Classification non supervisée de courbes basée sur l'information au second ordre : détection de la dégradation de l'état de pistes d'atterrissage

Stéphane Puechmorel, Florence Nicol, Cindie Andrieu, Baptiste Gregorutti

Classification non supervisée de courbes basée sur l'information au second ordre : détection de la dégradation de l'état de pistes d'atterrissage.

Stéphane Puechmorel [1] & Florence Nicol [1] & Cindie Andrieu [2] & Baptiste Gregorutti [2]

[1] *Université Fédérale de Toulouse, ENAC*
31055 Toulouse
prenom.nom@enac.fr

[2] *Safety Line*
75015 Paris
prenom.nom@safety-line.fr

Résumé. Sur la plupart des plateformes aéroportuaires, les déplacements des aéronefs sont enregistrés en continu par des radars de surveillance et les trajectoires ainsi obtenues peuvent être utilisées pour détecter ou prévenir des incidents lors du roulage, en particulier des dérapages. Cependant, l'exploitation de ces données est rendue difficile par le fait que les algorithmes statistiques connus, même ceux basés sur des données fonctionnelles, ne sont pas capables de distinguer les situations réellement dangereuses des déviations au comportement standard qui sont sans gravité. Dans cette étude, nous proposons un changement de paradigme en représentant les trajectoires comme des points dans un espace de formes. Dans ce nouveau cadre, il est possible de construire des distances de type Finsler, prenant explicitement en compte les dérivées secondes des trajectoires, et qui sont en mesure de bien séparer les trajectoires présentant un dérapage de celles dont la déviation par rapport au comportement nominal n'est pas due à un manque d'adhérence. Cette métrique est ensuite utilisée dans des méthodes de clustering pour détecter un mauvais état de la piste d'atterrissage. Des résultats sur des jeux de données de trajectoires synthétiques et réelles sont présentés ainsi qu'une comparaison avec des métriques existantes.

Mots-clés. Similarité entre courbes, espace de formes, données fonctionnelles, classification non supervisée de courbes, sécurité aéroportuaire.

Abstract. In air transportation, especially in airport safety, radar tracks are continuously recorded and may be used for detecting incidents on airport surface. However, all known statistical algorithms, even those based on functional data, are unable to distinguish between a safety critical flight and another one departing from standard behavior, but otherwise safe. In this work, we propose a change of paradigm by representing curves as points in a shape manifold. In this framework, it is possible to use Finsler distances between shapes that explicitly take into account the second derivative and can be able to correctly detect skid situations from deviant trajectories that cannot be considered as a slipped trajectory. This metric is next used in curve clustering for detecting bad runway conditions. Some results on datasets of synthetic and real trajectories are presented, as well as a comparison of existing metrics.

Keywords. Similarity measure, shape manifold, functional data analysis, curve clustering, airport safety.

1 Introduction

Les grands projets fédérateurs SESAR en Europe et Nextgen aux Etats-Unis ont été initiés afin de répondre à une augmentation conséquente du trafic aérien mondial jusqu'à l'horizon 2030. Le rapport d'Airbus Industries (Perspectives 2030) table sur un accroissement du trafic de passagers dans les vingt prochaines années (4,2% de 1986 à 2010 - 4,8% de 2010 à 2030), ce qui aboutirait à un doublement du nombre de passagers en 2030 par rapport à 2010. Cette augmentation du trafic aérien aura pour conséquence une saturation des aéroports et rendra plus complexe la gestion du trafic aussi bien en route qu'au sol. Dans l'optique du maintien d'un haut niveau de sécurité tout en augmentant la capacité, l'exploitation des données de flux de trafic sur les plateformes aéroportuaires devient une priorité.

Dans ce contexte, la détection et la prévention d'un état d'adhérence déficient des pistes est un des points à traiter. En effet, les mauvaises conditions météorologiques comme la pluie, la neige et le givre, peuvent considérablement diminuer la qualité de surface des pistes d'atterrissage, les rendant plus glissantes. Cette perte d'adhérence augmente alors la distance d'atterrissage et rend les manoeuvres plus difficiles à effectuer. Dans certains cas extrêmes, une perte de contrôle peut même se produire. Lorsque la piste est présumée en mauvais état, la procédure actuellement appliquée consiste à inspecter directement son revêtement, ce qui implique l'envoi d'un véhicule avec un outil dédié et une interruption du trafic d'environ 20 minutes. Cette interruption provoque alors des perturbations très importantes (mises en attente, retards, etc.), ce qui va interférer avec le trafic en cours et rendre la gestion de la situation future délicate. Pour optimiser le déclenchement des interventions sur les pistes et les voies de circulation (déneigement, traitement des pistes, etc.) et ainsi limiter le temps d'inactivité de la plateforme, de nouveaux outils, basés uniquement sur l'observation des trajectoires radar des aéronefs à l'atterrissage, sont développés. L'approche retenue consiste alors à détecter les trajectoires des avions à l'atterrissage qui s'écartent notablement d'une trajectoire nominale, par exemple une ligne moyenne ou encore le centre de la piste.

Les traces radar permettent de localiser précisément tout mobile sur l'aéroport, aussi bien sur les taxiways que sur les pistes. Elles contiennent les coordonnées latitude/longitude et la vitesse de chaque objet en mouvement. Le cadre des statistiques sur données fonctionnelles est particulièrement adapté pour traiter ce type de données car il permet de tenir compte de la relation temps-position sous-jacente aux données et de leur structure interne. Des travaux préliminaires ont déjà montré la performance des algorithmes de régression, d'analyse en composantes principales et de classification fonctionnelles sur les données aéronautiques [1]. Si les outils existants permettent de tenir compte de la nature fonctionnelle des données radar, la prise en compte de l'information à la fois temporelle et spatiale est un problème difficile [2, 3]. Pour les implémenter, il faut disposer d'une notion de distance entre trajectoires, qui constitue un thème actif de recherche dans la communauté des statistiques fonctionnelles. Le plus souvent, on utilise des méthodes de recalage basées sur une norme L^p et non sur des caractéristiques purement géométriques. Le problème qui se pose alors est de déterminer une fonction de distorsion temporelle, éventuellement non linéaire, ne modifiant pas la structure intrinsèque des données que l'on cherche à identifier. Les résultats sont généralement assez décevants.

Comme alternative, l'approche retenue dans [4] utilise des techniques issues de la géométrie différentielle et des espaces de formes pour la construction d'une mesure de similarité entre trajectoires. Elle consiste à représenter des données fonctionnelles comme des points dans une variété riemannienne, appelée espace des formes, qui est formellement définie comme un quotient de la variété des immersions par un groupe de reparamétrage. Dans ce nouvel espace de représentation, des métriques riemanniennes classiques satisfaisant une propriété d'invariance globale par reparamétrisation peuvent être construites pour détecter des différences de formes entre trajectoires [5, 6]. Bien que cette approche permette de détecter un écart latéral des trajectoires, elle n'est pas suffisante pour construire un indicateur de mauvais état de la piste. En effet, étant donné qu'un phénomène de dérapage de l'avion correspond à une perte d'adhérence de la piste, il y a lieu de tenir compte par ailleurs de l'accélération : un faible profil de décélération indique que la piste ne répond pas au freinage.

Dans ce travail, l'objectif est de développer des mesures de similarité spatio-temporelles pour les trajectoires qui seront ensuite utilisées dans des méthodes de clustering en vue de détecter un mauvais état de la piste d'atterrissage. Les outils issus de la géométrie différentielle et des espaces de formes ont été utilisés pour construire une métrique géométrique permettant de prendre en compte non seulement une différence dans la forme des trajectoires d'atterrissage mais aussi dans leur profil de décélération. Une application numérique a été faite sur des données simulées et des données réelles. Les résultats obtenus ont été comparés avec des métriques existantes [7].

2 Métrique géométrique dans les espaces de courbes adaptée à la détection de phénomènes de glissance

En se basant sur les outils développés dans le cadre des espaces de formes, une approche similaire à [4] a été utilisée pour construire une nouvelle métrique entre trajectoires permettant de détecter des situations de dérapage à l'atterrissage.

L'espace de représentation des trajectoires est $Imm([0,1];\mathbb{R}^d)$, l'ensemble des immersions de l'intervalle $[0,1]$ dans $\mathbb{R}^d$, dont les points sont des trajectoires lisses γ dont la vitesse ne s'annule jamais. En quotientant cet espace par le groupe des difféomorphismes sur $[0,1]$, les trajectoires sont des points d'une variété riemannienne, appelée espace des formes. La métrique riemannienne utilisée dans [4] permet alors de prendre en compte directement les caractéristiques géométriques mais pas les profils de décélération. Notons que, puisque nous devons tenir compte de la condition de glissance, il n'est plus souhaitable d'avoir une propriété d'invariance globale par reparamétrisation. Une nouvelle métrique satisfaisant une propriété d'invariance plus faible sera obtenue.

Pour construire une métrique qui tient compte des profils de décélération, nous utilisons l'indicateur de glissance suivant, basé sur l'angle θ entre les vecteurs d'accélération et de vitesse :

$$\sin(\theta(s)) = \frac{\kappa(s)\|D_s\gamma(s)\|^2}{\|D_{ss}\gamma(s)\|} = \frac{\det\left(D_s\gamma(s), D_{ss}\gamma(s)\right)}{\|D_s\gamma(s)\|\|D_{ss}\gamma\|}, \tag{1}$$

où γ représente la trajectoire de l'avion, κ sa courbure, $D_s\gamma$ et $D_{ss}\gamma$ sa dérivée première et sa dérivée seconde relativement à la variable temporelle s. Lorsque la piste d'atterrissage répond correctement au freinage, l'accélération tangentielle sera constante et élevée par rapport à l'accélération normale. L'indicateur de glissance devrait alors être proche de zéro. Dans le cas contraire, l'angle θ sera proche des valeurs limites $\pm\pi/2$ et l'indicateur sera proche de ± 1. On peut en déduire la semi-norme suivante en un point γ :

$$M^2_{\gamma(u)} = \int_0^1 \left(\frac{|D_s u|_N|}{|D_s\gamma|} - \frac{|D_{ss} u|_{\tilde{N}}|}{|D_{ss}\gamma|} \right)^2 ds, \tag{2}$$

où N (resp. $\tilde{N}$) représente le vecteur normal à γ (resp. $D_s\gamma$) et (γ, u) est le vecteur tangent au point γ, avec $u \in \mathcal{C}^\infty([0,1], \mathbb{R}^d)$. Pour un vecteur tangent (γ, u), $M^2_{\gamma(u)}$ mesure l'énergie de déformation associée au déplacement u.

La distance minimale, ou géodésique, entre deux trajectoires γ_1, γ_2, sera déterminée comme étant le chemin d'énergie minimale entre les deux courbes. Un chemin entre deux trajectoires γ_1, γ_2 de $Imm([0,1];\mathbb{R}^d)$ peut se représenter par une application continue $\Phi\colon [0,1] \times [0,1] \to \mathbb{R}^d$ telle que $\Phi(0,\bullet) = \gamma_1(\bullet), \Phi(1,\bullet) = \gamma_2(\bullet)$ avec $D_t(\Phi)$ non nulle sur tout le domaine. On impose ici une régularité C^∞ à Φ. Dans ce cadre, on remarquera que si $t \in [0,1]$ est fixé, le champ de vecteurs $D_t\Phi(t,\bullet)$ définit un vecteur tangent à la courbe $\Phi(t,\bullet)$. L'objectif étant de mesurer l'écart entre une trajectoire γ et une trajectoire de référence que l'on note γ_{ref}. Cela revient à déterminer un chemin $\Phi\colon [0,1] \times [0,1] \to \mathbb{R}^d$ tel que $\Phi(0,\bullet) = \gamma(\bullet), \Phi(1,\bullet) = \gamma_{ref}(\bullet)$ dont l'énergie de déformation $E(\Phi)$ est minimale. Cette énergie de déformation est calculée en se basant sur la semi-métrique (2) appliquée au vecteur tangent $D_t\Phi$:

$$E(\Phi) = \int_0^1 |D_t\Phi(t,0)|^2 + |D_t\Phi(t,1)|^2 dt \tag{3}$$

$$+ \int_0^1 \int_0^1 |D_s D_t \Phi|^2 |D_s\Phi| ds dt \tag{4}$$

$$+ \lambda \int_0^1 \int_0^1 \left(\frac{|D_s D_t \Phi|_N|}{|D_s\Phi|} - \frac{|D_{ss} D_t \Phi|_{\tilde{N}}|}{|D_{ss}\Phi|} \right)^2 |D_s\Phi| ds dt, \tag{5}$$

où le paramètre $\lambda > 0$ permet de régler l'importance de la forme et de la glissance dans l'énergie. Les trajectoires étant des courbes ouvertes, le premier terme (3) permet de tenir compte de leurs extrémités. Le second terme (4) correspond à une métrique riemannienne classique de type Sobolev qui détecte des différences de formes. Elle est rendue nécessaire pour éviter une dégénérescence issue de la semi-métrique (1). Enfin, le troisième terme (5) permet de prendre en considération à la fois la forme mais aussi la glissance. Notons que le terme $|D_s\Phi|$ permet l'indépendance vis à vis du représentant dans $Imm([0,1];\mathbb{R}^d)$ et que la définition de l'énergie n'est pas invariante par reparamétrisation mais possède la propriété d'invariance par reparamétrisation affine. Finalement, la métrique globale est de type Finsler, en raison de la présence du terme (5).

Pour finir, l'énergie de déformation permet de calculer la distance minimale, ou géodésique, entre les deux courbes γ et γ_{ref}, tout en tenant compte du profil de décélération

$$\Phi_0 = argmin_{\Phi \in I} E(\Phi), \tag{6}$$

où I est l'ensemble des chemins allant de γ to γ_{ref}. En pratique, une approximation du chemin Φ est donnée par ses valeurs Φ_{ij} sur une grille régulière de $[0,1] \times [0,1]$, les dérivées étant approchées par différences finies. Le problème de détermination du chemin géodésique Φ_0 se ramène à un programme d'optimisation non linéaire. Un algorithme de type LM-BFGS [8] permet une résolution efficace pour des programmes de grande taille. L'énergie minimale $E(\Phi_0)$ est ensuite implémentée dans une phase de clustering.

3 Application à des trajectoires d'atterrissage simulées et des observations de données radar

Dans la pratique, il est quasiment impossible de s'assurer qu'une trajectoire observée qui s'écarte de la trajectoire nominale est le résultat d'une mauvaise condition d'adhérence de la piste. En effet, il faudrait disposer de données radar juste après ou avant avoir effectué une mesure physique de l'état de la piste. Pour cette raison, il a été rendu nécessaire de simuler des trajectoires d'atterrissage de façon réaliste afin de valider les performances de la métrique proposée. Un simulateur a été codé en Java pour des pistes d'atterrissage virtuelles possédant un coefficient d'adhérence réglable. Toutes les trajectoires d'atterrissage générées sont alors labélisées comme étant normale ou en situation de dérapage. Les trajectoires d'atterrissage et de roulage sont simulées suivant un cercle représentant la trajectoire nominale. Un effet de dérapage est induit au début de la trajectoire d'atterrissage qui est ensuite corrigé par les régulateurs du simulateur, ainsi que cela se passe en réalité dans un avion. L'échantillon de trajectoires simulées est composé de 16 trajectoires en situation de dérapage et de 284 trajectoires normales. Après avoir calculé les distances entre trajectoires, nous avons appliqué l'algorithme de clustering des K-medoïdes pour détecter ces deux classes. Nous avons implémenté plusieurs types de métriques. Tout d'abord, deux métriques géométriques : la métrique de Mumford [9] et notre métrique de glissance développée dans la section précédente. Les résultats de ces deux métriques ont été comparés avec des métriques alternatives décrites dans [7] : des distances basées sur la déformation temporelle comme DTW, LCSS, EDR et ERP, mais aussi des métriques basées sur la forme comme la distance de Hausdorff, de Fréchet et la SSPD. Appliquées aux trajectoires simulées suivant un cercle, seules notre métrique de glissance et la distance ERP permettent de détecter correctement les deux classes. On remarque que les deux métriques EDR et LCSS produisent de très mauvais résultats : mise à part une seule trajectoire, toutes les autres sont classées ensemble. Quant aux autres métriques, elles produisent bien une classe ne contenant que des trajectoires normales mais la seconde classe caractérisant les trajectoires en situation de dérapage contient aussi des trajectoires normales avec un taux élevé de fausse alarme allant de 42% à 48%. Afin d'éprouver la robustesse des résultats aux déformations aléatoires apportées sur la forme des trajectoires simulées, nous avons légèrement déformé les trajectoires au moyen d'un coefficient aléatoire pour obtenir des trajectoires suivant des formes de type ellipsoïde. Alors que la métrique de glissance permet toujours de détecter correctement les deux classes de trajectoires, la métrique ERP n'est plus à même de le faire : le taux de fausse détection atteint maintenant 50%.

Par ailleurs, un ensemble de 357 trajectoires d'atterrissage a été observé et les deux métriques géométriques ont été implémentées pour déterminer quatre classes de trajectoires. Comme attendu, la métrique de Mumford permet de distinguer des différences de formes : une classe contenant trois trajectoires atypiques, deux classes représentant les trajectoires ayant un comportement non dangereux (suivant la trajectoire de référence ou présentant de faibles déviations) et enfin une classe reflétant une situation dangereuse avec des trajectoires présentant toutes un fort écart latéral à la trajectoire nominale. Quant à la nouvelle métrique de glissance, elle tient compte non seulement des différences de formes mais intègre aussi l'information sur la décélération. On retrouve bien les quatre classes de trajectoires identifiées par la métrique de Mumford. Cependant, nous observons qu'une des trajectoires déviantes n'est plus classée comme étant dangereuse. Après un examen du profil de décélération de cette trajectoire, il s'avère que la déviation par rapport au comportement nominal n'est pas due à un manque d'adhérence de la piste. Cette nouvelle métrique présente donc deux intérêts majeurs. Tout d'abord, elle permet de détecter des déviations latérales à une trajectoire nominale. Enfin, elle permet d'éviter les faux positifs en tenant compte de l'information sur la décélération.

Références

[1] B. Gregorutti, “Forêts aléatoires et sélection de variables : analyse des données des enregistreurs de vol pour la sécurité aérienne,” Thèse de Doctorat, Université Pierre et Marie Curie - Paris VI, 2015. [Online]. Available : https ://tel.archives-ouvertes.fr/tel-01146830

[2] J. Ramsay and B. Silverman, *Functional Data Analysis*, ser. Springer Series in Statistics. Springer, 2005. [Online]. Available : https ://books.google.co.uk/books ?id=mU3dop5wY_4C

[3] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis : Theory and Practice*, ser. Springer Series in Statistics. Springer New York, 2006. [Online]. Available : https ://books.google.fr/books ?id=lMy6WPFZYFcC

[4] C. Andrieu, B. Gregorutti, F. Nicol, and S. Puechmorel, “Espaces de courbes pour l’analyse de données aéronautiques,” in *48èmes Journées de Statistique de la SFDS*, 2016.

[5] M. Bauer, M. Bruveris, and P. W. Michor, “Overview of the geometries of shape spaces and diffeomorphism groups,” *J. Math. Imaging Vis.*, vol. 50, no. 1-2, pp. 60–97, Sep. 2014. [Online]. Available : http ://dx.doi.org/10.1007/s10851-013-0490-z

[6] P. W. Michor and D. Mumford, “Riemannian geometries on spaces of plane curves,” *J. Eur. Math. Soc. (JEMS)*, vol. 8, pp. 1–48, 2006.

[7] P. C. Besse, B. Guillouet, J.-M. Loubes, and F. Royer, “Review and perspective for distance-based clustering of vehicle trajectories,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 11, pp. 3306–3317, Nov 2016.

[8] D. C. Liu and J. Nocedal, “On the limited memory BFGS method for large scale optimization,” *Mathematical Programming*, vol. 45, no. 1, pp. 503–528, Aug 1989.

[9] P. W. Michor and D. Mumford, “Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms.” *Documenta Mathematica*, vol. 10, pp. 217–245, 2005. [Online]. Available : http ://eudml.org/doc/125727