

# CRÉATION D'UN MODÈLE Statistique DE PRÉDICTION DE POSITION D'UN AVION

## 1 Le contexte des outils CDM (Collaborative Decision Making)

L'un des objectifs principaux de la DSNA (Direction des services de la Navigation aérienne) est la modernisation de la navigation aérienne.

Le plan de modernisation technique prend en compte le programme de recherche européen " SESAR", volet technologique du projet de " ciel unique européen ", et en particulier le cadre réglementaire défini par le " Pilot Common Project " première étape de déploiement des travaux dès 2015.

Il s'appuie sur des outils de nouvelle génération (4-Flight, ERATO...) et des procédures opérationnelles optimisées pour toutes les phases du vol (EGNOS, Airport CDM...) et plus particulièrement le projet SALTO en lien avec la présente thèse.

Le projet SALTO s'inscrit dans un programme ambitieux de modernisation des outils pour la gestion du trafic aérien en France présenté dans la Figure 1.

## UNE STRATEGIE DE MODERNISATION TECHNIQUE AMBITIEUSE

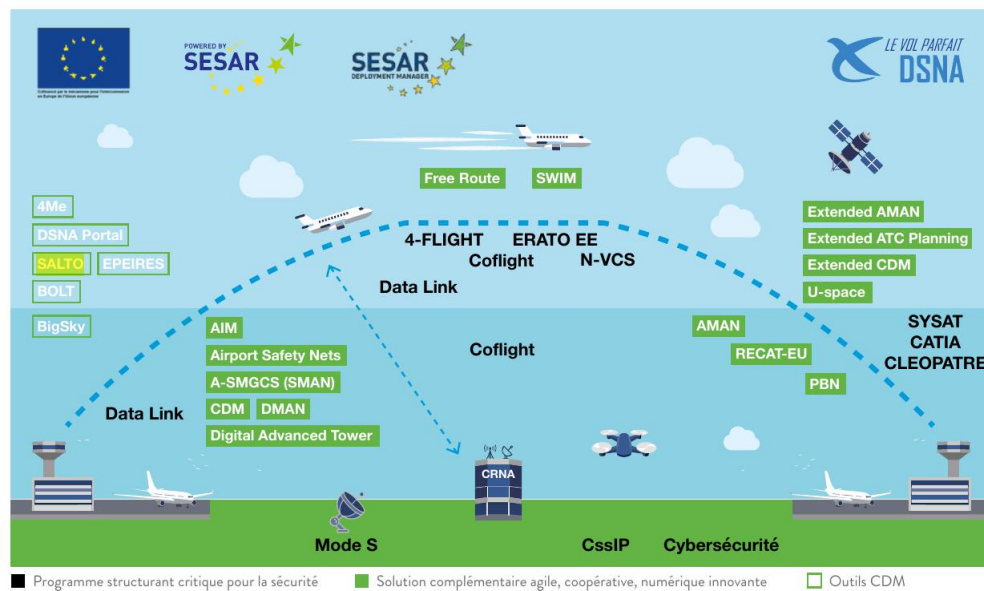


Figure 1: Projet de modernisation du trafic aérien en France

### 1.1 Présentation du projet

Le projet SALTO (Swift ATFCM/ASM Local Traffic Optimizer) est un projet dont la DGAC est maître d'ouvrage et Capgemini en est le maître d'œuvre.

SALTO permet de surveiller les secteurs de contrôles d'un CRNA.

C'est un outil adapté aux besoins des FMP (Flow Management Position). L'outil SALTO répond aux besoins des activités lors des phases stratégiques, pré-tactique, tactique et post-analyse. L'application SALTO propose 3 fenêtres principales :

1. L'ASD (Air Situation Display) représente les vols constituant le trafic et évoluant sur leur trajectoire prévue, ainsi que des outils spécifiques permettant de modifier les contextes de visualisation du trafic.
2. Le FMD (Flow Manager Display), qui à l'instar du CHMI installé sur les positions du FMP actuellement, visualise un ensemble d'informations permettant d'appréhender l'intensité du trafic sur des Traffic Volume (surveillance et visualisation du nombre de vols entrant dans le trafic volume considéré pendant un laps de temps selon un pas choisi ainsi que du nombre de vols présents dans le trafic volume considéré pendant un laps de temps et selon un pas donné).
3. La fenêtre de visualisation des vols et des régulations visualise différentes listes de vols: liste de l'ensemble des vols d'un centre, liste des régulations, fenêtre de listes de vols, de hot spot,...

Les exploitants de SALTO sont les ACDS (Assistants Chefs de Salle). Les Services Techniques assurent la supervision et l'exploitation technique du constituant.

## 1.2 Architecture logicielle du projet

La Figure 2 présente l'architecture logicielle du projet. Elle se décompose en trois éléments principaux : le serveur, le client et la supervision.

1. Le serveur : se connecte au NM (Network Manager) pour récupérer les données (vols, traffic volumes...) Adapte les données aux données au modèle SALTO Diffuse les données au client
2. Le client : permet l'affichage des informations reçus du serveur: l'occupation et les entrées dans les Traffic volumes, Liste de vols et de régulations Permet de visualiser les vols constituant le trafic avec leurs trajectoires
3. La supervision : permet de surveiller l'état des différents composants et du lien entre chacun d'eux.

## 1.3 Les fonctions

1. L'ACDS peut surveiller la charge de trafic de son centre afin de décider de l'ouverture ou fermeture de positions de contrôle pour adapter la charge de travail des contrôleurs aériens.
2. L'ACDS peut implémenter des mesures (poser des régulations, changer la trajectoire de certains vols...) afin d'atténuer une surcharge de travail avérée.
3. SALTO propose un environnement de simulation destiné à l'ACDS. Cet environnement permet de simuler l'effet des mesures à implémenter sur le trafic aérien.

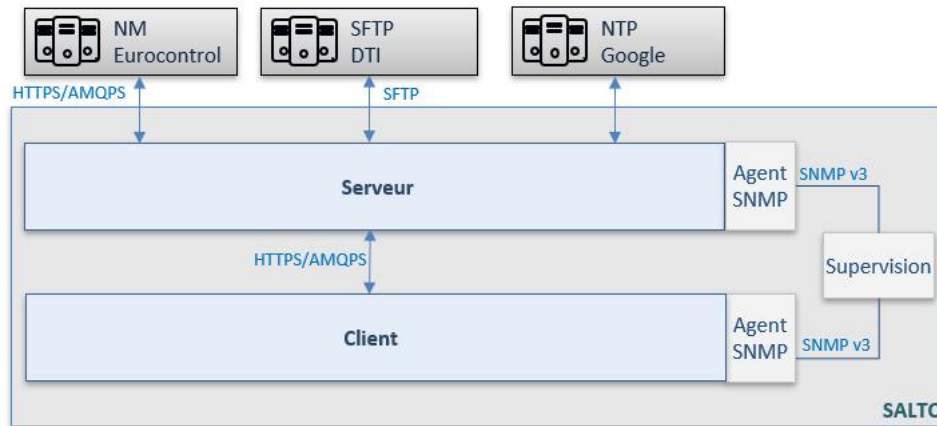


Figure 2: Architecture logique du projet Salto

#### 1.4 Les outils de visualisation du projet

Le projet possède deux environnements : un environnement opérationnel représenté dans la Figure 3 et un environnement de Simulation (la fonction "What-if") représenté dans la Figure 4

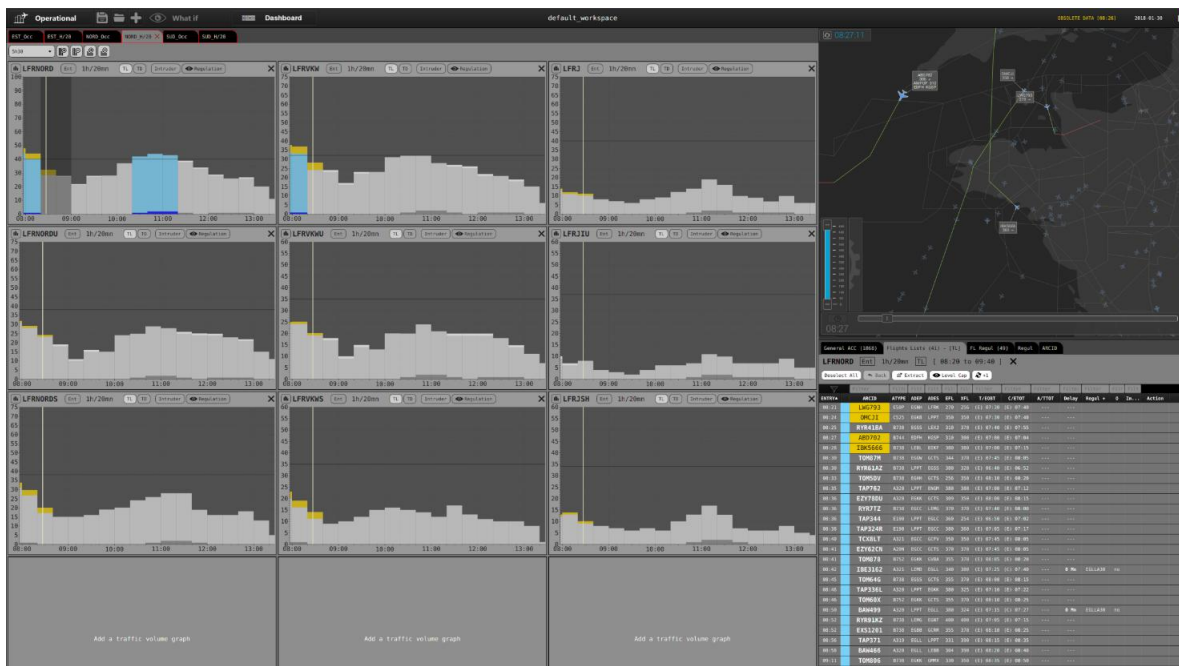


Figure 3: Environnement opérationnel du projet Salto

Ces outils opérationnels et de simulations reposent sur un calcul de prédiction de "présence" des vols dans le volume aérien considéré. La qualité de la prédiction est directement à la qualité des outils mis à disposition de l'ACDS. L'amélioration de cette prédiction aura donc des répercussions

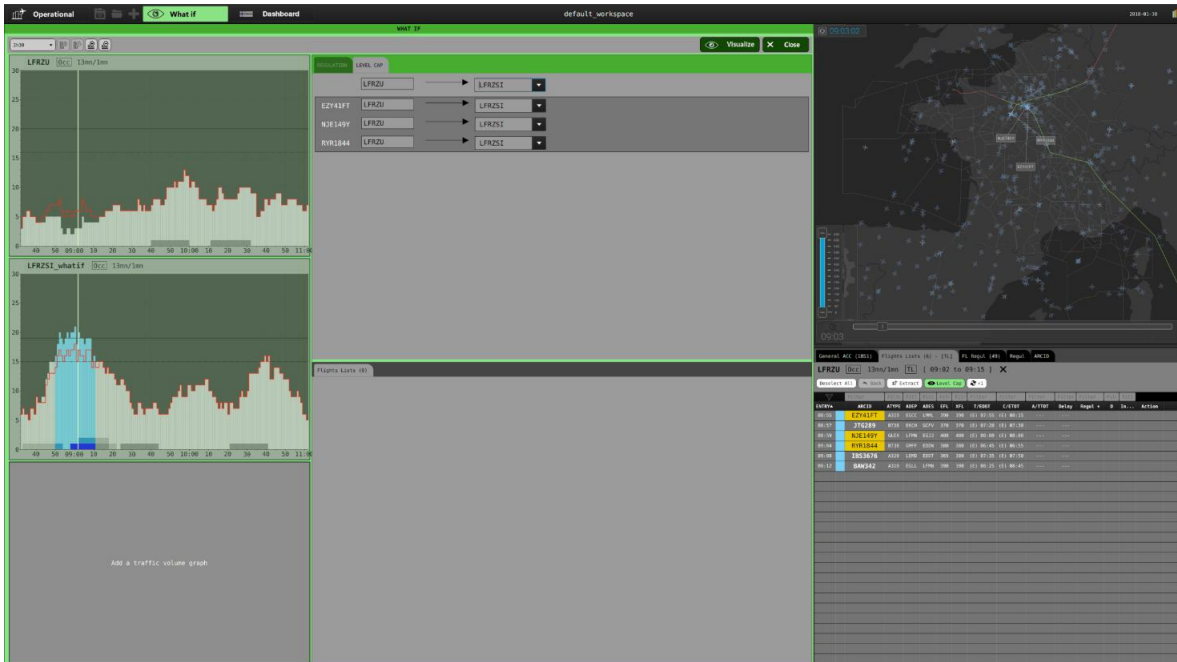


Figure 4: Environnement de simulation du projet Salto

directes sur l'efficacité des outils et la confiance des opérateurs pour les utiliser. Améliorer la prédiction est aussi un enjeu opérationnel et est un élément de leur acceptabilité des outils et ainsi favoriser la pérennité du projet SALTO.

Ajoutons que ce calcul de prédiction qui est au centre des nombreuses autres fonctionnalités cruciales notamment la fonctionnalité centrale qu'est la fonction de prédiction de trajectoire (TP). Nous espérons par les travaux menés ici améliorer le TP également?

Dans la figure 5 suivante on peut voir les prévisions de charge de trafic aérien dans le secteur aérien LFRNORD entre 8 heure et 13h30. Cette charge de trafic est exprimée en nombre de vols prévus suivant l'axe des ordonnées. On pourra remarquer la ligne à 40 vols qui représente le seuil fixé pour le paramétrage courant et la ligne verticale blanche qui indique l'heure courante.

D'après la Figure 4, l'ACDS actif doit faire attention à une monter de trafic prévu vers 10h20 jusqu'à 11h20. Cette fonction de prédiction est au centre de la fiabilité de cet outil. La confiance des opérateurs dépend en grande partie si cette prédiction est fiable.

Nous proposons dans le cadre de cette thèse d'améliorer cette fonction de prédiction. Nous allons proposer une modélisation des incertitudes afin d'associer à la prédiction un degré de confiance. Pour affiner ce degré de confiance nous utiliserons des techniques d'apprentissage pour prendre en compte l'historique de chacun des vols, capter ses habitudes et les événements récurrents de l'environnement aérien traversé (ZM activée / désactivée, phénomènes météorologiques récurrents, retards aériens spécifiques à la congestion du trafic aérien, etc.).

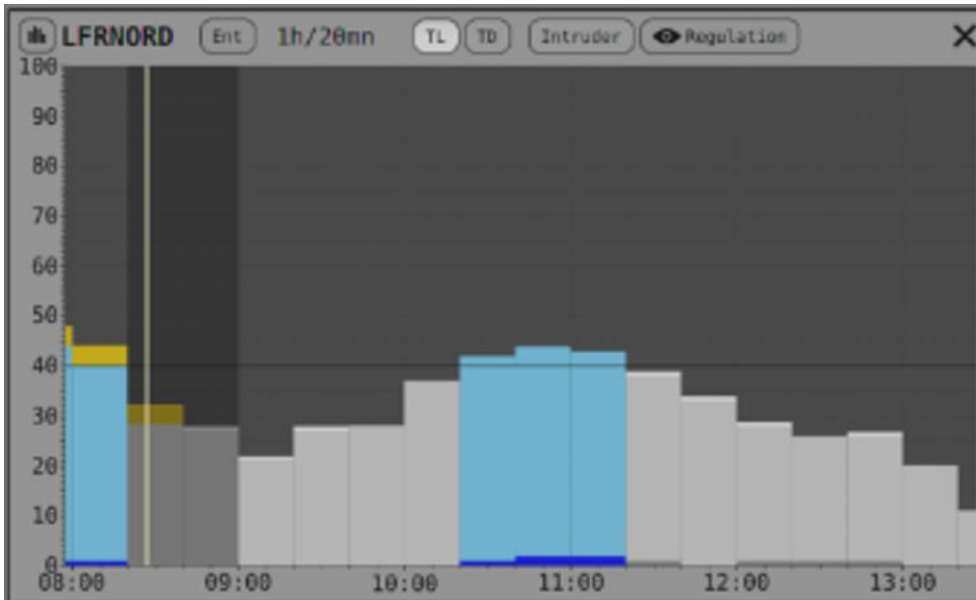


Figure 5: Prévisions de charge de trafic aérien dans le secteur aérien LFRNORD entre 8 heure et 13h30

## 1.5 Données et objectifs

Nous allons récupérer les trajectoires réelles enregistrées par le projet SALTO. Ce type de données spatio-temporelles soulève un ensemble de questions classiques en fouille de données mais plus originales lorsqu'elles sont regroupées pour constituer des trajectoires. Peut-on définir une trajectoire type ? Peut-on détecter des trajectoires anormales ? Comment regrouper les trajectoires similaires ?

Le principe d'une telle étude repose sur la création d'une distance adaptée aux propriétés particulières que présentent les données de trajectoires. Les trajectoires sont des suites de points du plan indexés par le temps. Pour pouvoir les comparer il faut prendre en compte non seulement le point de départ et le point d'arrivée, qui définissent l'itinéraire, mais également leur longueur et leur forme ainsi que la fenêtre temporelle associée.

De nombreuses distances ont déjà été développées en vue de cet objectif. Dans tous les cas les distances que nous utiliserons seront basées sur les critères suivants: la distance physique entre deux trajectoires, la forme des trajectoires (orientation, longueur) et la "dépendance temporelle".

## 2 Tâche 1: Construction d'un méta-modèle

Comme dit au paragraphe précédent la problématique majeure pour un contrôleur aérien est de pouvoir estimer à un instant  $t_0 + h$  la position d'un avion sur sa trajectoire curviligne connaissant sa position à l'instant  $t_0$ . Pour cela, nous disposons d'une base de données dans laquelle nous avons:

1.  $N$  trajectoires théoriques (celles déposées auprès d'Eurocontrol) notées:  $(P_{th}^i)_{i=1...N}$  ( $P_{th}^i$  est une fonction qui à un temps  $t$  renvoie l'abscisse curviligne de l'aéronef sur sa trajectoire)

2. les  $N$  trajectoires réellement effectuées: notées  $(P_{eff}^i)_{i=1\dots N}$
3. Des donnés météo pour chaque vol (carte de vents,...) notées  $(M^i)_{i=1\dots N}$
4. L'Etat du ciel aérien (ouverture, fermeture de certains espaces).  $(C^i)_{i=1\dots N}$ .

D'un point de vue théorique, il existe une fonction  $f$  telle que:

$$P_{eff}^i = f(P_{th}^i, M^i, C^i, \epsilon^i) \quad (2.1)$$

où  $\epsilon^i$  représente toutes les variables que nous ne contrôlons pas ou n'avons pas accès.

La première étape est de construire à partir de nos données une approximation de la fonction  $f$  classiquement appelée un méta-modèle. Il existe de nombreuses techniques de méta-modélisation [22] comme par exemple le krigeage [16] ou les méthodes basées sur des polynômes de chaos [15]. Le Krigeage a pour principale avantage d'être exact aux données d'observations et de fournir des bornes d'erreurs aux autres points. Cette méthode a été énormément étudiée et utilisée lorsque les données d'entrées et de sorties sont des réelles. Mais elle est encore à construire en présence de données fonctionnelles (ce qui est le cas dans notre problèmes). Récemment [17, 18, 1, 2] la construction de méta-modèle pour des données fonctionnelles a commencé à être investiguée dans des cas relativement simple mais jamais en présence de données très complexe comme des cartes de vents.

D'un point de vue opérationnel, il est fréquent que le controleur ait observé le début de la trajectoire réelle c'est à dire sur  $[0, t_0]$  et qu'il veuille la prédire à partir de  $t_0$ , pour cela nous devons non seulement construire un méta-modèle  $\hat{f}$  de (2.1) (qui correspond au cas  $t_0 = 0$ ) mais en fait toute une collection de méta-modèle  $(\hat{f}_t)_{t \geq 0}$  pour lesquelles nous avons une variable d'entrée supplémentaire  $P_{eff}^i$  observée uniquement sur  $[0, t_0]$  et nous voulons prédire  $P_{eff}^i$  sur le reste de sa trajectoire. En utilisant les techniques qui seront développées pour construire  $\hat{f}$ , il sera alors possible de construire  $(\hat{f}_t)$  pour un nombre fini de valeurs de  $t$ . L'idée est ensuite de construire une interpolation de ce nombre fini de  $(\hat{f}_t)$  pour obtenir la toute la collection  $(\hat{f}_t)_{t \geq 0}$ . En un certain sens, il s'agit donc de construire un méta-modèle de méta-modèles ce qui à notre connaissance n'a jamais été étudié.

Une fois ce méta-modèle construit, il sera alors possible d'effectuer une analyse de sensibilité pour détecter parmi les variables d'entrées celles qui influent le plus sur l'erreur de positionnement de l'aéronef (la différence entre sa position théorique et sa position réelle). Classiquement, lorsque la sortie est un nombre réel, l'importance d'une variable d'entrée est quantifiée par les indices de Sobol [21] dont l'étude théorique de leur estimateur classique a été largement étudié [13, 10], lorsque la sortie n'est plus réelle des travaux récents expliquent comment généraliser ces indices et les estimer [12, 9, 11]. De plus, comme la quantité d'intérêt pour un contrôleur ne sera pas l'erreur moyenne mais le fait que cette erreur dépasse un certain seuil  $s$ , il sera important ici dans un premier temps d'utiliser les résultats de [12] permettant de définir des indices lorsque la sortie est une fonction mais pour cela, nous aurons besoin d'une bonne notion de distance entre trajectoire

Puis dans un deuxième temps d'utiliser les travaux [7] qui permettent de définir des indices adaptés à des problématiques de dépassement de seuil.

### 3 Tâche 2: Distances entre trajectoires et Classification

Un deuxième aspect fondamental des travaux de thèse sera de construire un algorithme de classification automatique des trajectoires. En effet, toujours d'un point de vue opérationnel, le contrôleur a besoin de savoir si un vol doit être considéré "à risque"<sup>1</sup> ou non. Cela permettra d'allouer plus ou moins de contrôleur en fonction de la complexité du trafic.

Comme, nous venons de le voir, dans la section précédente, la notion de distance est fondamentale pour mettre en place une analyse de sensibilité pertinente, elle est tout autant si l'on désire faire une classification des trajectoires. En effet, tous les algorithmes classiques de classifications [4, 6], comme par exemple *k-means* [8], les arbres de régression ou même une classification basée sur une analyse en composantes principales reposent fortement sur une notion de distances. Ainsi, nous mettrons en place différentes notions de distances entre trajectoires que nous comparons entre elles et utiliserons pour mettre en place un algorithme de classification.

Le futur des outils de gestion du trafic aérien repose en grande partie sur des outils d'aides à la décision qui doivent permettre aux opérateurs de prendre en compte l'augmentation constante du trafic. Pour construire et évaluer ces nouveaux systèmes il est nécessaire de pouvoir quantifier les améliorations entre les nouvelles trajectoires et des trajectoires dites de références.

Avec la généralisation du concept des trajectoires 4D étudié dans le programme cadre Européen SESAR et SESAR 2, il y a un fort besoin de comparer avec précision les nouvelles trajectoires afin de valider les modèles (d'avion, de poursuites radars, etc.). Cette validation est généralement basée sur une comparaison entre la trajectoire réelle et une trajectoire de référence en utilisant la moyenne d'une distance de trajectoire.

La plupart des mesures existantes [14], [20], calculent une distance moyenne des positions correspondantes entre deux trajectoires de durées égales. Des statistiques supplémentaires telles que la variance, la médiane, les distances minimales et maximales sont également suggérées pour élargir la description de la similarité. Dans [19], Needman a proposé une métrique de distance basée sur l'alignement qui révèle la transition spatiale et le décalage temporel entre les trajectoires données, et a introduit une métrique d'aire qui calcule la surface fermée entre les trajectoires en utilisant l'intersection de trajectoires.

L'un des principaux inconvénients des approches existantes est qu'elles sont toutes limitées aux trajectoires à durée égale (durée de vie). Par durée, nous nous référons au nombre de points de coordonnées qui constituent la trajectoire. Ces coordonnées sont échantillonnées à différentes occasions. Comme les mesures existantes dépendent des correspondances de coordonnées mutuelles, elles ne peuvent pas être appliquées à des trajectoires ayant des durées différentes. Les mesures de distance conventionnelles supposent que les vitesses d'échantillonnage temporelles des trajectoires sont égales.

Les approches ne traitent pas les instances d'échantillonnage inégales, c'est-à-dire que la distance temporelle varie entre les coordonnées. Par conséquent, il est nécessaire de développer d'autres alternatives capables de mesurer efficacement la différence entre des trajectoires non contraintes. Par conséquent, il est nécessaire de développer d'autres alternatives capables de mesurer efficacement la différence entre les trajectoires non contraintes. Des nombreux travaux dans ce sens sont réalisés dans notre laboratoire de recherche de l'ENAC, par exemple on peut citer [5] qui

---

<sup>1</sup>On considérera qu'un vol est à risque si la distance entre une trajectoire déposée et celle effectuée est grande

propose une représentation mathématique de la trajectoire sous la forme d'un objet fonctionnel plongé dans un espace fonctionnel avec ses opérateurs associés qui permettent des calculs de distances. Nous disposons également d'algorithmes codés de cette méthode pour pouvoir calculer ces distances sur de gros volumes de trajectoires.

Nous proposons de définir une nouvelle distance basée sur la distance de Wasserstein. En effet comme nous travaillons avec les abscisses curvilignes, quitte à les renormaliser, chaque trajectoire peut-être vue comme une fonction de répartition. On pourra donc s'inspirer de l'ACP introduite par [3] pour faire une classification pertinente des trajectoires.

## 4 Tâche 3: Utilisation de la classification pour améliorer la méta-modélisation

Une fois, une classification des trajectoires disponible, il sera alors intéressant de construire un méta-modèle adapté à chacune des classes.

## 5 Tâche 4: Connection avec le projet Salto

## 6 Planning prévisionnel de la thèse

La durée prévue pour la thèse est de 36 mois. Nous prévoyons le planning suivant.

- De  $T_0$  à  $T_0+6$ : Etude bibliographique et prise en main des données
- De  $T_0 + 6$  à  $T_0+18$ : Tâche 1
- De  $T_0 + 18$  à  $T_0+24$ : Tâche 2
- De  $T_0 + 24$  à  $T_0+30$ : Tâche 3
- De  $T_0 + 30$  à  $T_0+33$ : Tâche 4
- De  $T_0 + 33$  à  $T_0+36$ : Rédaction du manuscrit

## References

- [1] Anestis Antoniadis, Céline Helbert, Clémentine Prieur, and Laurence Viry. Spatio-temporal metamodeling for West African monsoon. *Environmetrics*, 23(1):24–36, February 2012. Special Issue: Spatio-Temporal Stochastic Modelling (METMAV).
- [2] José Betancourt, François Bachoc, Thierry Klein, Déborah Idier, Rodrigo Pedreros, and Jérémy Rohmer. Gaussian process metamodeling of functional-input code for coastal flood hazard assessment. working paper or preprint, January 2019.
- [3] Jérémie Bigot, Raúl Gouet, Thierry Klein, Alfredo López, et al. Geodesic pca in the wasserstein space by convex pca. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 53, pages 1–26. Institut Henri Poincaré, 2017.



- [4] Stéphane Boucheron, Olivier Bousquet, and Gábor Lugosi. Theory of classification: a survey of some recent advances. *ESAIM Probab. Stat.*, 9:323–375, 2005.
- [5] Daniel Delahaye, Stéphane Puechmorel, Sameer Alam, and Eric Féron. Trajectory Mathematical Distance Applied to Airspace Major Flows Extraction. In *EIWAC 2017, 5th ENRI International Workshop on ATM/CNS*, volume Lecture Notes in Electrical Engineering of *ENRI International Workshop on ATM/CNS EIWAC 2017: Air Traffic Management and Systems III*, pages pp. 51–66., Tokyo, Japan, June 2018. ENRI, Springer.
- [6] Luc Devroye, László Györfi, and Gábor Lugosi. *A probabilistic theory of pattern recognition*, volume 31 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1996.
- [7] J.-C. Fort, T. Klein, and N. Rachdi. New sensitivity analysis subordinated to a contrast. *ArXiv e-prints*, May 2013.
- [8] Sébastien Gadat, Thierry Klein, Clément Marteau, et al. Classification in general finite dimensional spaces with the k-nearest neighbor rule. *The Annals of Statistics*, 44(3):982–1009, 2016.
- [9] Fabrice Gamboa, Alexandre Janon, Thierry Klein, Agnès Lagnoux, et al. Sensitivity analysis for multidimensional and functional outputs. *Electronic Journal of Statistics*, 8(1):575–603, 2014.
- [10] Fabrice Gamboa, Alexandre Janon, Thierry Klein, Agnès Lagnoux-Renaudie, and Clémentine Prieur. Statistical inference for sobol pick freeze monte carlo method. Preprint available at <http://hal.inria.fr/hal-00804668/en>, 2013.
- [11] Fabrice Gamboa, Thierry Klein, and Agnès Lagnoux. Sensitivity analysis based on cramér-von mises distance. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):522–548, 2018.
- [12] Fabrice Gamboa, Thierry Klein, Agnès Lagnoux, and Leonardo Moreno. Sensitivity analysis in general metric spaces. working paper or preprint, February 2019.
- [13] Alexandre Janon, Thierry Klein, Agnès Lagnoux, Maëlle Nodet, and Clémentine Prieur. Asymptotic normality and efficiency of two sobol index estimators. *ESAIM: Probability and Statistics*, 18:342–364, 1 2014.
- [14] Christopher Jaynes, Stephen Webb, R. Matt Steele, and Quanren Xiong. An open development environment for evaluation of video surveillance systems. In *Proceedings of the Third International Workshop on Performance Evaluation of Tracking and Surveillance (PETS'2002)*, pages 32–39, 2002.
- [15] Chu V. Mai and Bruno. Sudret. Surrogate models for oscillatory systems using sparse polynomial chaos expansions and stochastic time warping. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):540–571, 2017.
- [16] Amandine Marrel, Bertrand Iooss, François Van Dorpe, and Elena Volkova. An efficient methodology for modeling complex computer codes with gaussian processes. *Computational Statistics & Data Analysis*, 52(10):4731–4744, 2008.

- [17] Thomas Muehlenstaedt, Jana Fruth, and Olivier Roustant. Computer experiments with functional inputs and scalar outputs by a norm-based approach. *Statistics and Computing*, 27(4):1083–1097, July 2017.
- [18] Simon. Nanty, Céline. Helbert, Amandine. Marrel, Nadia. Pérot, and Clémentine. Prieur. Sampling, metamodeling, and sensitivity analysis of numerical simulators with functional stochastic inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):636–659, 2016.
- [19] Chris J. Needham and Roger D. Boyle. Performance evaluation metrics and statistics for positional tracker evaluation. In James L. Crowley, Justus H. Piater, Markus Vincze, and Lucas Paletta, editors, *Computer Vision Systems*, pages 278–289, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.
- [20] Andrew Senior, Arun Hampapur, Ying li Tian, Lisa Brown, Sharath Pankanti, and Ruud Bolle. Appearance models for occlusion handling. In *2nd IEEE Workshop on Performance Evaluation of Tracking and Surveillance*, 2001.
- [21] I.M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.
- [22] Michael L Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.