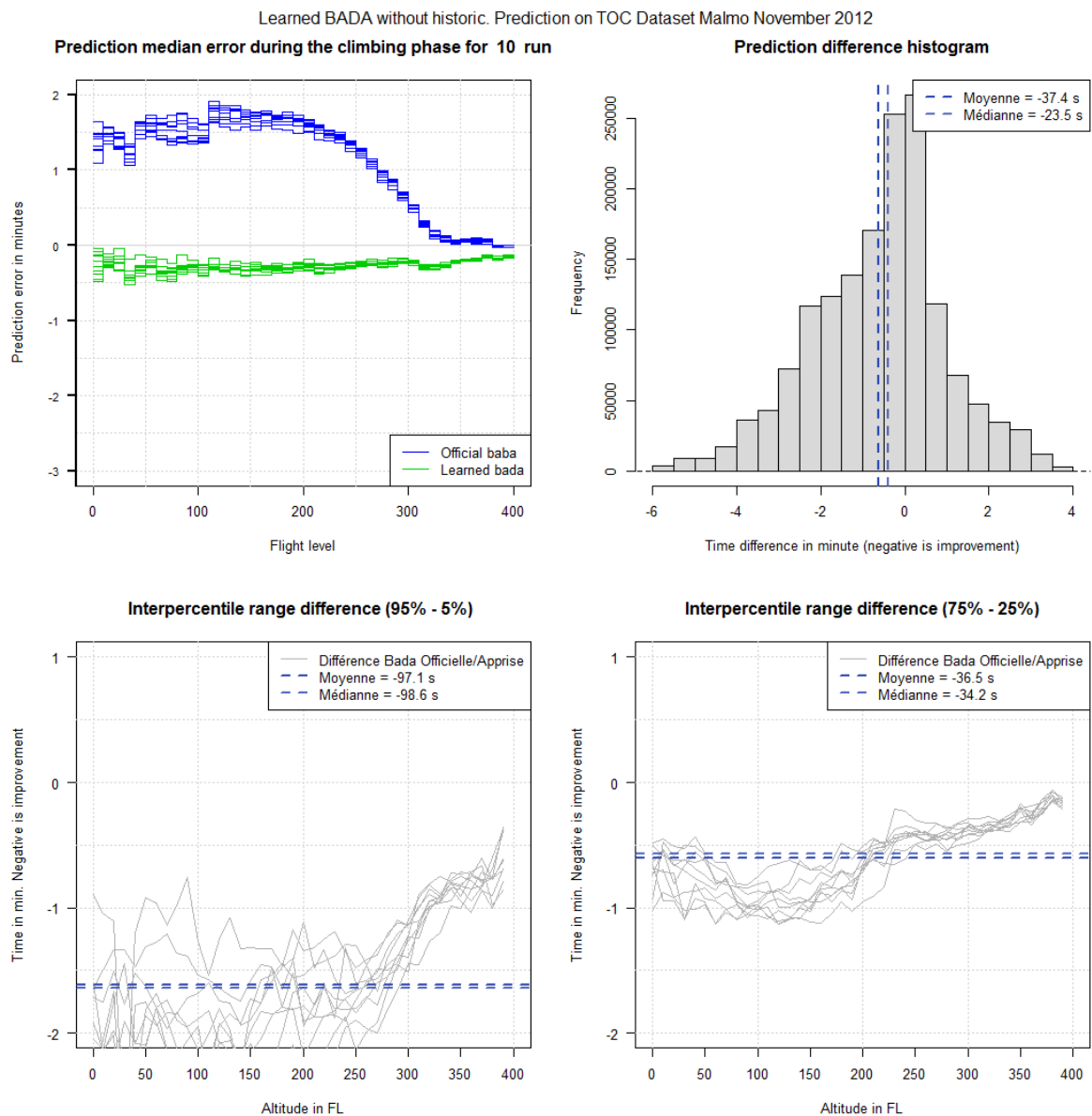


FIGURE 4.6 – Résumé des résultats obtenus par le modèle BADA appris sans historique par rapport au modèle BADA.

4. Une erreur négative signifie que le TOC a été prédit trop tôt.

Les deux graphiques du bas présentent la différence d'incertitude des prédictions entre la BADA officielle et la BADA apprise. Le graphique en bas à gauche représente la différence d'écart interquartiles entre le 5^e et le 95^e centile. Le graphique en bas à droite représente la différence d'écart entre le premier quartile et le troisième quartile. Là encore, on remarque un gros gain avec le nouveau modèle ce qui veut dire que les prédictions sont plus précises qu'avec le modèle BADA. L'amélioration est en de moyenne d'1 minute 30 sur l'écart entre le 5^e et le 95^e centile, et de 35 secondes sur l'écart entre le premier quartile et le troisième quartile.

Au delà du modèle BADA

Une fois ces premières améliorations constatées se pose la question de l'évolution de ce modèle. Comme dit précédemment, les modèles de type BADA permettent de personnaliser les prédictions uniquement de manière grossière, c'est à dire par le type d'avion. S'est alors posée la question de personnaliser la prédiction à un vol en particulier, en utilisant son historique. Grâce à l'historique, il est possible de mesurer si l'avion a volé plus lentement ou plus rapidement que la valeur nominale. Or, il a été remarqué que au cours des différents vols, un avion garde cette tendance de vol, plus lente ou plus rapide par rapport la valeur nominale. Il s'agit d'un invariant que l'on peut apprendre au cours du vol. Le but est alors de quantifier cet invariant puis de réussir à le répliquer lors du calcul des prédictions.

Pour quantifier cet invariant, un simple calcul d'écart entre la valeur observée et la valeur nominale ne suffit pas car ce n'est pas assez flexible et ne correspond pas à la réalité. Le plus efficace est de comparer le comportement de l'avion par rapport à l'ensemble des vols stockés et ainsi obtenir une évaluation du comportement relative aux comportements de l'ensemble des vols. C'est pourquoi au lieu de calculer la médiane du taux de montée, comme le fait la table BADA, l'ensemble de la distribution des taux de montée est calculé. Ainsi pour chaque taux de montée de l'historique, on stocke le placement de ce taux de montée vis à vis de la distribution des taux de montées. Il est alors possible d'estimer le placement des futurs taux de montées via la moyenne des placements stockés dans l'historique. Le placement ainsi calculé est utilisé sur chaque intervalle pour calculer les futurs taux de montée. Ainsi le comportement passé d'un avion sera pris en compte lors du calcul de la prédiction. Le graphique 4.7 illustre bien ce fonctionnement :

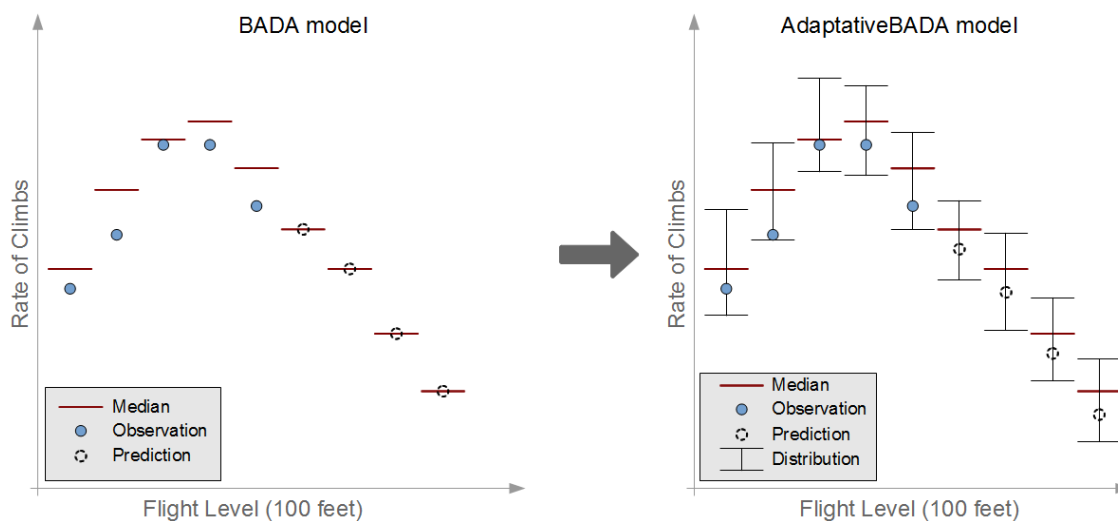


FIGURE 4.7 – Évolution d'un modèle BADA statique à un modèle adaptatif.

Tout comme précédemment, le modèle est évalué sur 10 tests successifs pour visualiser la stabilité des résultats. Les mêmes indicateurs sont visibles sur la figure 4.8, la seule différence est l'échelle utilisée pour les graphiques dans la partie inférieure. La médiane de l'erreur est similaire comme on peut le voir sur le graphique en haut à gauche. Cependant, on remarque que l'amélioration moyenne et médiane d'une prédiction par rapport au modèle BADA est bien meilleure grâce à l'historique sur le graphique en haut à droite. Il y a un gain moyen de 55 secondes sur une prédiction soit environ 18 secondes de mieux que le modèle sans historique. C'est sur l'incertitude que l'amélioration est la plus nette. L'écart entre le premier quartile et le troisième quartile est en moyenne réduit de 1 minute par rapport à la BADA, ce qui est aussi plus efficace qu'avec le précédent modèle. L'écart entre le 5^e et le 95^e centile est lui réduit de presque 3 minutes en moyenne, là encore il s'agit d'une nette amélioration par rapport aux modèles BADA officiel et appris.

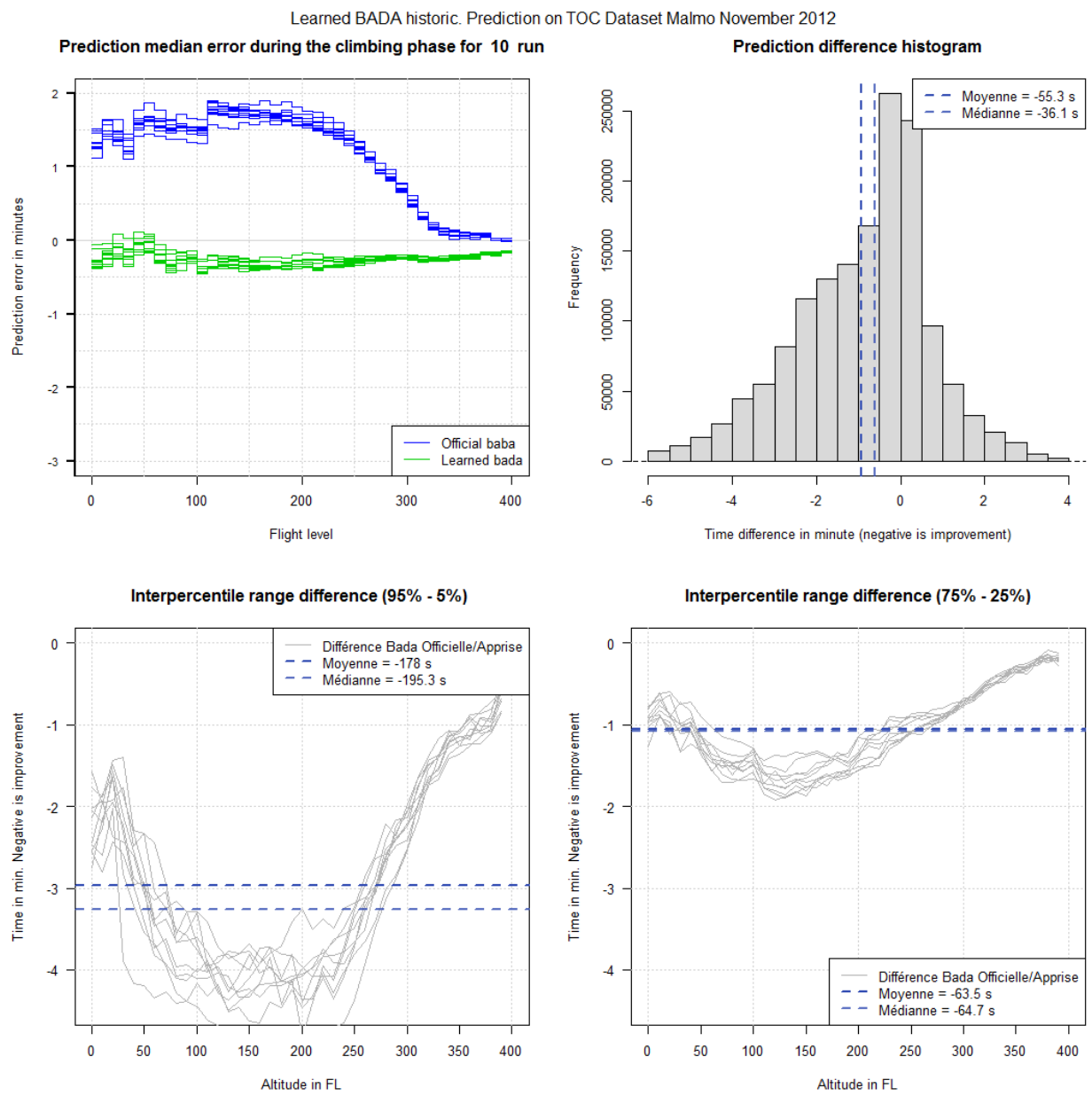


FIGURE 4.8 – Résumé des résultats obtenus par le modèle BADA appris avec historique par rapport au modèle BADA.

L'amélioration sur l'incertitude est moins importante pour les altitudes inférieures au FL 90. Cela est dû à la nature de l'apprentissage. Dans les premières minutes d'un vol, il n'y a que peu d'historique et donc les prédictions reposent sur les valeurs médianes des distributions, ce qui explique que le gain soit similaire au modèle sans historique. L'amélioration de l'incertitude dans les dernières minutes est également faible car il ne reste que peu de temps avant le TOC et donc l'incertitude est déjà faible.

Pour se rendre compte des résultats, il faut également examiner la représentation en boîtes à moustaches en annexe page 34. Cette figure est à comparer avec les performances du modèle BADA page 13.

4.3 Autres pistes explorées

En plus de la conception de nouveaux modèles, la compréhension des erreurs de prédiction a également pris une place importante durant le stage. En effet, les modèles présentés ne tiennent absolument pas compte des spécificités de la gestion de trafic aérien (météorologie, routes aériennes, pilote, ...). Or, les erreurs qui sont faites peuvent être liées à un certain contexte ou à certaines trajectoires. Si il est possible de mettre en avant une telle corrélation, alors les prédictions pourraient en tirer parti et être encore améliorées.

4.3.1 Corrélation de l'erreur avec la position géographique

Pour visualiser ces données géolocalisées, le logiciel Gephi a été utilisé[1]. C'est un logiciel de visualisation de graphes avec de nombreux outils. Il permet notamment de visualiser des points géolocalisés mais aussi d'y appliquer des filtres pour ne garder qu'une partie des points. Ce sont ces deux fonctionnalités qui ont permis d'étudier l'influence de la position de l'avion sur les prédictions.

La première chose qui a été visualisée est l'erreur de la prédiction faite par le modèle BADA adaptatif par rapport à la position géographique. Comme à chaque position une prédiction est faite par le modèle, il est possible de voir à quels instants la prédiction se détériore. Tout d'abord, l'ensemble des trajectoires de Dublin a été visualisé via les relevés de position. Puis ont été visualisées uniquement les positions qui ont donné lieu à une mauvaise prédiction, c'est à dire une prédiction se trompant de plus de 2 minutes. Aucune corrélation n'a été trouvée entre la position géographique et les mauvaises prédictions.

Il y a une hypothèse forte qui est faite lors des prédictions : les avions effectuent une montée continue jusqu'au TOC. Or, lors d'un vol, le pilote peut être amené à effectuer un palier dans la montée ce qui retarde le temps d'arrivée. Ces paliers sont, dans la majorité des cas, dus à des ordres provenant du contrôleurs afin d'éviter un conflit ou de faciliter la transition entre secteurs. Il est alors pertinent de mesurer l'existence de ces paliers et de visualiser leur répartition. La première conclusion de l'étude est la validation de l'hypothèse de montée continue des avions. Il y a en effet très peu de paliers effectués, par ailleurs une grande partie de ces paliers n'excède pas une durée de 30 secondes. Cependant la présence de ces paliers détériore effectivement les prédictions car ils entraînent systématiquement une prévision du TOC prématurée par rapport au TOC réel. Du fait du faible nombre de paliers et de leur répartition, aucune corrélation n'a été trouvée entre position géographique et existence de palier.

4.3.2 Correction des prédictions existantes

Comme il n'a pas été possible de trouver de corrélation géographique, il reste la possibilité de corriger les prédictions générées par un modèle de manière statistique. Il s'agit de mesurer le biais médian au cours du temps sur les prédictions puis d'appliquer une correction équivalente à ce biais. Ce processus nécessite encore une fois un fonctionnement en deux étapes : apprentissage puis validation. Ce type de correction permet d'améliorer la médiane de l'erreur mais n'influence que très peu l'incertitude. Le but est de centrer la médiane de l'erreur autour de 0. Ces corrections ont été appliquées aux prédictions issues du Ground TP car ce sont elles qui présentaient le plus gros biais. Les résultats sont visibles en annexe sur la figure A.2 page 31.

Cette méthode a effectivement permis d'améliorer les prédictions dans les cas où le biais était significatif, dans les premières minutes, mais ce processus a également détérioré de nombreuses prédictions à court terme comme à long terme. De plus la correction apprise est entièrement dépendante du système de prédiction ce qui pour le Ground TP pose des questions sur la stabilité et sur la possible généralisation de cette méthode. Pour les méthodes développées, le biais est faible et il était peu pertinent d'ajouter une nouvelle phase d'apprentissage plutôt que de tenter d'améliorer le modèle directement.

4.4 Conclusion technique

L'étude sur les modèles déjà existants a permis de mesurer la qualité des nouveaux modèles avec le Ground TP comme référence. Il a été remarqué que le Ground TP ne permettait pas d'établir des prédictions de manière certaine. Il y a une forte incertitude des prédictions qui ne baisse que tardivement au cours du vol. Pour palier à ce problème, différentes méthodes ont été testées. Tout d'abord le modèle BADA tabulaire a été implémenté. Ce modèle a plusieurs avantages par rapport au Ground TP. Il est plus simple car il n'est constitué que de tableaux à lire et il permet d'obtenir des prédictions légèrement meilleures en termes d'incertitude.

Les méthodes d'extrapolation développées par la suite étaient prometteuses car elles ne nécessitaient aucun apprentissage. Cependant pour les deux méthodes d'extrapolations développées, les résultats obtenus n'égalent pas ceux du modèle BADA. Bien que l'incertitude soit meilleure que le Ground TP, il existe un biais trop important sur la médiane de l'erreur. Malgré les tentatives pour mieux informer ce type de modèle, aucune solution n'a été trouvée pour améliorer ces modèles.

Vu le succès du modèle BADA, l'autre modèle développé par la suite avait le même fonctionnement puis il a été modifié pour essayer de l'améliorer. La création du modèle s'est déroulée en deux étapes. Tout d'abord, les données des tableaux du modèle ont été apprises sur la base de données réelles. À partir de ce modèle, de meilleurs résultats que pour le modèle BADA ont été obtenus, aussi bien en terme de précision que d'incertitude. Fort de ce succès, des capacités d'adaptations ont été ajoutées au modèle. Le but est de pouvoir personnaliser une prédiction au vol en cours en fonction de son historique. Ces capacités d'adaptations ont permis d'obtenir de très nettes améliorations en terme d'exactitude et d'incertitude, tout en gardant un modèle simple.

En parallèle du développement de nouveaux modèles, l'influence de la position géographique de l'avion sur les prédictions a été étudiée afin d'améliorer encore les prédictions. Cependant aucun lien direct n'a été trouvé. Une méthode d'amélioration de prédictions a posteriori a également été développée mais elle n'améliore que peu les prédictions du Ground TP et l'application aux autres modèles paraît peu pertinente. L'ensemble des résultats obtenus est également résumé dans le tableau 4.9 visible page suivante.

Nom de la méthode	Points négatifs	Points positifs
Prédiction issue du Ground TP	<ul style="list-style-type: none"> • Incertitude toujours élevée • Précision faible sur le long terme 	Bonne exactitude sur le court terme
BADA officielle	<ul style="list-style-type: none"> • Incertitude toujours élevée • Léger biais sur la médiane de l'erreur 	<ul style="list-style-type: none"> • Méthode simple • Un peu moins incertaine que le Ground TP
Extrapolation linéaire	Très mauvaise précision	<ul style="list-style-type: none"> • Utilise uniquement les relevés de position • Incertitude relativement faible
Extrapolation quadratique	<ul style="list-style-type: none"> • Très mauvaise précision • Prédiction instable au cours du temps 	Utilise uniquement les relevés de position
BADA apprise adaptative	Nécessite une phase d'apprentissage	<ul style="list-style-type: none"> • Très bonne précision • Incertitude améliorée • Résultat obtenu facile à comprendre • Capable de s'adapter à un vol en particulier
Correction de prédiction	<ul style="list-style-type: none"> • Gains très faibles sur la précision • Seule l'application sur les prédictions issues de Topsyky semble utile 	Applicable à toutes les méthodes

FIGURE 4.9 – Tableau récapitulatif des performances des différentes méthodes.

Organisation de mon stage

Étant inscrit au le parcours recherche proposé par l'UTC, mon stage de fin d'étude se devait de comporter une forte composante recherche. Au début du stage, seules les premières taches étaient imposées ce qui m'a permis d'être très libre par la suite quant aux pistes explorées. Ainsi, une fois l'étude sur le Ground TP faite, la seule consigne fut de développer des méthodes permettant d'obtenir de meilleurs résultats. Grâce à cette liberté, il a été possible d'envisager beaucoup de solutions différentes et d'effectuer des travaux de recherche.

L'élaboration d'une solution se déroulait globalement en trois phases :

1. Phase de questionnement technique
2. Phase de développement
3. Phase d'étude des résultats obtenus

Tout d'abord a lieu une phase de questionnement technique. Il s'agit de penser la structure de la méthode et sa possible évolution en plus des algorithmes sous-jacents. Pour valider les algorithmes utilisés, l'expertise de Gaétan Marceau-Caron, actuellement en thèse dans l'Innovation Lab a été très utile. Il m'a permis de me guider dans la recherche d'informations sur les différents outils disponibles, notamment les algorithmes de *machine learning*. Après cette phase d'étude, le développement commence, suivi par une discussion des résultats. Lors de l'étude des résultats, le but est double : d'une part évaluer l'efficacité de la méthode et d'autre part décider de la marche à suivre à l'avenir. L'efficacité est mesurée via les graphiques présentés dans la partie précédente 4. À partir de ces résultats, les différentes évolutions possibles sont discutées : soit il existe des améliorations qui semblent prometteuses pour la méthode, soit une autre méthode doit être envisagée. Il s'agit d'un processus itératif qui pourrait se rapprocher des méthodes agiles de management.

Tout au long du stage, j'ai présenté les résultats obtenus à mon tuteur Areski Hadjaz. Il s'agit de faire le point toute les 4 semaines environ pour connaître les résultats obtenus, les pistes explorées et les perspectives d'évolution. Dans la suite est présentée la chronologie des travaux réalisés ainsi que les temps forts qui ont marqués le stage.

5.1 Déroulement chronologique

Dans cette partie, le but est de retracer l'évolution des travaux effectués ainsi que le temps affecté à ces travaux. Lors de mon arrivée, l'Innovation Lab était en pleine préparation pour le *World ATM Congress* qui se tient à Madrid tous les ans. L'Innovation Lab y a présenté plusieurs projets sur des interactions homme-machines innovantes. Dans ce contexte, j'ai pu participer à un projet pendant 2 semaines sur les interactions avec un kinect¹. Le but de ma participation était d'améliorer la reconnaissance en temps réel d'un Ipad dans les images capturées par le kinect. Un algorithme était déjà implémenté par un ancien stagiaire et il s'agissait d'en améliorer les performances et de proposer de nouvelles techniques de détections. Malheureusement, malgré les améliorations apportées, le projet n'est pas arrivé à la hauteur des espérances à cause de problèmes de précision et de stabilité.

1. Le kinect est un périphérique développé par MicrosoftTM composé de microphones, d'une caméra classique et d'un capteur de profondeur 3D.

Une fois ce premier travail terminé, le travail sur les prédictions de trajectoires a pu commencer. La première étape du stage fut l'étude des prédictions faites par le Ground TP via l'enregistrement de Dublin juin 2011. L'enregistrement se trouvant sous la forme d'un fichier de format *xml*, j'ai cherché un parseur en C++ et j'ai trouvé Pugixml[4]. Il a fallu environ 3 semaines pour mettre en place l'algorithme de parcours et d'extraction des données et pouvoir visualiser les premiers résultats sous Excel. Suite à cela, la méthode d'amélioration de prédictions existantes a été mise en place.

Après l'amélioration des prédictions, ce sont les techniques d'extrapolation qui ont été développées. Les extrapolations linéaire et quadratique non contrainte ont été implémentées grâce à la bibliothèque Armadillo[8]. L'ajout de contraintes a été nécessaire car les résultats obtenus n'étaient pas satisfaisant. Ceci a nécessité un long travail de recherche pour trouver un solveur capable de résoudre ce genre de problème. En même temps que ce travail, des scripts R permettant la visualisation des résultats ont été créés. Tout ceci a duré environ 4 semaines.

L'étape suivante a été l'implémentation du modèle BADA tabulaire pour effectuer des prédictions. Le modèle tabulaire étant simple, ce travail n'a duré qu'une semaine. Après l'implémentation du modèle BADA tabulaire, c'est l'apprentissage du modèle BADA qui a été développé. Cette phase, de 3 semaines, a été relativement longue car il a été nécessaire de se familiariser avec les algorithmes d'apprentissages. Une fois la structure d'apprentissage mise en place, ce n'est pas la séparation sur le type d'avion qui a été mise en place mais le modèle BADA adaptatif. En effet, à ce moment du stage, l'enregistrement de Malmö n'était toujours pas disponible. La conception du modèle BADA adaptatif a été longue, environ 5 semaines, pour plusieurs raisons. Tout d'abord, il était nécessaire d'étudier en précision les taux de montée observés, puis une restructuration profonde du code a été nécessaire avant de pouvoir tester la méthode.

Puis, j'ai pu avoir accès à l'enregistrement de Malmö. S'est alors posée la question de la séparation du type d'avion et de la validation entre aéroports. Il a également fallu apporter quelques modifications mineures dans le *parseur xml*. L'ajout de la séparation par type d'avion et des différents tests de validations entre aéroports a duré environ 3 semaines.

Enfin s'est posée la question de la corrélation de l'erreur avec une position géographique. À cette occasion, j'ai utilisé le logiciel Gephi. Ce fut un travail assez rapide, d'une semaine, car je connaissais déjà Gephi et il n'y a eu que l'export des données qui a posé problème.

Lors des dernières semaines, mon temps a été consacré à la passation de la connaissance et des travaux réalisés. Il a fallu finaliser la documentation du projet pour en faciliter la prise en main, en plus des commentaires dans le code. Enfin quelques corrections mineures ont été apportées au projet.

5.2 Temps forts du stage

Tout au long du stage, plusieurs événements ponctuels se sont déroulés et ont permis d'explorer d'autres facettes du travail d'ingénieur. Il y a eu les réunions internes à l'équipe, appelées LabCom qui ont lieu tous les 2 mois environ. Ce sont de courtes réunions de maximum 1h30 où chaque personne présente de manière succincte l'avancement du travail sur son ou ses projet ainsi que les futurs travaux. Ce type de réunion permet d'éviter l'isolement d'une personne dans son projet et

favorise les interactions entre les différents membres de l'équipe. Ce fut l'occasion pour moi de découvrir les projets de toute l'équipe et également de présenter le mien.

Le domaine de l'ATM est un domaine qui est déroutant lorsqu'on l'aborde pour la première fois. Les premières semaines nécessitèrent un temps d'adaptation. Pour mieux comprendre comment fonctionne le contrôle aérien, il m'a été permis de visiter l'environnement des contrôleurs à l'aéroport Charles-de-Gaulle. Ce fût une visite sur une journée avec d'autres stagiaires de l'Innovation Lab encadrée par un contrôleur/formateur. Grâce à cette journée, j'ai découvert le découpage en secteurs de l'espace aérien ainsi que la charge de travail d'un contrôleur. Cela m'a également permis de mesurer l'utilité des prédictions pour les contrôleurs et de mieux comprendre l'intérêt de mon sujet de stage.

Le dernier événement marquant de ce stage fut un exercice de communication. Thales encourage les différents services à présenter les travaux effectués à travers des sessions de présentation appelés "Tell me about". Une fois encore, le but est d'encourager les interactions et de donner plus de visibilité à de petites structures comme l'Innovation Lab. Ces sessions de présentation durent en moyenne deux heures durant lesquelles les visiteurs, des employés de Thales, découvrent les différents projets. Pour le "Tell me about ATM innovation" dédié à l'Innovation Lab, les visiteurs étaient accueillis par une présentation générale de l'Innovation Lab puis étaient libres de circuler parmi les stands le reste du temps. Chaque membre de l'équipe présentait son projet sur son stand. L'objectif est d'avoir une présentation assez courte, une vingtaine de minutes, pour laisser le temps aux visiteurs de poser des questions et surtout de voir les autres stands. Les deux sessions furent une réussite avec environ 30 personnes à chaque session.

Pour ma présentation, j'ai choisi de présenter le principe de fonctionnement des méthodes développées. Pour la forme, j'ai utilisé reveal.js[3] qui permet de réaliser des diapositives plus interactives que Powerpoint. Un des défis de ces présentations est de réussir à s'adapter aux visiteurs qui ne sont pas forcément familiers du domaine.

Conclusion

Lors de ce stage, j'ai eu, avec l'aide de mon tuteur, à conduire une étude sur les performances des méthodes existantes de prédiction des points caractéristiques d'un plan de vol d'avion. Une fois l'étude finie, de nouvelles méthodes ont été développées. L'objectif du stage était de proposer des méthodes innovantes capables de rivaliser avec l'existant ou même de surpasser l'existant.

Au début de ce stage, deux méthodes existantes ont été étudiées pour ainsi permettre d'avoir une référence. Suite à cette étude et à la phase d'adaptation au domaine, j'étais donc en mesure d'imaginer de nouvelles solutions et de les comparer à ce qui existe déjà. De nombreuses méthodes ont été testées ce qui a permis d'approcher un même problème sous différents angles. Toutes les techniques ne sont pas aussi performantes que ce qui existe déjà mais le modèle BADA adaptatif permet déjà d'obtenir de meilleurs résultats.

Bien entendu il reste des pistes à explorer pour améliorer encore la solution développée. Les prédictions sur le temps d'arrivée méritent d'être plus approfondies. Par ailleurs, il faudrait étudier plus en détail l'influence de l'aéroport sur le comportement des avions. Peut-être que cette étude montrera la nécessité d'ajouter des paramètres lors de l'apprentissage. Cependant pour faire cela, il est nécessaire d'avoir des données provenant d'autres aéroports. Il pourrait aussi être judicieux d'étudier les données d'un aéroport sur une période plus longue pour vérifier que le modèle s'adapte bien aux différentes conditions météorologiques selon les saisons. Enfin il faudrait étudier la précision des prédictions à plus longs termes, plus d'une heure à l'avance. Là encore, le manque de données à l'heure actuelle empêche ces travaux.

Au final, le stage en lui-même a été une expérience très enrichissante. Tant techniquement avec des problématiques liées à l'ATM, domaine que je ne connaissais pas avant le stage, qu'humainement. En effet, les échanges constants avec tous les membres de l'Innovation Lab m'ont permis d'appliquer mes connaissances à un projet sur le long terme et de m'ouvrir à d'autres problématiques. Ce stage chez Thales m'a également fait connaître un environnement stimulant propice à l'innovation.

Bibliographie

- [1] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi : An open source software for exploring and manipulating networks, 2009.
- [2] L. D. Gaspero. Quadprog++ 1.2. : C++ library for quadratic programming, 2009. <http://sourceforge.net/projects/quadprog/>.
- [3] Hakim El Hattab. reveal.js : Html presentations made easy, 2012. <http://lab.hakim.se/reveal-js/#/>.
- [4] Arseny Kapoulkine. pugixml : Light-weight, simple and fast xml parser for c++ with xpath support, 2012. <http://pugixml.org/>.
- [5] Chen-Pan Liao. Boxplot and a probability density function of a normal population, 2011. http://en.wikipedia.org/wiki/File:Boxplot_vs_PDF.svg.
- [6] Équipe de Mathématiques Appliquées. MT09-Analyse numérique élémentaire. UTC, 2007.
- [7] R Core Team. R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. <http://www.R-project.org>.
- [8] Conrad Sanderson. Armadillo : An Open Source C++ Linear Algebra Library for Fast Prototyping and Computationally Intensive Experiments. Technical report, NICTA, 2010. <http://arma.sourceforge.net/>.

Table des figures

2.1	Capture d'écran de TopSky	6
3.1	Exemple de profil de plan de vol avec les points caractéristiques.	8
3.2	Exemple de boîte à moustaches[5].	9
4.1	Graphique de l'erreur de la prédiction faite par le Ground TP au cours du temps. . .	11
4.2	Graphique de l'erreur de la prédiction du TOC faite par la BADA au cours du temps. 12	
4.3	Graphique de l'erreur de la prédiction du TOC faite par la BADA en fonction de l'altitude.	13
4.4	Exemple de prédictions faites par extrapolations.	15
4.5	ROC au cours de la montée, pour l'ensemble des vols de l'enregistrement de Dublin juin 2011.	17
4.6	Résumé des résultats obtenus par le modèle BADA appris sans historique par rapport au modèle BADA.	18
4.7	Évolution d'un modèle BADA statique à un modèle adaptatif.	19
4.8	Résumé des résultats obtenus par le modèle BADA appris avec historique par rapport au modèle BADA.	20
4.9	Tableau récapitulatif des performances des différentes méthodes.	24
A.1	Répartition du temps d'atteinte du TOC en fonction de son altitude en FL. Chaque point représente une trajectoire observée.	31
A.2	Évolution de l'erreur de la prédiction faite par le Ground TP corrigé au cours du temps.	31
A.3	Processus de prédiction par extrapolations. Début de la montée.	32
A.4	Processus de prédiction par extrapolations. Milieu de la montée.	32
A.5	Processus de prédiction par extrapolations. Fin de la montée.	33
A.6	Évolution de l'erreur de la prédiction faite par extrapolation linéaire au cours du temps.	33
A.7	ROC au cours de la montée pour deux types d'avions. Enregistrement de Malmö novembre 2012	34
A.8	Évolution de l'erreur de la prédiction faite par le modèle BADA adaptatif au cours de la montée.	34

Annexes

Temps d'atteinte du TOC en fonction de son altitude

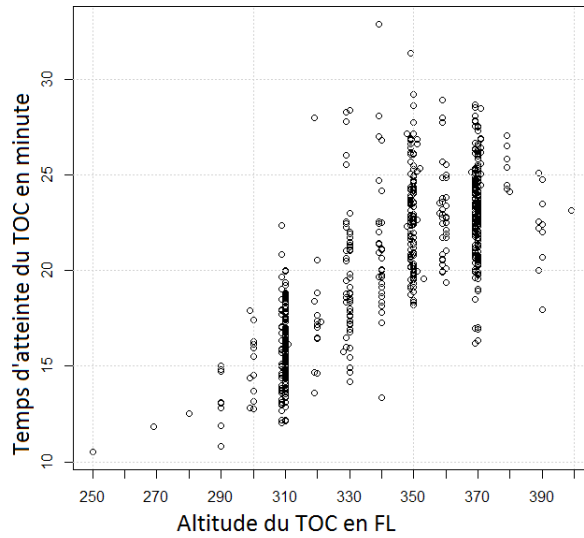


FIGURE A.1 – Répartition du temps d’atteinte du TOC en fonction de son altitude en FL. Chaque point représente une trajectoire observée.

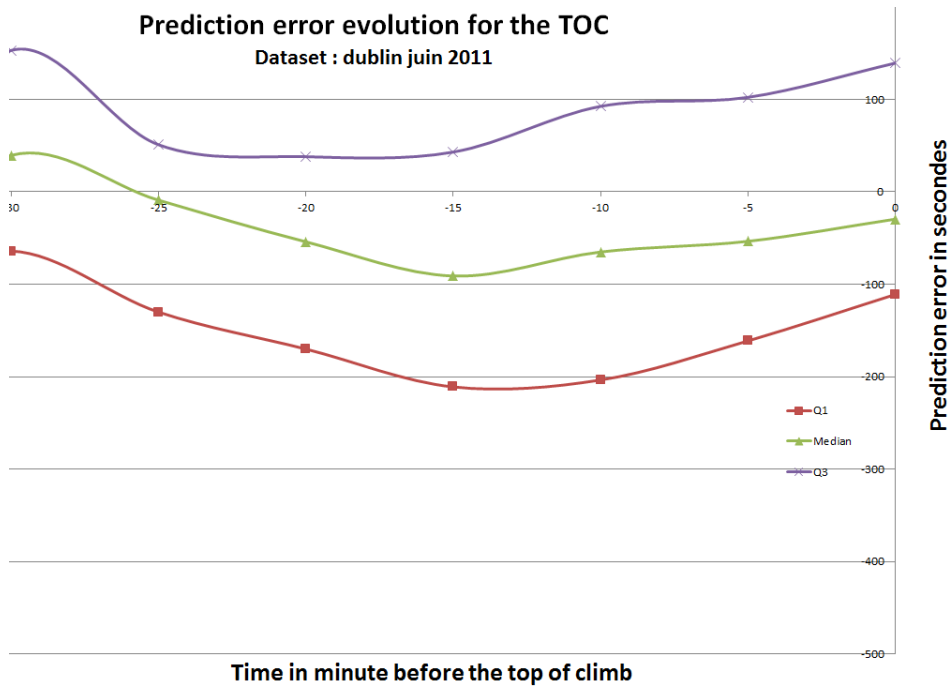


FIGURE A.2 – Évolution de l’erreur de la prédiction faite par le Ground TP corrigé au cours du temps.

Processus de prédiction par extrapolation

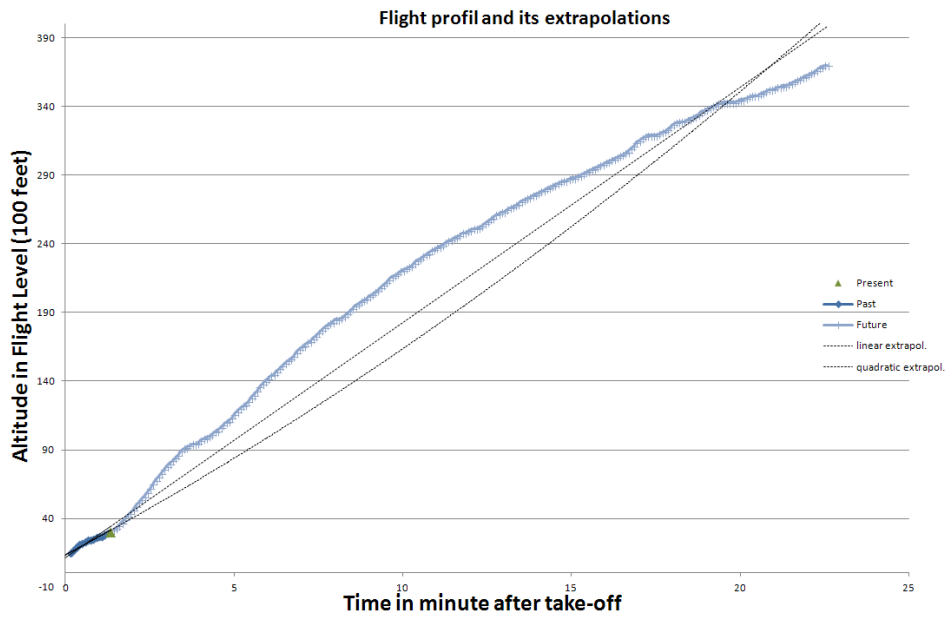


FIGURE A.3 – Processus de prédiction par extrapolations. Début de la montée.

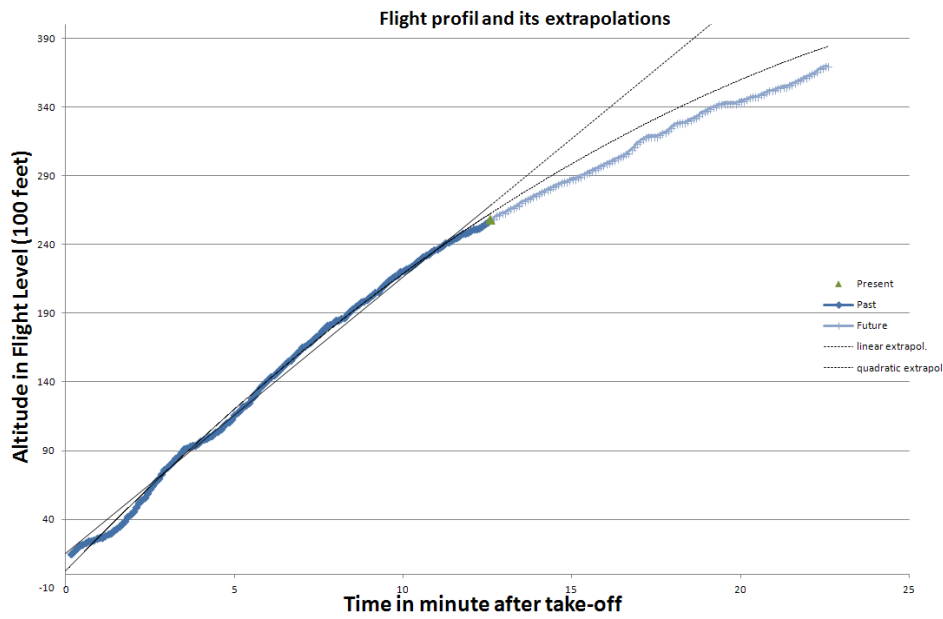


FIGURE A.4 – Processus de prédiction par extrapolations. Milieu de la montée.

Comparaison du ROC par avion

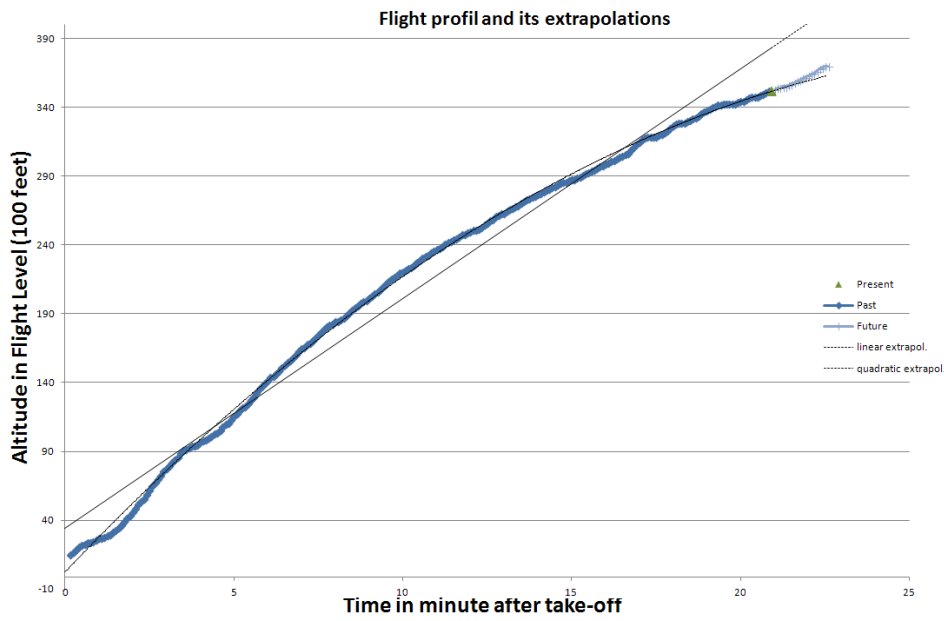


FIGURE A.5 – Processus de prédiction par extrapolations. Fin de la montée.

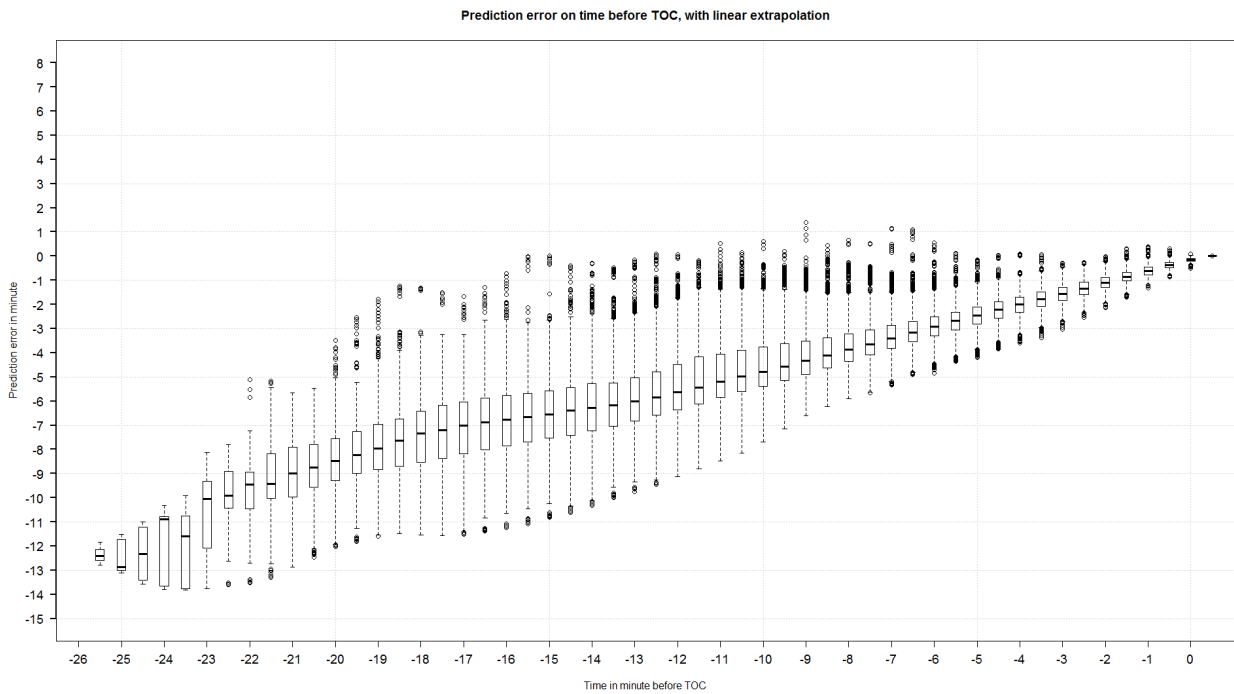


FIGURE A.6 – Évolution de l'erreur de la prédiction faite par extrapolation linéaire au cours du temps.

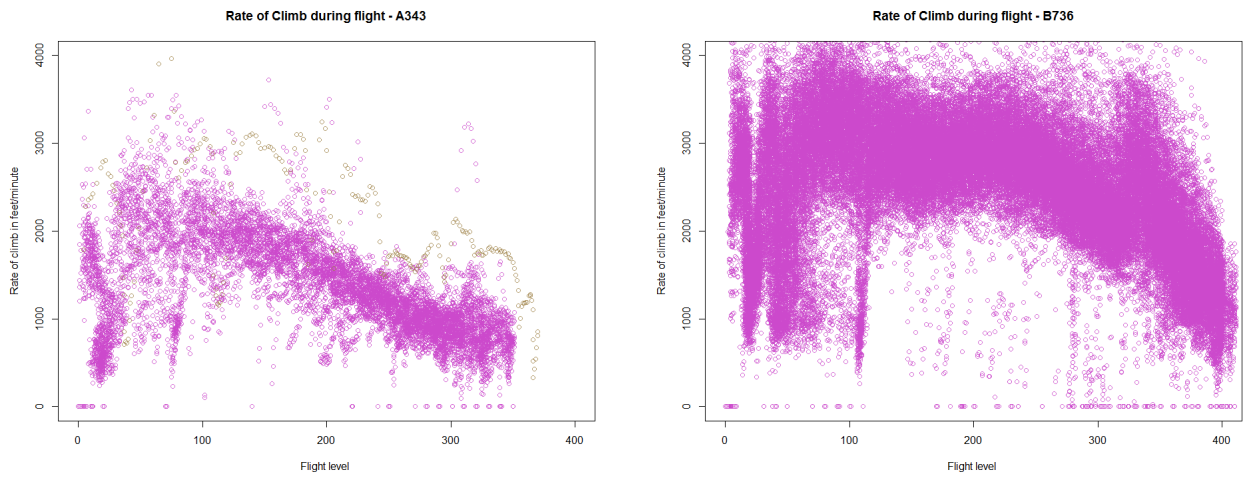


FIGURE A.7 – ROC au cours de la montée pour deux types d’avions.
Enregistrement de Malmö novembre 2012

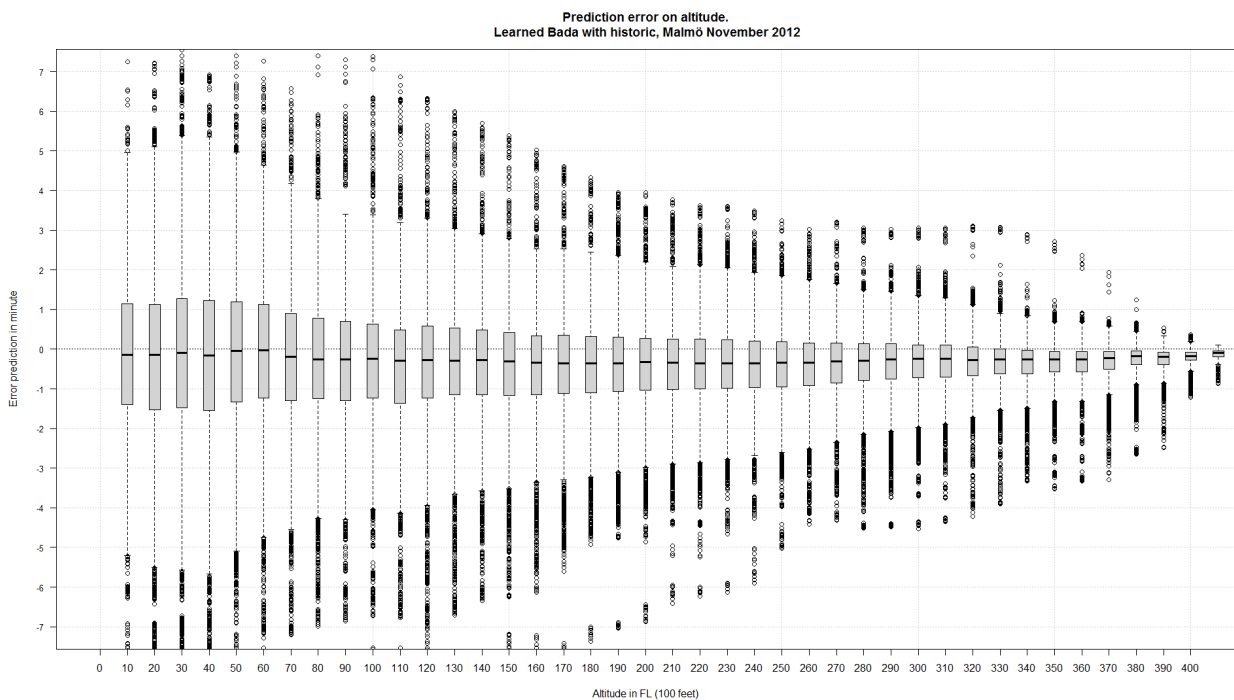


FIGURE A.8 – Évolution de l’erreur de la prédiction faite par le modèle BADA adaptatif au cours de la montée.