Contributions à la gestion du trafic aérien : prédiction des trajectoires par réseaux de neurones et cartes de complexité par maximum d'entropie

Thi Lich Nghiem

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

Université de Toulouse

Doctorat de l'Université de Toulouse

préparé à l'Université Toulouse III - Paul Sabatier

Contributions à la gestion du trafic aérien: prédiction des trajectoires par réseaux de neurones et cartes de complexité par maximum d'entropie.

Thèse présentée et soutenue, le 16 septembre 2024 par

Thi-Lich NGHIEM

École doctorale
EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse

Spécialité
Informatique et Télécommunications

Unité de recherche
IMT : Institut de Mathématiques de Toulouse

Thèse dirigée par
Pierre MARÉCHAL et Thi-Lan LE

Composition du jury
M. Eric Marie FERON, Président, King Abdullah University of Science and Technology
M. Duc-Dung NGUYEN, Rapporteur, Vietnam Academy of Science and Technology
Mme Valérie GIRARDIN, Rapporteure, Université Caen Normandie
M. Denis KOUAMÉ, Examinateur, Université Toulouse III - Paul Sabatier
M. Pierre MARÉCHAL, Directeur de thèse, Université Toulouse III - Paul Sabatier
Mme Thi-Lan LE, Co-directrice de thèse, Hanoi University of Science and Technology

Membres invités
M. Daniel Delahaye, Ecole Nationale de l'Aviation Civile

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *16/09/2024* **par :**

Thi Lich NGHIEM

Contributions à la gestion du trafic aérien: prédiction des trajectoires par réseaux de neurones et cartes de complexité par maximum d'entropie.

JURY

Valérie Girardin	Professeure des Universités	Rapporteur
Duc-Dung Nguyen	Maître de Conférences	Rapporteur
Eric Marie Feron	Professeur des Universités	Examinateur
Denis Kouamé	Professeur des Universités	Examinateur
Daniel Delahaye	ENAC	Membre invité
Pierre Maréchal	Professeur des Universités	Directeur de Thèse
Thi-Lan Le	Maîtresse de Conférences	Codirectrice de Thèse

École doctorale et spécialité :
MITT : Informatique et Telecommunications

Unité de Recherche :
Laboratoire de Recherche ENAC

Directeur(s) de Thèse :
Pierre Maréchal et *Thi-Lan Le*

Rapporteurs :
Valérie Girardin et *Duc-Dung Nguyen*

Résumé

Dans le contexte de la gestion du trafic aérien (ATM), la demande croissante d'espace aérien a exercé une pression considérable sur les systèmes existants. Cette demande croissante nécessite l'élaboration et la mise en œuvre urgentes de mesures de sécurité renforcées et d'une efficacité optimisée de l'espace aérien pour faire face à l'augmentation des volumes de trafic. La complexité du trafic aérien est déterminée par le flux du trafic en temps réel dans un espace aérien donné, y compris le nombre de vols et leurs positions relatives. En raison des contraintes liées à la conception de l'espace aérien, les vols doivent modifier leurs trajectoires pour s'adapter à ces limites, ce qui entraîne des changements continus dans la complexité du trafic. En conséquence, mesurer avec précision la complexité dynamique du trafic aérien est une tâche difficile. Cette thèse vise à explorer les difficultés liées à la prédiction de trajectoires de vol et à l'estimation de la complexité du trafic aérien en tirant parti des méthodologies d'apprentissage automatique, du principe d'entropie maximale et de la géométrie de l'information.

La prévision des trajectoires de vol est essentielle dans l'aviation, car elle favorise une gestion efficace du trafic aérien et garantit des opérations aériennes sûres et fluides. Les méthodes de prévision traditionnelles ne parviennent souvent pas à saisir les dépendances spatio-temporelles complexes et les incertitudes spécifiques au secteur de l'aviation. Nous proposons un nouveau modèle nommé modèle BayesCoordLSTM. Ce modèle hybride intègre la transformation de coordonnées et le méthode Bayésien dans les modèles ConvLSTM. En combinant les caractéristiques spatiales extraites par le réseau neuronal convolutif (CNN) avec les dépendances temporelles identifiées par la mémoire à long terme (LSTM), le modèle améliore considérablement la précision de la prédiction de trajectoire. L'incorporation du méthode bayésien offre des prévisions de trajectoire probabilistes avec des niveaux de confiance, tandis que la transformation des coordonnées améliore la conscience spatiale et les performances prédictives. Nous utilisons un ensemble de données de vol réel pour la mise en œuvre. Nos résultats démontrent la remarquable supériorité du modèle BayesCoordLSTM sur les méthodes existantes.

Dans divers scénarios et conditions de vol, ce modèle surpasse systématiquement les modèles de base en termes de précision de prédiction de trajectoire. L'utilisation de l'optimisation bayésienne rationalise non seulement le réglage des hyperparamètres, mais donne également des résultats prometteurs en termes d'amélioration des performances prédictives.

De plus, dans l'estimation de la complexité du trafic aérien, la reconstruction des fonctions de densité de probabilité angulaire est cruciale car elle permet de modéliser et d'évaluer la répartition spatiale du flux de trafic aérien. Cette approche permet d'évaluer la congestion de l'espace aérien et de prévoir des schémas de trafic complexes, facilitant ainsi des stratégies de gestion du trafic aérien plus efficaces. Dans cette thèse, nous proposons d'utiliser le principe d'entropie maximale pour générer des cartes détaillées de complexité du trafic à partir de données de trajectoires, en exploitant les propriétés statistiques des coefficients de Fourier. La méthode est validée par des simulations numériques, démontrant sa capacité à représenter avec précision des distributions angulaires de scénarios simples à complexes.

À l'aide de données historiques sur les vols et le trafic aérien, la reconstruction des fonctions de densité de probabilité angulaire par maximum d'entropie est proposée, puis utilisée pour construire des cartes de complexité du trafic aérien. La géométrie de l'information est à la base de l'évaluation de la complexité locale. En discrétisant l'espace aérien et en appliquant la dualité de Fenchelpour reconstruire les densités angulaires, nous obtenons une densité locale dans la classe de densité de Von Mises Généralisées. L'indice de complexité local est alors calculés sur la base des distances géodésiques entre la distribution obtenue et la distribution uniforme (qui est bien sûr associée à la plus grande compliexité). L'utilisation de la divergence Kullback-Leibler symétrisée garantit l'efficacité des calculs, permettant une application pratique dans les systèmes ATM. Ces cartes de complexité fournissent des informations essentielles sur les modèles de trafic et les réponses systémiques, soutenant la gestion stratégique.

En conclusion, cette thèse propose des approches efficaces pour résoudre certains problèmes en ATM, notamment en matière de prédiction de trajectoire de vol et d'évaluation de la complexité du trafic aérien.

Mots clés: Gestion du trafic aérien, Complexité du trafic aérien, Réseaux de neuronnes, Prédiction des trajectoire, Géométrie de l'information, Entropie maximale.

Abstract

In the context of air traffic management (ATM), the increasing demand for airspace has exerted substantial pressure on existing systems. This growing demand necessitates the urgent development and implementation of enhanced safety measures and optimized airspace efficiency to accommodate escalating traffic volumes. The complexity of air traffic is determined by the real-time flow of traffic within a given airspace, including the number of flight and their relative positions. Due to the constraints of airspace design, flight must alter their trajectories to fit within these boundaries, resulting in ongoing changes in traffic complexity. As a result, precisely measuring the dynamic complexity of air traffic is a challenging task. This thesis aims to explore the difficulties in flight trajectory prediction and air traffic complexity estimation by leveraging machine learning methodologies, Maximum Entropy principle and information geometry.

Predicting flight trajectories is essential in aviation, as it supports efficient air traffic management and ensures safe and smooth flight operations. Traditional prediction methods often fall short in capturing the complex spatio-temporal dependencies and uncertainties specific to the aviation sector. We propose a novel model named BayesCoordLSTM model. This hybrid model integrates coordinate transformation and Bayesian method into ConvLSTM models. By combining the spatial features extracted by the Convolutional Neural Network (CNN) with the temporal dependencies identified by Long Short-Term Memory (LSTM), the model significantly enhances trajectory prediction accuracy. The incorporation of Bayesian method offers probabilistic trajectory forecasts with confidence levels, while coordinate transformation improves spatial awareness and predictive performance. We use real flight dataset to implement, our results demonstrate the remarkable superiority of the BayesCoordLSTM model over existing methods. Across various flight scenarios and conditions, this model consistently outperforms baseline models in terms of trajectory prediction accuracy. The utilization of Bayesian optimization not only streamlines hyperparameter tuning but also yields promising results in terms of predictive performance enhancement.

In addition, in air traffic complexity estimation, reconstructing angular probability density functions is crucial as it helps model and assess the spatial distribution of air traffic flow. This approach helps assess airspace congestion and predict complex traffic patterns, facilitating more effective air traffic management strategies. In this dissertation, we propose to use the Maximum Entropy principle generating detailed traffic complexity maps from trajectory data, leveraging the statistical properties of Fourier coefficients. The method is validated through numerical simulations, demonstrating its ability to accurately represent angular distributions from simple to complex scenarios.

Using historical flight and airspace traffic data, reconstructing angular probability density functions improves predictions of congestion areas and complexity levels. And then, we illustrate a novel methodology for constructing air traffic complexity maps is developed using information geometry. By discretizing the airspace and applying the Generalized von Mises distribution (GvM) alongside Fenchel duality, local traffic complexity indices are calculated based on geodesic distances between distributions. The use of symmetrized Kullback-Leibler divergence ensures computational efficiency, enabling practical application in ATM systems. These complexity maps provide critical insights into traffic patterns and systemic responses, supporting strategic management.

In conclusion, this dissertation propose efficient approaches to solve some problems in ATM, especially in flight trajectory prediction, and air traffic complexity assessment.

Keywords: Air traffic management, Air traffic complexity, Neural networks, Prediction of trajectories, Information geometry, Maximum Entropy.

Acknowledgments

I would like to express my deepest and most heartfelt gratitude to those who have accompanied and supported me throughout the challenging journey of completing my PhD. This process has been an experience filled with both intellectual and emotional trials, and during the most difficult moments, I have been fortunate to receive unwavering encouragement and support from many people.

First and foremost, I extend my profound thanks to my advisors, Prof. Pierre Maréchal and Assoc. Prof. Thi-Lan Le. They have been more than supervisors; they have been true mentors and steadfast companions on this journey. Their guidance, encouragement, and belief in my work helped me navigate every obstacle and bring this thesis to completion.

I am deeply grateful to my referees, Prof. Valérie Girardin and Assoc. Prof. Duc-Dung Nguyen. Their thorough evaluation of my work, their insightful critique, and their detailed, thoughtful feedback have had an immense impact not only on this thesis but on shaping my future work. Their contributions have been invaluable, and I am deeply thankful for the time and effort they devoted to this process.

I also sincerely appreciate the other esteemed members of my thesis jury, Prof. Eric Marie Feron, Prof. Denis Kouamé, and Prof. Daniel Delahaye, for their time, insightful comments, and constructive suggestions, which have enriched my work and inspired new ideas for future endeavors.

Special thanks to Prof. Stéphane Puechmorel and Assoc. Prof. Andrija Vidosavljevic for their invaluable support and mentorship. I would also like to thank my friends in Z108/Z109, who have shared countless memorable moments over coffee and lunch breaks, bringing joy and balance to my days.

I am deeply thankful to ENAC (École Nationale de l'Aviation Civile), University of Toulouse III - Paul Sabatier and HUST (Hanoi University of Science and Technology) for providing exceptional research facilities, and particularly to ENAC for the financial support that made this work possible.

I extend my appreciation to all members of the Computer Vision Department, MICA Research Institute, HUST for their willingness to assist and offer guidance whenever needed.

I would also like to thank my Vietnamese friends in France: Thinh, Nga, Hung and Lam, who stood by me during the intense pressures of thesis completion. My heartfelt thanks go to my colleagues at Thuongmai University for their support.

A special note of gratitude goes to the families of Christine, Anne, and An. Their kindness and hospitality made me feel like a part of their families, providing warmth and comfort during some of the most difficult times.

Lastly, but most importantly, I express my deepest love and appreciation to my parents, my husband Khiem, and my children Minh and Ngoc. Their unconditional love, patience, and belief in me have been the foundation upon which I have built this achievement. I am also grateful to my sisters for their constant understanding and companionship on this long journey.

Each of these people has played an irreplaceable role in my life, and I will forever cherish their support and kindness.

Contents

Résumé **i**

Abstract **iii**

Acknowledgments **v**

Abbreviations **xv**

Publications **xvii**

1 Introduction **1**

1.1 Motivation 1

1.2 Background 3

1.3 Air traffic complexity 7

1.3.1 Factors contributing to air traffic complexity 11

1.3.2 Air traffic complexity metrics 12

1.4 Contributions 16

1.5 Structure of the thesis 18

2 Applying the Bayesian method to improve flight trajectory prediction **21**

2.1 Introduction 21

2.2 Literature review 24
2.3 Long Short-Term Memory 30
2.3.1 LSTM architecture 30
2.3.2 LSTM operation 31
2.4 Proposed model 32
2.4.1 Coordinate transformation 34
2.4.2 CNN layer 35
2.4.3 Bayesian optimization 36
2.5 Experimental results 39
2.5.1 Data preparation 39
2.5.2 Hyperparameter tuning 42
2.5.3 Evaluation metrics 42
2.5.4 Comparison of experimental results 43
2.6 Conclusions 47

3 Reconstructing angular probability density by using Maximum Entropy 49

3.1 Introduction 49
3.2 Literature review 51
3.3 Setting 54
3.3.1 Computing moments 55
3.3.2 Formulating the Maximum Entropy problem 56
3.4 Duality in action 58
3.4.1 Review of useful convex analysis 59
3.4.2 Computing Maximum Entropy densities 64
3.4.3 Deriving an algorithm 66
3.5 Simulations 70

3.5.1 Validation with Dirac distributions 70
3.5.2 Numerical implementations 72
3.6 Towards air traffic complexity estimation 79
3.7 Conclusions 81

4 Building complexity maps of airspace via information geometry 83

4.1 Introduction 83
4.1.1 Motivation 83
4.1.2 Overview 84
4.2 Deriving a complexity index via information geometry 85
4.2.1 Maximum Entropy angular densities 86
4.2.2 The Fisher information metric 89
4.3 Validation 97
4.3.1 Simulation 97
4.3.2 Implementation 99
4.4 Conclusions 107

5 Maximum Entropy multivariate density estimation 109

5.1 Introduction 109
5.2 Statement of the problem 110
5.3 Maximum Entropy solutions 113
5.4 Relaxation 119
5.5 Conclusion 121

6 Conclusion and further work 123

6.1 Conclusion 123

6.2 Further work 124

References **138**

Appendix

A Computing conjugates . . . . . .

List of Figures

1.1 One-day traffic density of trajectories across Europe [96]. . . . 4

1.2 Some proposed approaches to evaluate the air traffic complexity. 8

1.3 An example of three traffic scenarios classified in order of increasing complexity [38]. 14

2.1 Different stages of a flight [135]. 25

2.2 LSTM architecture [56]. 30

2.3 The BayesCoordLSTM framework to predict flight trajectory. 33

2.4 Optimizing hyperparameters using the Bayesian method. . . . 37

2.5 An example of addressing missing data before (left) and after (right) using cubic spline interpolation. 41

2.6 Sliding time window. 41

2.7 A comparison of three configurations of the CNN-LSTM model. 46

3.1 An example map of trajectories through a selected cell and the angular histogram of the selected cell. 55

3.2 Kullback–Leibler relative entropy versus highest frequency. . . 68

3.3 Simulate angular samples based on a original probability $p_\circ$. . 71

3.4 Gradient norm when $N = 40$. 72

3.5 Residual values when α values from 1 to 10000. 73

3.6 Results on the Original and Optimal densities with two peaks. 74

3.7 Results on the Original and Optimal densities with five peaks. 75

3.8 Comparison of computational time by seven optimal models with five peaks. 79

4.1 Example of a cell around a grid point in the airspace, with trajectories across ATL on December 26, 2018. 87

4.2 The $\mathbf{GvM}_N$ with different β values and peak numbers: Top to bottom, $\beta = 0.01$ and $\beta = 0.3$; Left: 1 peak, Right: 20 peaks. 99

4.3 Traffic density of trajectories across ATL on December 26, 2018.101

4.4 All flight trajectories across ATL on December 26, 2018. . . . 102

4.5 Complexity maps on ATL airport dataset. 103

4.6 High complexity cell: many flights, diverse directions. 105

4.7 Low complexity cell: few flights, few directions. 105

4.8 Medium complexity cell: moderate flights, moderate directions.106

4.9 High complexity cell: fewer flights, diverse directions. 106

5.1 Plots of $\varepsilon\psi_\beta(\eta)$ and its conjugate for $\beta = 1$ and various values of ε. 121

List of Tables

2.1 4D Trajectory data features 39
2.2 An example of 4D Flight trajectory information. 40
2.3 Basic parameters of proposed model. 43
2.4 Comparison with state-of-the-art methods. 44

4.1 The simulation results 98

Abbreviations

Abbreviation	Meaning
ADS-B	**A**utomatic **D**ependent **S**urveillance **B**roadcast
ATCOs	**A**ir **T**raffic **C**ontrollers
ATFM	**A**ir **T**raffic **F**low **M**anagement
ATL	**ATL**anta International Airport
ATM	**A**ir **T**raffic **M**anagement
BCNN	**B**ayesian **C**onvolutional **N**eural **N**etwork
BFGS	**B**royden **F**letcher **G**oldfarb **S**hann
BNN	**B**ayesian **N**eural **N**etwork
CCI	**C**omposite **C**omplexity **I**ndex
CG	**C**onjugate **G**radient
CG3D	**C**NN-**G**RU with a **3D**imensional Convolution
CNN	**C**onvolutional **N**eural **N**etwork
COBYLA	**C**onstrained **O**ptimization **BY** **L**inear **A**pproximations
DNN	**D**eep **N**eural **N**etwork
FIM	**F**isher **I**nformation **M**etric
GvM	**G**eneralized **v**on **M**ises
GRU	**G**ated **R**ecurrent **U**nit
LSTM	**L**ong **S**hort- **T**erm **M**emory
MAE	**M**ean **A**bsolute **E**rror
NEXTGEN	**NEXT GEN**eration Air Transportation System
PDFs	**P**robability **D**ensity **F**unctions
QC	**Q**ualification of **C**onstraints
ReLU	**Re**ctified **L**inear **U**nit
RMSE	**R**oot **M**ean **S**quare **E**rror
RNN	**R**ecurrent **N**eural **N**etwork
SESAR	**S**ingle **E**uropean **S**ky **A**TM **R**esearch
SKL	**S**ymmetrized **K**ullback-**L**eible
TNC	**T**runcated **N**ewton **C**onjugate-Gradient
Trust-Constr	**Trust** Region **Constr**aint

Publications

1 **Thi-Lich Nghiem**, Viet-Duc Le, Thi-Lan Le, "Quantitative evaluation of robustness of Bayesian CNN models for image classification", National Conference on Some selected issues of information and communication technology, Thai Nguyen, Vietnam, 2021, pp. 454-460, ISBN 978-604-67-1744-7.

2 **Thi-Lich Nghiem**, Viet-Duc Le, Thi-Lan Le, Pierre Maréchal, Daniel Delahaye, Andrija Vidosavljevic, "Applying Bayesian inference in a hybrid CNN-LSTM model for time-series prediction", International Conference on Multimedia Analysis and Pattern Recognition (MAPR), Phu Quoc, Vietnam, 2022, pp. 1-6, DOI: 10.1109/MAPR56351.2022.9924783.

3 **Thi-Lich Nghiem**, Viet-Duc Le, Thi-Lan Le, Pierre Maréchal, Daniel Delahaye, "Improving flight trajectory predictions with Bayesian-optimized ConvLSTM models," International Conference on Intelligent Systems and Networks (ICISN), Hanoi, Vietnam, 2024, pp. 604-614, DOI: 10.1007/978-981-97-5504-2_70.

4 **Thi-Lich Nghiem**, Viet-Duc Le, Thi-Lan Le, Daniel Delahaye, Pierre Maréchal, "Angular probability density reconstruction by maximum entropy - Towards air traffic complexity estimation," submitted, 2024.

5 Pierre Maréchal, **Thi-Lich Nghiem**, "A note on maximum entropy multivariate density estimation," submitted, 2024.

6 Pierre Maréchal, **Thi-Lich Nghiem**, Stéphane Puechmorel, "Building complexity maps of airspace via information geometry," submitted, 2024.

Chapter 1

Introduction

This chapter illustrates a comprehensive exploration of air traffic complexity, delving into the intricate operational dynamics of air traffic management. It includes a detailed analysis of the factors that contribute to the heightened complexity of air traffic, as well as an intensive discussion of the metrics used to assess this complexity. This discussion is the foundational motivation for our study, which is presented together with the objectives and scope outlined in the dissertation. The significance of our contributions is briefly addressed here, with a more detailed examination to follow in subsequent chapters. Additionally, we provide an overview of the thesis structure to clarify its organization and progression.

1.1 Motivation

In the context of air traffic management (ATM), the growing demand for airspace has placed unprecedented pressure on existing systems. This exponential growth in airspace demands not only highlights the critical need for enhanced safety measures but also requires efficient optimization of airspace to accommodate increasing air traffic volumes.

Air traffic management faces numerous complex challenges that impact operational efficiency and safety. The structure of current airspace configurations, characterized by rigid layouts and fixed sector divisions, significantly contributes to congestion and inefficiencies. This complexity encompasses both physical aspects, such as airport layouts and airway structures, and operational dynamics involving the intricate interactions of aircraft within these frameworks [68].

Human factors further exacerbate these challenges, as air traffic controllers (ATCOs) manage traffic within designated sectors while grappling with cognitive limitations and the increasing complexity of traffic patterns [36]. The uneven distribution of traffic and the inflexible configuration of airspace sectors create congestion hotspots, necessitating strategic management strategies to optimize airspace usage and mitigate safety risks [63].

Within the broader context of air traffic complexity, accurate prediction of aircraft trajectories is crucial. ATM systems must ensure precise trajectory forecasts to maintain safe separation standards and prevent potential collisions [98]. However, trajectory prediction is affected by inherent uncertainties, particularly influenced by dynamic meteorological conditions such as wind. These uncertainties arise from incomplete data and model inaccuracies in numerical weather prediction, which affect the velocity and precision of trajectory forecasts [40].

The complexity of trajectory prediction is further compounded by the nonlinear dynamics of atmospheric interactions, making it challenging to reliably predict aircraft paths under varying conditions. Existing conflict resolution methodologies often struggle to adequately account for these uncertainties, risking violations of safety standards during real-time operations.

Addressing uncertainties in trajectory prediction is essential for enhancing conflict resolution methodologies and ensuring robust air traffic management. By developing methods that accurately estimate and incorporate uncertainty into trajectory forecasts, more resilient and adaptive solutions can be implemented to maintain safety and efficiency in increasingly congested airspace environments.

The complexities of ATM systems demand innovative methodologies to effectively address the challenges posed by escalating air traffic congestion. Recognizing the critical role of accurate air traffic complexity analysis in enhancing operational efficiency and safety, this thesis endeavors to introduce novel approaches for analyzing and managing air traffic complexity.

Addressing the uncertainties inherent in ATM is imperative for developing robust and adaptive solutions that ensure safe and efficient air traffic operations. By accurately assessing air traffic complexity, it becomes possible to identify congestion hotspots, optimize airspace utilization, and mitigate potential safety risks.

Through empirical analysis and theoretical inquiry, this study seeks to provide insights into the complexities of ATM and inform the development of strategies to address operational challenges in ATM systems. By offering innovative methodologies for analyzing air traffic complexity, this thesis aims to contribute to the advancement of ATM practices, ensuring the safety and efficiency of air travel in increasingly congested airspace environments.

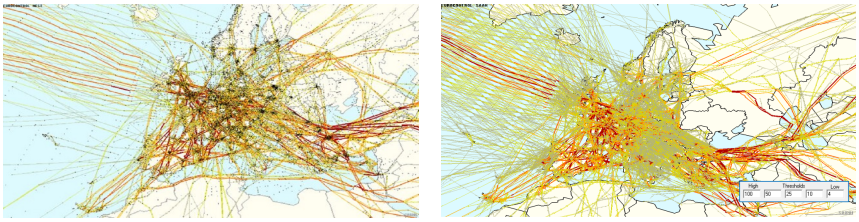
1.2 Background

Air traffic management (ATM) is a critical system that ensures the safety and efficiency of air travel globally. ATM encompasses a wide range of services including air traffic control, which manages the movement of aircraft on the ground and in the air; air traffic flow management (ATFM), which regulates the flow of aircraft to prevent congestion; and ATM, which involves the strategic organization of airspace to optimize its use [40]. Together, these components work to maintain safe separation between aircraft and manage the complex dynamics of air traffic.

The growth in air traffic has led to increased congestion and complexity within airspace systems. This is particularly evident in regions with dense air traffic, such as Europe. The dense network of flight paths in these areas highlights the challenges faced by modern ATM systems in managing high volumes of traffic. Such congestion necessitates the implementation of advanced methodologies to ensure aircraft maintain safe separation distances and to prevent conflicts.

Figure 1.1 illustrates the one-day traffic density of trajectories across Europe on two specific years: 2019 (see Figure 1.1a) and 2024 (see Figure 1.1b). A comparative analysis of these images reveals a significant increase in traffic density over the five-year period from 2019 to 2024, demonstrating the increased complexity and congestion within European airspace.

Effective trajectory prediction is essential for managing air traffic and preventing collisions. Predicting the future positions of aircraft allows air traffic control to anticipate potential conflicts and take preemptive measures. However, trajectory prediction is inherently complex due to various sources of uncertainty, particularly those related to meteorological conditions. Winds and other atmospheric factors can significantly affect aircraft velocity and



(a) 2019 (b) 2024

Figure 1.1: One-day traffic density of trajectories across Europe [96].

direction, leading to deviations from planned paths [98]. Understanding and mitigating these uncertainties are crucial for accurate trajectory prediction. Meteorological conditions, especially winds, are a major source of uncertainty in trajectory prediction. Numerical weather prediction models often suffer from inaccuracies due to incomplete atmospheric data and model limitations. These uncertainties are further amplified by the nonlinear and chaotic nature of atmospheric dynamics, making simple statistical methods inadequate for accurate representation [40]. This can lead to errors in predicting the velocity and trajectory of aircraft, complicating the task of maintaining safe separation.

Numerous methods have been developed to resolve potential conflicts between aircraft. These approaches generally involve predictive modeling to forecast aircraft positions and algorithms to suggest corrective actions. However, many traditional methods do not fully account for the dynamic uncertainties inherent in trajectory predictions, resulting in solutions that may be sensitive to perturbations and increasing the risk of separation standard violations [98].

Air traffic complexity refers to the intricate interactions and dynamics of aircraft within airspace. It includes factors such as the number and types of aircraft, their velocities, trajectories, and the various environmental and operational conditions that affect their movements. In contrast, airspace complexity involves the structural aspects such as the number and location of airports, airways, and air traffic control sectors [36]. Both forms of complexity are interrelated and must be managed effectively to ensure efficient and safe airspace operations. In this study, we focus how to estimate the air traffic complexity.

Air traffic complexity is a multifaceted concept crucial for ATM, involving various interdependent parameters. Despite extensive research, no consensus on a single definition exists due to the numerous and highly correlated factors involved [89]. These parameters include airspace structure, weather conditions, and technological factors. Complexity assessment traditionally relies on the subjective evaluations of air traffic controllers, but there is a shift towards objective indicators based on various air traffic parameters [36, 89]. Key indicators include *Aircraft Density*, *Dynamic Density*, *Interval Complexity*, *Fractal Dimension*, *Input/Output Approach*, and *Lapunov Exponents*, each with distinct strengths and limitations [68].

Traditional airspace structures are often rigid and lack the flexibility to adapt to the dynamic demands of modern air traffic. This rigidity can lead to inefficiencies and potential safety hazards, particularly in high-density traffic areas. ATCOs manage traffic within these structures, but their ability to handle large volumes of traffic is limited by cognitive constraints and workload [68]. This highlights the need for more adaptable and responsive ATM strategies.

Complexity maps have emerged as valuable tools for visualizing and managing air traffic complexity. These maps provide detailed insights into the operational dynamics within airspace, helping to identify areas of high traffic density and potential conflicts [93]. By analyzing complexity maps, ATCOs and other stakeholders can make more informed decisions about aircraft routing and sector management, potentially reducing conflict risks and improving operational efficiency.

Recent research has focused on developing advanced metrics and methodologies for evaluating air traffic complexity. For example, geodesic metrics and other mathematical models have been used to quantify the level of disorder in aircraft trajectories. While various approaches are proposed to deal with the air traffic complexity, only a few papers manage air traffic and reduce the risk of conflicts between aircraft. In 2009, Lee et al. [68] proposed the strategy for optimizing the entry points of aircraft before reaching the sector boundary based on a complexity map. They showed that complexity map can provide a visual representation of how a system responds to disturbances and can be used to identify key drivers of complexity and inform decision-making. In the context of ATM, complexity map can provide valuable insights into the behavior of the system under different traffic con-

ditions (such as the entry point angle and heading of the entering aircraft) and environmental factors (like convective weather), and help identify areas where improvements can be made. By analyzing the complexity map, the air traffic control system can identify areas of high traffic density or other factors that may increase the risk of conflicts between aircraft [93]. In this research, complexity maps are considered as an air traffic evaluation approach which illustrate how to manage the sector's conflicts based on the entering aircraft. Unlike other approaches, Hong et al. proposed a new method called Variance Quadtree-Based Map Building to manage the conflict by determining the best strategy when adding an aircraft into a sector. This can help improve the safety and efficiency of air traffic control operations, as well as reduce delays and improve overall airspace capacity [57]. In 2020, Juntama et al. [63] used König metric to build a complexity map to reduce and optimize the air traffic for strategic 4D trajectory. Delahaye et al. [36] quantify air traffic complexity of a set of aircraft trajectories by approximating a metric based on linear dynamical system which can fit a vector field as closely as possible to the observed aircraft positions and speeds. Therefore, this system can evaluate the local disorder of a set of trajectories in the neighborhood of a given aircraft at a given time based on the eigenvalues of the matrix A associated with organized traffic situations, including the vertical strip around the imaginary axis. It provides a nuanced understanding of how aircraft interact within congested airspace.

This background provides a comprehensive overview of the key challenges and considerations in ATM. The increasing demand for airspace, the complexities of trajectory prediction, and the rigid structure of traditional ATM systems underscore the need for innovative approaches. This research aims to address the critical gaps in trajectory prediction, heading angle distribution reconstruction, and air traffic complexity management. By tackling these issues, the primary objective of this research is to introduce novel methodologies aimed at enhancing the understanding and management of air traffic complexity in ATM systems. Specifically, the research seeks to address the following objectives:

- Propose a novel approach to enhance the accuracy and robustness of flight trajectory prediction.
- Investigate methodologies for reconstructing heading angle distributions accurately to improve ATM.

- Propose a new metric to build and evaluate air traffic complexity map, thereby improving safety and efficiency in ATM.
- Develop a method to estimate instantaneous aircraft density at the pre-strategic stage and adapt it for broader applications, including species distribution modeling.

By achieving these objectives, this research aims to contribute to the advancement of knowledge and practices in ATM, thereby promoting a safer, more efficient, and more resilient airspace environments. Through empirical analysis and theoretical inquiry, this study attempts to provide insights into the complexity of air traffic and inform the development of strategies to address operational challenges in ATM systems.

1.3 Air traffic complexity

Air traffic complexity refers to the intricacy and difficulty involved in managing air traffic within a given airspace. This complexity arises from various factors including the number of aircraft, their performance characteristics, the structure of airspace, weather conditions, human factors, and technological advancements. Complexity in air traffic can be spatial, temporal, and influenced by dynamic interactions between different elements.

Understanding air traffic complexity is crucial for ensuring the safety and efficiency of ATM. High complexity can lead to increased workload for air traffic controllers, higher chances of conflicts, and potential delays. By understanding and managing complexity, it is possible to optimize air traffic flow, enhance safety, and improve overall operational efficiency.

Air traffic complexity is an objective and critical measure for determining the operational state of a given airspace. It is a concept used to assess the difficulty of handling the air traffic situation. Air traffic complexity is a multi-faceted subject that has received a great deal of attention in the aviation industry. Many researchers have identified several factors that contribute to airspace's inherent complexity, such as the diversity of aircraft types and capabilities, the various types of airspace classifications, the presence of numerous air traffic control systems, unpredictable weather patterns, and varying levels of air traffic volume. In addition, the frequency

of aircraft travelling through a specific sector over a given time period is referred to as air traffic density or congestion. Higher air traffic congestion can make possible air traffic accidents, so that more stringent monitoring becomes necessary. The increase in complexity might have a negative affect on the controller's decision-making ability, resulting in more errors. Thus, so-called hot-spots of air traffic congestion necessitate controllers to determine whether conflict avoidance or modification is required for any planned trajectories. The measuring of air traffic complexity is a significant issue in this regard.

In recent years, there has been a growing interest in developing methods to assess the complexity of air traffic. This has led to a variety of approaches to evaluate air traffic complexity which can be broadly categorized into three major groups, including geometry-based approaches, agent-based models, and machine learning techniques listed in Figure 1.2.

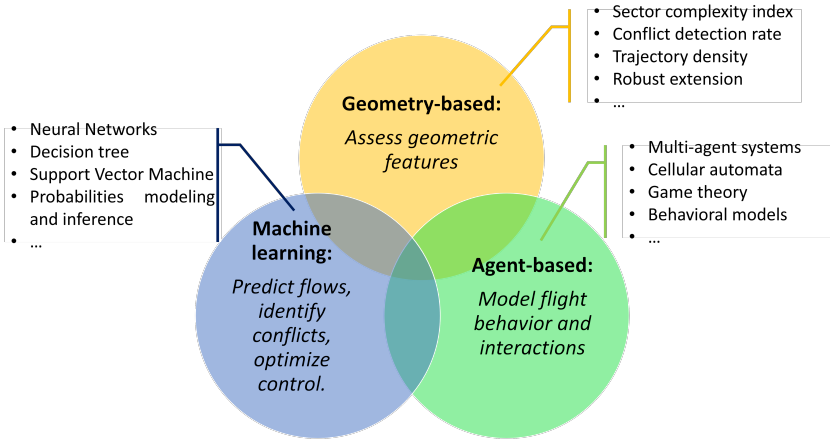


Figure 1.2: Some proposed approaches to evaluate the air traffic complexity.

Geometry-based approaches focus on quantifying the geometric features of airspace, such as the number of sectors, conflicts, and aircraft trajectories. Typically, geometry-based approaches utilize the geometric properties of airspace sectors, such as their size, shape, and orientation. Some common geometry-based approaches include sector complexity index, conflict detection rate, and trajectory density. The specific number of metrics used in geometry-based approaches such as proximity [37], convergence [37], Grassmannian metric [38], König [63], fractal dimension [83], Robust extension

[119]. While these metrics have proven effective in simple scenarios, they are unsuitable for complex or large-scale applications. The approaches also fail to account for spatial-temporal data [124]. The current work attempted to overcome these challenges by implementing an air traffic complexity metric based on linear dynamical systems. Numerous studies have examined traffic structures and airspace geometries to demonstrate the efficacy of this approach to estimate metric for local disorder and interact the trajectory sets in ATM. Its applicability derives from its efficiency, relative simplicity, and mathematically predictable behavior. For example, topological entropy can evaluate the air traffic complexity by capturing the connectivity and structure of the airspace [129, 39, 32]. Or Shannon entropy and Kullback-Leibler divergence were proposed to capture the complexity of air traffic in both spatial and temporal dimensions, and can be used to evaluate complexity at different levels of air traffic control [132, 65].

Agent-based models use a simulation-based approach to model the behavior of individual aircraft and the interaction between them. This approach considers various factors, such as weather, airspace structure, and traffic demand, to assess the complexity of air traffic. Agent-based models have been used to analyze ATFM and air traffic control procedures. Dynamic density [11, 67] only listed a few factors for dynamic density, and each one is assigned a subjective weight. Aspects such as the quantity of aircraft, the number of aircraft with heading changes greater than 15 degrees or speed changes greater than 10 knots, the sector size, and others are taken into account. Some papers are related to the application of agent-based simulation in various aspects of air traffic management such as simulating aircraft taxiing operations [74], assessing unmanned aircraft systems integration in the national airspace system [27, 127], evaluating and simulating air traffic flow management strategies by [126, 128, 44, 46].

Machine learning techniques have become increasingly popular in the field of air traffic complexity evaluation due to their ability to process large amounts of data and extract patterns and relationships that may not be easily identifiable through traditional analytical methods. These approaches have been used in air traffic management to predict traffic flows, identify conflicts, and optimize air traffic control procedures. For example, some researches developed machine learning techniques to evaluate air traffic complexity using flight delay, cancellation, rerouting, and rescheduling [137], or using traffic volume, density, and complexity level [69]. Other researches used machine

learning approach to predict and manage air traffic complexity by considering air traffic volume, flow, weather conditions, and airport capacity [21, 4, 28, 121, 137]. In 2023, Alharbi et al. (2023) [5] used CNN and LSTM to predict the air traffic flow as well as compute the air traffic complexity. The metrics used in the studies varied, including flight delay, cancellation, rerouting, rescheduling, traffic volume, density, complexity level, air traffic flow, weather conditions, and airport capacity. The papers highlighted the potential of machine learning in accurately predicting and managing air traffic complexity. However, limited data availability and potential biases in the data were identified as potential limitations in some studies.

Hybrid approaches for air traffic complexity evaluation typically combine two or more different techniques, such as geometric metrics, agent-based models, or machine learning, in order to obtain a more comprehensive and accurate understanding of the complexity of the airspace system. These approaches have become increasingly popular in recent years, as they allow for a more nuanced evaluation of air traffic complexity that takes into account both the geometric and dynamic aspects of the system. One notable example is the hybrid approach proposed by Marwala and Nelwamondo, which combines machine learning and agent-based methods for air traffic flow management. This method considers metrics such as delay, capacity, and safety, providing a holistic view of air traffic complexity. However, it requires a significant amount of data to train the machine learning models effectively, highlighting the importance of robust data collection and management in air traffic complexity assessment [81]. Another significant contribution is the work of Arribas et al., they integrated geometric metrics (such as distance, density, angle, and altitude difference) with machine learning techniques. Their approach was validated using real air traffic data, demonstrating its effectiveness in capturing the multifaceted nature of air traffic complexity. By leveraging both geometric and data-driven insights, this hybrid method enhances the accuracy and reliability of complexity evaluations, providing valuable information for air traffic controllers and management systems [9].

Incorporating hybrid approaches into air traffic complexity assessment offers several advantages. They enable the integration of multiple data sources and methodologies, resulting in a more detailed and robust analysis. This, in turn, supports the development of more adaptive and responsive air traffic management strategies, crucial for handling the increasing demands of modern airspace operations.

1.3.1 Factors contributing to air traffic complexity

Air traffic complexity is influenced by various factors that affect the management and coordination of aircraft movements. Understanding these factors is essential for improving the efficiency and safety of air traffic operations.

The design and organization of air routes are critical in determining air traffic complexity. Air routes are predetermined paths that aircraft follow, and their configuration can significantly influence traffic flow, congestion, and the potential for conflicts. Efficiently designed air routes help manage complexity by reducing bottlenecks and distributing traffic more evenly. Additionally, airspace is divided into control sectors, each managed by a specific air traffic control unit. The size, shape, and traffic volume of these sectors greatly impact complexity. Smaller sectors with high traffic volumes or complex geometries can increase the difficulty for controllers to manage aircraft movements safely and efficiently.

Traffic density refers to the number of aircraft within a specific volume of airspace at a given time. High-density areas, such as those near major airports or busy air corridors, significantly increase complexity. Managing high-density traffic requires precise coordination and efficient use of available airspace to minimize conflicts and delays. Controllers must be adept at handling a large number of aircraft simultaneously, ensuring safe separation and smooth flow.

Different aircraft types have varying performance characteristics, including speed, maneuverability, and size. These differences contribute to air traffic complexity as controllers must consider the performance limitations and capabilities of each aircraft when planning and coordinating movements. Mixed traffic, involving both fast and slow aircraft, adds to the challenge, requiring careful sequencing and spacing to maintain safety and efficiency.

Weather conditions, such as storms, turbulence, and low visibility, have a substantial impact on air traffic complexity. Adverse weather can force changes in flight paths, increase separation requirements, and lead to delays. Seasonal variations also play a role, with certain times of the year posing more significant weather-related challenges than others. Controllers and pilots must adapt quickly to changing conditions to maintain safe and efficient operations.

Human factors, including the workload, decision-making processes, and communication between air traffic controllers and pilots, are critical elements of air traffic complexity. High complexity can lead to increased stress and fatigue among controllers, potentially impacting their performance and decision-making abilities. Effective training, communication protocols, and support systems are essential for managing these human factors. Ensuring that controllers are well-prepared and supported can mitigate the impact of high complexity on human performance.

Technological advancements, such as Automatic Dependent Surveillance-Broadcast (ADS-B), Next Generation Air Transportation System (NextGen), and the Single European Sky ATM Research (SESAR) initiative, profoundly impact managing air traffic complexity. These technologies enhance situational awareness, improve communication, and provide more accurate and timely information, helping to reduce complexity and improve efficiency. By leveraging these advancements, air traffic management can become more proactive and adaptive, ensuring smoother and safer operations.

Given these factors, accurately determining the complexity of airspace is crucial for effective air traffic management. Understanding complexity helps in resource allocation, workload distribution, and strategic planning, ensuring that both safety and efficiency are maintained. Among the factors discussed, traffic density is often the most influential, as it directly impacts the number of aircraft that must be managed simultaneously, thus driving the need for precise coordination and effective use of airspace. Identifying and addressing areas of high complexity allows for targeted improvements and innovations in air traffic control strategies.

1.3.2 Air traffic complexity metrics

In a control sector, the control workload increases non-linearly with the number of aircraft. There exists a threshold beyond which controllers can no longer accept additional aircraft, necessitating rerouting through less congested neighboring sectors. This state, known as sector saturation, must be avoided as it leads to cumulative overloading in preceding sectors, potentially backing up to departure airports. Estimating the saturation threshold is complex and depends on factors such as route geometry, sector layout, aircraft distribution, and controller performance. A widely accepted threshold is 3

conflicts and 15 aircraft per sector, with a maximum duration of 10 minutes to avoid excessive controller stress, ensuring optimal safety conditions.

Control workload measurement is critical in many areas of ATM, as it underpins optimization processes. Applications include airspace comparison (United States-US/Europe), validation of future concepts SESAR, NEXTGEN, analysis of traffic control action modes, sectorization optimization, and traffic assignment optimization. Additionally, it is essential for determining congestion pricing zones, developing organic control assistance tools, generating 4D trajectories, and predicting congested zones.

Currently, the operational capacity of a control sector is measured by the maximum number of aircraft traversing the sector within a given time frame. This metric does not account for traffic orientation, treating structured and disordered traffic equally. Thus, controllers may accept traffic exceeding operational capacity in structured scenarios while rejecting traffic in disordered scenarios even if capacity is not reached. Therefore, measuring the number of aircraft per unit time is an insufficient metric for representing the difficulty of a traffic situation.

Ideally, a metric should precisely measure the mental effort required to manage a set of aircraft. While challenging, complexity metrics extending beyond simple aircraft counts are possible. Two essential notions are:

- **Control Workload**: Measures the difficulty for the traffic control system, whether human operators or automated processes, to manage a situation. This workload is linked to cognitive processes such as conflict prediction and resolution, and trajectory monitoring.
- **Traffic Complexity**: Intrinsically measures the complexity of a traffic situation, independent of the managing system, based solely on trajectory geometry. This complexity is linked to sensitivity to initial conditions and conflict interdependencies. Uncertainty in positions and speeds complicates future trajectory predictions, sometimes exponentially increasing, making reliable extrapolation difficult. Conflict resolution can generate new conflicts, influenced by trajectory mixing levels. As illustrated in Figure 1.3, traffic situations vary in difficulty based on predictability and trajectory interdependency levels. For example, in a context where radar imaging equipment fails, the situation on the right would attract immediate attention due to its difficulty

in conflict prediction and high trajectory interdependency. The middle situation, though presenting significant conflict risk, is manageable by giving a uniform direction to all aircraft to place them in safe roundabout trajectories. The left situation, with its non-challenging trajectories and stable relative distances, presents minimal immediate difficulties.

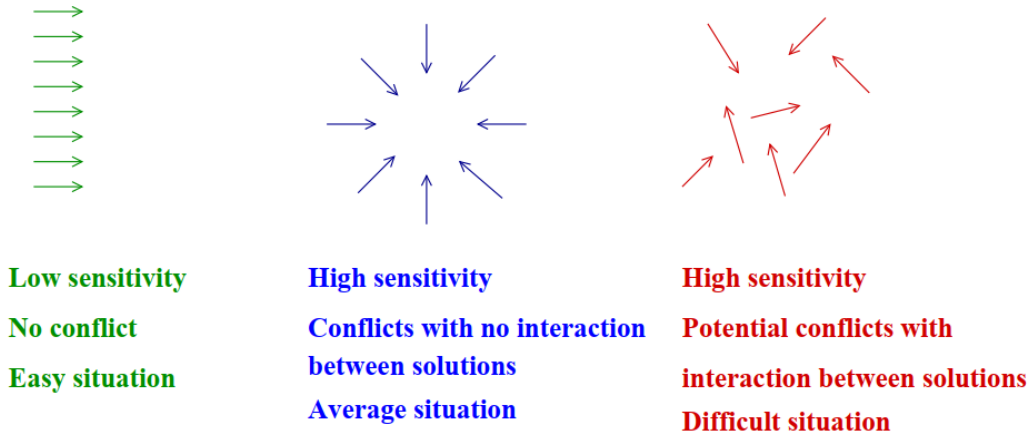


Figure 1.3: An example of three traffic scenarios classified in order of increasing complexity [38].

Research into air traffic complexity metrics has gained considerable attention in recent years, particularly in the US and Europe. The first projects began in Germany in the 1970s, and the subject has since developed. Currently, National Aeronautics and Space Administration (NASA), Massachusetts Institute of Technology, and Georgia Tech are working on the NextGen project in the US, while in Europe, Direction des Services de la Navigation (DSNA), Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Center), and Nationaal Lucht- en Ruimtevaartlaboratorium (Netherlands Aerospace Centre) are involved in similar activities related to SESAR.

Thus, metrics are essential for measuring and managing air traffic complexity. They provide quantitative data that can be analyzed to understand the level of complexity in a given airspace and to develop strategies for managing it. Different metrics focus on various aspects of complexity, offering a comprehensive view when used together. In this chapter, we will present the following three approaches to measure air traffic complexity:

- Flow-based metrics.
- Geometry-based metrics.
- Dynamic-based metrics.

Flow-based metrics

The flow-based metrics measures complexity based on traffic flow and density. It focuses on the number of aircraft passing through a specific area over a period and the interactions between these aircraft. Traffic density metrics measure the number of aircraft within a defined airspace volume, while the traffic flow rate assesses the rate at which aircraft enter and exit a specific sector or route. Flow-based metrics are particularly useful in high-density airspaces, such as those around major airports. Monitoring traffic density can help identify potential bottlenecks and enable proactive management to prevent congestion. Flow-based metrics are straightforward and easy to interpret, making them useful for quick assessments. However, they may not capture the full complexity of interactions between aircraft or account for dynamic changes in traffic patterns.

Geometry-based metrics

The geometry-based metrics (based on the geometric distribution of speed vectors in the airspace) focuses on the spatial aspects of air traffic complexity. It considers the positions, trajectories, and separation distances between aircraft. Proximity measures evaluate horizontal and vertical separation between aircraft, while conflict probability assesses the likelihood of potential conflicts based on current trajectories. Geometrical metrics are particularly useful in sectors with complex airspace structures. For example, evaluating proximity measures can help ensure safe separation distances are maintained, reducing the risk of conflicts. Geometrical metrics provide detailed insights into spatial relationships between aircraft, offering a clear picture of potential conflicts. However, they can be complex to calculate and may require advanced tools and software for accurate assessment.

Dynamic-based metrics

The dynamic-based metrics using a dynamic system (linear or non-linear) to model air traffic focuses on the temporal aspects and changes in traffic patterns over time. It considers how traffic evolves and the stability of traffic flows. Temporal stability assesses how stable traffic patterns are over time, while the rate of change in traffic patterns measures the speed at which traffic patterns change, indicating potential volatility. Dynamical metrics are useful in managing air traffic during peak periods or unexpected disruptions. For example, assessing temporal stability can help predict and manage traffic surges, ensuring smoother operations. Dynamical metrics provide insights into the temporal dynamics of air traffic, helping to anticipate and manage fluctuations. However, they can be challenging to interpret without comprehensive data and may require sophisticated analytical tools.

Accurately measuring air traffic complexity is crucial for effective air traffic management. Metrics enable air traffic controllers and planners to allocate resources efficiently, distribute workloads effectively, and implement strategic planning to maintain safety and efficiency. Among the various metrics, those related to traffic density often have the most significant impact, as they directly influence the number of aircraft that need to be managed simultaneously. Understanding and utilizing these metrics allows for targeted improvements in air traffic control strategies, helping to address areas of high complexity and enhance overall ATM.

1.4 Contributions

Here, we contribute to resolve the following challenges: First and foremost, in the context of flight trajectory prediction, ensuring high accuracy is pivotal for effective air traffic management. Existing methods often struggle to capture the intricate spatial and temporal dependencies inherent in flight data, leading to sub-optimal prediction performance.

Additionally, reconstructing the heading angle distribution plays a crucial role in understanding the directional dynamics of airspace traffic. However, traditional approaches fail to provide a comprehensive framework for reconstructing these distributions accurately from observed or simulated data.

Moreover, assessing spatial traffic complexity is essential for optimizing ATM strategies. Current methodologies lack a systematic approach to quantify the intricate spatial relationships within airspace, hindering efforts to improve overall airspace efficiency and safety.

The four key contributions of the thesis are:

1. A method named BayesCoordLSTM for enhancing flight trajectory prediction performance. In BayesCoordLSTM method, a coordinate transformation step was proposed to convert the flight trajectory coordinates into cylindrical coordinates. Then, we proposed to use a hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architectures to capture spatial features and temporal dependencies of flight data. Finally, Bayesian theorem that provides a powerful framework for incorporating prior knowledge and updating predictions based on new information was employed. Bayesian theorem allows for a more nuanced and probabilistic approach to trajectory prediction. Using Bayesian theorem can better estimate the uncertainties inherent in trajectory predictions and develop more effective strategies for managing air traffic in increasingly congested airspace. Experimental results on a real dataset confirm that the proposed model enhances predictive capabilities, providing probabilistic trajectory forecasts with increased reliability.

2. A method for constructing spatial traffic complexity maps grounded in the Maximum Entropy principle: This contribution addresses the challenge of reconstructing heading angle distributions from observed or simulated data. By employing the principle of Maximum Entropy, this methodology incorporates statistical estimation of Fourier coefficients to assess density and employs Mahalanobis distance to optimize density via partially finite convex programming. Our approach enables the precise reconstruction of heading angle distributions, offering valuable insights into the directional behavior and dynamics of air traffic flow. The resulting probability distributions are then used to produce spatial complexity maps, which provide a visual and analytical representation of air traffic complexity. These maps can be instrumental in identifying areas of potential congestion and in guiding traffic management strategies to ensure the safe and efficient movement of air traffic. The detailed information are to be exploited to produce complexity maps,

as explained in the next point. The resulting probability distributions are to be exploited to produce complexity maps, as explained in the next point.

3. A method for building a spatial complexity map: This contribution presents a systematic framework for assessing spatial traffic complexity within airspace. This methodology to assess local air traffic complexity using the Maximum Entropy principle and information geometry on generalized von Mises distributions. Utilizing information geometry, it computes geodesic distances to the uniform distribution and creates complexity maps by using Composite Complexity Index that evaluate an index combining angular distribution and traffic intensity. These maps offer a nuanced understanding of air traffic complexity, enabling more informed ATM strategies. Our approach demonstrates the numerical efficiency and effectiveness of the strategy with real data.

4. A method for reconstructing spatial density based once again on the Maximum Entropy principle. This is a fairly general method, which could be used in ATM for the reconstruction of instantaneous aircraft density in airspace, at the pre-strategic stage. This method largely utilizes the tools from the chapter on the reconstruction of angular densities, but their use differs significantly. The developed approach not only addresses the unique requirements of spatial density reconstruction in aviation but is also versatile and adaptable, making it suitable for a broad range of other applications in statistics. For instance, it could be effectively employed in environmental studies for the modeling and prediction of species distributions across different habitats, thereby highlighting its versatility beyond aviation.

1.5 Structure of the thesis

The thesis is organized as follows: Chapter 2 introduces an new method to predict flight trajectories, named BayesCoordLSTM. It is a hybrid model that transforms coordinate and applies Bayesian theorem into the ConvLSTM models. The proposed model leverages the spatial features gleaned by the CNN architecture and the temporal dependencies captured by LSTM to enhance the accuracy of trajectory predictions. By incorporating Bayesian the-

orem, our model provides a probabilistic trajectory forecasts and associated confidence levels while coordinate transformation enhances spatial awareness and predictive capabilities. We test on the flight data collected from the Hartsfield–Jackson Atlanta International Airport (ATL). Our experimental results demonstrates the effectiveness of the proposed BayesCoordLSTM-based approach in improving flight trajectory prediction accuracy, with a focus on Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values. The integration of the Bayesian theorem and Coordinate transformation into ConvLSTM models represents a substantial advancement in the field of flight trajectory prediction.

An important issue in air traffic management is the construction of traffic complexity maps. Observed or simulated data provide a statistical sample of trajectory angles. The underlying angular density is a crucial component in assessing local complexity. In Chapter 3, we propose a Maximum Entropy method for reconstructing the angular densities. These densities are constrained by the estimation of there Fourier coefficients. Accounting for the statistical properties of moment estimation is achieved by using the square of the Mahalanobis distance between statistical data and Fourier coefficients of the sought density as the data fitting term. Partially finite convex programming is then implemented to compute the optimal density. The cornerstone of this approach, enabling a rigorous treatment of variational aspects of the chosen method, is an infinite-dimensional version of Fenchel's duality theorem, coupled with results on the conjugacy of integral functionals. We also discuss the numerical aspects of our approach and present some simulations and reconstructions from simulated data.

In order to define and compute the air traffic complexity map, we use information geometry. Chapter 4 focuses on the methodology for building air traffic complexity maps by using the geodesic distance between the Maximum Entropy densities obtained in the previous chapter and the uniform distribution, which corresponds for a given density of traffic to the maximal complexity. We also test on real flight data, demonstrates robustness, with alternative criteria possible by comparing heading distributions at grid cell boundaries. Our strategy for building complexity maps is both numerically efficient and effective in capturing air traffic complexity.

In Chapter 5, we develop a Maximum Entropy approach to the evaluation of spatial density. This methodology extends upon previous contri-

butions in the field, particularly those related to the modeling of species distribution. However, it introduces significant refinements by avoiding the discretization of the sample space, thus allowing for a more continuous and precise representation of spatial density. This continuous approach provides a more accurate framework for analyzing complex spatial patterns, which is particularly valuable in various scientific and engineering applications. While this chapter focuses primarily on the theoretical development and demonstration of the methodology, its potential use in ATM, particularly for optimizing airspace utilization and enhancing safety, will be explored and expanded upon in future work.

Finally, Chapter 6 provides a summary of the thesis, gives conclusions and potential research directions based on the presented results and methodologies that were explored.

Chapter 2

Applying the Bayesian method to improve flight trajectory prediction

Flight trajectory prediction is essential for aviation, playing a key role in efficient air traffic management and ensuring safe flight operations. Traditional methods often struggle to manage the complex spatio-temporal dependencies and uncertainties inherent in aviation data. In this study, we propose a novel approach, BayesCoordLSTM, for trajectory prediction. This hybrid model combines Convolutional LSTM (ConvLSTM) with the Bayesian method and coordinate transformation. ConvLSTM leverages CNN architecture to extract spatial features, while LSTM captures temporal dependencies. By incorporating the Bayesian method, our model offers probabilistic trajectory forecasts with associated confidence levels. Additionally, coordinate transformation enhances spatial awareness and predictive accuracy. Experimental results show that the BayesCoordLSTM model significantly improves prediction accuracy, as indicated by MAE and RMSE metrics. The integration of the Bayesian method and coordinate transformation into ConvLSTM models represents a significant advancement in flight trajectory prediction.

2.1 Introduction

Deep neural networks (DNNs) are extensively applied in various domains, including safety-critical areas such as self-driving automobiles and security systems that use face and image recognition. However, DNNs are considered as deterministic models because all unknown parameters, such as net-

work weights, biases, and model predictions, are single point-estimates. This determinism necessitates a large dataset for training, which makes convolutional neural networks (CNNs) vulnerable to overfitting on small datasets. Consequently, DNNs may perform well on training data but fail to generalize to new data, leading to overconfidence or uncertainty in their predictions. This issue can be addressed using the Bayesian method.

The Bayesian method has become crucial for enhancing DNNs. In a Bayesian model, the posterior distribution provides comprehensive information about unknown parameters given the data. Various techniques, including Markov Chain Monte Carlo (MCMC), Laplace approximation, expectation propagation, and variational inference, have been employed to approximate the posterior distribution. These methods make Bayesian models robust to overfitting, capable of estimating uncertainty, and effective at learning from smaller datasets. Variational inference, in particular, minimizes the Kullback-Leibler divergence between a proposed approximate posterior and the true unknown posterior distribution of the weights, addressing the intractability of the exact Bayesian method.

In air traffic management (ATM), the increasing airspace demand puts pressure on current systems to ensure aircraft maintain safe separation standards to avoid collisions [40]. This necessitates accurate predictions of airplane orbits in the near future [98]. However, this is challenging due to the uncertainty inherent in predicting aircraft trajectories. Therefore, understanding and managing this uncertainty is crucial in ATM, where meteorological factors, especially winds, represent significant sources of uncertainty. Uncertainty in weather forecasts arises from inadequate or inaccurate knowledge of the atmosphere's state at the time of prediction and the uncertainties of the model [40]. These uncertainties propagate through nonlinear and chaotic atmospheric dynamics, making simple statistical characterization inadequate for depicting prediction uncertainty. Consequently, these factors can affect aircraft velocity and the accuracy of trajectory predictions. Most methods for flight conflict resolution do not account for this uncertainty. As a result, they produce solutions like collision-free flight routes that are vulnerable to disturbances and increase the risk of violating separation standards.

As air transport data volumes increase, novel analysis techniques and operational strategies are crucial in ATM for more precise trajectory prediction models. However, uncertainties, including weather and unknown vari-

ables from un-modeled actors, remain significant challenges in predicting flight trajectories [88]. Researchers have explored many approaches, from physics-based methods relying on flight dynamics to data-driven methods using machine learning algorithms to analyze historical data [70]. As demands on ATM grow and the need for more accurate predictions intensifies, there is a rising interest in data-driven approaches to overcome the limitations of traditional methods. One significant challenge in predicting aircraft trajectories is the uncertainty arising from factors such as changing weather conditions. Many researchers utilize the Bayesian method to quantify uncertainty in predicting time-series data, including flight trajectory prediction in ATM.

Despite significant advancements in trajectory prediction, data-driven approaches remain susceptible to overfitting and often struggle to generalize well to new and unseen flight patterns. Furthermore, information about an aircraft position, altitude, and heading is typically processed relative to a reference point, making coordinate transformation essential for accurately understanding and predicting aircraft trajectories. In this chapter, we aim to improve flight trajectory prediction accuracy by integrating Bayesian theory and coordinate transformation in a proposed model called BayesCoordLSTM. The primary contributions of this study can be summarized as follows:

- **Coordinate transformation**. By adopting coordinate transformation, the model gains a significant advantage in spatial awareness and predictive capabilities. This transformation enhances the model's understanding of the spatial aspects of aviation trajectories, allowing it to effectively relate positions and headings to a reference point. As a result, the model is able to make more accurate predictions.

- **Bayesian integration**. We apply the Bayesian method to update the weights of hyperparameters in a hybrid prediction method that combines CNN and LSTM. The 1D convolutional layer of the CNN extracts spatial features, while the LSTM captures the temporal features of the flight dataset. By incorporating Bayesian principles, our model effectively handles uncertainty in predictions, offering probabilistic trajectory forecasts along with confidence levels. This comprehensive approach not only enhances predictive accuracy but also optimizes hyperparameters.

- **Empirical validation**. We test our model on a real dataset and compare its prediction performance with three models, including 3D-CNN [95], CNN-Gated Recurrent Unit (GRU) [106], and a combination of CNN-GRU with 3D Convolution (CG3D) [107]. The analysis highlights the superior accuracy and robustness of our proposed model in flight trajectory prediction, demonstrating its practical utility in ATM.

The remainder of this chapter follows this structure: Section 2.2 presents a brief literature review on flight trajectory prediction. Some basic information of LSTM are presented in Section 2.3. Section 2.4 illustrates our model and the experimental results of a real flight dataset are presented in Section 2.5. Finally, Section 2.6 summarizes the conclusions and outlines future works in this field.

2.2 Literature review

In recent years, the demand for air travel has significantly increased, leading to congestion in the control sectors. With the rapid increase in global air traffic, particularly in high-density airspace regions, ensuring the safety and efficiency of air transportation has become a critical challenge. Air traffic control is a service provided by ground-based controllers who direct aircraft on the ground and through controlled airspace, and can provide advisory services to aircraft in non-controlled airspace.

The primary purpose of air traffic control is to prevent collisions, organize and expedite the flow of air traffic, and provide information and other support for pilots. This involves enforcing traffic separation rules to ensure that each aircraft maintains a minimum amount of empty space around it at all times. Despite advancements, tactical air traffic control decisions to ensure safe separation between aircraft are still made by human controllers, similar to practices from 50 years ago [113, 15, 33].

According to the Federal Aviation Administration (FAA) and Euro Control Action Plan 16 (AP16), a trajectory is defined in four dimensions: latitude, longitude, altitude, and time. Due to factors such as airspace congestion, weather conditions, temporary military activities, and airspace restrictions, the actual and planned flight paths often differ (see Figure 2.1).

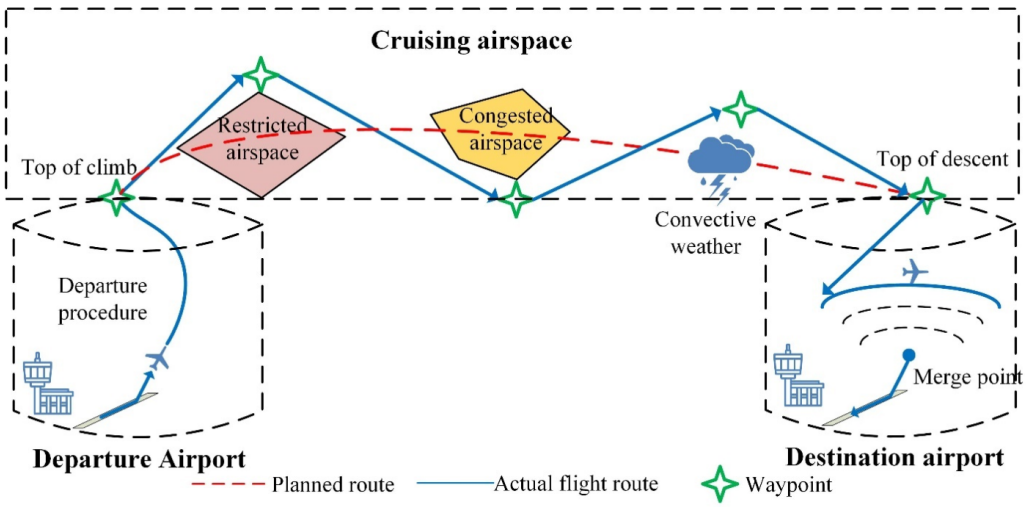


Figure 2.1: Different stages of a flight [135].

As part of ATM, ATFM provides services that complement air traffic control. ATFM aims to ensure an optimal flow of air traffic through areas where demand exceeds capacity, thus protecting air traffic control from overload situations that could compromise safety [3, 72]. Accurate flight trajectory prediction is vital for preventing traffic jams and collisions, optimizing airspace utilization, and mitigating congestion challenges [136].

Flight trajectory prediction involves estimating a flight's future position, altitude, heading, and speed based on historical data and current state. There are two main types of trajectory prediction: strategic and tactical predictions. Strategic prediction forecasts possible future trajectories based on flight plans, weather forecasts, and historical patterns, while tactical prediction incorporates dynamic information like current flight status and airspace congestion. Additionally, trajectory prediction can be categorized into short-term and medium to long-term based on the time scale [125].

- **Short-term prediction**. This involves predicting the flight trajectory over a period of a few minutes or less. Due to the brief forecast period, detailed information about weather conditions and long-term flight intentions is not required. However, assumptions such as fixed aircraft control and constant rotational speed are necessary. These assumptions are valid for a shorter propagation period, enhancing pre-

diction accuracy while reducing the size of the prediction interval. This type of prediction facilitates immediate conflict risk detection, allowing for effective conflict resolution. Additionally, accurate trajectory predictions can support the generation of viable alternative trajectories, accommodating existing constraints.

- **Medium to long-term prediction**. This involves predicting the flight trajectory over a period of ten minutes or more. Due to the extended prediction horizon, it requires essential information such as long-term flight intentions, environmental data, flight performance data, and navigation data. The inherent uncertainties in these inputs can reduce prediction accuracy and extend the prediction period. The primary purpose of this type of prediction is to assess airspace flow and develop flight plans by the aircraft operations center or flight control center.

Trajectory prediction methods can generally be categorized into two primary approaches: physics-based methods and data-driven approaches.

Physics-based methods rely on classical mechanics and aerodynamics principles to model flight motion. These methods use mathematical equations to describe the dynamics of a flight, taking into account forces like lift, drag, thrust, and gravity. In this approach, the trajectory prediction is based on the flight dynamics model, motion characteristics, and the Markov property of its state [136]. This approach is highly solvable and performs well in short-term predictions, especially for large flights with relatively stable motion parameters. The next moment's state of the flight is closely linked to its current state, following the principles of Markov theory [123]. For instance, researchers frequently apply the Kalman Filter to estimate future trajectories based on noisy measurements [138]. These approaches provide valuable insights into flight dynamics and are often employed for trajectory estimation under controlled and predictable conditions. However, as the prediction time span increases, the error in these state estimation models tends to escalate rapidly [136].

In the context of air traffic control, short-term trajectory predictions (typically up to 6 seconds on radar screens) rely on extrapolating speed trends using simple state estimation models. Ongoing efforts aim to improve accuracy in long-term trajectory predictions by addressing challenges related to uncertainties and nonlinearities in flight dynamics over extended periods.

Therefore, in recent years, data-driven approaches have gained popularity, leveraging historical flight data and other relevant variables to learn patterns and relationships. These patterns are used to create a data-driven model, which is trained to represent the collective behavior of the trajectories based on input features (e.g., past positions, velocities) and future trajectory predictions using machine learning models. According to [35], although machine learning is one of the most researched topics in computer science, it has not yet been widely adopted by end-users in ATM domain. This research demonstrates that the number of papers on machine learning in ATM has significantly increased in the recent years, especially in predicting and optimizing hotspots to avoid collisions and managing traffic flows - one of the most consistent subjects of machine learning in ATM. Using the machine learning model, it can extract the common traits by learning from extensive historical trajectory data. These techniques can also incorporate additional inputs such as aircraft performance, air route structures, and environmental factors (data-to-data). Machine learning models consistently provide accurate predictions and benefit from a wide range of trainable parameters. Supervised learning, in particular, is a well-known approach used by many researchers for trajectory prediction. By learning from large datasets of historical trajectory data, these models can generalize and predict future trajectories effectively. Furthermore, these methods can utilize flight performance metrics, flight path structures, and environmental parameters as additional inputs for enhanced trajectory prediction.

Machine learning algorithms consistently provide accurate predictions and profit from a wide range of trainable parameters [34]. These approaches often use filter-based methods [94, 130] or deep learning-based methods [107] to extract the properties of data changes through data analysis to forecast the trajectory. Recently, significant advancements have been made in flight trajectory prediction through the use of advanced machine learning techniques, particularly CNN and Recurrent Neural Network (RNN). Among the various types of RNNs, LSTM networks have proven particularly effective for handling sequences and time series data. In 2021, Shi et al. proposed a constrained LSTM model specifically for predicting flight trajectories [111]. In their research, they defined three types of constraints: climbing, cruising, and descending/approaching phases. Their constrained LSTM model incorporates these phase-specific constraints to account for the dynamic nature of flight, ensuring both long-term dependencies and realistic physical

constraints. Compared to a standard LSTM, this model significantly enhances prediction accuracy by maintaining trajectory continuity and effectively managing sparse way points in flight data. Other methods applied the Bayesian method in machine learning models: Graph Convolutional Network to forecast the traffic flow in the next 15, 30, 45, and 60 minutes (SZtaxi and Los-loop) [45], a deep Gaussian process [29], LSTM and the LSTM's variants, including gated LSTM [31], convolutional LSTM [110, 87, 108], CNN-LSTM [73, 76], deep LSTM [87, 131, 30, 1], and sequence-to-sequence LSTM [114]. CNNs, RNNs, and LSTMs were used to extract spatial features from local regions and long-range dependencies in images, respectively. These studies provide valuable references for the model architecture used in this work.

LSTM networks are designed to handle sequential data and capture temporal dependencies effectively, making them well-suited for tasks involving time series data. However, these models are not inherently equipped to capture spatial features unless they are modified or combined with other methods. In applications such as flight trajectory prediction, where both temporal and spatial features are crucial, a standard LSTM may fall short in adequately addressing spatial aspects. This is why additional constraints or hybrid models that incorporate spatial information are often necessary. In this context, CNN is more suitable for extracting spatial features. Hence, a hybrid CNN and LSTM [76, 71] or combined CNN and GRU [107, 116] approaches have been widely used in classification and prediction tasks. For example, Shafienya et al. (2022) [106] compared various CNN-GRU models for predicting 4D flight trajectories using ADS-B data. CNNs extract spatial features from flight data, while GRUs capture temporal dependencies. Combining CNNs with GRUs enhances prediction accuracy by effectively capturing both spatial and temporal patterns. They subsequently extended their research by combining a CNN-GRU with a 3D-CNN, which effectively extracts spatial-temporal features for prediction [107]. However, a potential drawback of their models is their reliance on large datasets for training and evaluation, as well as the complexity involved in parameter tuning for the combined CNN-GRU and 3D-CNN architectures.

Additionally, these approaches often learn to predict only specific types of flight trajectories. To handle different trajectory types or routes, researchers must retrain the model or adjust its parameters. Although machine learning has achieved significant advancements in trajectory prediction,

it still has limitations, such as learning only specific types of trajectories. Therefore, if the model needs to predict different types of flights or routes, it must be retrained by adding or adjusting hyperparameters. This process increases both the cost and the time required for retraining.

Moreover, in flight trajectory prediction, uncertainty can arise from factors such as weather conditions or unpredictable aircraft behaviors, as well as from the data itself or the optimal values of model parameters. Failure to capture these uncertainties can result in models that do not provide probabilistic predictions or associated confidence levels, ultimately diminishing prediction accuracy. This can have significant implications for safety and performance in the aviation industry.

To address this problem, the Bayesian framework offers several advantages: it accommodates various prior ranges, avoids costly computations, integrates well with LSTM models for continuous-time responses, and effectively handles large datasets. Many researchers have applied the Bayesian method to various models, including Graph Transformers [88] and CNN-GRU architectures [59]. This approach not only estimates model uncertainties but also optimizes predictions to enhance performance [10, 2].

In a 3D space, as aircraft move, information about their position, altitude, and heading is processed relative to a reference point. Without coordinate transformation, the model may not effectively capture the real spatial understanding or accurately depict the complex variations in flight trajectories [118]. This limitation diminishes the accuracy and reliability of aircraft position predictions. Therefore, integrating the Bayesian method with coordinate transformation into these models enhances their capability to handle flight trajectory predictions. This combination improves the model's ability to manage uncertainty, benefiting ATM by enhancing safety and efficiency, and optimizing model parameters for better and more stable performance.

By learning from historical data, data-driven approaches can adapt to changing conditions and improve the accuracy and reliability of flight path estimation. However, challenges such as generalizing from sparse data, dealing with unpredictable weather conditions, navigating dynamic airspace structures, and achieving real-time prediction remain. To address these challenges, integrating physics-based knowledge with data-driven techniques presents a promising solution. This hybrid approach leverages the strengths of both methods, offering enhanced prediction accuracy and reliability in ATM [70].

2.3 Long Short-Term Memory

2.3.1 LSTM architecture

Long Short-Term Memory (LSTM) is an advanced variant of RNN introduced by Hochreiter and Schmidhuber in 1997 [56]. The main purpose of LSTM is to overcome the limitations of traditional RNNs, such as difficulties in learning and remembering long-term dependencies due to the vanishing gradient or exploding gradient problems. LSTM allows neural networks to retain information for longer periods and learn long-term dependencies, which are important for tasks involving sequential data, such as natural language processing, speech recognition, and time series prediction. Figure 2.2 illustrates the LSTM architecture.

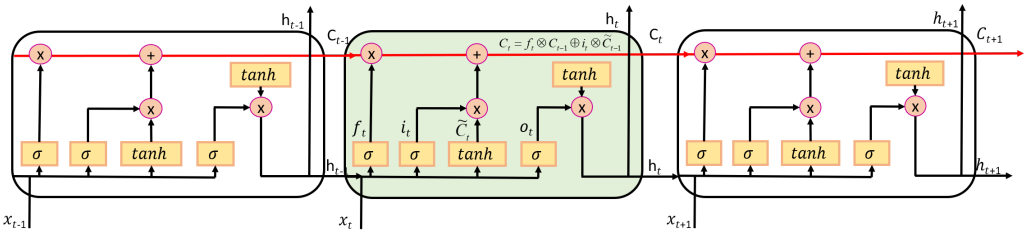


Figure 2.2: LSTM architecture [56].

The architecture of LSTM consists of a series of cells connected sequentially, where each cell C at time step t is responsible for storing and updating information from the previous time step $(t-1)$. The key improvement of LSTM over traditional RNNs lies in the structure of these cells, which include three primary gates that manage and regulate the flow of information:

- Forget gate f: decides which information from the previous cell state should be discarded.
- Input gate i: determines which new information should be added to the cell state.
- Output gate o: controls which part of the cell state should be output as the current hidden state.

These gates consist of sigmoid layers, with output values between 0 and 1. These values decide how much information to retain or discard, enabling LSTMs to better control the flow of information over time.

2.3.2 LSTM operation

LSTM works through a series of operations to update the cell state and compute the hidden state at each time step t. We denote by x_t, C_t and h_t as the input, the cell state and the hidden state at time t, respectively.

The operations in LSTM can be divided into the following steps:

First, the LSTM determines which information from the previous cell state (C_{t-1}) should be forgotten or retained. This decision is made through the *forget gate*, which takes the previous hidden state h_{t-1} and the current input x_t as inputs:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \tag{2.1}$$

The output f_t is a value between 0 and 1, representing the fraction of information to retain from the previous cell state. If $f_t = 0$, all the information is discarded, and if $f_t = 1$, all the information is retained.

Next, the LSTM decides which new information will be added to the current cell state. This is done by the *input gate*, where two main operations take place:

- A sigmoid layer determines which values should be updated:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i). \tag{2.2}$$

- Then, a tanh layer creates a vector of candidate values $\tilde{C}_t$, representing potential new information to be added to the cell state:

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c). \tag{2.3}$$

The new cell state C_t is updated by combining the filtered previous state and the new candidate values. This update is performed as follows:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \tag{2.4}$$

where f_t helps in discarding or retaining the old information while i_t and $\tilde{C}_t$ add new information to the cell state.

Finally, the LSTM decides what to output as the hidden state at the current time step through the *output gate*. The output gate uses a sigmoid layer to filter the cell state and determine which parts of it should be used to compute the hidden state:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o). \tag{2.5}$$

The final hidden state h_t is computed by multiplying o_t by the tanh of the updated cell state C_t:

$$h_t = o_t * \tanh(C_t).$$

2.4 Proposed model

While LSTMs excel at capturing temporal dependencies, they may not adequately capture spatial features. On the other hand, CNNs are well-suited for extracting spatial information. Consequently, the combination of CNNs and LSTMs has been widely adopted for various classification and prediction tasks [109, 58].

Moreover, applying the Bayesian method to LSTM models has been proposed by many researchers [82, 122, 133, 105]. The Bayesian framework is flexible with various prior ranges, avoids expensive computations, and integrates with LSTM models to develop continuous-time responses and handle large datasets. This approach allows hyperparameters to be tailored to the specific characteristics of the data. Therefore, it not only estimates the model's uncertainty but also optimizes predictions [8, 10, 2, 112].

To enhance the performance of flight trajectory prediction, we propose a hybrid model named BayesCoordLSTM. This model combines coordinate transformation and the Bayesian method within a CNN-LSTM framework. As illustrated in Figure 2.3, the BayesCoordLSTM comprises two main components: a Convolutional Neural Network (CNN) and a Bayesian LSTM network.

The first component of the BayesCoordLSTM is a CNN, which extracts spatial features from the input data, allowing the model to effectively capture

and process intricate spatial patterns in flight trajectories. The second component is a Bayesian LSTM network, designed to capture long-term temporal dependencies. By incorporating the Bayesian method, the LSTM network enhances its ability to manage uncertainties and provide more robust predictions over extended time sequences. Together, these components enable the BayesCoordLSTM to deliver superior performance in flight trajectory prediction by integrating advanced spatial and temporal data processing with probabilistic reasoning.

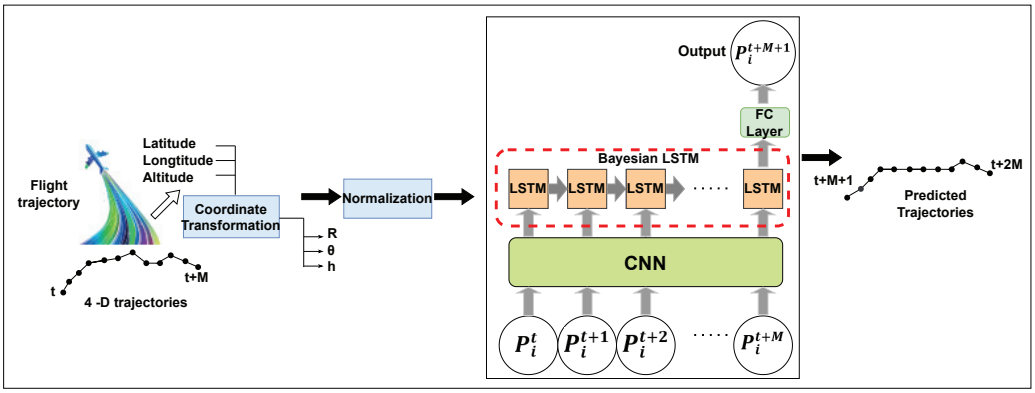


Figure 2.3: The BayesCoordLSTM framework to predict flight trajectory.

In the context of this research, the inputs and outputs of our model for flight trajectory prediction are defined as follows:

- **Input**: $\mathscr{X}$,X_i^t, P_i^t.

  Where:

  - $\mathscr{X} = \{X_i\}$ denotes the set of flight trajectories. Here i represents the trajectory number ranging from 1 to N, where N is the total number of flight trajectories. Each X_i represents the flight path of trajectory i.
  - A trajectory includes a series of multidimensional trajectory points ordered by time: $X_i^t = \left\{P_i^t, P_i^{t+1}, \ldots, P_i^{t+M}\right\}$ is a list of M points of flight trajectory. P_i^t denotes the point of the observed flight trajectory i at time t.
  - Each trajectory point P_i^t is represented as

$$P_i^t = (long_i^t, lat_i^t, alt_i^t, v_i^t, \theta_i^t),$$

where each point has five components, corresponding to longitude, latitude, altitude, speed, and heading of the flight at time t, respectively.

- **Output**: $\left(P_i^{t+M+1}, P_i^{t+M+2}, \ldots, P_i^{t+2M}\right)$ represents the predicted flight trajectories.

For example, assume that we use 100 points to predict five points in the future, so that $M = 100$. For flight i, starting at time $t = 1$, we have $P_i^t = (P_i^1, P_i^2, \ldots, P_i^{100})$. The output then becomes $(P_i^{101}, P_i^{102}, \ldots, P_i^{105})$, corresponding to five points in the future.

2.4.1 Coordinate transformation

In the trajectory prediction framework, both the input and output trajectories are converted into discretized vectors to ensure numerical stability in deep learning models. Flight trajectories involve latitudes, longitudes, and altitudes that span different value ranges. To achieve uniformity, these variables are normalized based on their specific minimum and maximum values for a given airport. However, this normalization approach lacks generalizability across different airports due to varying latitude and longitude ranges. To address this issue, trajectories are transformed into a cylindrical coordinate system, aligning flight states relative to the airport's position [24]. Given the flight position $\mathbf{x}_{fp} = (long_{fp}, lat_{fp}, alt_{fp})$ and the airport position $\mathbf{x}_{ap} = (long_{ap}, lat_{ap}, alt_{ap})$, we can transform coordinates as follows:

$$R = \frac{\|\mathbf{x}_{\text{fp}} - \mathbf{x}_{\text{ap}}\|}{D_{\max}}, \tag{2.6}$$

$$\theta = \arcsin\left(\frac{long_{\text{fp}} - long_{\text{ap}}}{\|\mathbf{x}_{\text{fp}} - \mathbf{x}_{\text{ap}}\|}\right), \tag{2.7}$$

$$h = \frac{|alt_{\text{fp}} - alt_{\text{ap}}|}{A_{\max}}, \tag{2.8}$$

where:

- D_{max} is the distance extending from the airport and encompassed by the predictive scope of the framework;

- A_{max} corresponds to the highest achievable altitude within the predictive domain;
- R, h: are normalized distances to the airport and altitude difference;
- θ: is the angle (heading) with the north direction.

In this thesis, we choose $D_{max} = 328084$ feet and $A_{max} = 39370$ feet to predict aircraft trajectories, balancing computational efficiency and practicality. These values provide sufficient coverage while conserving computational resources and align with typical aircraft operating ranges and altitudes, ensuring relevance to real scenarios.

2.4.2 CNN layer

Using the Conv1D layer plays a crucial role in high-level feature extraction, enabling the network to effectively understand and process temporal dependencies within the input data. In this layer, the input consists of *6 channels*, representing different features (timestamp, longitude, latitude, altitude, speed, and heading). The layer utilizes *1-dimensional filters* that slide across the time dimension to capture relevant patterns within the data.

1. Convolution operation: Each filter performs a convolution operation, which involves multiplying the filter weights with the input values and summing the results to produce a single output for each position. This process highlights important features by learning spatial hierarchies.
2. Activation function: After convolution, the *ReLU (Rectified Linear Unit)* activation function is applied. The ReLU function introduces non-linearity by retaining positive values while setting negative values to zero, allowing the model to learn complex relationships within the data.
3. Output: The output from the Conv1D layer consists of *32 feature maps*, each representing a learned pattern from the input data. This enriched representation retains critical information for subsequent layers in the model.

2.4.3 Bayesian optimization

Bayesian optimization is a powerful technique for optimizing complex, high-dimensional functions with costly evaluations, such as the hyperparameters of deep learning models. Incorporating the Bayesian methods into LSTM models, as proposed by several researchers [82, 99, 139, 88, 59], involves creating a probabilistic model to capture the uncertainty in the function being optimized. This approach is advantageous because it enables a more efficient and informed optimization process, leveraging prior knowledge to minimize the number of evaluations required. The Bayesian framework's flexibility with various prior ranges, its avoidance of expensive computations, and its compatibility with LSTM models for continuous-time responses make it well-suited for handling large datasets. By utilizing Bayes' theorem, the hyperparameters of the LSTM model can be tuned to specific characteristics of the data, enhancing model performance and providing insights into uncertainty.

In the context of Bayesian LSTM modeling, where uncertainty is critical for parameter estimation, we use a general formula for sampling weights W and biases b at the t^{th} time step of the n^{th} layer. These are given by Equation (2.9) and Equation (2.10), respectively:

$$W_{(n)}^{(t)} = \mathcal{N}(0,1) * \log\left(1 + \rho_{(w)}^{(t)}\right) + \mu_{(w)}^{(t)}, \tag{2.9}$$

and

$$b_{(n)}^{(t)} = \mathcal{N}(0,1) * \log\left(1 + \rho_{(b)}^{(t)}\right) + \mu_{(b)}^{(t)}, \tag{2.10}$$

where $\mathcal{N}(0,1)$ represents sampling from a standard Gaussian distribution, introducing randomness in weight and bias initialization, and ρ and μ are discussed hereafter. This randomness is vital for effectively exploring the solution space during training.

The term $\log(1+\rho)$ transforms the estimated standard deviation ρ of the input feature into a non-negative space, ensuring the non-negativity of sampled weights and biases, which is essential for numerical stability during both training and inference.

Moreover, incorporating the mean value μ derived from the training data into the sampling process enables the model to utilize prior knowledge about the distribution of weights and biases. This enhances parameter estimation efficiency and improves the model's performance to capture complex temporal dependencies.

In this phase, Bayesian optimization leverages the Bayesian methods to compute the posterior distribution of the objective function, assisting in the selection of hyperparameters for LSTM models. It iteratively refines the model based on past evaluations to optimize performance. If the results are unsatisfactory, the hyperparameters and architecture are reassessed; otherwise, successful outcomes guide future predictions using the optimized weights. In our proposed model, Bayesian optimization plays a crucial role in identifying key LSTM hyperparameter values (see Figure 2.4 and Algorithm 1).

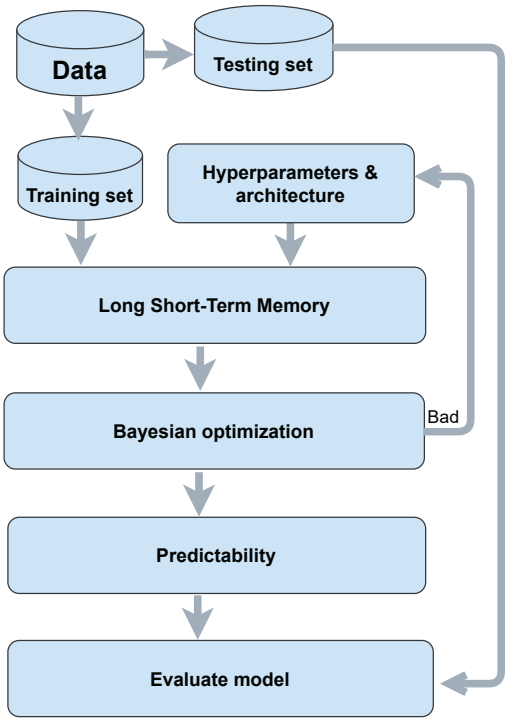


Figure 2.4: Optimizing hyperparameters using the Bayesian method.

Algorithm 1 Bayesian Optimization for LSTM Hyperparameters

Input: Set of flight trajectories $\mathscr{X}$, Total number of iterations N, Length of trajectory points M, Objective function f, Acquisition function A.
Output: Θ (Set of observed pairs of hyperparameters and objective function values).
Initialize $\Theta \leftarrow \emptyset$; $n \leftarrow 0$
Define $\mathscr{X} = \{X_i\}$
while $n \leq N$ **do**
    $x^* \leftarrow \text{argmax}_x A(x \mid \Theta)$
    $y \leftarrow f(x^*)$
    $\Theta \leftarrow \Theta \cup \{(x^*, y)\}$
    Train the LSTM model using the updated Θ
    Compute the weights W and biases b based on Equation (2.9) and Equation (2.10).
    $n \leftarrow n + 1$
end while
Return Θ

The selection function is the criteria by which the next set of hyperparameters are chosen from the surrogate function. The most common choice of criteria is Expected Improvement (EI) [62]:

$$EI(x) = \int_{-\infty}^{\infty} \max(y - y^*, 0) p(y \mid x)\, dy, \tag{2.11}$$

where:

- y^* is a threshold value of the objective function.
- x is the proposed set of hyperparameters.
- y is the actual value of the objective function using hyperparameters x.
- $p(y \mid x)$ is the surrogate probability model expressing the probability of y given x.

If $p(y \mid x)$ is zero for all $y < y^*$, it indicates that the hyperparameters x are unlikely to produce improvements. Conversely, if the integral is positive, it suggests that the hyperparameters x are expected to perform better than the threshold value.

2.5 Experimental results

2.5.1 Data preparation

Dataset

Automatic Dependent Surveillance-Broadcast (ADS-B) system is one of the most popular tools used to collect the flight data. It utilizes satellite navigation to determine and broadcast a flight position to air traffic control and other nearby aircraft. This capability enhances situational awareness by making flights visible to each other, thereby facilitating self-separation. Each trajectory includes essential information such as timestamp, flight ID, position (longitude, latitude, altitude), speed, heading, hour, etc.

In this thesis, our analysis uses the flight data collected from the Hartsfield–Jackson Atlanta International Airport (ATL) [107], with a sample row of 4D trajectory data provided in Table 2.2.

Table 2.1: 4D Trajectory data features

Feature	**Unit**	**Description**
Timestamp	Unix	Unix timestamp indicating the time when data was recorded.
Flight ID	Icao24	Unique identifier (Icao24) representing each individual flight.
Longitude	Degree	Geographic coordinate indicating the east-west position of the aircraft in degrees.
Latitude	Degree	Geographic coordinate indicating the north-south position of the aircraft in degrees.
Altitude	Feet	Altitude above sea level measured in feet.
Speed	Knot	Speed of the aircraft relative to the ground, measured in knots (nautical miles per hour).
Heading	Degree	Direction of aircraft travel relative to true north, expressed in degrees.
Hour	Unix	Unix timestamp indicating the specific hour when data was recorded.

It is noteworthy that records with the same identifier correspond to

the same aircraft, and a collection of these records together forms the trajectory X_i. In this research, we utilize the complete set of trajectory data sourced from geographical coordinates associated with ATL airport. Each trajectory is recorded at one-second intervals, providing detailed track points for analysis.

Table 2.2: An example of 4D Flight trajectory information.

Feature	Value
Timestamp	1537009474
Flight ID	e8044e
Longitude	-84.402633978396
Latitude	33.7542572021484
Altitude	12275.82
Speed	260.53122347432
Heading	12.1974800580645
Hour	1537009200

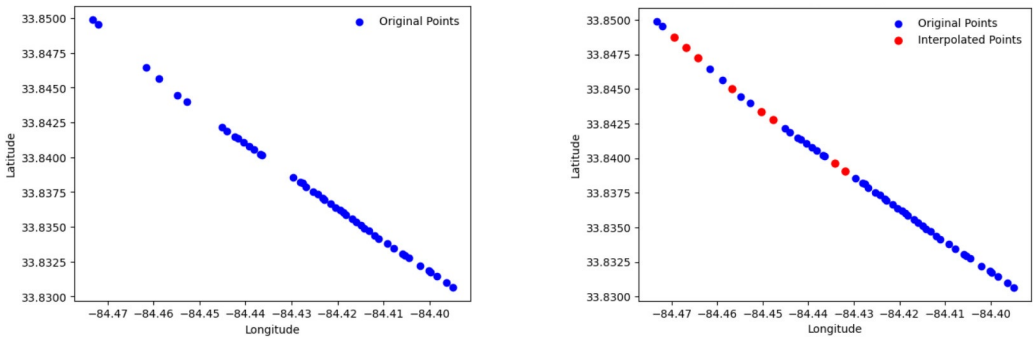
The dataset includes both static and dynamic information, with the focus on dynamic data such as heading, speed, and 4D data (timestamp, longitude, latitude, and altitude). Updates to this data occur every 5 seconds. To address the continuity and smoothness of flight trajectories [76], missing points are filled using the cubic spline interpolation, which offers a smooth and accurate method for filling in gaps in the dataset (see Figure 2.5). To manage the spatial range of the input data and enhance forecast accuracy and smoothness, we employ a sliding window approach with a window size of 100 seconds and a step size of 1 second. This method segments the data into 105-second intervals, where the first 100 seconds are used as historical data and the last 5 seconds serve as the prediction window.

Given a series of trajectory points $P_1, P_2, P_3, \ldots, P_n$, with a time window size of 100, the trajectory at time t can be expressed as

$$X_i^t = \{P_i^{t-99}, P_i^{t-98}, P_i^{t-97}, \ldots, P_i^t\}.$$

Each trajectory segment contains 100 consecutive trajectory points. The sliding window is moved forward continuously to perform the segmentation operation until the last trajectory point is reached. The segmented trajectory is denoted by

$$\mathscr{X} = \{X_i^t, X_i^{t+1}, X_i^{t+2}, \ldots, X_i^n\},$$



(a) Before interpolation (b) After interpolation

Figure 2.5: An example of addressing missing data before (left) and after (right) using cubic spline interpolation.

where $t = 100, 101, \ldots, n$. The experimental process is shown in Figure 2.6.

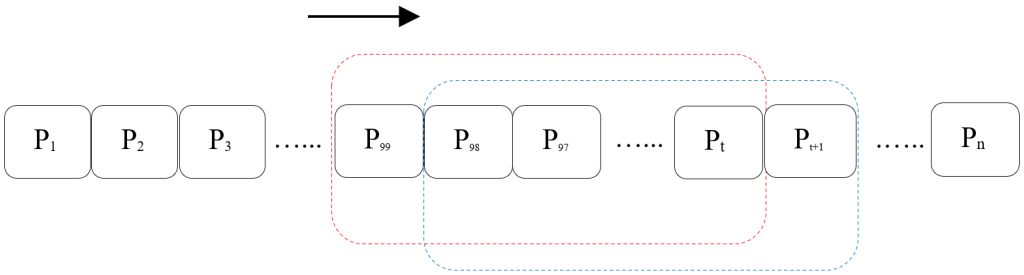


Figure 2.6: Sliding time window.

This results in 21258 trajectories, which are then divided into training and test sets, including 17006 and 4252 samples in training and testing sets, respectively.

- Training set: LSTM model is trained using Bayesian optimization to adjust parameters iteratively, minimizing prediction error, and leveraging probability distributions to represent uncertainty and enhance effectiveness, particularly with outliers.

- Testing set: Trained LSTM model is evaluated on a testing dataset to assess performance on unseen data. Predictions are compared with

ground truth values using metrics like RMSE and MAE to quantify accuracy, helping gauge the model's ability to generalize and providing insights into overall performance.

Normalization

In this thesis, we use Z-Score normalization, features are transformed to become standardized as follows:

$$X_{\text{new}} = \frac{X - \min(X)}{\max(X) - \min(X)}, \tag{2.12}$$

where:

- X_{new} is the normalized value, obtained by scaling the original value X to fit within a standardized range.
- $\min(X)$ and $\max(X)$ denote the minimum and the maximum values in the dataset X, respectively.

2.5.2 Hyperparameter tuning

Hyperparameters are crucial for shaping the architecture of a machine learning model. Optimal hyperparameters can markedly enhance the model's performance. In our proposed model, we outline the dimensional adjustments of both input and output data for every layer. Initially, the input data comprises a 3D tensor with dimensions (None, 100, 6), where "None" denotes the number of batch samples during model training. The detailed information is described in Table 2.3.

2.5.3 Evaluation metrics

To evaluate model performance, we employ two common metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

- RMSE is used to measure the magnitude of errors in the predictions. It is calculated as the square root of the average of the squared differences

Table 2.3: Basic parameters of proposed model.

Layer	**Input size**	**Output size**	**Parameter**
Convolutional	(None, 100, 6)	(None, 100, 32)	1x3 kernels, 32 filters
ReLU	(None, 100, 32)	(None, 100, 32)	-
Max Pooling	(None, 100, 32)	(None, 50, 32)	Window size: 2
Dropout	(None, 50, 32)	(None, 50, 32)	Dropout rate: 0.3
LSTM	(None, 50, 32)	(None, 50)	50 units
Fully Connected	(None, 50)	(None, 4)	4 nodes

between the predicted and observed values, providing a measure of the prediction error.

- MAE is used to evaluate the model's performance by capturing the average of the absolute differences between the predicted and observed values across all instances in the test set, providing a measure of prediction accuracy.

RMSE and MAE are respectively defined by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\widehat{P}_i^t - P_i^t\right)^2}, \tag{2.13}$$

and

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|\widehat{P}_i^t - P_i^t\right|, \tag{2.14}$$

in which n represents the number of predicted points, while $\widehat{P}_i^t$ and P_i^t correspond to the predicted and actual values at time step t of flight i.

2.5.4 Comparison of experimental results

To assess the effectiveness of our proposed BayesCoordLSTM method, we evaluate its performance on the ATL dataset [107], as detailed in Section 2.5.1. This evaluation comprises two primary comparative analyses:

(1)) Comparison with state-of-the-art methods. We compare the BayesCoordLSTM with three state-of-the-art models: 3D-CNN [95], CNN-GRU [106], and CG3D [107]. This comparative analysis highlights the

advancements and performance improvements of the BayesCoordLSTM relative to these established alternatives.

(2) Impact of the Bayesian method and coordinate transformation. We assess, through controlled experiments, three configurations of the CNN-LSTM model: the baseline CNN-LSTM, CNN-LSTM with coordinate transformation, and CNN-LSTM with both the Bayesian method and coordinate transformation. This analysis provides how each enhancement improves the accuracy and reliability of flight trajectory predictions.

Comparison with state-of-the-art methods

We first compare the BayesCoordLSTM with the state-of-the-art models: 3D-CNN [95], CNN-GRU [106], and CG3D [107], using a system configured with an Intel i7-8700 CPU, GTX 1080 Ti GPU, and 31 GB of RAM. Our experimental results focus on the performance and training time of each model illustrated in Table 2.4.

Table 2.4: Comparison with state-of-the-art methods.

Predicting model	RMSE	MAE	Training time (s)
CNN-GRU [106]	0.3728	0.2164	11503
3D CNN [95]	0.2646	0.1785	10060
CG3D [107]	0.2626	0.1776	**9042**
Our proposed method	**0.2154**	**0.1624**	9450

Among the evaluated models, the CNN-GRU model demonstrated the highest RMSE value of 0.3728, signifying a larger discrepancy between predicted and actual values. Similarly, its MAE of 0.2164 indicates relatively higher prediction errors. The 3D CNN model showed improvement with an RMSE of 0.2646 and MAE of 0.1785, indicating a more accurate predictions than the CNN-GRU. Notably, the CG3D model achieved an even lower RMSE of 0.2626 and MAE of 0.1776, positioning it as one of the top-performing models in terms of prediction accuracy. In our proposed model, it can be attributed to two key factors: using coordinate transformation and the Bayesian method. By implementing coordinate transformation, which facilitates consistent data normalization, we effectively reduced the impact

of outliers. Furthermore, the incorporation of the Bayesian method helped prevent overfitting by using probability distribution during training, thereby improving the model's generalization ability. This combination of techniques enabled our proposed model to achieve more accurate predictions. As seen in Table 2.4, our method achieved an RMSE of 0.2154 and MAE of 0.1624, demonstrating its superior predictive accuracy compared to the other models.

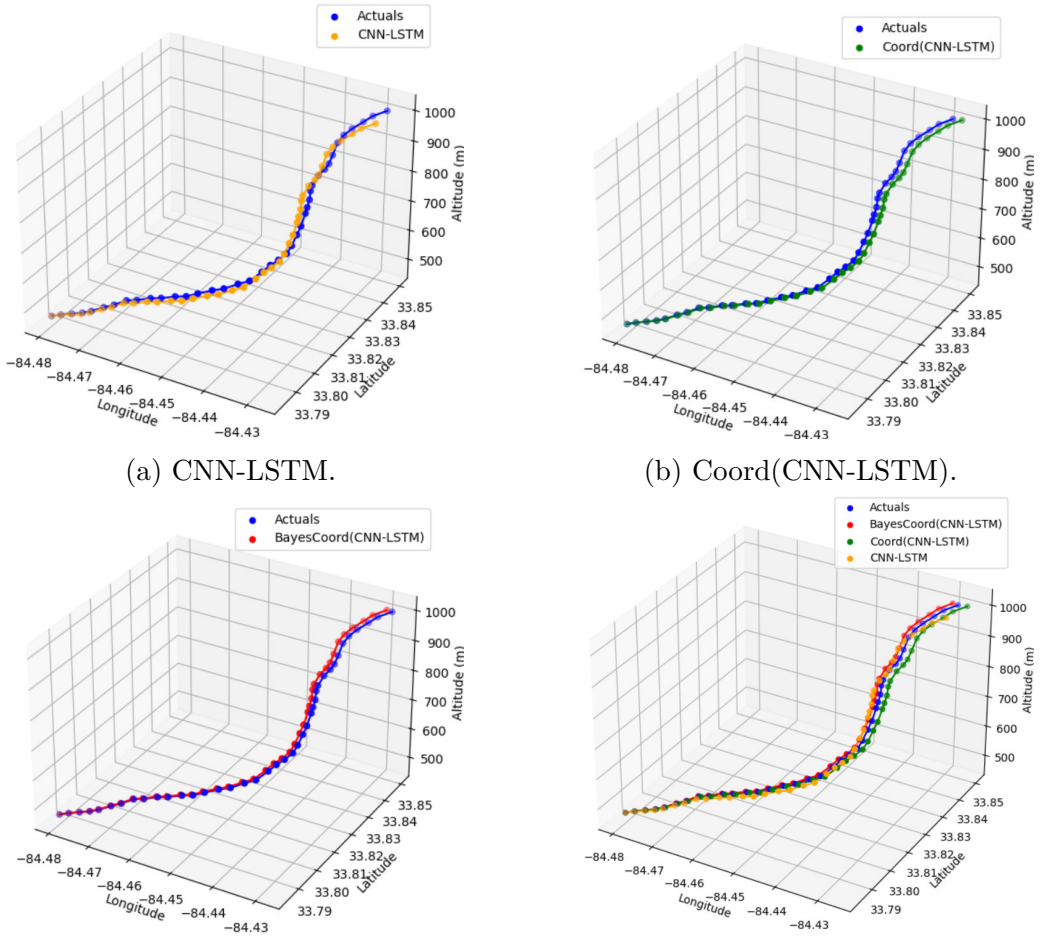
Although the training time of the BayesCoordLSTM, at 9450 seconds, is slightly longer than that of CG3D, at 9042 seconds, this marginal increase is outweighed by the significant improvements in both RMSE and MAE. Therefore, the BayesCoordLSTM not only achieves superior predictive performance but also maintains competitive computational efficiency, underscoring its overall effectiveness in the comparative analysis.

Impact of the Bayesian method and coordinate transformation

We then explore the individual and combined impact of integrating the Bayesian method and coordinate transformation within the BayesCoordLSTM framework. Through a series of controlled experiments, we evaluate three configurations of the CNN-LSTM model: the baseline CNN-LSTM, CNN-LSTM with coordinate transformation, and CNN-LSTM with both the Bayesian method and coordinate transformation. This analysis provides insights into how each enhancement contributes to improving the accuracy and reliability of flight trajectory predictions.

The prediction comparisons for a single aircraft, as shown Figure 2.7a, Figure 2.7b, and Figure 2.7c, further illustrate these improvements. The results demonstrate that the BayesCoord(CNN-LSTM) model consistently produces the most accurate trajectory predictions, closely aligning with the actual trajectory (see Figure 2.7d).

In Figure 2.7, the predicted and actual trajectories in 3D space are depicted. It can be observed that the predicted trajectories of the three models exhibit similar trends to the actual trajectories, but notably, the CNN-LSTM model's predicted curve deviates significantly from the other two models. The use of coordinate transformation in the Coord(CNN-LSTM) and the proposed models helps reduce errors in predicting latitude, longitude, and especially altitude. As shown in Figure 2.7d, the predicted trajectory by the proposed model closely matches the actual trajectory with the small-



(a) CNN-LSTM.

(b) Coord(CNN-LSTM).

(c) Our proposed model.

(d) Three configurations.

Figure 2.7: A comparison of three configurations of the CNN-LSTM model.

est prediction error, followed by the Coord(CNN-LSTM) and CNN-LSTM models.

Based on these results, the evaluation and comparison of the CNN-LSTM, Coord(CNN-LSTM), and BayesCoord(CNN-LSTM) models in predicting trajectories in 3D space reveal significant differences in prediction accuracy. The Coord(CNN-LSTM) model shows improvement in latitude and longitude prediction compared to the baseline CNN-LSTM, highlighting the effectiveness of coordinate transformation. Moreover, the BayesCoord(CNN-LSTM) model achieves the best performance, producing trajectories closest to reality with the smallest prediction errors among the evaluated models.

2.6 Conclusions

CNNs and LSTMs provide state-of-the-art performance in various tasks. However, these models often face issues with overfitting on small datasets and lack the capability to measure uncertainty, which negatively affects their generalization abilities. Additionally, prediction tasks, especially those involving time-series datasets, can encounter challenges due to complex long-term fluctuations. Recently, the application of the Bayesian method in deep learning has been introduced to estimate uncertainty in model predictions. This approach is highly robust to overfitting and allows for uncertainty estimation. In air traffic management, prediction is crucial for preventing traffic jams and collisions. It typically involves predicting aspects such as delays, flow, and trajectories. Therefore, improving performance in flight trajectory prediction is of significant interest to many researchers.

In this chapter, we introduced a novel approach to enhance flight trajectory prediction, named BayesCoordLSTM. This model efficiently integrates Bayesian optimization and coordinate transformation into a hybrid CNN-LSTM framework, achieving accurate flight path forecasts with improved precision and efficiency. Our experiments, conducted using a real flight trajectory dataset, clearly demonstrate the remarkable superiority of the BayesCoordLSTM model over existing methods. Across various flight scenarios and conditions, this model consistently outperforms baseline models in terms of trajectory prediction accuracy. The utilization of Bayesian optimization not only streamlines hyperparameter tuning but also significantly

enhances predictive performance. While our study provides a solid foundation for predicting flight trajectories, future research could explore dynamic hyperparameter adaptation techniques. These techniques could involve real-time adjustments to hyperparameters in response to evolving flight scenarios or changing weather conditions, further enhancing the BayesCoordLSTM model's adaptability and real-time predictive capabilities.

Chapter 3

Reconstructing angular probability density by using Maximum Entropy

An important issue in air traffic management is the construction of traffic complexity maps. Observed or simulated data provide a statistical sample of trajectory angles. The underlying angular density is a crucial component in assessing local complexity. Our approach involves applying the Maximum Entropy principle, constrained by the statistical estimation of Fourier coefficients for the desired density. Accounting for the statistical properties of moment estimation is achieved by using the square of the Mahalanobis distance between statistical data and Fourier coefficients of the sought density as the data fitting term. Partially finite convex programming is then implemented to compute the optimal density. The cornerstone of this approach, enabling a rigorous treatment of variational aspects of the chosen method, is an infinite-dimensional version of Fenchel's duality theorem, coupled with results on the conjugacy of integral functionals. We also discuss the numerical aspects of our approach and present some simulations and reconstructions from simulated data.

3.1 Introduction

In contemporary data analysis, the reconstruction of probability density functions stands as a pivotal challenge across various domains, ranging from physics to social sciences [25, 90]. The quest for robust methodologies capable of faithfully capturing underlying distributions from sparse data remains

relevant, particularly in complex systems such as those found in ATM [41, 47]. This chapter addresses this challenge by proposing a framework for reconstructing angular probability density functions through the lens of the Maximum Entropy principle.

The driving force behind our endeavor stems from the pressing need to estimate the complexity of air traffic from flight trajectory data [63, 104]. With the exponential growth in air travel, understanding and managing the intricate dynamics of air traffic flows has become paramount [36]. By harnessing the power of Maximum Entropy reconstruction, we aim to unveil underlying patterns and structures within flight trajectory data, thereby shedding light on the complexity of air traffic systems.

Our approach hinges on several key pillars, each meticulously designed to navigate a specific aspect related to angular density reconstruction. First and foremost, we formulate the reconstruction problem within the framework of the Maximum Entropy principle. By leveraging this principle, we seek solutions that strike a balance between fidelity to observed constraints and maximization of entropy, thus ensuring maximal neutrality while remaining consistent with available data.

A distinguishing feature of our methodology lies in the formulation of constraints in terms of Fourier coefficients derived from statistical samples. This not only allows us to encapsulate essential characteristics of the angular distribution but also enhance the integration of empirical data into the reconstruction process. Moreover, to account for the inherent variability within the data, we introduce a fit term expressed via the squared Mahalanobis distance, thereby, weighing the evidence of each Fourier coefficient according to its uncertainty expressed through the covariance term.

Critically, we provide a rigorous treatment of the underlying optimization problem, associated with the infinite-dimensional space. Drawing upon concepts from partially finite convex programming and leveraging insights into the conjugate of convex integral functionals, we devise a robust methodology for tackling this challenge. Importantly, our approach avoids the need for discretization, thereby offering an efficient numerical solution for computing Maximum Entropy solutions.

3.2 Literature review

The Maximum Entropy principle, deeply embedded in statistical mechanics and information theory, originates from Ludwig Boltzmann's pioneering efforts in the late 19^{th} century. This foundation work led to the entropy concept, which Jaynes (1957) pivotalized in the context of probabilistic reasoning [60]. It is considered as a powerful tool for reconstructing probability distributions from limited information. It seeks to maximize the entropy of a probability distribution while satisfying given constraints derived from observed data. Jaynes further elaborated on this concept, advocating for the adoption of the probability distribution that maximizes entropy, consistent with the available information [61]. Over the decades, its utility has expanded in various domains from traditional physics to applications in biology, economics, and machine learning [25, 61].

The principle's utility extends beyond its foundational domains, finding significant application in fields such as image processing, pattern recognition, and machine learning. Advancements in computational techniques over time have increased its practical applicability, enabling its integration into diverse scientific disciplines [26]. Today, it stands as a cornerstone of probabilistic reasoning, providing a powerful framework for inference under uncertainty [16]. Therefore, the Maximum Entropy principle continues to inspire new methodologies and applications, ensuring its enduring relevance in modern science.

In computational biology, Boomsma et al. (2014) effectively used Maximum Entropy to bridge computational models with experimental data, enhancing the reliability of biological simulations [17]. His research highlighted that its ability can bridge gaps between computational models and experimental data, indicating its broad applicability in sequence analysis, structural modeling, and molecular simulations. Meanwhile, Wan et al. (2019) applied the principle to refine data interpretation methods through Principal Component Analysis, noting that the method's increased computational demands could limit its widespread application [120].

Furthermore, considering the inherent complexities and frequent errors in real data, numerous researchers have demonstrated that their algorithms can accurately reconstruct probability densities. A significant strength of the Maximum Entropy principle is its capability to effectively manage data

uncertainty. Gomes et al. (2014) developed a methodology to estimate probability densities even in environments with noisy data, proving robust in simulations but needing further validation in real scenarios [51]. This capability is particularly relevant in fields where data integrity may be compromised.

Lately, with the introduction of new methods, advancements in entropy estimation techniques have played a vital role in enhancing the effectiveness of the Maximum Entropy principle. In 2011, Behmardi et al. presented a parametric entropy estimator that applied the Maximum Entropy principle for approximating distributions without relying on local density estimation. The estimator showed significant accuracy improvements over traditional methods, especially in complex sensor network data scenarios, providing a robust tool for various applications in signal processing and machine learning [13]. Although their parametric entropy estimator can avoid the pitfalls of local density estimation, providing a more reliable tool for data analysis in complex environments like sensor networks, it might not be as flexible in handling diverse or non-standard data distributions without adjustments or extensions to the model. This is where Lakshmanan and Pichler's recent study in 2023 introduced a novel approach to soft quantization using entropic regularization, providing a promising alternative to traditional hard quantization methods that often lead to significant information loss. Their method, grounded in robust theoretical formulations, demonstrates improved robustness and signal fidelity across various data sets, especially in the presence of noise and data variability. However, the practical implementation details and computational efficiency of this approach are not extensively covered, which could be crucial for real applications where computational resources and processing time are limited [66].

Recent developments in this domain have shown varied applications and improvements to the principle. From a theoretical perspective, one notable application of the Maximum Entropy principle is in the reconstruction of angular probability density functions (PDFs) from observed or simulated data. In 2017, Gresele and Marsili explored both the philosophical and practical aspects of Maximum Entropy principle, highlighting its applicability across various scientific domains. However, the study only focused on theoretical aspects only, lacking examples of direct empirical application [52].

A few years later, Amari et al. (2018) and Gzyl et al. (2021) have provided deeper insights into the mathematical underpinnings of the Maximum

Entropy principle, connecting it with significant statistical concepts like the Wasserstein distance and Kullback-Leibler divergence [7, 53]. Due to lacking of experimental results on simulation, these studies offer a comprehensive understanding of the principle's integration with broader probability theory, though their complexity may pose challenges for those without a strong mathematical background. Wan et al. in 2019 [120] utilized Maximum Entropy to enhance the accuracy of Principal Component Analysis, providing a clear advantage in data interpretation, although the method's increased computational demand may limit its applicability for large or real-time datasets.

These studies collectively underscore the importance and potential of Maximum Entropy in a range of applications, from enhancing computational algorithms to addressing real data challenges. The principle's adaptability and robustness, demonstrated across different scenarios and data complexities, support its continued use and development for complex systems analysis, including air traffic complexity estimation.

In air traffic management, the precise reconstruction of angular distributions is crucial for assessing traffic complexity and managing congestion [75]. The Maximum Entropy principle is particularly well-suited for this task, offering a robust method for dealing with the inherent uncertainties and variabilities of air traffic data. This principle allows us to derive the most generalized form of a probability distribution by maximizing entropy, subject to the empirical constraints provided by observed data, such as angular means or variances.

Prior research has explored various approaches to angular density reconstruction, with a particular focus on incorporating the principles of Maximum Entropy to enhance the fidelity of the reconstructed PDFs.

The motivation for employing the Maximum Entropy principle in angular density reconstruction lies in its ability to provide a principled and data-driven approach to probability distribution estimation. Unlike parametric methods that rely on specific functional forms, Maximum Entropy reconstruction allows for greater flexibility and adaptability to the underlying data distribution. This flexibility is particularly advantageous in the context of air traffic complexity estimation, where traffic patterns can exhibit significant variability and non-linearity.

In simulation settings, the Maximum Entropy principle is highly effec-

tive for reconstructing angular distributions, especially in air traffic management. This method involves setting constraints based on available data, such as average angles and variances, to develop the most unbiased probability distributions possible. By maximizing entropy within these constraints, the principle ensures that the resulting angular distributions in simulations accurately capture the diverse and unpredictable nature of air traffic patterns. This approach not only improves the realism of the simulations but also provides deeper insights into the complexities of air traffic dynamics.

3.3 Setting

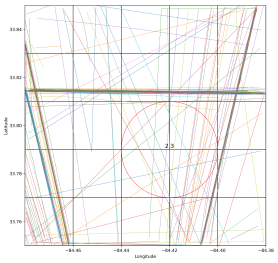
In this chapter, we address the complexity of air traffic by modeling airspace as a two-dimensional (2D) image, segmented into a Cartesian grid where each intersection point denotes the center of a defined zone, referred to as a *cell*. For the sake of simplicity, we employ circular cells. For each cell we apply the followings steps:

(1) Select the trajectories that pass through the cell and determine, for each one, the angles corresponding to entry and exit points, with respect to a given fixed direction. From such a *statistical sample* estimate, via empirical means, the Fourier coefficients of the angular probability distributions on the circle.

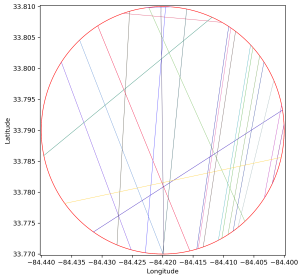
(2) Solve a maximum entropy problem to obtain probability distributions that are as neutral as possible. From the mathematical viewpoint, this step uses a *Fenchel duality* approach, that enables to solve some infinite dimensional entropy problem via the resolution of a finite dimensional concave smooth maximization problem.

(3) Use the optimal probabilities to estimate the *local complexity*.

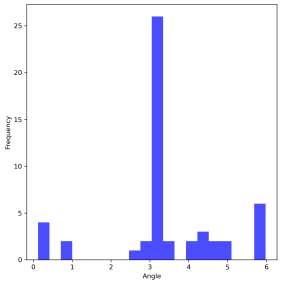
Figure 3.1 gives an example to visualize a map of full trajectories and the angular histogram of a selected cell. In Figure 3.1 (a), an airspace map shows an array of intersecting flight trajectories within a vast grid, highlighting the dense navigation challenges. A specific circular cell with numerous trajectories is illustrated in Figure 3.1 (b). At this cell, the angular histogram



(a) A map of full trajectories.



(b) All trajectories crossing a selected cell.



(c) Angular histogram of the selected cell.

Figure 3.1: An example map of trajectories through a selected cell and the angular histogram of the selected cell.

of these paths is showed the diversity of movement through the concentrated airspace as shown in Figure 3.1 (c).

Here we focus solely on the reconstruction of probability densities (the first two steps). The actual application to the construction of complexity maps is deferred to a subsequent chapter, currently being written. We now specify the notations used and clarify our entropy model.

3.3.1 Computing moments

The first task is that of detecting trajectories that intersect each circular cell S. We assume that all cells have the same radius δ. A cell can then be written as

$$S = \{(x, y) \mid (x - u_1)^2 + (y - u_2)^2 \leq \delta^2\}, \tag{3.1}$$

in which $\mathbf{u} = (u_1, u_2)$ is the coordinate-vector of its center.

Further, we denote by n the number of trajectories $\mathscr{T}_j$ that intersect the cell, and the set of indices of these trajectories is J:

$$n = \operatorname{card} J \quad \text{where} \quad J = \{j \in \mathbb{N} \mid \mathscr{T}_j \cap S \neq \emptyset\}. \tag{3.2}$$

Each trajectory that intersects the cell enters with an angle (with respect to a fixed direction) which we call *heading* and denote by θ_j. The complex

number $e^{i\theta_j}$ is the corresponding point on the unit circle in $\mathbb{C}$.

We now regard the θ_j's as realizations of a random angle variable θ. We are then interested in estimating the probability density $p(\theta)$ of such a variable. From the angular sampling θ_j, we may build a set of empirical moments. The Fourier coefficients of p are defined as

$$a_l = \frac{1}{\pi} \int_0^{2\pi} p(\theta) \cos(l\theta) \, \mathrm{d}\theta, \quad l \in \mathbb{N}, \tag{3.3}$$

and

$$b_l = \frac{1}{\pi} \int_0^{2\pi} p(\theta) \sin(l\theta) \, \mathrm{d}\theta, \quad l \in \mathbb{N}^*. \tag{3.4}$$

The empirical coefficients

$$x_l = \frac{1}{\pi n} \sum_{j \in J} \cos(l\theta_j) \quad \text{and} \quad y_l = \frac{1}{\pi n} \sum_{j \in J} \sin(l\theta_j) \tag{3.5}$$

are regarded as statistical estimators of a_l and b_l, respectively. Note in passing that the estimator x_0 gives the exact value $1/\pi$ of a_0. Clearly, in practice, these estimators will be meaningful for values of l limited to some range $\{1, \ldots, N\}$. We are then led to consider the data vector

$$\mathbf{z} = (x_1, y_1, \ldots, x_N, y_N) \in \mathbb{R}^{2N}.$$

The parameter N corresponds to the highest frequencies that are meaningful and useful, depending on the number of trajectories intersecting the cell S.

3.3.2 Formulating the Maximum Entropy problem

We discuss the selection of a probability for angle. If we could dispose of exact Fourier coefficients, we would then maximize the entropy of p under the constraint that the moments of p match these exact values. We would

then aim at solving the following constrained optimization problem:

$$(\mathscr{P}_\circ)\quad \left|\begin{array}{ll} \text{Min} & H(p) \\ \text{s.t.} & p \in L^1([0,2\pi)), \\ & 1 = \displaystyle\int_0^{2\pi} p(\theta)\,\mathrm{d}\theta, \\ & x_l = \dfrac{1}{\pi}\displaystyle\int_0^{2\pi} p(\theta)\cos(l\theta)\,\mathrm{d}\theta,\ l \in \{1,\dots,N\}, \\ & y_l = \dfrac{1}{\pi}\displaystyle\int_0^{2\pi} p(\theta)\sin(l\theta)\,\mathrm{d}\theta,\ l \in \{1,\dots,N\}. \end{array}\right.$$

Here, $H(p)$ denotes the Shannon neg-entropy of p:

$$H(p) = \int_0^{2\pi} p(\theta)\ln p(\theta)\,\mathrm{d}\theta.$$

However, we only dispose of approximate values of the Fourier coefficients, except for $l = 0$ where the value is exact. It is customary to relax the inexact constraints, and we therefore aim at solving the following optimization problem:

$$(\mathscr{P})\quad \left|\begin{array}{ll} \text{Minimize} & H(p) + \dfrac{\alpha}{2}\|\mathbf{z} - \mathbb{A}p\|^2_{\Sigma^{-1}} \\ \text{s.t.} & 1 = \displaystyle\int_0^{2\pi} p(\theta)\,\mathrm{d}\theta, \end{array}\right.$$

in which α balances the intensity of relaxation. Here, we use the following notation:

- $\|\cdot\|_{\Sigma^{-1}}$ denotes the function given by

  $$\|\mathbf{z}'\|_{\Sigma^{-1}} = \sqrt{\langle \mathbf{z}', \Sigma^{-1}\mathbf{z}'\rangle},$$

  in which Σ denotes the covariance matrix of random vector $\mathbf{z}$. In this optimization problem, x_l and y_l are considered as random variables because they are computed from independent samples of the random variable θ. As θ is random, so are x_l and y_l, which are the averages of the cosine and sine functions evaluated at these angles, respectively.

- $\mathbb{A}\colon L^1([0,2\pi)) \to \mathbb{R}^{2N}$ is the linear mapping defined by

  $$\mathbb{A}p = \int_0^{2\pi} p(\theta)\mathbf{T}(\theta)\,\mathrm{d}\theta$$

  in which $\mathbf{T}(\theta) := (\cos\theta, \sin\theta, \dots, \cos N\theta, \sin N\theta)$.

Therefore, in Problem ($\mathscr{P}$), the squared Mahalanobis distance between $\mathbb{A}p$ and $\mathbf{z}$ is penalized, as a *model fitting* requirement. Note that Σ is not known in practice. It will be necessary to estimate it. Using the Mahalanobis distance requires, in principle, to dispose of a positive definite estimate of Σ. However, when the estimated covariance matrix is singular, the inverse does not exist, and then we have to resort to a degenerate version of the Mahalanobis distance, given by

$$\|\mathbf{z}'\|^2_{\Sigma^\dagger} = \begin{cases} \left\langle \mathbf{z}', \Sigma^\dagger \mathbf{z}' \right\rangle & \text{if } \mathbf{z}' \in \operatorname{ran} \Sigma, \\ \infty & \text{otherwise,} \end{cases} \tag{3.6}$$

in which $\Sigma^\dagger$ denotes the pseudo-inverse of Σ and $\operatorname{ran} \Sigma$ denotes the range of Σ. The infinite value of the corresponding penalization is, of course, equivalent to a sharp constraint in problem ($\mathscr{P}$).

Problem ($\mathscr{P}$) pertains to *partially finite convex programming* [19, 20]. In Section 3.4 below, we give a detailed account of the dual approach to solving such problems. A remarkable aspect is that Fenchel duality will enable us to solve the *infinite dimensional constrained problem* ($\mathscr{P}$) via the corresponding *finite dimensional dual problem*, which turns out to be convex, smooth and unconstrained. It should be noticed that the dual treatment of problem ($\mathscr{P}$) will avoid having to compute the inverse (or pseudo-inverse) of the covariance matrix. Problem ($\mathscr{P}$) will be solved via the so-called dual problem, which turns out to be much better behaved, as will appear in the sequel.

3.4 Duality in action

Our aim in this subsection is to show how Fenchel duality can help us solve our optimization problems via the tractable dual problem. We first review a few basic definitions and results. The reader unfamiliar with convex analysis may refer to [100, 134].

3.4.1 Review of useful convex analysis

Let L be any real vector space. A function $f\colon L \to [-\infty, \infty]$ is said to be *convex* if its *epigraph*, the set

$$\operatorname{epi} f := \left\{ (x, \alpha) \in L \times \mathbb{R} \;\middle|\; f(x) \leq \alpha \right\},$$

is a convex subset of $L \times \mathbb{R}$. It is said to be *proper convex* if it never takes the value $-\infty$ and it is not identically equal to ∞. A function $g\colon L \to [-\infty, \infty]$ is said to be *concave* if $-g$ is convex, and *proper concave* if $-g$ is proper convex. Notice that g is concave if and only if its *hypograph*

$$\operatorname{hypo} g := \left\{ (\mathbf{x}, \alpha) \in L \times \mathbb{R} \;\middle|\; g(\mathbf{x}) \geq \alpha \right\}$$

is convex. The *effective domain* of a convex function f is the set

$$\operatorname{dom} f = \left\{ x \in L \;\middle|\; f(x) < \infty \right\}.$$

The *effective domain* of a concave function g is the set

$$\operatorname{dom} g = \left\{ x \in L \;\middle|\; g(x) > -\infty \right\}.$$

The only functions that are both convex and concave are the affine functions and the functions identically equal to $\pm\infty$. Clearly, the domain of each affine function is equal to L, as a convex or as a concave function. The other remaining functions would give rise to specific notation if they were not so rare.

It is customary to use *indicator functions* to encode constraints in optimization problems. Recall that the indicator function of a subset $C \subset L$ is the function

$$\delta_C(x) := \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{otherwise.} \end{cases}$$

Now, let L and Λ be vector spaces paired (in the algebraic sense) by a bilinear mapping

$$\begin{array}{rccc} \langle \cdot, \cdot \rangle : & L \times \Lambda & \longrightarrow & \mathbb{R} \\ & (x, \xi) & \longmapsto & \langle x, \xi \rangle . \end{array}$$

An standard example is $L = \mathbb{R}^d = \Lambda$ with the usual Euclidean scalar product. The *convex conjugate* of a function f (that is convex or not) is the function

$$f^\star(\xi) := \sup \left\{ \langle x, \xi \rangle - f(x) \;\middle|\; x \in X \right\}, \quad \xi \in \Lambda.$$

The *concave conjugate* of a function f (that is concave or not) is the function

$$f_\star(\xi) := \inf \Big\{ \langle x, \xi \rangle - f(x) \;\Big|\; x \in X \Big\}, \quad \xi \in \Lambda.$$

It is remarkable that convex conjugacy acts as an involution on certain classes of functions. For example, if $f \colon \mathbb{R}^d \to [-\infty, \infty]$ is a lower-semicontinuous proper convex function, then

$$f^{\star\star} := (f^\star)^\star = f.$$

Given a convex subset $C \in \mathbb{R}^d$, we call *relative interior* of C the interior of C with respect to its *affine hull* $\operatorname{aff} C$. Recall that $\operatorname{aff} C$ is the smallest affine subspace that contains C. The relative interior of C is denoted by $\operatorname{ri} C$. For example, if C is a closed segment in $\mathbb{R}^2$, its interior is empty while its relative interior is the segment without its ends. It can be shown that the relative interior of a nonempty convex set is always nonempty.

Theorem 1 (Fenchel)**.** *Let f and g be functions on $\mathbb{R}^d$ respectively proper convex and proper concave such that*

$$\operatorname{ri} \operatorname{dom} f \cap \operatorname{ri} \operatorname{dom} g \neq \emptyset. \tag{3.7}$$

Then

$$\eta := \inf_{\mathbf{x} \in \mathbb{R}^d} \Big\{ f(\mathbf{x}) - g(\mathbf{x}) \Big\} = \sup_{\boldsymbol{\xi} \in \mathbb{R}^d} \Big\{ g_\star(\boldsymbol{\xi}) - f^\star(\boldsymbol{\xi}) \Big\}$$

and the supremum is attained.

The above theorem asserts equality between the optimal values of two problems, together with attainment in the second one. It is customary to call these underlying optimization problems the *primal problem* and the *dual problem*, respectively. In the above theorem, both the primal and dual are finite dimensional. Moreover, in order to tackle problems such as $(\mathscr{P})$, we must put some linear mapping in the picture. This is the purpose of the next theorem.

Theorem 2. *Let be given:*

1. *L and Λ, real vector spaces;*
2. *$\langle \cdot, \cdot \rangle$, a bilinear form on $L \times \Lambda$;*

3. $\mathbb{A}\colon L \to \mathbb{R}^d$, *a linear mapping;*

4. $F\colon L \to (-\infty, \infty]$, *a proper convex function;*

5. $g\colon \mathbb{R}^d \to [-\infty, \infty)$, *a proper concave function.*

Assume that $\mathbb{A}$ *admits a formal adjoint mapping* $\mathbb{A}^\star$*, that is, a linear mapping* $\mathbb{A}^\star\colon \mathbb{R}^d \to \Lambda$ *such that* $\langle \mathbb{A}x, \mathbf{y}\rangle = \langle x, \mathbb{A}^\star \mathbf{y}\rangle$ *for every* $x \in L$ *and every* $\mathbf{y} \in \mathbb{R}^d$*. Then, under the qualification condition*

$$(QC) \quad \operatorname{ri}(\mathbb{A}\operatorname{dom} F) \cap \operatorname{ri}(\operatorname{dom} g) \neq \emptyset,$$

one has

$$\eta := \inf_{x\in L}\Big\{F(x) - g(\mathbb{A}x)\Big\} = \max_{\boldsymbol{\lambda}\in\mathbb{R}^d}\Big\{g_\star(\boldsymbol{\lambda}) - F^\star(\mathbb{A}^\star\boldsymbol{\lambda})\Big\}.$$

This theorem is the corner stone of what is referred to as *partially finite convex programming*. Various forms appeared in the literature (see in particular [19, 20]). The selected form is as in [79], where no topological structure on the infinite dimensional side is requested. The optimization problems

$$\text{Minimize } \; F - g\circ\mathbb{A} \quad \text{and} \quad \text{Maximize } \; g_\star - F^\star\circ\mathbb{A}^\star$$

are respectively referred to as the *primal* and *dual* problems. The function $D := g_\star - F^\star \circ \mathbb{A}^\star$ appearing in the dual problem is referred to as the *dual function*. The theorem asserts the equality between the optimal values of the primal and dual problems, together with *dual attainment*. The next result will provide conditions that will guarantee *primal attainment* as well.

Theorem 3 (Primal attainment)**.** *With the notation and assumptions of the previous theorem, assume in addition that*

$$(QC^\star) \quad \operatorname{ri}\operatorname{dom} g_\star \cap \operatorname{ri}\operatorname{dom}(F^\star\circ\mathbb{A}^\star) \neq \emptyset.$$

Suppose further that

(a) $F^{\star\star} = F$ *and* $g_{\star\star} = g$*;*
(b) *there exists* $\bar{\boldsymbol{\lambda}}$ *dual optimal and* $\bar{x} \in \partial F^\star(\mathbb{A}^\star\bar{\boldsymbol{\lambda}})$ *such that* $F^\star\circ\mathbb{A}^\star$ *has gradient* $\mathbb{A}\bar{x}$ *at* $\bar{\boldsymbol{\lambda}}$*.*

Then $\bar{x}$ *is primal optimal.*

The latter result provides not only a condition for primal attainment, but it also makes appear as a watermark the possibility of a link between primal and dual solutions. The bi-conjugate relationships in Assumption (a) are central in the theorem, and the difficulty in our problem is to prove that the entropy satisfies this property. It turns out that, in our context, it is possible to compute the conjugate of the Shannon entropy by *conjugating through the integral sign*.

Generally speaking, an integral functional is a functional of the form

$$H(u) = \int_\Omega h(u(\omega), \omega)\, \mathrm{d}\mu(\omega), \quad u \in L.$$

Here, Ω is assumed to be endowed with a σ-algebra of measurable sets and with a measure denoted by μ; the function h is called the integrand, and the argument u is assumed to pertain to some space of measurable functions L. In our context, it is enough to consider such a functional on the familiar space $L = L^1([0, 2\pi))$, implicitly endowed with the Borel σ-algebra and the Lebesgue measure. Moreover, dependence of h in its second argument is not vital here, and we are only interested, in this chapter, in the case where $h(u(\omega), \omega) = h_\circ(u(\omega))$, with

$$h_\circ(t) = \begin{cases} t \ln t & \text{if } t > 0, \\ 0 & \text{if } t = 0, \\ \infty & t < 0. \end{cases}$$

Clearly, $h_\circ$ is a lower semi-continuous convex proper function and, as such, satisfies $h_\circ^{\star\star} = h_\circ$. Conjugating H is elegantly performed by conjugating the integrand, as we shall see now. Following Rockafellar, we say that:

> *A space L of measurable functions is decomposable if it is stable under bounded alteration on sets of finite measure.*

Otherwise expressed, L is decomposable if and only if it contains all functions of the form

$$\mathbb{1}_T u_\circ + \mathbb{1}_{T^c} u,$$

in which T has finite measure, $u_\circ$ is a measurable function such that the set $u_\circ(T)$ is bounded, and u is any member of L. Here, $\mathbb{1}_T$ denotes the characteristic function of T: $\mathbb{1}_T(x)$ equals 1 if $x \in T$ and $\mathbb{1}_T(x)$ equals zero otherwise; and T^c denotes the complement of T. It is easy to see that L^p-spaces are decomposable, which includes our workspace $L^1([0, 2\pi))$.

Theorem 4 (Rockafellar)**.** Let L and Λ be spaces of measurable functions on Ω paired by means of the standard integral bilinear form

$$\langle f, \varphi \rangle = \int_\Omega f(\omega)\varphi(\omega)\,\mathrm{d}\omega.$$

Let H be the functional of integrand $h_\circ$, that is,

$$H(u) = \int_\Omega h_\circ(u(\omega))\,\mathrm{d}\omega,$$

with $h_\circ$ proper convex and lower semi-continuous. Assume that L is decomposable and that H has nonempty effective domain. Then

$$H^\star(\varphi) = \int h^\star(\varphi(\omega))\,\mathrm{d}\omega$$

for every $\varphi \in \Lambda$, and $H^\star$ is convex on Λ.

Applying the latter theorem with $L = L^1([0, 2\pi))$, $\Lambda = L^\infty([0, 2\pi))$ and $h_\circ$ the above defined integrand of the Shannon entropy we see that, in our case,

$$H^\star(\varphi) = \int_0^{2\pi} h_\circ^\star(\varphi(\omega))\,\mathrm{d}\omega.$$

As can be seen from the appendix, the function $h_\circ^\star$ is given by

$$h_\circ^\star(\tau) = \exp(\tau - 1), \quad \tau \in \mathbb{R}.$$

Finally, since $\Lambda = L^\infty([0, 2\pi))$ is also decomposable, we obtain that

$$H^{\star\star}(u) = \int_0^{2\pi} h_\circ^{\star\star}(u(\omega))\,\mathrm{d}\omega = \int_0^{2\pi} h_\circ(u(\omega))\,\mathrm{d}\omega = H(u).$$

We conclude that our entropy satisfies the bi-conjugacy relationship requested in Theorem 3.

Before returning to our specific problem, let us state one more theorem, in which an explicit relationship between primal and dual solutions is obtained.

Theorem 5 (Primal-dual relationship)**.** *With the notation and assumptions of Theorem 2 assume in addition that* dom D *has nonempty interior, that H is an integral functional of integrand h such that conjugacy through the integral*

sign is permitted. Assume that, as in Theorem 3, $H^{\star\star} = H$ and $g_{\star\star} = g$. Assume finally that the conjugate integrand $h^\star$ is differentiable over $\mathbb{R}$, and that there exists some dual-optimal vector $\bar{\boldsymbol{\lambda}}$ in $\operatorname{int}\operatorname{dom} D$*. If*

$$\bar{u}(\omega) := h^{\star\prime}\Big([\mathbb{A}^\star\bar{\boldsymbol{\lambda}}](\omega), \omega\Big) \in L,$$

then $\bar{u}$ is a primal solution.

We are now ready to get back to our specific entropy problem.

3.4.2 Computing Maximum Entropy densities

On denoting by $\mathbb{I}$ the linear mapping given by

$$\mathbb{I}p = \int_0^{2\pi} p(\theta)\,\mathrm{d}\theta, \quad p \in L^1([0, 2\pi)),$$

Problem $(\mathscr{P})$ can be written as

$$\text{Minimize} \quad H(p) - g_\circ(\mathbb{A}_\circ p)$$

over the space $L^1([0, 2\pi))$, in which

$$\mathbb{A}_\circ p = (\mathbb{I}p; \mathbb{A}p) \in \mathbb{R} \times \mathbb{R}^{2N} = \mathbb{R}^{1+2N}$$

and $g_\circ$ is the function on $\mathbb{R}^{1+2N}$ given by

$$g_\circ(\eta_\circ, \boldsymbol{\eta}) = -\frac{\alpha}{2}\|\mathbf{z} - \boldsymbol{\eta}\|^2_{\Sigma^{-1}} - \delta_{\{1\}}(\eta_\circ).$$

Clearly, the use of the indicator function $\delta_{\{1\}}(\cdot)$ enables to encode the (unrelaxed) normality constraint $\mathbb{I}p = 1$. Straightforward computations show that the adjoint mapping $\mathbb{A}^\star_\circ\colon \mathbb{R}^{1+2N} \to L^\infty([0, 2\pi))$ is given by

$$\mathbb{A}^\star_\circ(\lambda_\circ, \boldsymbol{\lambda})(\theta) = \lambda_\circ + \langle \boldsymbol{\lambda}, \mathbf{T}(\theta)\rangle\,.$$

Moreover,

$$(g_\circ)_\star(\lambda_\circ, \boldsymbol{\lambda}) = \lambda_\circ + \langle \mathbf{z}, \boldsymbol{\lambda}\rangle - \frac{1}{2\alpha}\|\boldsymbol{\lambda}\|^2_\Sigma.$$

Accounting for the fact that, as we have seen above, the entropy can be conjugated by conjugating through the integral sign, the dual problem reads:

$$\begin{array}{rcl} \text{Maximize} \quad D(\lambda_\circ, \boldsymbol{\lambda}) & := & \lambda_\circ + \langle \boldsymbol{\lambda}, \mathbf{z} \rangle \\ & & -\dfrac{1}{2\alpha}\|\boldsymbol{\lambda}\|_\Sigma^2 - \exp(\lambda_\circ - 1)\displaystyle\int_0^{2\pi} \exp\langle \boldsymbol{\lambda}, \mathbf{T}(\theta)\rangle \, \mathrm{d}\theta. \end{array}$$

The function D is concave, finite and differentiable on $\mathbb{R}^{1+2N}$. Its stationary points must satisfy the system

$$\left\{ \begin{array}{rcl} 0 & = & 1 - \exp(\bar{\lambda}_\circ - 1)\displaystyle\int_0^{2\pi} \exp\left\langle \bar{\boldsymbol{\lambda}}, \mathbf{T}(\theta)\right\rangle \, \mathrm{d}\theta, \\ \mathbf{0} & = & \mathbf{z} - \dfrac{1}{\alpha}\Sigma\bar{\boldsymbol{\lambda}} - \exp(\bar{\lambda}_\circ - 1)\displaystyle\int_0^{2\pi} \mathbf{T}(\theta)\exp\left\langle \bar{\boldsymbol{\lambda}}, \mathbf{T}(\theta)\right\rangle \, \mathrm{d}\theta, \end{array} \right.$$

which reduces to

$$\mathbf{0} = \mathbf{z} - \frac{1}{\alpha}\Sigma\bar{\boldsymbol{\lambda}} - \frac{\displaystyle\int_0^{2\pi} \mathbf{T}(\theta)\exp\left\langle \bar{\boldsymbol{\lambda}}, \mathbf{T}(\theta)\right\rangle \, \mathrm{d}\theta}{\displaystyle\int_0^{2\pi} \exp\left\langle \bar{\boldsymbol{\lambda}}, \mathbf{T}(\theta)\right\rangle \, \mathrm{d}\theta}.$$

It should be noticed that the above system is also the optimality system of the problem

$$(\tilde{\mathscr{D}}) \quad \left| \begin{array}{rl} \text{Maximize} & \langle \boldsymbol{\lambda}, \mathbf{z} \rangle - \dfrac{1}{2\alpha}\|\boldsymbol{\lambda}\|_\Sigma^2 - \ln\displaystyle\int_0^{2\pi} \exp\langle \boldsymbol{\lambda}, \mathbf{T}(\theta)\rangle \, \mathrm{d}\theta \\ \text{s.t.} & \boldsymbol{\lambda} \in \mathbb{R}^{2N}. \end{array} \right.$$

Proposition 6. *The function*

$$\tilde{D}(\boldsymbol{\lambda}) := \langle \boldsymbol{\lambda}, \mathbf{z} \rangle - \frac{1}{2\alpha}\|\boldsymbol{\lambda}\|_\Sigma^2 - \ln\int_0^{2\pi} \exp\langle \boldsymbol{\lambda}, \mathbf{T}(\theta)\rangle \, \mathrm{d}\theta$$

to be maximized in Problem $(\tilde{\mathscr{D}})$ *is concave, smooth and everywhere finite. Its gradient is given by*

$$\nabla\tilde{D}(\boldsymbol{\lambda}) = \mathbf{z} - \frac{1}{\alpha}\Sigma\boldsymbol{\lambda} - \frac{\displaystyle\int_0^{2\pi} \mathbf{T}(\theta)\exp\langle \boldsymbol{\lambda}, \mathbf{T}(\theta)\rangle \, \mathrm{d}\theta}{\displaystyle\int_0^{2\pi} \exp\langle \boldsymbol{\lambda}, \mathbf{T}(\theta)\rangle \, \mathrm{d}\theta}.$$

The function $h_\circ^\star(\tau) = \exp(\tau - 1)$ obviously meets the requirements of Theorem 5. Provided we can obtain a dual solution $(\bar{\lambda}_\circ, \bar{\boldsymbol{\lambda}})$, the optimal density is then given by

$$\bar{p}(\theta) = \exp\Big[\bar{\lambda}_\circ - 1 + \Big\langle \bar{\boldsymbol{\lambda}}, \mathbf{T}(\theta) \Big\rangle\Big] = \frac{\exp\Big\langle \bar{\boldsymbol{\lambda}}, \mathbf{T}(\theta) \Big\rangle}{\displaystyle\int_0^{2\pi} \exp\Big\langle \bar{\boldsymbol{\lambda}}, \mathbf{T}(\theta) \Big\rangle \, \mathrm{d}\theta}, \tag{3.8}$$

where $\bar{\boldsymbol{\lambda}}$ maximizes the function $\tilde{D}$.

Remark 7. The primal dual relationship obtained above shows that the solution to our Maximum Entropy problem exists and belongs to the *exponential family*. The particular form in Equation (3.8) is well known. In the *unrelaxed* form of Problem ($\mathscr{P}$), it can be obtained using classical inequalities (see e.g. [64], cited in [49]). It is worth noticing that, in [50], similar results are obtained without resorting to Fenchel duality. However, we wish to stress some advantages one may have to resort to Partially Finite Convex Programming. First, we notice that the latter formalism can deal in a very elegant way with problems in which constraints are relaxed. More importantly, by using our approach, we learn that the appropriate value of the Lagrange multiplier maximizes the well behaved dual function, which offers a powerful way to derive a computationally efficient algorithm. Finally, we see that the proposed framework may be used with a variety of entropy functional (such as Burg [23], Renyi [97], Tsallis [117, 80], etc.). Although this goes beyond the scope of this chapter, it certainly deserves to be mentioned.

3.4.3 Deriving an algorithm

Main computational aspects

The above developments yield the following algorithm for the computation of Maximum Entropy densities.

Algorithm 2 Computing Maximum Entropy densities

1: **Input**: $N \in \mathbb{N}^*$, $\mathbf{z} \in \mathbb{R}^{2N}$, $\alpha > 0$.
2: **Output**: The Maximum Entropy probability angular distributions
3: Maximize the dual function

$$\tilde{D}(\boldsymbol{\lambda}) := \langle \boldsymbol{\lambda}, \mathbf{z} \rangle - \frac{1}{2\alpha} \|\boldsymbol{\lambda}\|_{\Sigma}^2 - \ln \int_0^{2\pi} \exp \langle \boldsymbol{\lambda}, \mathbf{T}(\theta) \rangle \, \mathrm{d}\theta.$$

4: From the dual optimal solution $\bar{\boldsymbol{\lambda}}$ obtained in the previous step, form the optimal density

$$\bar{p}(\theta) = \frac{\exp \left\langle \bar{\boldsymbol{\lambda}}, \mathbf{T}(\theta) \right\rangle}{\displaystyle\int_0^{2\pi} \exp \left\langle \bar{\boldsymbol{\lambda}}, \mathbf{T}(\theta) \right\rangle \, \mathrm{d}\theta}.$$

Maximizing the above dual functions is easy, since $\tilde{D}$ is concave, smooth and $\mathbb{R}$-valued. Such maximization can be efficiently performed by means of some quasi-Newton method such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [22], which requires the evaluation of $\tilde{D}$ and its gradient only.

How to choose N and α?

In the Maximum Entropy problem, the *attach term*

$$\frac{\alpha}{2} \|\mathbf{z} - \mathbb{A}p\|_{\Sigma^{-1}}^2$$

depends on the choice of

- the highest frequency N to be accounted for;
- the estimated covariance matrix Σ;
- the *relaxation parameter* α.

We now describe our strategy for their empirical determination.

It is intuitively clear that the information provided by higher frequencies tend to be irrelevant: in the spirit of the Maximum Entropy strategy, we

are looking for the most unbiased information on the underlying random variable (the angle θ) so that fine (high resolution) details are unwanted; moreover, the empirical Fourier coefficients are expected to become poorer approximations of the true ones as the frequency l increases, in particular when the number of angular samples is low. It is therefore only natural to put a limit N on the highest frequency to be accounted for. The dimension of the dual variable will then be $2N$. In the context of simulated random angles, we observe that the Kullback-Leibler entropy of the true density $p_\circ$ relative to the reconstructed density p decreases as N increases, and stabilizes beyond some value of N. Otherwise expressed, beyond a certain value of N, the gain in information provided by additional Fourier coefficients is negligible or non-existent.

For example, Figure 3.2 illustrates that there is no gain beyond $N = 50$ in our simulation, indicating that higher frequencies do not substantially improve the approximation of the true distribution. This finding supports the principle of using a limited number of Fourier coefficients to avoid overfitting and unnecessary computational complexity in the estimation process.

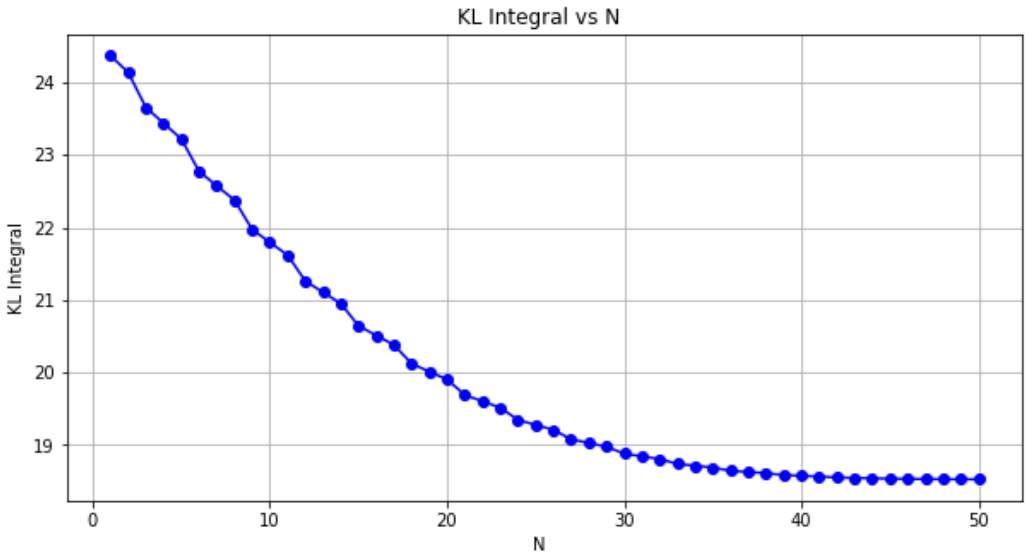


Figure 3.2: Kullback–Leibler relative entropy versus highest frequency.

It now remains to select the value of the *relaxation parameter* α. In the literature on inverse problems, one can find a variety of parameter se-

lection rules. Among these rules, the Morozov discrepancy principle states that α should be such that the corresponding solution $\bar{p}$ should give a residual $\|\mathbf{z} - \mathbb{A}\bar{p}\|$ equal to a number strictly greater than, but close to, the estimated size ρ of the error on the data. Therefore, in order to implement the Morozov principle, a reasonable upper bound on the error should be known, an assumption that we make in the context of this chapter.

Our strategy is as follows: starting from an initial guess $\alpha_\circ > 0$, we compute the corresponding Maximum Entropy solution, which we now denote by $\bar{p}_{\alpha_\circ}$. We then compute the residual $\|\mathbf{z} - \mathbb{A}\bar{p}_{\alpha_\circ}\|$. If the latter residual is above the value 1.105ρ, then set $\alpha_1 > \alpha_\circ$; otherwise set $\alpha_1 < \alpha_\circ$. We then iterate on this process, until α gives a residual approximately equal to 1.105ρ. In practice, a dichotomy strategy can be used so as to converge to the appropriate value of α. The detailed information can be seen in Algorithm 3.

Algorithm 3 Determining α by means of the Morozov discrepancy principle

1: **Input**: ρ, condition (e.g. $1.095\rho \leq$ residual $\leq 1.105\rho$), μ (e.g. μ =1.2), maximum number of iterations $N_{\max}$
2: **Output**: Morozov value of α
3: Set $i = 0$, $\alpha_{\min} = 0$, $\alpha_{\max} = \infty$, $\alpha_0 = 1$
4: **while** condition is not satisfied and $i < N_{\max}$ **do**
5: Compute the Maximum Entropy solution $\bar{p}_{\alpha_i}$
6: Compute the residual $\|\mathbf{z} - \mathbf{A}\bar{p}_{\alpha_i}\|$
7: **if** residual $< 1.095\rho$ **then**
8: Set $\alpha_{\max} = \alpha_i$
9: Set $\alpha_{i+1} = \frac{\alpha_{\min} + \alpha_{\max}}{2}$
10: **end if**
11: **if** residual $> 1.105\rho$ **then**
12: Set $\alpha_{\min} = \alpha_i$
13: Set $\alpha_{i+1} = \begin{cases} \frac{\alpha_{\min} + \alpha_{\max}}{2} & \text{if } \alpha_{\max} < \infty \\ \mu\alpha_i & \text{otherwise} \end{cases}$
14: **end if**
15: Set $i = i + 1$
16: **end while**
17: Return last value of α

3.5 Simulations

3.5.1 Validation with Dirac distributions

We conduct testing on a single cell to assess the fundamental functioning of the model. The results from this step provide essential insights into accuracy and basic performance.

We start by performing simulations. We shall:

(a) generate an angular sample following a probability $p_\circ$;

(b) compute the corresponding empirical Fourier coefficients;

(c) compute the Maximum Entropy density that is compatible, in the relaxed setting described above, with our empirical Fourier coefficients.

This simulation will allow us to get an idea of the sample size that is necessary for obtaining relevant information, merely by comparing the *true* density with the one obtained through the maximization of entropy.

We start with the design of densities on $[0, 2\pi)$. To this end, we recall the definition of the *bump function*:

$$\psi_\circ(\theta) = \begin{cases} \exp\left(-\dfrac{1}{1-\theta^2}\right) & \text{if } \theta \in (-1,1), \\ 0 & \text{otherwise.} \end{cases}$$

Recall that $\psi_\circ$ is a $\mathscr{C}^\infty$-function, whose support is indeed the compact interval $[-1, 1]$. We can emulate the behaviour of a Dirac distribution by letting

$$\psi_\beta(\theta) := \frac{1}{\beta c}\psi_\circ\left(\frac{\theta}{\beta}\right) \quad \text{in which} \quad c := \int \psi_\circ(\theta)\,\mathrm{d}\theta.$$

Here, β is a strictly positive parameter. The closer β to zero, the more ψ_β is *concentrated* about the origin. It is readily seen that the integral of ψ_β is equal to 1 whatever may be the value of β, and that the support of ψ_β is the interval $[-\beta, \beta]$.

We may now define, for the purpose of simulation, densities of the form

$$p_\circ(\theta) = \sum_{k=1}^{p} c_k \psi_{\beta_k}(\theta - \theta_k), \tag{3.9}$$

in which the c_k's are positive numbers that sum up to 1. The *translation* parameters θ_k are chosen in such a way that the support of $p_\circ$ is contained in $[0, 2\pi]$. We then simulate angular samples that follow such a distribution.

This simulation will allow us to get an idea of the sample size that is necessary for obtaining relevant information, merely by comparing the true density with the one obtained through the maximization of entropy.

To visually demonstrate the effectiveness of our simulation in capturing the original distribution's characteristics, we proceed with a specific example.

Figure 3.3 is one of a simulation example based on the original probability $p_\circ(\theta)$ with 2 peaks and $\beta = 0.1$. It shows the underlying (original) probability density $p_\circ(\theta)$ and the histogram obtained by simulation.

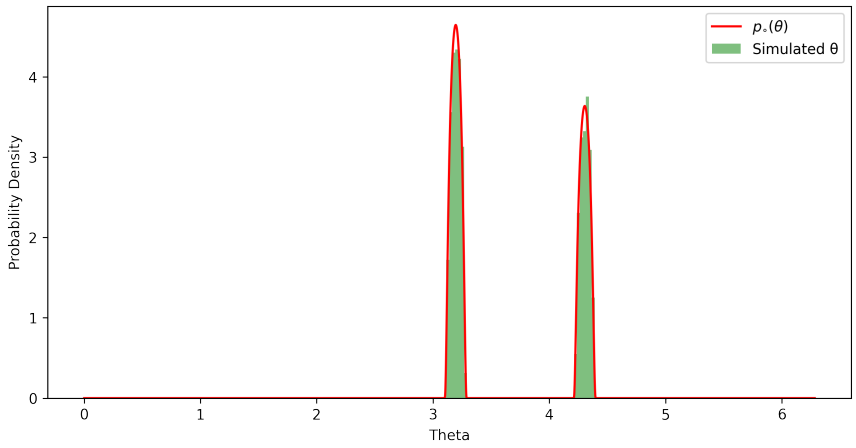


Figure 3.3: Simulate angular samples based on a original probability $p_\circ$.

Following the successful validation of the simulation, we proceed to refine our model to enhance its performance further.

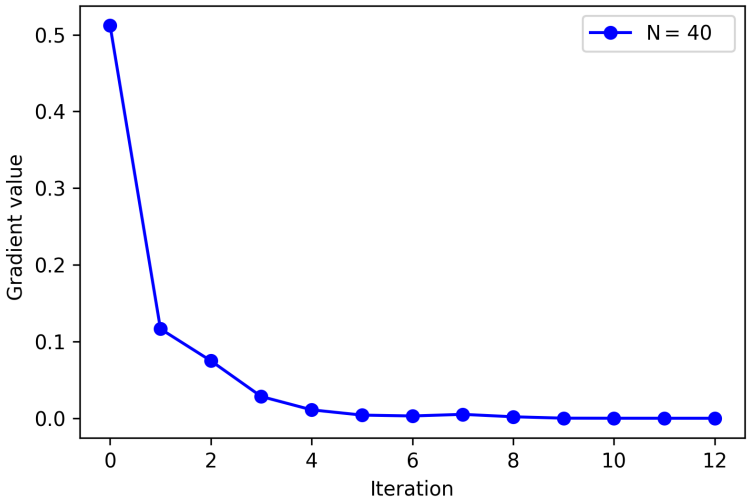


Figure 3.4: Gradient norm when $N = 40$.

Figure 3.4 illustrates the convergence behavior of an optimization process with the gradient norm plotted against the iteration number. Initially, there is a sharp decrease from the first to the second iteration, indicating a rapid improvement in the optimization process. Following this, the curve shows a rapid convergence of the gradient norm to zero.

3.5.2 Numerical implementations

Morozov principle for the selection of α

In this section, we investigate the impact of the regularization parameter α on the selection of the optimal solution in the inverse problem.

We have conducted an algorithmic search to choose the appropriate value for α, using the Morozov principle as described in Algorithm 3. The results of the algorithm reveal the relationship between α and the residual as shown in Figure 3.5.

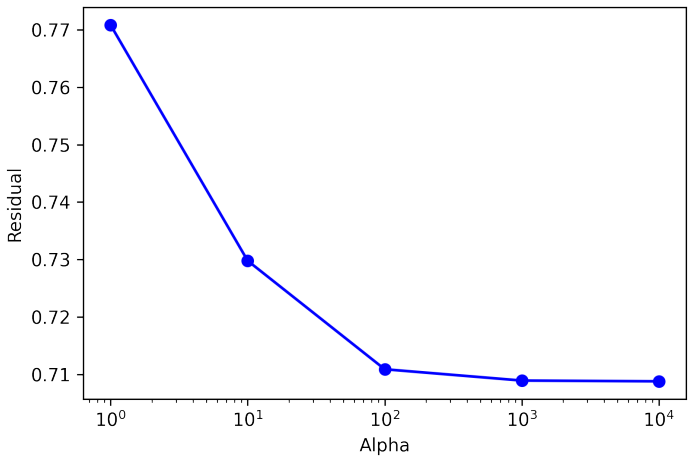


Figure 3.5: Residual values when α values from 1 to 10000.

As α increases from 1 to 100, the residual gradually decreases from approximately 0.771 to about 0.711. However, once α surpasses the value of 100, further increases do not significantly reduce the residual. This indicates that selecting an appropriate value for α is vital to achieve the desired residual.

The findings from this analysis are crucial for understanding how the value of α impacts the algorithm's performance in real scenarios, suggesting that value of α around 100 is likely optimal for minimizing the residuals without unnecessary computational expenditure.

Reconstructions

After determining the optimal parameter values, we compute the optimal density by restructuring the original distribution. Figure 3.6 and Figure 3.7 present a comparative analysis between the original and optimized probability density functions, reflecting the outcome of an optimization process aimed at reconstructing the original distribution.

In the two-peak scenario (as shown in Figure 3.6), the optimal density closely aligns with the original density, demonstrating the algorithm's success in capturing the essential features of the distribution, namely the peak locations and heights. It indicates that the optimization process is robust in

simpler scenarios with fewer peaks.

In the more complex five-peak scenario (as shown in Figure 3.7), the reconstructed density still captures the general shape and peak positions of the original density, which is commendable given the increased complexity.

In all situations, the reconstructed density closely mirrors the shape of the original density successfully capturing the positions and general form of the peaks. However, there are minor differences, particularly in the central peak where the optimal density does not perfectly align, hinting at slight deviations. Despite these small deviations, the optimized distribution is a close approximation of the original distribution.

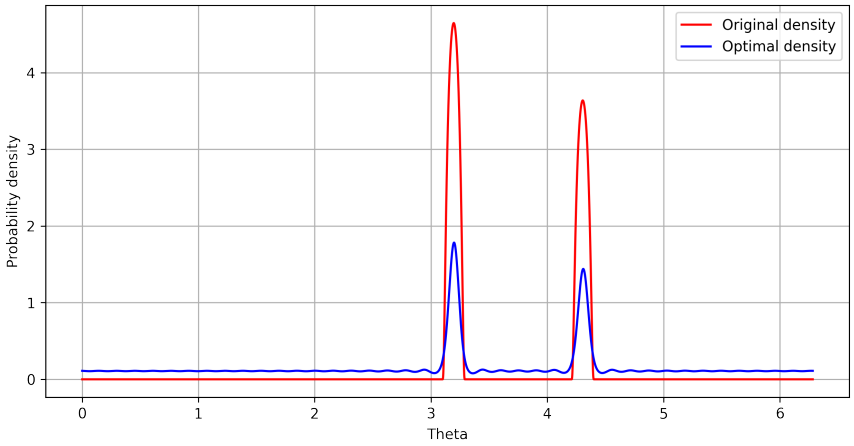


Figure 3.6: Results on the Original and Optimal densities with two peaks.

The slight deviations in peak heights and widths provide avenues for refinement, but overall, the optimization algorithm appears to be performing effectively. It is successful in identifying and adapting to the multiple modes of the distribution, which is a positive outcome, a particularly positive development within the challenging multi-peak environment. Further fine-tuning of the algorithm might lead to even closer alignment, enhancing the model's precision.

The effectiveness of the BFGS algorithm in maximizing concave, smooth, and $\mathbb{R}$-valued dual functions is further explored by comparing its performance with several other optimization methods. These include Powell [92], Nelder-

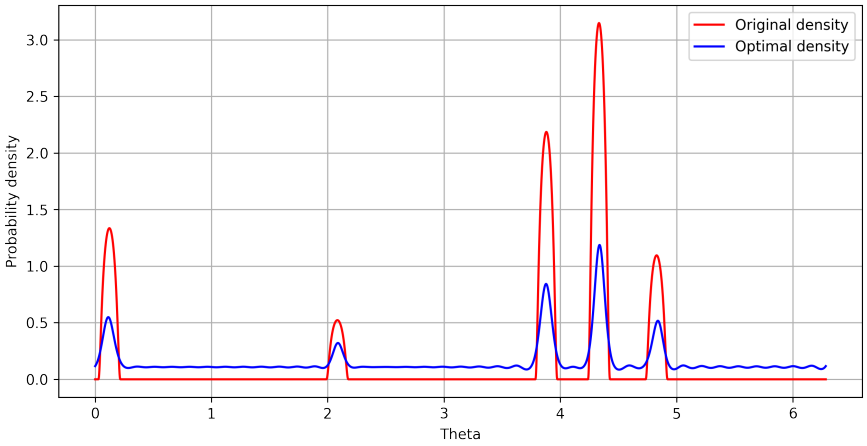


Figure 3.7: Results on the Original and Optimal densities with five peaks.

Mead [85], Conjugate Gradient (CG) [54], Truncated Newton Conjugate-Gradient (TNC) [43], Constrained Optimization BY Linear Approximations (COBYLA) [91], and Trust Region Constraint (Trust-Constr) [84].

Some basic information of these approaches are presented as follows:

1. BFGS method. It is a quasi-Newton method used to solve unconstrained optimization problems by approximating the Hessian matrix of the objective function. Instead of directly computing the Hessian matrix, BFGS uses gradient-based updates to build the inverse Hessian matrix approximation. The update formula for the inverse Hessian matrix B_{k+1} at the $(k+1)^{th}$ iteration is:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}$$

where:

- $s_k = x_{k+1} - x_k$ is the difference between successive points.
- $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ is the change of gradients.

Next step, BFGS computes x_{k+1} by using B_{k+1}:

$$x_{k+1} = x_k - \alpha_k B_k^{-1} \nabla f(x_k).$$

2. Powell method. It is a derivative-free optimization method using a series of linear searches to find the minimum value of a function. The optimization process begins with a set of initial directions (usually the unit vectors of the coordinate axes). At each iteration, the method performs a linear search along each of these directions to reduce the value of the objective function $f(x)$.

   Given a set of directions $\{d_1, d_2, \ldots, d_n\}$, this method finds the optimal point x_{k+1} by performing a minimization along each direction d_i as:

   $$x_{k+1} = x_k + \lambda_i d_i,$$

   where λ_i is a scalar chosen to minimize $f(x)$ along the direction d_i. After each line search sequence, the method updates the optimal direction by adding the direction from the starting point to the ending point of the search sequence and replacing the direction that caused the least reduction in the objective function. The new direction is computed as:

   $$d_{\text{new}} = x_n - x_0.$$

3. Nelder-Mead method. It is a derivative-free optimization method that keep track of $n+1$ interest points in $\mathbb{R}^n$, whose convex hull forms a simplex to find the minimum of the objective function $f(x)$. Given a simplex S with vertices $\{x_1, x_2, \ldots, x_{n+1}\}$, we compute $f(x)$ at each vertex of the simple and determine x_l, x_h, x_s corresponding to the best, worst and second-worst vertices based on the values of $f(x)$. Operations on the simplex include reflection, expansion, contraction, and shrinkage.

   - **Reflection**: After computing the centroid x_c of the simplex, move the worst point x_h through x_c to generate a new point x_r:

     $$x_r = x_c + \alpha(x_c - x_h)$$

     where $\alpha > 0$ is the reflection coefficient.
   - **Expansion**: If reflection reduces the function value, try expanding further to find a better point x_e:

     $$x_e = x_c + \gamma(x_r - x_c)$$

     with $\gamma > 1$.

- **Contraction**: If reflection fails, try contracting the point x_r towards x_c:

$$x_c = x_c + \rho(x_h - x_c)$$

with $0 < \rho < 1$.

- **Shrinkage**: If contraction also fails, shrink the entire simplex towards the best point x_l:

$$x_i = x_l + \sigma(x_i - x_l)$$

with $0 < \sigma < 1$.

4. CG method. It is a powerful optimization technique, particularly effective for large-scale problems where the objective function is a quadratic function. At each iteration k, the method finds a conjugate direction d_k by combining the current gradient with the previous direction:

$$d_k = -\nabla f(x_k) + \beta_{k-1} d_{k-1}$$

where the scalar β_{k-1} is computed as:

$$\beta_{k-1} = \frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_{k-1})^T \nabla f(x_{k-1})}$$

The next point x_{k+1} is updated as:

$$x_{k+1} = x_k + \alpha_k d_k$$

where α_k is the optimal step size, usually determined by a line search.

5. TNC method. It is a hybrid method that combines the conjugate gradient method with the Newton method, particularly useful for large-scale optimization problems. Instead of using the full Hessian matrix, TNC only uses part of it, reducing computational complexity. The Newton step p_k is computed by solving:

$$B_k p_k = -\nabla f(x_k)$$

where B_k is the Hessian matrix at point x_k.

In TNC, p_k is approximated by truncating the iterations of the conjugate gradient method, finding a near-optimal solution. The next point is updated as:

$$x_{k+1} = x_k + \alpha_k p_k$$

where α_k is the optimal step size.

6. COBYLA method. It is a derivative-free optimization method that handles constraints by approximating both the objective function and the constraints using linear models. At each step, a linearized subproblem within a trust region is solved to find the optimal solution while ensuring that the constraints are satisfied. Suppose the original objective function is $f(x)$ and the constraints are $g_i(x) \leq 0$. COBYLA approximates them with linear functions:

$$\hat{f}(x) = f(x_0) + \nabla f(x_0)^T (x - x_0),$$

$$\hat{g_i}(x) = g_i(x_0) + \nabla g_i(x_0)^T (x - x_0).$$

The optimization is performed on these linearized functions, and the result is used to update the current point.

7. Trust-Constr method. It optimizes a function within a trust region around the current point, using a local model of the objective function. Given a local model $m_k(p)$ and a trust-region radius Δ_k, the subproblem to solve is:

$$\min_p m_k(p) \quad \text{subject to} \quad \|p\| \leq \Delta_k..$$

The model $m_k(p)$ is typically a quadratic approximation of the objective function:

$$m_k(p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p$$

where B_k is an approximation of the Hessian matrix. If p_k is the solution to the subproblem, the new point is updated as:

$$x_{k+1} = x_k + p_k.$$

The trust-region radius Δ_k may be adjusted based on the accuracy of the model.

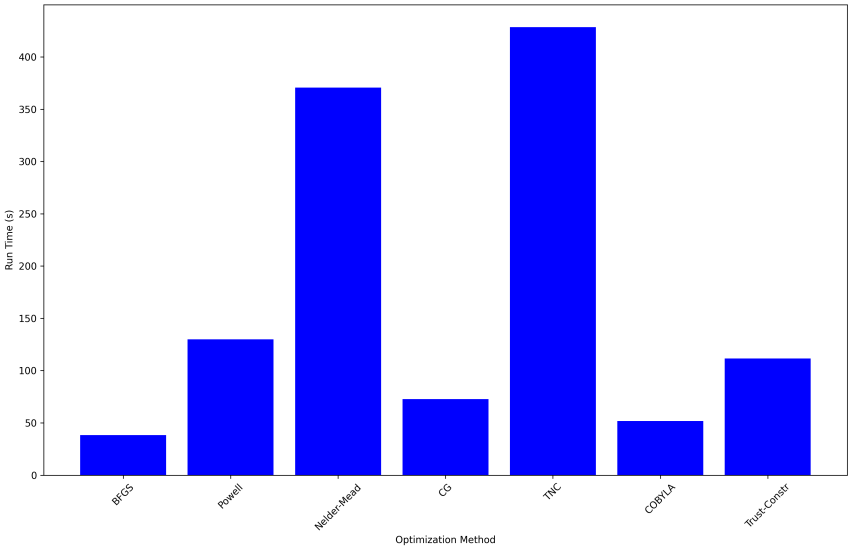


Figure 3.8: Comparison of computational time by seven optimal models with five peaks.

Each of these methods, like BFGS, primarily relies on evaluations of the function and its gradient to find optimal solutions. The assessment of these methods focuses on computational time (see Figure 3.8). In this figure, COBYLA and BFGS methods are noted for their rapid execution times, making them preferable for applications requiring quick data processing. In contrast, Powell and TNC methods require longer computational times, which could be a disadvantage in time-sensitive applications.

3.6 Towards air traffic complexity estimation

The application of the Maximum Entropy method offers a novel avenue for modeling and visualizing the complexities inherent in air traffic management. By reconstructing angular probability densities from simulated trajectory data, this method allows for a nuanced analysis of traffic dynamics, particularly in terms of predictability and potential trajectory conflicts within specific airspace sectors.

Various measures of complexity may be envisaged from the obtained probability density. In the context of aircraft trajectories, complexity measures are used to understand and quantify the level of predictability and potential conflicts in these trajectories. It can be mentioned by some key points:

- *Divergence between entrance and exit density*: This concept suggests that by comparing the angular density at the entrance of a specific region (cell S) with the angular density at the exit, we can determine if aircraft tend to change their paths within that region. If there is a significant difference between the two densities, it indicates that aircraft are altering their trajectories in that region. This divergence signifies a lack of predictability and contributes to the complexity of the overall air traffic system.

- *Counting peaks in entrance density*: Another method for assessing complexity is to count the number of peaks in the entrance density. Each peak may represent a concentration of aircraft in a specific direction or route. If there are multiple peaks, it suggests that different aircraft are following distinct paths when entering the region. A single peak indicates that most aircraft tend to follow the same direction, which results in lower conflict and, consequently, lower complexity.

- *Weighting complexity by traffic intensity*: It is important to consider the intensity of air traffic when evaluating complexity. The number of trajectories (n) within the specific region (cell S) is used as a measure of traffic intensity. Complexity should be weighed or adjusted based on the volume of traffic. A high level of complexity might be more concerning if it occurs in regions with heavy air traffic compared to regions with low traffic.

In this context, the goal is to use these measures to understand the patterns and behaviors of aircraft trajectories. Higher complexity, indicated by divergence, multiple peaks, or significant traffic, could imply a need for more sophisticated air traffic management strategies to ensure safety and efficiency. These measures help in assessing and optimizing air traffic control systems.

3.7 Conclusions

In this chapter, we envisage a novel approach to construct traffic complexity maps for air traffic management by focusing on the reconstruction of angular probability density functions from simulated trajectory data. The core of our methodology is to use the Maximum Entropy principle, rigorously incorporating the statistical properties of the Fourier coefficients of the underlying angular density. This approach ensures that the local complexity is accurately assessed.

Our application of partially finite convex programming, grounded in an infinite-dimensional form of Fenchel's duality theorem and the conjugacy of integral functionals, effectively computes optimal densities. The implementation of this method presents a robust solution to the data fitting challenge within the framework of angular density estimation.

The validation of our method through numerical simulations and the reconstruction of simulated angular data have demonstrated its proficiency in precisely capturing the original angular probability densities. From two-peak distributions to more complex five-peak scenarios, our algorithm consistently achieved an optimal density that closely resembles the original distribution, effectively preserving the defining characteristics of the distribution — such as peak locations and heights — even as the complexity increases. This consistency ensures that our method can provide reliable information for air traffic complexity maps, enhancing the predictive capabilities and operational efficiency of air traffic management systems.

The increasing number of peaks, correlating with the number of aircraft trajectory directions within a sector, further validates our method's capability to significantly improve the evaluation of air traffic complexity. This is instrumental for real-time decision-making and strategic planning in air traffic control, where understanding complex flight patterns is crucial for safety and efficiency.

The potential applications of this methodology in real air traffic management are promising. By integrating real traffic data into this framework, future research could refine and validate the models to enhance their practical utility. Additionally, expanding this approach to incorporate other variables such as altitude and speed could provide a more comprehensive tool

for complexity assessment, supporting more sophisticated air traffic control strategies.

This research not only establishes a robust mathematical framework for estimating angular density but also paves the way for significant advancements in managing the complexities of growing air traffic. As the volume of air traffic continues to increase, the relevance and necessity of deploying advanced analytical tools such as those developed in this study will become ever more critical.

Chapter 4

Building complexity maps of airspace via information geometry

Building air traffic complexity maps is crucial for enhancing air traffic management and ensuring safety. By understanding the local complexity, air traffic controllers can better manage aircraft flow, reduce congestion, and anticipate potential conflicts, ultimately leading to more efficient and safer airspace operations. In this chapter, we introduce a methodology to assess air traffic complexity through three steps. First, we estimate the probability density of aircraft headings on a Cartesian grid using Maximum Entropy. Next, we approximate the geodesic distance to the uniform distribution in the Maksimov manifold. Finally, we compute a complexity index combining angular distribution and local traffic intensity to create a complexity map. This index, which has been tested on real data, demonstrates reliability, with alternative criteria possible by comparing heading distributions at cell boundaries. Our approach is both numerically efficient and effective in capturing air traffic complexity.

4.1 Introduction

4.1.1 Motivation

Air traffic complexity involves the dynamic interactions and trajectories of aircraft within the airspace. This chapter presents a novel methodology for constructing complexity maps that strategically anticipate and manage air

traffic challenges. By developing advanced complexity measures, we aim to enhance ATM systems and ensure efficient and safe operations through pre-emptive planning.

Current airspace management faces significant constraints due to the rigid structures and the cognitive limitations of ATCOs. They are tasked with safely managing an increasing number of aircraft, especially within congested sectors. Despite technological advancements in ATM, the combination of fixed airspace configurations and controller workload constraints lead to inefficiencies and potential safety risks. As air traffic continues to grow, these challenges become more pronounced, necessitating innovative solutions to maintain and improve safety and efficiency.

Proactive development of complexity measures and maps is essential for effective ATM. Complexity maps provide valuable insights into traffic patterns and system responses to various conditions, facilitating better strategic planning. Lee et al. (2009) [68] demonstrated how complexity maps can optimize aircraft entry points and manage sector boundaries, while Hong et al. (2014) [57] proposed a Variance Quadtree-based Map for resolving conflicts by identifying optimal strategies for aircraft entries into sectors.

Further, studies by Juntama et al. (2020) [63] and Delahaye et al. (2022) [36] showed that complexity maps are instrumental in optimizing 4D trajectories and evaluating local traffic disorder. These tools are crucial for enhancing air traffic control operations, reducing delays, and improving overall airspace capacity by anticipating and managing complexity before it arises. By leveraging complexity maps, ATM systems can better handle the intricacies of air traffic, ensuring that safety and efficiency are maintained even as traffic volumes increase.

4.1.2 Overview

Our methodology for constructing air traffic complexity maps begins with discretizing the airspace using a Cartesian grid. At each grid point, we define a circular cell with a radius compliant with safety regulations. For each cell, we compute an angular distribution of trajectories using the principle of Maximum Entropy.

This distribution thus obtained belongs to the family of Generalized von

Mises (GvM) distributions. The determination of the probability density of the underlying angular variable is achieved using Fenchel duality, where empirical moment constraints are relaxed in accordance with the estimated covariance matrix of the data, implying the use of Mahalanobis distance. The details of this strategy were proposed in Chapter 3, and the key points are summarized in Subsection 4.2.1. We then calculate the geodesic distance between this distribution and the uniform distribution, where a smaller distance indicates higher local traffic complexity. The local complexity index takes into account this distance and the local intensity of the traffic.

To manage the computational intensity of evaluating the geodesic distance, we employ a practical approximation using the symmetrized Kullback-Leibler (SKL) divergence. This allows for efficient estimation of local traffic complexity across the grid, making the methodology feasible for real-time applications.

4.2 Deriving a complexity index via information geometry

Our objective, in this section is to obtain a local complexity index using concepts from information theory. This construction consists of two parts.

In the first part, we construct an angular probability density at a given point on the grid. The trajectories crossing the corresponding cell provide a statistical sample of the angular variable, and we apply the principle of Maximum Entropy to determine an angular density consistent with the empirical Fourier coefficients obtained. The constraints corresponding to the empirical coefficients are relaxed based on the quality of the estimation, particularly the number of samples (i.e., the number of trajectories crossing the cell). By applying Fenchel duality, we can prove that the density maximizing entropy belongs to the GvM density family. Since it is expressed via the maximizer of the dual function, the solution actually appears in the Maksimov form.

The second part, which is the main focus of this chapter, involves evaluating the geodesic distance between the obtained distribution and the uniform angular distribution that corresponds to the less organized traffic hence the most difficult one to control.

We will see in Subsection 4.3.1 that the natural parametrization of the GvM distribution manifold suggested by Fenchel duality, which corresponds to the Maksimov form, theoretically allows for the calculation of this geodesic distance by solving a relatively high-dimensional system of ordinary differential equations. For reasons of algorithmic and numerical simplicity, we will then use the classical approximation of the geodesic distance by the symmetrized Kullback-Leibler divergence.

4.2.1 Maximum Entropy angular densities

Consider a cell centered at a grid point in the airspace, an example of which is given in Figure 4.1. The empirical Fourier coefficients

$$x_l = \frac{1}{\pi n} \sum_{j \in J} \cos(l\theta_j) \quad \text{and} \quad y_l = \frac{1}{\pi n} \sum_{j \in J} \sin(l\theta_j) \tag{4.1}$$

approximate the unknown Fourier coefficients of the searched angular distribution $p(\theta)$. We may assume to have access to the quality of this approximation, under the form of the covariance matrix Σ of the vector

$$\mathbf{T}(\theta) = \Big(\cos\theta, \sin\theta, \ldots, \cos(N\theta), \sin(N\theta) \Big).$$

We define $\mathbf{z} = (x_1, y_1, \ldots, x_N, y_N)$. The choice of the *highest frequency* N has been discussed in Chapter 3. In our Maximum Entropy approach, adequacy of the moments of the searched density $p(\theta)$ is requested, which involves the linear operator $\mathbb{A} \colon L^1(\mathbb{S}) \to \mathbb{R}^{2N}$ given by

$$\mathbb{A}p = \int_{\mathbb{S}} p(\theta) \mathbf{T}(\theta) \, \mathrm{d}\theta,$$

where $\mathbb{S} := \mathbb{R}/2\pi\mathbb{Z}$. The *attach term* in our Maximum Entropy problem accounts for the covariance matrix Σ whenever an estimate of it is available. It is defined as

$$\|\mathbf{z} - \mathbb{A}p\|^2_{\Sigma^\dagger} = \begin{cases} \Big\langle \mathbf{z} - \mathbb{A}p, \Sigma^\dagger(\mathbf{z} - \mathbb{A}p) \Big\rangle & \text{if } \mathbf{z} - \mathbb{A}p \in \operatorname{ran}\Sigma, \\ \infty & \text{otherwise,} \end{cases}$$

in which $\Sigma^\dagger$ is the Moore-Penrose generalized inverse of Σ and $\operatorname{ran}\Sigma$ denotes the range of Σ. The infinite value forces the moment vector $\mathbb{A}p$ to lie in

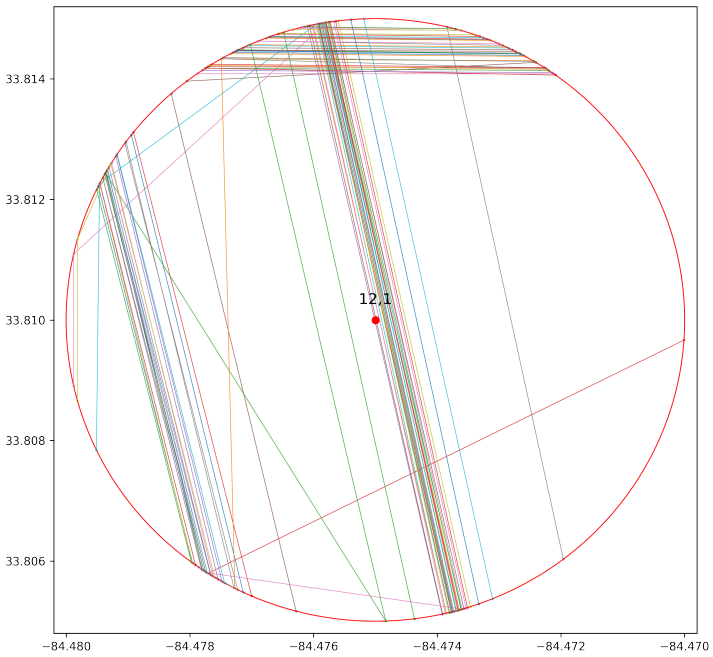


Figure 4.1: Example of a cell around a grid point in the airspace, with trajectories across ATL on December 26, 2018.

the affine space $\{\mathbf{z}\} + \operatorname{ran}\Sigma$. If Σ is positive definite, then $\Sigma^\dagger$ is merely the inverse of Σ, and $\|\mathbf{z} - \mathbb{A}p\|_{\Sigma^\dagger}$ is the Mahalanobis distance between $\mathbf{z}$ and $\mathbb{A}p$. The Maximum Entropy problem reads

$$(\mathscr{P}) \quad \left| \begin{array}{rl} \text{Minimize} & H(p) + \dfrac{\alpha}{2}\|\mathbf{z} - \mathbb{A}p\|_{\Sigma^\dagger}^2 \\ \text{s.t.} & 1 = \displaystyle\int_{\mathbb{S}} p(\theta)\,\mathrm{d}\theta. \end{array} \right.$$

Here, the entropy functional H is defined on $L^1(\mathbb{S})$ by

$$H(p) = \int_{\mathbb{S}} h\Big(p(\theta)\Big)\,\mathrm{d}\theta \quad \text{with} \quad h(t) := \begin{cases} t\ln t & \text{if } t > 0, \\ 0 & \text{if } t = 0, \\ \infty & \text{if } t < 0 \end{cases}$$

and $\alpha > 0$ fixes the balance between the entropy and the attach term. It was shown in [86] that a Fenchel duality approach enables to solve the infinite dimensional constrained problem $(\mathscr{P})$ via a finite dimensional unconstrained smooth concave maximization problem, which moreover involves Σ rather than $\Sigma^\dagger$. The solution to Problem $(\mathscr{P})$ is given by

$$\bar{p}(\theta) = \frac{\exp\left\langle \bar{\boldsymbol{\lambda}}, \mathbf{T}(\theta) \right\rangle}{\displaystyle\int_{\mathbb{S}} \exp\left\langle \bar{\boldsymbol{\lambda}}, \mathbf{T}(\theta) \right\rangle \, \mathrm{d}\theta}, \tag{4.2}$$

in which $\bar{\boldsymbol{\lambda}}$ maximizes the function $\tilde{D}$ given by

$$\tilde{D}(\boldsymbol{\lambda}) := \langle \boldsymbol{\lambda}, \mathbf{z} \rangle - \frac{1}{2\alpha} \|\boldsymbol{\lambda}\|_\Sigma^2 - \ln \int_{\mathbb{S}} \exp \langle \boldsymbol{\lambda}, \mathbf{T}(\theta) \rangle \, \mathrm{d}\theta.$$

Observe that:

- the *dual function* $\tilde{D}$ is everywhere finite, smooth and strongly concave, so that it's maximization is numerically rather easy;
- the solution $\bar{p}$, which is strictly positive and smooth on $\mathbb{S}$ by construction, pertains to the family of GvM distribution of degree N.

The generalized von Mises distribution of degree N, hereafter denoted by $\mathbf{GvM}_N$ was studied in [49, 48]. Its canonical form is:

$$p(\theta) = \frac{1}{G_0} \exp\left(\sum_{k=1}^{N} \kappa_k \cos k(\theta - \mu_k) \right), \tag{4.3}$$

where

$$G_0 = \int_{\mathbb{S}} \exp\left(\sum_{k=1}^{N} \kappa_k \cos k \, (\theta - \mu_k) \right) \, \mathrm{d}\theta \tag{4.4}$$

The $\mathbf{GvM}_N$ distribution can be put in the Maksimov form [77]:

$$p(\theta) = \frac{1}{G_0} \varphi(\theta) \tag{4.5}$$

where

$$\varphi(\theta) = \exp\left(\sum_{k=1}^{N} \kappa_k \cos \mu_k \cos k\theta + \kappa_k \sin \mu_k \sin k\theta \right) \tag{4.6}$$

which is the canonical form of an exponential distribution:

$$p(\theta) = \exp\Big[\, \langle \boldsymbol{\lambda}, T(\theta) \rangle - \ln G_0 \Big] \tag{4.7}$$

with $\boldsymbol{\lambda} := (\kappa_1 \cos\mu_1, \kappa_1 \sin\mu_1, \ldots, \kappa_N \cos\mu_N, \kappa_N \sin\mu_N)$, so that $\lambda_{2k-1} = \kappa_k \cos\mu_k$ and $\lambda_{2k} = \kappa_k \sin\mu_k$ for $k = 1, \ldots, N$. Finally, for a discussion on the choices of the *highest frequency* N and the *relaxation parameter* α, the reader is invited to refer to Chapter 3 or the companion paper [86].

4.2.2 The Fisher information metric

In this section, we recall some basic definitions and properties in information geometry. The reader is referred to [6] for a comprehensive introduction.

Definitions and properties

Definition 8. *A statistical model is a pair $(\mathcal{M}, p)$ where $\mathcal{M}$ is an oriented N-dimensional smooth manifold and $(p_\theta)_{\theta\in\mathcal{M}}$ is a parameterized family of probability densities on a measure space $(\Omega, \mathcal{T}, \mu)$ such that, putting $p(\theta, \omega) = p_\theta(\omega)$:*

- *For μ-almost all $\omega \in \Omega$, the mapping $\theta \mapsto p(\theta, \omega)$ is smooth.*
- *For any $\theta \in \mathcal{M}$, there exists an open neighborhood U_θ of θ and an integrable mapping $h\colon \Omega \to \mathbb{R}^+$ such that, for any $\xi \in U_\theta$, $|\partial_\theta p(\xi, \omega)| \leq h$.*
- *The mapping $\theta \to p_\theta \in L^1(\Omega, \mu)$ is one-to-one.*
- *The support of p_θ does not depend on θ.*

Definition 9. *Assuming p never vanishes, one can define the score $l\colon \mathcal{M} \times \Omega \to \mathbb{R}$ as:*

$$l(\theta, \omega) = \log p(\theta, \omega) \tag{4.8}$$

or in brief:

$$l_\theta = \log p_\theta. \tag{4.9}$$

Proposition 10. *The score satisfies:*

$$E\left[\partial_i l_\theta\right]_{p_\theta} = 0,\ i = 1 \dots N. \tag{4.10}$$

in which E denotes the expectation operator.

A simple computation shows that:

$$E\left[\partial_i l_\theta \partial_j l_\theta\right] = \int_\Omega \frac{\partial_i p_\theta}{\sqrt{p_\theta}} \frac{\partial_j p_\theta}{\sqrt{p_\theta}} d\mu(\omega) = 4 \int_\Omega \partial_i\left(\sqrt{p_\theta}\right) \partial_j\left(\sqrt{p_\theta}\right) d\mu(\omega),\ i,j = 1 \dots N \tag{4.11}$$

proving that:

$$g_{ij} = E\left[\partial_i l_\theta \partial_j l_\theta\right] = \left\langle \partial_i\left(\sqrt{p_\theta}\right), \partial_j\left(\sqrt{p_\theta}\right)\right\rangle_{L^2(\Omega,\mu)} \tag{4.12}$$

Let g be the section of $T\mathcal{M}^* \otimes T\mathcal{M}^*$ defined by [1]:

$$g = g_{ij} d\theta^i \otimes d\theta^j \tag{4.13}$$

Now, given any tangent vector $X = X^i \partial_i \in T_\theta \mathcal{M}$:

$$\begin{aligned} g(\theta; X, X) = g_{ij} X^i X^j &= \left\langle \partial_i\left(\sqrt{p_\theta}\right), \partial_j\left(\sqrt{p_\theta}\right)\right\rangle_{L^2(\Omega,\mu)} && (4.14)\\ &= \left\langle X^i \partial_i\left(\sqrt{p_\theta}\right), X^j \partial_j\left(\sqrt{p_\theta}\right)\right\rangle_{L^2(\Omega,\mu)} && (4.15)\\ &= \left\langle Z, Z\right\rangle_{L^2(\Omega,\mu)} && (4.16) \end{aligned}$$

with $Z = X^i \partial_i\left(\sqrt{p_\theta}\right)$. Given the assumptions made on the family p_θ, g is a thus a positive definite symmetric section of $T\mathcal{M} \otimes T\mathcal{M}$, hence a Riemannian metric on $\mathcal{M}$ called the Fisher Information Metric (FIM).

Remark 11. *The mapping $\mathcal{I}\colon \theta \mapsto \sqrt{p_\theta}$ embeds $\mathcal{M}$ as a submanifold of the unit sphere in $L^2_{\Omega,\mu}$ and the Fisher information metric is just the pullback of the ambient metric in $L^2_{\Omega,\mu}$ with respect to $\mathcal{I}$.*

The next proposition is well known in statistics, provided partial derivatives and expectation commute.

Proposition 12. *The Fisher information satisfies:*

$$E\left[\partial_i l_\theta \partial_j l_\theta\right] = -E\left[\partial_i \partial_j l_\theta\right]. \tag{4.17}$$

[1] The summing convention on repeated indices is used through the document.

Proof. Taking the second derivative of the score yields:

$$\partial_i \partial_j l_\theta = \frac{\partial_i \partial_j p_\theta}{p_\theta} - \frac{\partial_i p_\theta \partial_j p_\theta}{p_\theta^2}. \tag{4.18}$$

Taking the expectation and using commutation assumption between derivatives and expectations:

$$\partial_i \partial_j E[p_\theta] = 0 = E\left[\partial_i \partial_j p_\theta\right] = \int_\Omega \frac{\partial_i \partial_j p_\theta}{p_\theta} p_\theta d\omega. \tag{4.19}$$

Now, the claim follows from:

$$E\left[\partial_i l_\theta \partial_j l_\theta\right] = \int_\Omega \frac{\partial_i p_\theta \partial_j p_\theta}{p_\theta} d\omega. \tag{4.20}$$

□

The form of Fisher information given by Proposition 12 is particularly suitable for distributions belonging to the exponential family. This fact will be used in the next section.

Geodesic distances

We now turn to the computation of geodesic distances. Clearly, for a given number of trajectories, the closer the optimal distribution $\bar{p}$ to the uniform distribution, the more complex the traffic. Note that the uniform distribution on $\mathbb{S}$ can be regarded as the limit of $\mathbf{GvM}_N$ distributions as the parameter $\boldsymbol{\lambda}$ goes to zero. We therefore proceed to compute the geodesic distance between $\mathbf{GvM}_N$ distributions in Maksimov form. We recall without proof basic facts that can be found in any textbook dealing with Riemannian geometry.

Proposition 13. *Let (M, g) be a Riemannian manifold. There exists a unique Koszul connection ∇ on TM, called the Levi-Civita connection, such that, for any vector fields X, Y, Z:*

- $\nabla_X Y - \nabla_Y X = [X, Y]$.
- $Zg\left(X, Y\right) = g\left(\nabla_Z X, Y\right) + g\left(X, \nabla_Z Y\right)$.

Proposition 14. *Under the same assumptions as in Proposition 13, if*

$$\gamma \colon [a, b] \to M$$

is a smooth path, its length is defined as:

$$l(\gamma) = \int_a^b g^{1/2} \left(\dot{\gamma}(t), \dot{\gamma}(t) \right) dt. \tag{4.21}$$

Remark 15. *Since the length is invariant by a change of parametrization, one can assume that the domain of γ is the interval $[0, 1]$.*

The geodesic distance between two points p, q on M is defined as

$$\inf \left\{ l(\gamma),\, \gamma(0) = p, \gamma(1) = q \right\}.$$

A path γ realizing the infimum is called a minimizing geodesic.

Proposition 16. *A minimizing geodesic γ satisfies the equation:*

$$\nabla_{\dot{\gamma}} \dot{\gamma} = 0. \tag{4.22}$$

Definition 17. *Let $(e_1, \ldots, e_N)$ be a basis of T_pM. Then:*

$$\nabla_{e_i} e_j = \Gamma_{ij}^k e_k. \tag{4.23}$$

The coefficients Γ_{ij}^k are called the Christoffel symbols of the connection ∇.

Proposition 18. *In coordinates, Equation* (4.22) *becomes:*

$$\ddot{\gamma}^k + \Gamma_{ij}^k \dot{\gamma}^i \dot{\gamma}^j = 0,\ k = 1 \ldots N. \tag{4.24}$$

This proposition gives a practical way of computing a geodesic by integrating an ordinary differential equation. Canned routines solving boundary value problems can be used for that purpose. In the sequel, an explicit form for the Christoffel symbols is given.

Let p_η belong to the exponential family and parameterized by some vector $\eta \in \mathbb{R}^N$:

$$p_\eta(x) = \exp \Big[\langle \eta, x \rangle - \psi(\eta) \Big]. \tag{4.25}$$

Its Fisher information matrix is defined as:

$$g_{i,j}(\eta) = -E \left[\partial^2_{\eta_i \eta_j} l_\eta \right], \tag{4.26}$$

in which $l_\eta(x) := \ln p_\eta$. Using Equation (4.25), it comes:

$$g_{i,j}(\eta) = \partial^2_{\eta_i \eta_j} \psi(\eta). \tag{4.27}$$

In the case of the $\mathbf{GvM}_N$ density, it is convenient to use Maksimov parametrization as in Equation (4.5). Since $\psi = \ln G_0$, the derivatives to be considered are:

$$\partial^2_{\lambda_i \lambda_j} \log G_0 = \frac{\partial^2_{\lambda_i \lambda_j} G_0}{G_0} - \frac{\partial_{\lambda_i} G_0 \cdot \partial_{\lambda_j} G_0}{G_0^2}. \tag{4.28}$$

Exchanging derivative and integral is valid, so it comes:

$$\partial_{\lambda_i} G_0 = \int_{\mathbb{S}} T_i(\theta) \exp \langle \boldsymbol{\lambda}, \mathbf{T}(\theta) \rangle \, \mathrm{d}\theta \tag{4.29}$$

and

$$\partial^2_{\lambda_i \lambda_j} G_0 = \int_{\mathbb{S}} T_i(\theta) T_j(\theta) \exp \langle \boldsymbol{\lambda}, \mathbf{T}(\theta) \rangle \, \mathrm{d}\theta, \tag{4.30}$$

in which $T_{2k-1}(\theta) = \cos k\theta$ and $T_{2k}(\theta) = \sin k\theta$, $k = 1, \ldots, N$.

Introducing the Fourier coefficients of φ, namely

$$a_k(\varphi) = \frac{1}{\pi} \int_{\mathbb{S}} \varphi(\theta) \cos(k\theta) \, \mathrm{d}\theta, \ l \in \mathbb{N},$$

and

$$b_k(\varphi) = \frac{1}{\pi} \int_{\mathbb{S}} \varphi(\theta) \sin(k\theta) \, \mathrm{d}\theta, \ k \in \mathbb{N}^*.$$

we see that $G_0 = \pi a_0(\varphi)$ and that the derivatives are given by

$$\partial_{\lambda_i} G_0(\boldsymbol{\lambda}) = \begin{cases} \pi a_k(\varphi) & \text{if } i = 2k-1 \\ \pi b_k(\varphi) & \text{if } i = 2k \end{cases}$$

and

$$
\begin{aligned}
&\partial^2_{\lambda_i\lambda_j} G_0 \\
&= \begin{cases}
\displaystyle\int_{\mathbb{S}} \frac{\cos(k+l)\theta + \cos(k-l)\theta}{2}\varphi(\theta)\,\mathrm{d}\theta & \text{if } i = 2k-1,\ j = 2l-1, \\
\displaystyle\int_{\mathbb{S}} \frac{\sin(k+l)\theta - \sin(k-l)\theta}{2}\varphi(\theta)\,\mathrm{d}\theta & \text{if } i = 2k-1,\ j = 2l, \\
\displaystyle\int_{\mathbb{S}} \frac{\sin(k+l)\theta + \sin(k-l)\theta}{2}\varphi(\theta)\,\mathrm{d}\theta & \text{if } i = 2k,\ j = 2l-1, \\
\displaystyle\int_{\mathbb{S}} \frac{\cos(k+l)\theta + \cos(k-l)\theta}{2}\varphi(\theta)\,\mathrm{d}\theta & \text{if } i = 2k,\ j = 2l
\end{cases} \\
&= \frac{\pi}{2}\begin{cases}
a_{k+l}(\varphi) + a_{k-l}(\varphi) & \text{if } i = 2k-1,\ j = 2l-1, \\
b_{k+1}(\varphi) - b_{k-l}(\varphi) & \text{if } i = 2k-1,\ j = 2l, \\
b_{k+l}(\varphi) + b_{k-l}(\varphi) & \text{if } i = 2k,\ j = 2l-1, \\
-a_{k+1}(\varphi) + a_{k-l}(\varphi) & \text{if } i = 2k,\ j = 2l.
\end{cases}
\end{aligned}
$$

The Fisher information metric then reads:

$$
\begin{aligned}
g_{ij} &= \frac{\partial^2_{\lambda_i\lambda_j} G_0}{G_0} - \frac{\partial_{\lambda_i} G_0 \cdot \partial_{\lambda_j} G_0}{G_0^2} \\
&= \begin{cases}
\dfrac{1}{2}\dfrac{a_{k+l} + a_{k-l}}{a_0} - \dfrac{a_k a_l}{a_0^2} & \text{if } i = 2k-1,\ j = 2l-1 \\
\dfrac{1}{2}\dfrac{b_{k+l} - b_{k-l}}{a_0} - \dfrac{a_k b_l}{a_0^2} & \text{if } i = 2k-1,\ j = 2l, \\
\dfrac{1}{2}\dfrac{b_{k+l} + b_{k-l}}{a_0} - \dfrac{b_k a_l}{a_0^2} & \text{if } i = 2k,\ j = 2l-1, \\
\dfrac{1}{2}\dfrac{-a_{k+l} + a_{k-l}}{a_0} - \dfrac{b_k b_l}{a_0^2} & \text{if } i = 2k,\ j = 2l
\end{cases}
\end{aligned}
$$

with $k, l \in \{1, \ldots, N\}$, in which we omitted the dependence on φ when writing the Fourier coefficients to simplify the notation.

The Christoffel coefficients needed in the geodesic equation can be com-

puted pretty much the same way using the well-known formula:

$$\Gamma_{ij}^{p} = \frac{1}{2} g^{pl} \left(\frac{\partial g_{li}}{\partial x_j} + \frac{\partial g_{lj}}{\partial x_i} - \frac{\partial g_{ij}}{\partial x_l} \right) \tag{4.31}$$

where the convention of summation on repeated indices is taken and the notation g^{pl} stands for the (p, l) element of the inverse matrix g^{-1}. The third partial derivative of $\ln G_0$ with respect to $\lambda_i, \lambda_j, \lambda_p$ is readily obtained from Equation (4.28):

$$\begin{aligned} \partial^3_{\lambda_i \lambda_j \lambda_p} \log G_0 &= \frac{\partial^3_{\lambda_i \lambda_j \lambda_k} G_0}{G_0} + \frac{\partial_{\lambda_i} G_0 \partial_{\lambda_j} G_0 \partial_{\lambda_p} G_0}{G_0^3} \\ &\quad - \frac{\partial^2_{\lambda_i \lambda_j} G_0 \partial_{\lambda_p} G_0 + \partial^2_{\lambda_p \lambda_i} G_0 \partial_{\lambda_j} G_0 + \partial^2_{\lambda_j \lambda_p} G_0 \partial_{\lambda_i} G_0}{G_0^2}. \end{aligned}$$

All terms in Equation (4.32) are already known except for the first one. Starting with Equation (4.29), all third partial derivatives can be obtained using the Fourier coefficients of φ as indicated below:

$$\begin{aligned} \partial^3 \lambda_{2i} \lambda_{2j} \lambda_{2k} G_0 &= \frac{1}{8} \left(a_{i+j+k} + a_{i+j-k} + a_{i-j+k} + a_{i-j-k} \right), \\ \partial^3 \lambda_{2i} \lambda_{2j} \lambda_{2k+1} G_0 &= \frac{1}{8} \left(b_{i+j+k} - b_{i+j-k} + b_{i-j+k} - b_{i-j-k} \right), \\ \partial^3 \lambda_{2i} \lambda_{2j+1} \lambda_{2k+1} G_0 &= \frac{1}{8} \left(-a_{i+j+k} + a_{i+j-k} + a_{i-j+k} - a_{i-j-k} \right), \\ \partial^3 \lambda_{2i+1} \lambda_{2j+1} \lambda_{2k+1} G_0 &= \frac{1}{8} \left(-b_{i+j+k} + b_{i+j-k} + b_{i-j+k} - b_{i-j-k} \right). \end{aligned}$$

The geodesic distance between two $\mathbf{GvM}_N$ distributions p, q is the infimum of the lengths of smooth paths joining p and q. It is the length of a minimizing geodesic γ with endpoints p, q. Such a curve is the solution of a differential equation with boundary conditions:

$$\begin{cases} \nabla_{\dot\gamma} \dot\gamma = 0 \\ \gamma(0) = p,\ \gamma(1) = q \end{cases} \tag{4.32}$$

where ∇ is the Levi-Civita connection. Expanding γ in λ coordinates, that is $\gamma(t) = \left(\gamma^1(t), \ldots, \gamma^{2N}(t) \right)$, gives:

$$\begin{cases} \ddot\gamma^i(t) + \Gamma^i_{jk} \dot\gamma^j(t) \dot\gamma^k(t) = 0,\ i = 1 \ldots 2N \\ \gamma^i(0) = p^i,\ \gamma^i(1) = q^i,\ i = 1 \ldots 2N. \end{cases} \tag{4.33}$$

with $p = (p^1, \ldots, p^{2N})$, $q = (q^1, \ldots, q^{2N})$ in λ coordinates. Finding the geodetic distance between two $\mathbf{GvM}_N$ distributions can be accomplished numerically by a boundary value problem solver or a shooting algorithm. The computational cost is quite high, but it turns out on some examples that a pretty good approximation is given by the SKL divergence. It was therefore decided to validate the concept using this value instead of the true one, leaving the design of an efficient geodetic algorithm for future developments.

Remark 19. *The asymptotic behavior of the $\boldsymbol{GvM}_N$ distribution in the limit of large $\kappa_i, i = 1 \ldots N$ arguments can be found using Laplace's approximation for integrals* [14]. *Let $f \in C^\infty([0, 2\pi]; \mathbb{R})$ then:*

$$\begin{aligned} &\int_0^{2\pi} f(\theta) \exp \lambda \left(\sum_{i=1}^{N} \kappa_i \cos i \left(\theta - \mu_i \right) \right) d\theta \\ &= \left(\sum_{\eta} \sqrt{\frac{2\pi}{\lambda |Q''(\eta)|}} f(\eta) e^{\lambda Q(\eta)} \right) \left(1 + O(\lambda^{-1}) \right) \end{aligned} \tag{4.34}$$

where Q is the mapping:

$$Q(\theta) = \sum_{i=1}^{N} \kappa_i \cos i \left(\theta - \mu_i \right) \tag{4.35}$$

and η ranges over the roots of Q' such that $Q''(\eta) < 0$, assuming that all roots are of multiplicity one. Introducing the complex polynomial P :

$$P(z) = -\sum_{j=1}^{N} j \kappa_j e^{-ij\mu_j} z^j \tag{4.36}$$

the roots of $Q'(\theta)$ are exactly the intersection points of the closed curve $\gamma \colon t \in [0, 1] \to P\left(e^{i2\pi t}\right)$ with the real axis and there is exactly N of them corresponding to maxima of Q.

Now, using Formula (4.34) *again for the normalizing factor in Equation* (4.3), *it comes:*

$$\lim_{\lambda \to +\infty} E[f]_{\lambda\kappa_1, \ldots, \lambda\kappa_N; \mu_1, \ldots, \mu_N} = \frac{\sum_\xi |Q''|^{-1/2}(\xi) f(\xi)}{\sum_\xi |Q''|^{-1/2}(\xi)} \tag{4.37}$$

where ξ ranges over the η realizing the global maximum of Q. So, in the sense of distributions:

$$\lim_{\lambda \to +\infty} p_{\lambda\kappa_1,\ldots,\lambda\kappa_N;\mu_1,\ldots,\mu_N} = \frac{\sum_\xi |Q''|^{-1/2}(\xi)\delta_\xi}{\sum_\xi |Q''|^{-1/2}(\xi)} \tag{4.38}$$

The extra degree of freedom gained from the introduction of λ calls for a normalization condition. One possible choice is to set $\kappa_N = 1$.

4.3 Validation

4.3.1 Simulation

We conduct testing on a single cell to assess the fundamental functioning of the model. The results from this step provide essential insights into accuracy and basic performance.

We start by performing simulations. We shall:

(a) Generate an angular sample following a probability $p_\circ$;

(b) Compute the corresponding empirical Fourier coefficients;

(c) Compute the Maximum Entropy density that is compatible, in the relaxed setting described above, with our empirical Fourier coefficients.

(d) Compute the $\mathbf{GvM}_N$ distribution corresponding to optimal density.

(e) Compute the SKL divergence between $\mathbf{GvM}_N$ distribution $q(\theta)$ and uniform distribution $u(\theta)$ as follow:

$$\mathrm{SKL}\,(q\|u) = \int_0^{2\pi} u(\theta) \ln \frac{u(\theta)}{q(\theta)}\,\mathrm{d}(\theta) + \int_0^{2\pi} q(\theta) \ln \frac{q(\theta)}{u(\theta)}\,\mathrm{d}\theta. \tag{4.39}$$

This simulation will allow us to get an idea of the sample size that is necessary for obtaining relevant information, merely by comparing the *true* density with the one obtained through the maximization of entropy. And then based on the obtained SKL divergence, we can estimate the complexity.

To better understand this relationship, Table 4.1 presents the SKL divergence between $u(\theta)$ and the $\mathbf{GvM}_N$ $q(\theta)$ with varying numbers of peaks and varying peak widths β. The goal is to understand how these parameters affect the SKL divergence and their implications for calculating complexity maps in ATM.

Table 4.1: The simulation results

Peaks-β	0.01	0.1	0.2	0.3
1	14.93	10.82	9.44	8.76
2	12.92	9.42	8.14	7.45
3	11.96	8.78	7.54	6.95
4	10.65	8.54	7.46	6.79
5	10.33	7.82	6.87	6.49
10	8.80	6.84	6.06	5.75
20	7.31	5.98	5.14	5.06

The simulation results indicate that as the number of peaks of the $\mathbf{GvM}_N$ increases, the SKL divergence between the uniform distribution and the $\mathbf{GvM}_N$ decreases. Specifically, with β=0.01, the SKL divergence decreases dramatically from 14.93 with one peak to 7.31 with 20 peaks. This trend is consistent across all values of β. This can be explained by the fact that increasing the number of peaks makes the $\mathbf{GvM}_N$ resemble the uniform distribution more closely, as multiple peaks create a flatter distribution. In the context of ATM, an increase in the number of peaks corresponds to an increase in the number of aircraft headings. As the number of headings increases, the distribution of aircraft becomes more uniform, reducing the difference between the actual and uniform distributions, thereby increasing air traffic complexity.

Furthermore, the results show that as the value of β increases, the SKL divergence decreases. For instance, increasing the values of β from 0.01 to 0.3, the SKL divergence decreases from 14.93 to 8.76 with one peak and from 7.31 to 5.06 with 20 peaks. A larger value of β causes the peaks of the $\mathbf{GvM}_N$ to become flatter, making the distribution less concentrated and more similar to the uniform distribution (see Figure 4.2). In ATM, increasing the peak width β corresponds to more dispersed aircraft headings.

Special cases, such as when the number of peaks is one and the value of β

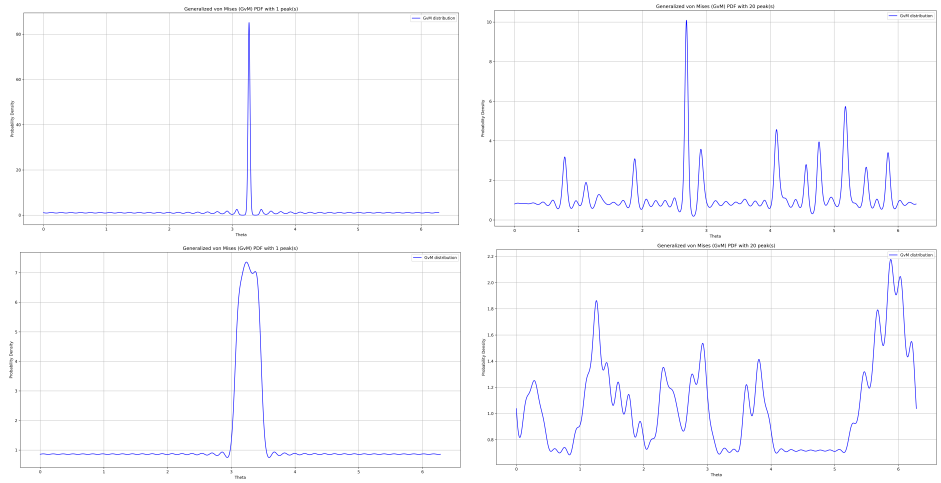


Figure 4.2: The $\mathbf{GvM}_N$ with different β values and peak numbers: Top to bottom, $\beta = 0.01$ and $\beta = 0.3$; Left: 1 peak, Right: 20 peaks.

increases from 0.01 to 0.3, show a significant decrease in SKL divergence from 14.93 to 8.76. This indicates that even with a single peak, making the peak flatter and more spread out (larger β) can increase the complexity. Similarly, with 20 peaks, the SKL divergence is the lowest among all peak numbers across all values of β, emphasizing that multiple peaks in the $\mathbf{GvM}_N$ make it closer to the uniform distribution, especially when β value is also large.

These results not only provide insights into how the parameters of the $\mathbf{GvM}_N$ affect its divergence from the uniform distribution but also lay the foundation for calculating complexity maps in ATM. Understanding these relationships may assist air traffic managers in predicting, planning, and adjusting air traffic flows to ensure optimal safety and efficiency. Making the distribution of headings more concentrated can help reduce air traffic complexity, thereby improving safety and efficiency in air traffic coordination.

4.3.2 Implementation

In the simulation phase, we investigated the behavior of the SKL divergence under varying distributions of flight headings (θ) within individual cells. Each cell has different concentrations and spreads of flight headings. Specifically, we tested the metric on a synthetic dataset where flight headings were dis-

tributed according to the number of peaks (concentration points) and their widths.

Building upon the insights gained from simulations, in order to validate our findings on a real dataset, we proposed a new complexity index, called *Composite Complexity Index*, in short, CCI (see Equation (4.40)). It is weighted by the SKL divergence and the number of flights.

Our analysis utilizes flight data collected from ATL [107]. The dataset provides a snapshot of flight trajectories on a specific date, enabling a detailed examination of spatial distributions and flight path characteristics within the airspace. The heading information from each trajectory is particularly leveraged in our computation of the complexity map, which is based on the SKL divergence determined in our simulation phase. These distances quantify the directional dissimilarities between flight paths, offering insights into the spatial complexities of aircraft movements around ATL.

Composite Complexity Index (CCI)

Local complexity depends on the intensity of air traffic and on the angular distribution of trajectories. Each individual quantity is, of course, insufficient. Intense traffic with a predominantly unidirectional distribution or a spread distribution with low-intensity traffic is not very complex. Therefore, we choose, as a complexity index, a criterion that increases with both the spread of the distribution and the intensity of traffic, named CCI. Since the spread of the distribution is measured using the SKL divergence, it is natural to choose, as a complexity criterion, the ratio

$$\mathrm{CCI} = \frac{n}{a + \mathrm{SKL}\,(q\|u)}, \tag{4.40}$$

in which n is the number of flights through the cell under consideration and $\mathrm{SKL}\,(q\|u)$ is the SKL divergence from q to u. The positive parameter a avoids singularity of the formula in the case where $q = u$. It is chosen empirically, in such a way that the contrast in the complexity map is satisfactory.

Dataset

Based on the dataset described in Subsection 2.5.1 of Chapter 2, we build complexity maps of air traffic by focusing on records of one day of air traffic at

ATL airport on December 26, 2018, which includes 215 flights. The detailed attributes of each trajectory are outlined in Table 2.1 and Table 2.2.

Figure 4.3 illustrates the one-day traffic density of trajectories across ATL on December 26, 2018. The highest number of flights occurs around 13:00 during these times. In contrast, the lowest flight density is observed at around 06:00 to 08:00. There are smaller peaks observed in the afternoon and evening, notably around 18:00 and 21:00. Overall, there is a noticeable decline in flight activity from midnight to early morning, followed by a sharp increase starting at 11:00, peaking in the mid-afternoon, and then stabilizing at a relatively high level until the late evening.

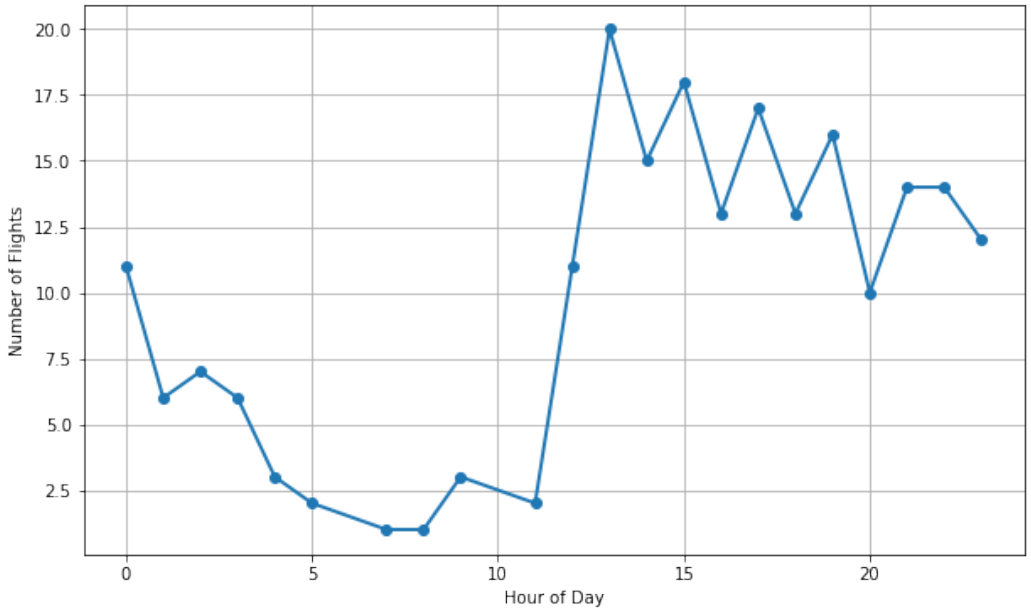


Figure 4.3: Traffic density of trajectories across ATL on December 26, 2018.

All flight trajectories on this day can be seen in Figure 4.4. From the figure, we can observe several prominent, thick lines suggesting popular flight corridors, particularly in the northeast-southwest and northwest-southeast directions.

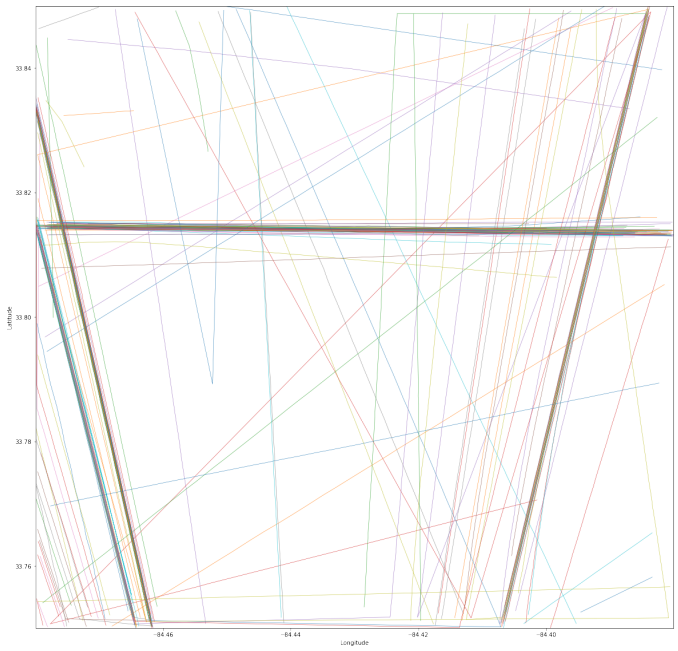


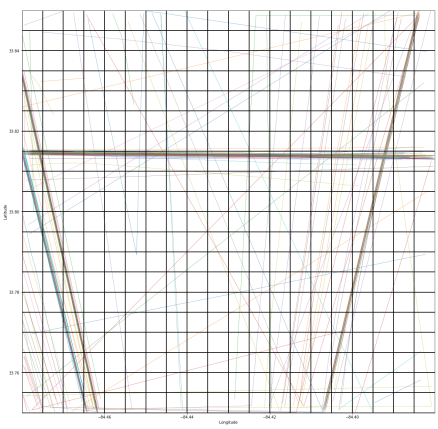
Figure 4.4: All flight trajectories across ATL on December 26, 2018.

Experimental Results

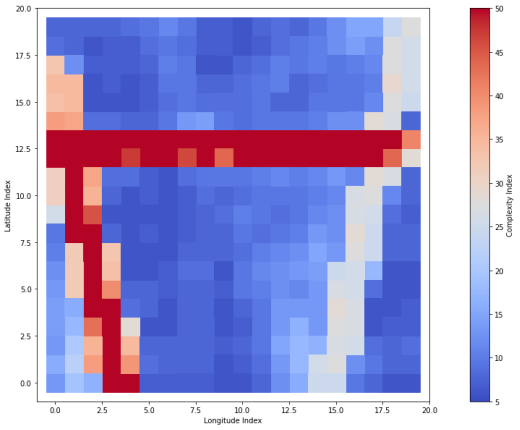
In order to build a complexity map, we determine the radius of the cell, and then we count the number of flights entering in each cell as well as the heading values. After obtaining the values of the SKL divergence (following the steps illustrated in Subsection 4.3.1). Then we employ a color map where blue denotes the lowest complexity indices and red the highest ones. This gradient allows for clear visualization of spatial complexity variations, aiding in immediate interpretation of relative intensity levels across the mapped area.

To analyze the impact of the number of flights and the SKL divergence on the CCI of the airspace, we performed experiments on a dataset from ATL airport described in Subsection 4.3.2. The cell radius was chosen equal to 10 kms. We obtained the complexity maps in Figure 4.5.

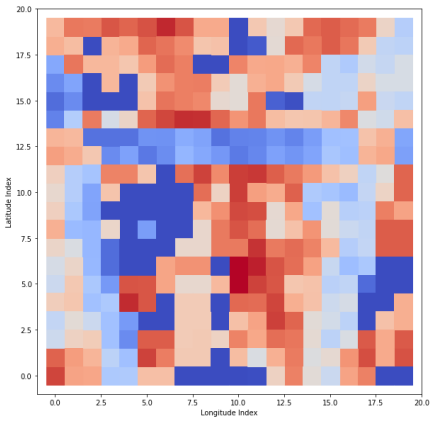
Figure 4.5b presents a complexity map based on the traffic intensity. Row 12 and columns from 2 to 18 show very high complexity: indeed a large number of flights traverse these areas. The regions with the lowest complexity



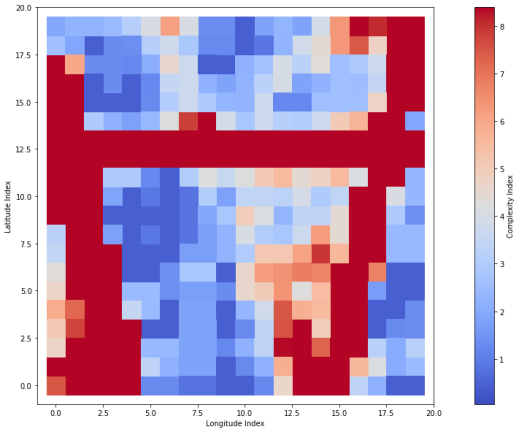
(a) Trajectories



(b) Based on traffic intensity



(c) Based on SKL divergence



(d) Based on CCI

Figure 4.5: Complexity maps on ATL airport dataset.

are rows from 0 to 6 and columns from 6 to 20 having fewer flights.

Figure 4.5c displays a complexity map based on SKL divergence. The rows from 5 to 10 and columns from 7 to 15 show high complexity. Conversely, rows from 0 to 4 and columns from 0 to 6 exhibit lower complexity. This suggests that the central region experiences significant variability in distances between points, while the northwestern region has lower variability.

Figure 4.5d is a composite of the previous two factors, providing a more comprehensive complexity map. The cells at row 12 and columns from 2 to 18 stand out again, indicating high complexity where both SKL divergence and the number of flights are high. The least complex regions are rows from 0 to 6 and columns from 6 to 20, showing that when both factors are low, the overall complexity decreases. This combination offers a thorough understanding of complexity distribution, aiding in identifying areas that require attention in ATM.

Overall, the complexity map clearly indicates that the point of highest complexity is at the intersection of the main flight paths. This position represents the densest air traffic and highest workload for ATCOs. The points with the lowest complexity correspond to minimal workload for controllers.

For the detailed information, we compare the distance between the GvM distribution (blue line) and the uniform distribution (green dashed line) in three scenarios, including high, medium, and low complexities.

While a cell with index (12,1) has high complexity with 103 flights cross this cell, the low complexity index is found in cell (7,5) having only one aircraft flying over. The obtained GvM distributions in 2 cells are illustrated in Figure 4.6 and Figure 4.7, respectively.

Figure 4.6 shows three predominant headings (near π, $\frac{3\pi}{2}$, and 2π). This distribution deviates significantly from the uniform distribution.

The cell (16,16) exhibits a moderate complexity index (see Figure 4.8), the GvM distribution has a significant deviation from the uniform distribution, indicating an average level of complexity in this region. This is illustrated by the mixture of various headings (concentrating on two directions) in the inset diagram, which suggests a moderate level of traffic diversity.

An interesting observation is that although cell (15,7) has only 5 flights concentrated in the three directions close to $\frac{\pi}{6}$, π, and 2π, which increases

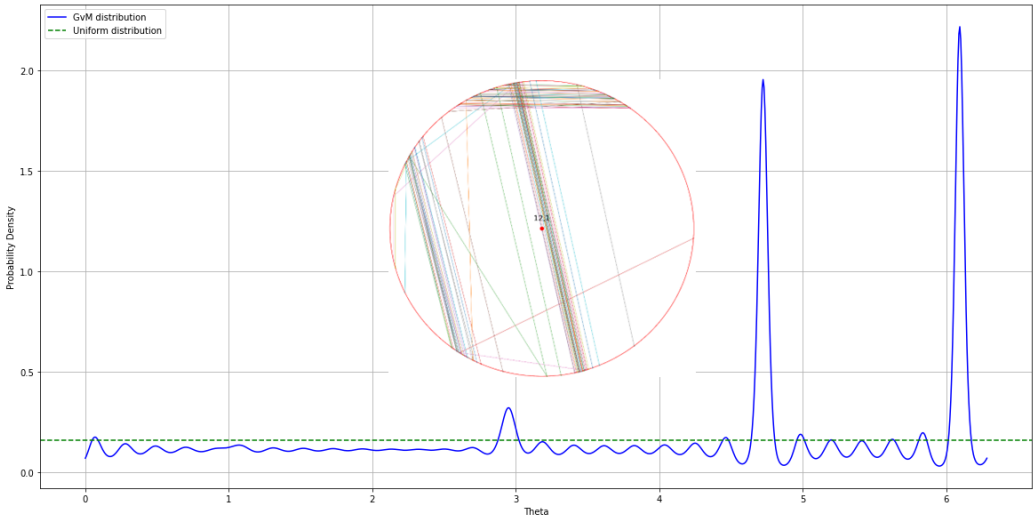


Figure 4.6: High complexity cell: many flights, diverse directions.

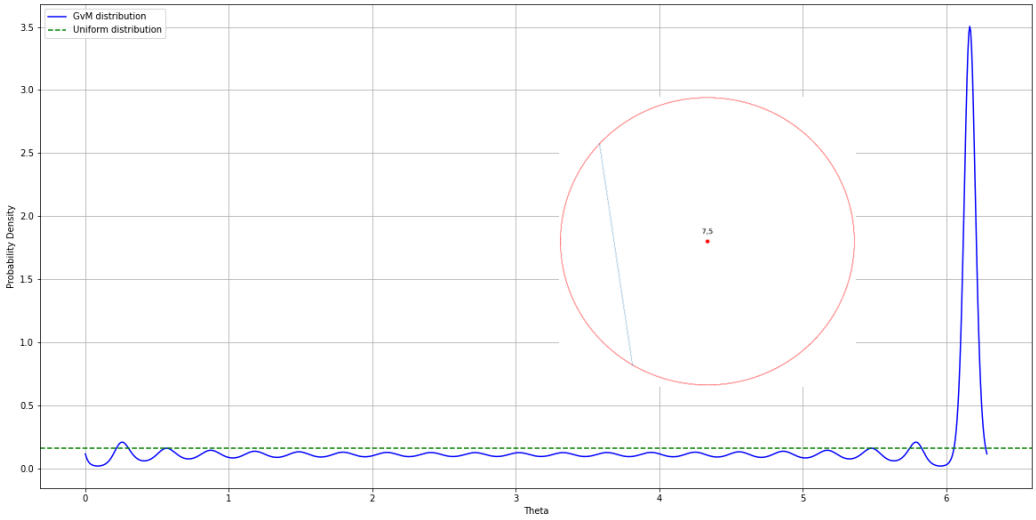


Figure 4.7: Low complexity cell: few flights, few directions.

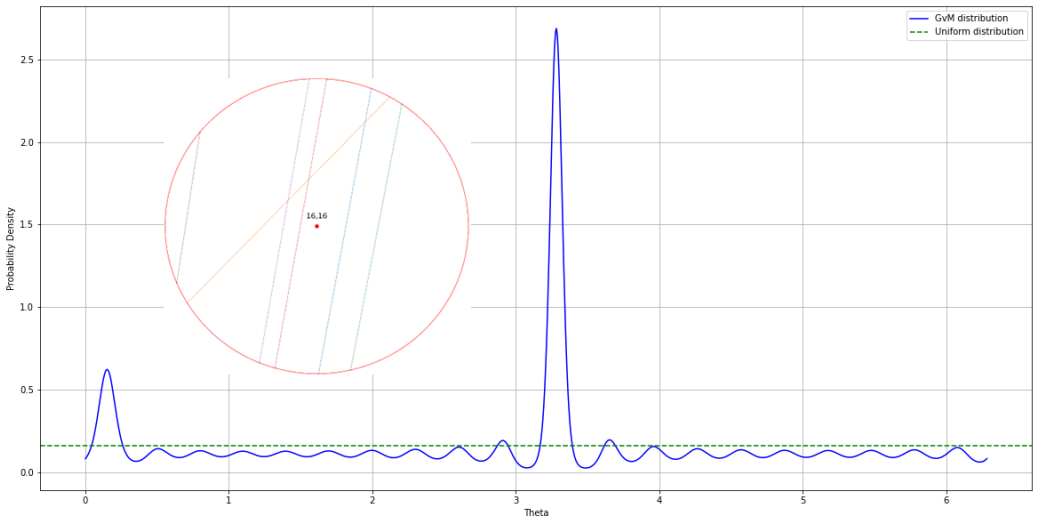


Figure 4.8: Medium complexity cell: moderate flights, moderate directions.

the complexity index, resulting in a high complexity level for this cell (see Figure 4.9).

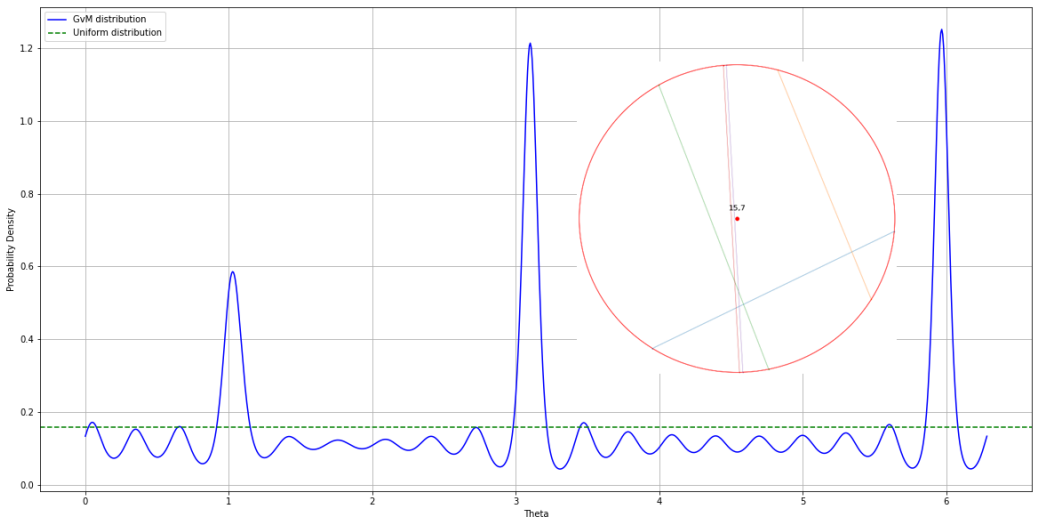


Figure 4.9: High complexity cell: fewer flights, diverse directions.

4.4 Conclusions

We have developed a methodology for assessing the local complexity of air traffic. This strategy is decomposed in three steps: we first estimate, at each point of a Cartesian grid, the probability density of aircraft headings in corresponding cell, using the principle of Maximum Entropy (this was detailed in Chapter 3); then we compute an approximation of the geodesic distance to the uniform distribution, considered for obvious reasons as the distribution of maximum complexity; finally, we define and compute an index of complexity that combines both the angular distribution and the intensity the local traffic. This index of complexity is evaluated at each point of our grid, yielding a complexity map.

The chosen formula for the index of complexity is one choice among others. However, our construction on real data seems to give what one may expect. Other criteria may be obtained by considering the discrepancy between the heading distributions at entry and exit points of each cell. At all events, it appears that our strategy for building complexity maps is both numerically tractable and efficient. The problem of efficient geodesic distance computation will be addressed in a future work.

Chapter 5

Maximum Entropy multivariate density estimation

This chapter explores the computation of multivariate probability distributions using the Maximum Entropy principle. The statistics associated with a family of feature functions provide linear constraints on the unknown density. Discretization of the sought distribution is avoided, and the problem then falls under partially finite convex programming. The solution is expressed through a primal-dual relationship, where the dual variable is obtained by maximizing a concave function of a finite number of variables, whose properties depend on the choice of the relaxation function. Various types of constraint relaxations are considered, and in all cases, solving the dual problem can be achieved using now-classical optimization algorithms. The proposed formalism could, incidentally, optimize entropies other than the Kullback-Leibler one. It extends previous results in which the sampling space was assumed to be finite.

5.1 Introduction

In a number of applications, as in the modeling of animal or vegetal species distributions or in the evaluation of air traffic density, one is facing the problem of estimating a multivariate probability distribution from observed data in the form of samples, which are considered as being drawn independently according to the unknown distribution. In this chapter, we develop a non-parametric approach based on a combination of Maximum Entropy and the method of moments.

In [42], such a problem was considered for discrete finite sample spaces,

and tackled by means of the Maximum Entropy principle leading to a finite dimensional optimization problem under linear moment constraints. Each moment constraint is obtained via some *statistic*, that is, by estimating the expectation of some *feature function*. Such a problem is tackled by means of Fenchel duality, which provides a flexible framework for introducing the relaxation of the above mentioned constraints. A *potential function* encodes the type of relaxation (named regularization in the above reference). While the *equality constraint* (no relaxation at all) is encoded by means of the indicator function of a singleton, box constraints are also considered, as well as least-square relaxation.

The Fenchel duality formalism provides a machinery that contains all cases. Our main purpose in this chapter is to show that the nice framework proposed in [42] can be extented to the more general setting where the sample space is no longer discrete. We therefore consider the optimization of the so-called *differential entropy* or of its *relative* counterpart, the Kullback-Leibler divergence between probability measures.

Such an extended framework relies on extensions of the Fenchel duality theorem, as well as on Rockafellar's theory of convex integral functionals.

5.2 Statement of the problem

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain. We assume throughout that Ω is endowed with the σ-algebra of its Borel sets. The space of Borelian functions on Ω is denoted by $\mathrm{Bor}(\Omega)$.

Points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are observed in Ω. These points are assumed to be drawn according to some unknown probability P on Ω, and the fundamental problem is that of recovering P. Feature functions $f_1, \ldots, f_m \colon \Omega \to \mathbb{R}$ are chosen, from which statistics are derived:

$$\langle f_k \rangle = \frac{1}{n} \sum_{j=1}^{n} f_k(\mathbf{x}_j), \quad k = 1, \ldots, m. \tag{5.1}$$

Recall that the above equation expresses the expectation of f_k under the *empirical measure*

$$P_{\mathrm{emp}} = \frac{1}{n} \sum_{j=1}^{n} \delta_{\mathbf{x}_j},$$

in which $\delta_{\mathbf{x}}$ denotes the Dirac measure at $\mathbf{x}$. The Maximum Entropy principle states, in its native form, that among all probability measures P that are consistent with the moments constraints

$$\langle f_k \rangle = \int f_k \, \mathrm{d}P, \qquad k = 1, \ldots, m, \tag{5.2}$$

one should select the probability $\bar{P}$ which minimizes the Kullback-Leibler relative entropy $\mathscr{K}(P\|P_\circ)$ of P with respect to $P_\circ$. Here, $P_\circ$ is a *reference measure*, supported in Ω, which encodes the prior knowledge one may have about P. If no prior knowledge is available, it is customary to choose $P_\circ$ to be the uniform measure on Ω. Recall that the Kullback-Leibler relative entropy of P with respect to $P_\circ$ is defined as

$$\mathscr{K}(P\|P_\circ) := \begin{cases} \displaystyle\int \ln \frac{\mathrm{d}P}{\mathrm{d}P_\circ} \, \mathrm{d}P & \text{if } P \ll P_\circ, \\ \infty & \text{otherwise.} \end{cases}$$

Here, $\mathrm{d}P/\,\mathrm{d}P_\circ$ denotes the Radon-Nikodym derivative of P with respect to $P_\circ$, and the notation $P \ll P_\circ$ means that P is *absolutely continuous* with respect to $P_\circ$. We are facing the following optimization problem:

$$(\mathscr{P}_\circ) \quad \left| \begin{array}{ll} \text{Minimize} & \mathscr{K}(P\|P_\circ) \\ \text{s.t.} & 1 = \displaystyle\int_\Omega \mathrm{d}P, \; \mathbf{y} = \int_\Omega \mathbf{f} \, \mathrm{d}P. \end{array} \right.$$

The shorthand *s.t.* stands for *subject to*, while $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$ is the vector whose k-th component is equal to $\langle f_k \rangle$ and $\mathbf{f}$ is the $\mathbb{R}^m$-valued function whose k-th component is equal to f_k.

Problem $(\mathscr{P}_\circ)$ is an infinite dimensional convex problem with finitely many linear constraints. Although the optimized variable P lies in the space of Radon measures on Ω, it is equivalent to search for its density with respect to $P_\circ$. As a matter of fact, minimizing the Kullback-Leibler divergence discards measures that are not absolutely continuous with respect to $P_\circ$.

Moreover, since the statistics in Equation (5.1) are inaccurate values of the moments under consideration, the constraints in Equation (5.2) must be relaxed. Following [42] we introduce, for the purpose of relaxation, a convex *potential function* $U \colon \mathbb{R}^m \to (-\infty, \infty]$.

We therefore obtain the relaxed, functional counterpart of Problem $(\mathscr{P}_\circ)$, namely

$$(\mathscr{P}) \quad \left| \begin{array}{ll} \text{Minimize} & H(p) + U(\mathbb{F}p) \\ \text{s.t.} & 1 = \mathbb{I}p, \end{array} \right.$$

in which the variable p is the Radon-Nikodym derivative of the searched measure P with respect to $P_\circ$. The *neg-entropy* functional H is defined by

$$H(p) := \int h_\circ\Big(p(\mathbf{x})\Big)\, \mathrm{d}P_\circ(\mathbf{x}) \quad \text{with} \quad h_\circ(t) := \left\{ \begin{array}{ll} t \ln t & \text{if} \quad t > 0, \\ 0 & \text{if} \quad t = 0, \\ \infty & \text{if} \quad t < 0 \end{array} \right.$$

and the linear operators $\mathbb{F}$ and $\mathbb{I}$ are respectively defined by

$$\mathbb{F}p = \int_\Omega \mathbf{f}(\mathbf{x})\, p(\mathbf{x})\ \mathrm{d}P_\circ(\mathbf{x}) \quad \text{and} \quad \mathbb{I}p = \int_\Omega p(\mathbf{x})\ \mathrm{d}P_\circ(\mathbf{x}).$$

The potential function U is generally a convex non-negative function which attains its minimum value at $\mathbf{y}$. The case of non-relaxed constraint [42] corresponds to the choice $U_\mathbf{y}(\mathbf{s}) = U_\mathbf{y}^{(0)}(\mathbf{s}) = \delta_{\{\mathbf{y}\}}(\mathbf{s})$. Recall that the *indicator function* of a set S is the function

$$\delta_S(\mathbf{s}) := \left\{ \begin{array}{ll} 0 & \text{if } \mathbf{s} \in S, \\ \infty & \text{otherwise.} \end{array} \right.$$

Clearly, this choice enforces the equality $\mathbb{F}p = \mathbf{y}$. In [42], the case of box constraints was also considered. It corresponds to the choice

$$U_\mathbf{y}(\mathbf{s}) = U_\mathbf{y}^{(1)}(\mathbf{s}) = \delta_{B_\mathbf{y}}(\mathbf{s}) \quad \text{in which} \quad B_\mathbf{y} := \prod_{k=1}^{m} [y_k - \beta_k, y_k + \beta_k]$$

with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m) \in (\mathbb{R}_+^*)$. Since the convex conjugate of $\delta_{B_\mathbf{y}}(\cdot)$ fails to be differentiable, one may consider an approximation of the box constraint that yields a differentiable conjugate as in [42, Section 5.2]. This will also be considered in the present chapter. The *least square* relaxation corresponds to the choice

$$U_\mathbf{y}(\mathbf{s}) = U_\mathbf{y}^{(2)}(\mathbf{s}) = \frac{1}{2\alpha} \|\mathbf{y} - \mathbf{s}\|^2,$$

in which $\|\cdot\|$ denotes the Euclidian norm and α is some positive constant. We may also consider, more generally, the choice

$$U(\mathbf{s}) = \frac{1}{2\alpha} \Big\|\mathbf{y} - \mathbf{s}\Big\|_Q^2$$

in which $\|\mathbf{z}\|_Q^2 = \langle \mathbf{z}, \mathbf{z} \rangle_Q = \langle \mathbf{z}, Q\mathbf{z} \rangle$ with Q a positive definite symmetric matrix. An interesting example of this is obtained if we let $Q = \Sigma^{-1}$ with Σ an estimate of the covariance matrix of the vector $\mathbf{y}$: the relaxation then accounts for the variance of each *principal component* of the data vector regarded as a random vector.

Implicitly, the above problem is stated in the space of integrable functions having well-defined integrals against all feature functions. In general, such a space is a subspace of $L^1_{P_\circ}(\Omega)$, the space of measurable functions on Ω that are integrable with respect to $P_\circ$. We note that, in the case where all feature functions are bounded, our natural workspace is $L^1_{P_\circ}(\Omega)$ itself. In general, however, the natural workspace for Problem $(\mathscr{P})$ is a proper subspace of $L^1_{P_\circ}(\Omega)$. The reason for this is that the linear mapping $\mathbb{F}$ is well defined on the space

$$\left\{ p \in L^1_{P_\circ}(\Omega) \;\middle|\; pf_k \in L^1_{P_\circ}(\Omega),\ k = 1, \ldots, m \right\}.$$

Generally speaking, the implicit workspace involved in any moment problem that involves feature functions taken in a set Φ is the *Köthe space*

$$L_{P_\circ}(\Omega; \Phi) := \left\{ p \in \mathrm{Bor}(\Omega) \;\middle|\; \forall f \in \Phi,\ pf \in L^1_{P_\circ}(\Omega) \right\}. \tag{5.3}$$

In the next section, we shall explore the dualization of problem $(\mathscr{P})$ and show how this dualization provides algorithms for effective computation of Maximum Entropy solutions in our infinite dimensional setting, without having to discretize the statistical problem under consideration.

5.3 Maximum Entropy solutions

We now consider the dual approach for solving Probem $(\mathscr{P})$. The cornerstone of the analysis is an extension of Fenchel's duality theorem for which the *primal variable* lies in an infinite dimensional space. For convenience, we define on $\mathbb{R} \times \mathbb{R}^m$ the function

$$G(\eta_\circ, \boldsymbol{\eta}) := -\delta_{\{1\}}(\eta_\circ) - U(\boldsymbol{\eta})$$

and rewrite $(\mathscr{P})$ is the equivalent form

$$(\mathscr{P}) \quad \text{Minimize } \ H(p) - G(\mathbb{A}p),$$

in which $\mathbb{A}p := (\mathbb{I}p, \mathbb{F}p)$.

Theorem 20 (Fenchel duality)**.** *Suppose we are given L and $L^\star$ two vector spaces, algebraically paired by the bilinear mapping $\langle\cdot,\cdot\rangle$; $\mathbb{A}\colon L \to \mathbb{R}^d$, a linear mapping and $\mathbb{A}^\star\colon \mathbb{R}^d \to L^\star$, its adjoint; $H\colon L \to (-\infty,\infty]$, a proper convex functional and $H^\star\colon L^\star \to (-\infty,\infty]$, its convex conjugate; $G\colon \mathbb{R}^d \to [-\infty,\infty)$, a proper concave function and $G_\star\colon \mathbb{R}^d \to [-\infty,\infty)$, its concave conjugate. If the constraint qualification condition*

$$(QC) \quad \operatorname{ri}(\mathbb{A}\operatorname{dom} H) \cap \operatorname{ri}(\operatorname{dom} G) \neq \emptyset$$

is satisfied, then

$$\inf_{p\in L}\Big\{H(p) - G(\mathbb{A}p)\Big\} = \max_{\boldsymbol{\lambda}\in\mathbb{R}^d}\Big\{G_\star(\boldsymbol{\lambda}) - H^\star(\mathbb{A}^\star\boldsymbol{\lambda})\Big\}.$$

The notation $\operatorname{ri} S$ stands for the relative interior of the set S, that is, the interior of S in the *induced topology* of its affine hull. The effective domain $\operatorname{dom} H$ of a convex function H is the set of points p for which $H(p) < \infty$, and the effective domain $\operatorname{dom} G$ of a concave function is the set of points η for which $G(\eta) > -\infty$. The function $G_\star$ is the *concave conjugate* of G, while $H^\star$ denotes the *convex conjugate* of H. The reader unfamiliar with convex analysis is invited to refer to [100, 55, 18].

Theorem 20 establishes the equality between the optimal value in the *primal problem* (the minimization of the *primal function* $H - G\circ\mathbb{A}$) and that of the *dual problem* (the maximization of the *dual function* $G_\star - H^\star\circ\mathbb{A}^\star$). It also says that the dual problem is *attained*, meaning that the dual problem has at least one solution.

In order to tackle our generic Maximum Entropy problem, it remains to:

1. find condition for existence of a primal solution;
2. compute the conjugate of H, a highly non-trivial thing when H is an integral functional;
3. link the possible primal solution to a dual solution.

Theorem 21 (Primal attainment)**.** *Together with the notation and assumptions of Theorem 20, suppose that*

$$(QC^\star) \quad \operatorname{ri}\operatorname{dom} G_\star \cap \operatorname{ri}\operatorname{dom}(H^\star\circ\mathbb{A}^\star) \neq \emptyset.$$

Suppose in addition that

(a) $H^{\star\star} = H$ *and* $G_{\star\star} = G$*;*

(b) *there exists* $\bar{\boldsymbol{\lambda}}$ *maximizing the dual function and* $\bar{p} \in \partial H^{\star}(\mathbb{A}^{\star}\bar{\boldsymbol{\lambda}})$ *such that* $H^{\star} \circ \mathbb{A}^{\star}$ *has gradient* $\mathbb{A}\bar{p}$ *at* $\bar{\boldsymbol{\lambda}}$*.*

Then $\bar{p}$ *minimizes the primal function.*

The notation $\partial F(x)$ stands for the *subdifferential* of the function F at x. The condition $G_{\star\star} = G$ is satisfied whenever G is an upper-continuous proper concave function, a mild assumption satisfied by all examples of interest. The condition $H^{\star\star} = H$ is more difficult to verify when H is an integral functional defined on a functional space such as $L_{P_\circ}(\Omega; \Phi)$.

We owe to Rockafellar a series of works [103, 101, 102] allowing us to compute the convex conjugate of an integral functional in a rather simple way. Under certain conditions (concerning the integrand h of H but also the functional workspace) it is possible to *conjugate through integral sign*, that is to say, to obtain the conjugate of the integral as the integral with conjugate integrand.

An integral functional on Ω is a functional of the form

$$H(p) = \int_\Omega h(p(\mathbf{x}), \mathbf{x}) \, \mathrm{d}P_\circ(\mathbf{x}), \quad p \in L.$$

Here, we endow Ω with a σ-algebra and with a positive σ-finite measure $P_\circ$, and L is a space of measurable functions.

We note that the dependence of h in its second argument could be ignored here, since we focus on the case where $h(p(\mathbf{x}), \mathbf{x}) = h_\circ(p(\mathbf{x}))$.

A space L of measurable functions is *decomposable* if it is stable under bounded alteration on sets of finite measure, that is to say, if for every $p \in L$, every set T such that $P_\circ(T)$ is finite and every measurable function q that is bounded on T, the function

$$\tilde{p}(\mathbf{x}) := \begin{cases} p(\mathbf{x}) & \text{if } \mathbf{x} \in \Omega \setminus T, \\ q(\mathbf{x}) & \text{if } \mathbf{x} \in T \end{cases}$$

belongs to L. In the context of this chapter, we are mostly interested in the vector space $L = L_{P_\circ}(\Omega; \Phi)$ with $\Phi = \left\{\mathbb{1}_\Omega, f_1, \ldots, f_m\right\}$, where $\mathbb{1}_\Omega$ denotes the function identically equal to 1 on Ω.

Theorem 22 (Conjugacy through the integral sign)**.** *Let L and $L^\star$ be spaces of measurable functions on Ω paired by means of the integral bilinear form*

$$\langle p, \phi \rangle = \int_\Omega p(\mathbf{x})\phi(\mathbf{x}) \, \mathrm{d}P_\circ(\mathbf{x}).$$

Let $h\colon \mathbb{R} \times \Omega$ be a measurable integrand that is proper convex and lower semi-continuous in its first argument, and let H be the functional of integrand h:

$$H(p) = \int_\Omega h\Big(p(\mathbf{x}), \mathbf{x}\Big) \, \mathrm{d}P_\circ(\mathbf{x}), \quad p \in L.$$

Assume that L is decomposable and that H has nonempty effective domain. Then

$$H^\star(\phi) = \int h^\star(\phi(\mathbf{x}), \mathbf{x}) \, \mathrm{d}P_\circ(\mathbf{x})$$

for every $\phi \in L^\star$, in which $h^\star(\cdot, \mathbf{x}) := \Big(h(\cdot, \mathbf{x})\Big)^\star$. Moreover, $H^\star$ is convex on $L^\star$.

In our context, L and $L^\star$ are respectively the Köthe space $L_{P_\circ}(\Omega; \Phi)$, as defined in Equation (5.3), and its *dual* Köthe space

$$L^\star_{P_\circ}(\Omega; \Phi) := \Big\{ \varphi \in \mathrm{Bor}(\Omega) \;\Big|\; \forall p \in L_{P_\circ}(\Omega; \Phi),\ p\varphi \in L^1_{P_\circ}(\Omega) \Big\},$$

with $\Phi = \Big\{\mathbb{1}_\Omega, f_1, \ldots, f_m\Big\}$. It was shown that if a family Φ of *feature functions* is such that every $f \in \Phi$ is integrable over every set of finite measure, then $L_{P_\circ}(\Omega; \Phi)$ is decomposable, an moreover that if $\mathbb{1}_\Omega \in \Phi$, then $L^\star_{P_\circ}(\Omega; \Phi)$ is also decomposable. See Lemma 1 and Proposition 1 in [78].

The entropy kernel $h(t, \mathbf{x}) \equiv h_\circ(t)$ is clearly a measurable integrand that is proper convex and lower semi-continuous in its first argument. Its convex conjugate is given by

$$h^\star_\circ(\tau) = \exp(\tau - 1), \quad \tau \in \mathbb{R}.$$

Since our pair of functional spaces $L_{P_\circ}(\Omega; \Phi), L^\star_{P_\circ}(\Omega; \Phi)$ satisfies the requirements of Theorem 22, it comes that, for every $\varphi \in L^\star_{P_\circ}(\Omega; \Phi)$,

$$H^\star(\varphi) = \int_\Omega \exp\Big(\varphi(\mathbf{x}) - 1\Big) \, \mathrm{d}P_\circ(\mathbf{x}).$$

Now, $h_\circ^{\star\star} = h_\circ$ since $h_\circ$ is proper convex and lower semi-continuous. Moreover, it is readily seen that

$$L_{P_\circ}^{\star\star}(\Omega;\Phi) := \Big\{ p \in \mathrm{Bor}(\Omega) \;\Big|\; \forall \varphi \in L_{P_\circ}^{\star}(\Omega;\Phi),\ p\varphi \in L_{P_\circ}^{1}(\Omega) \Big\} = L_{P_\circ}(\Omega;\Phi).$$

See Property (f) page 207 in [78]. It follows that conjugacy through the integral sign applies one more time: for all $p \in L_{P_\circ}(\Omega;\Phi)$,

$$H^{\star\star}(p) = \int_\Omega h_\circ^{\star\star}\Big(p(\mathbf{x})\Big)\ \mathrm{d}P_\circ(\mathbf{x}) = H(p).$$

Theorem 23 (Primal-dual relationship)**.** *Together with the notation and assumptions of Theorem 20, assume that H is an integral functional satisfying the assumptions of Theorem 22. Assume further that G is proper concave and upper semi-continuous (so that it equals its concave bi-conjugate) and that the effective domain of the dual function $D := G_\star - H^\star \circ \mathbb{A}^\star$ has nonempty interior. Assume finally that the conjugate integrand $h^\star$ is differentiable over $\mathbb{R}$, and that there exists some dual-optimal vector $\bar{\boldsymbol{\lambda}}$ in $\operatorname{int}\operatorname{dom} D$. If*

$$\bar{p}(\mathbf{x}) := h^{\star\prime}\Big([\mathbb{A}^\star\bar{\boldsymbol{\lambda}}](\mathbf{x}), \mathbf{x}\Big) \in L,$$

then $\bar{p}$ is a primal solution.

Let us now get back to our Maximum Entropy problem. Most potential functions U are of the form $U(\boldsymbol{\eta}) = V(\mathbf{y} - \boldsymbol{\eta})$ with V a proper convex lower semi-continuous (and even) function. For example, the standard least-square relaxation corresponds to the case where

$$V(\boldsymbol{\eta}') = \frac{1}{2\alpha}\left\|\boldsymbol{\eta}'\right\|^2, \quad \boldsymbol{\eta}' \in \mathbb{R}^m.$$

An easy computation shows that, in this case,

$$G_\star(\lambda_\circ, \boldsymbol{\lambda}) = \lambda_\circ + \langle \mathbf{y}, \boldsymbol{\lambda}\rangle - V^\star(\boldsymbol{\lambda}).$$

The dual problem to Problem ($\mathscr{P}$) then consists in maximizing the function

$$D(\lambda_\circ, \boldsymbol{\lambda}) := \lambda_\circ + \langle \mathbf{y}, \boldsymbol{\lambda}\rangle - V^\star(\boldsymbol{\lambda}) - \exp(\lambda_\circ - 1)\int_\Omega \exp\langle \boldsymbol{\lambda}, \mathbf{f}(\mathbf{x})\rangle\ \mathrm{d}P_\circ(\mathbf{x}), \quad (5.4)$$

in which we account for conjugacy through the integral sign (Theorem 22). We note that the effective domain of D coincides with that of $V^\star$ whenever the feature function $\mathbf{f}$ has bounded components.

Provided the *qualification of constraints* (QC) and its dual counterpart ($QC^\star$) are satisfied, Theorem 23 tells us that the solution to the Problem ($\mathscr{P}$) is given by

$$\bar{p}(\mathbf{x}) = \exp\left[\bar{\lambda}_\circ - 1 + \left\langle \bar{\boldsymbol{\lambda}}, \mathbf{f}(\mathbf{x}) \right\rangle\right] = \exp(\bar{\lambda}_\circ - 1) \cdot \exp\left\langle \bar{\boldsymbol{\lambda}}, \mathbf{f}(\mathbf{x}) \right\rangle,$$

in which $(\bar{\lambda}_\circ, \bar{\boldsymbol{\lambda}})$ maximizes the dual function in Equation (5.4). Applying Fermat's principle, we find that $(\bar{\lambda}_\circ, \bar{\boldsymbol{\lambda}})$ must satisfy the condition

$$(0, \mathbf{0}) \in \partial D(\bar{\lambda}_\circ, \bar{\boldsymbol{\lambda}}).$$

Here, the subdifferential is taken is the sense of concavity. It is readily checked that the latter condition is equivalent to

$$\exp(\bar{\lambda}_\circ - 1) = \left(\int_\Omega \exp\left\langle \bar{\boldsymbol{\lambda}}, \mathbf{f}(\mathbf{x}) \right\rangle \, \mathrm{d}\mathbf{x}\right)^{-1} \quad \text{and} \quad \mathbf{0} \in \partial \tilde{D}(\bar{\boldsymbol{\lambda}}),$$

in which

$$\tilde{D}(\boldsymbol{\lambda}) := \langle \boldsymbol{\lambda}, \mathbf{y} \rangle - V^\star(\boldsymbol{\lambda}) - \ln \int_\Omega \exp \langle \boldsymbol{\lambda}, \mathbf{f}(\mathbf{x}) \rangle \, \mathrm{d}\mathbf{x}.$$

Therefore, maximizing the dual function in Equation (5.4) boils down to maximizing the function $\tilde{D}$. We observe that $\tilde{D}$ is concave a upper semi-continuous, as a consequence of the convexity and lower semi-continuity of the log-Laplace transform. We note that evaluating $\tilde{D}(\boldsymbol{\lambda})$ relies on some numerical integration in dimension m. This is by no means a limitation, due to the existence of powerful numerical integration algorithms such as Monte Carlo integration, for example.

Let us summarize the above achievements. Provided one can check Conditions (QC) and ($QC^\star$), which depend on the choice of the potential V and on that of the feature functions $f_1, \ldots, f_m$, we compute the optimal vector $\bar{\boldsymbol{\lambda}}$ by maximizing $\tilde{D}$, and we then obtain the Maximum Entropy solution $\bar{p}$ via the formula

$$\bar{p}(\mathbf{x}) = \frac{\exp\left\langle \bar{\boldsymbol{\lambda}}, \mathbf{f}(\mathbf{x}) \right\rangle}{\displaystyle\int_\Omega \exp\left\langle \bar{\boldsymbol{\lambda}}, \mathbf{f}(\mathbf{x}) \right\rangle \, \mathrm{d}\mathbf{x}}, \quad \mathbf{x} \in \Omega.$$

Some potentials V give rise to nice differentiable concave unconstrained maximization problems. This happens whenever the conjugate $V^\star$ has full domain and is differentiable. A standard example is provided by the least square relaxation. However, in general, the function $V^\star$ may suffer from

drawbacks such as failing to have full domain (so that the dual problem is constrained) or being non-differentiable. A standard example is provided by the case of box constraints.

In the smooth unconstrained case, quasi-Newtonian methods such as the BFGS algorithm are expected to provide fast and accurate solution. If $V^\star$ is not differentiable, then the situation is not hopeless at all, thanks to the *proximal gradient algorithm*. As a matter of fact, in the latter case, the function to be maximized is then a *composite* concave function, that is to say, a sum of two concave functions one being differentiable, the other being non-differentiable. The reader interested in the optimization of composite functions and the proximal gradient algorithm may refer to [12, 115] and the references therein.

5.4 Relaxation

Following [42], we explore a number of potential functions to be used for the reconstruction of multivariate distributions. We assume throughout that the domain of $H^\star \circ \mathbb{A}^\star$ is nonempty, a mild assumption on the family of feature functions.

No relaxation at all. In this case $V(\boldsymbol{\eta}) = \delta_{\{\mathbf{0}\}}(\boldsymbol{\eta})$. The convex conjugate $V^\star(\boldsymbol{\lambda})$ is then the function identically equal to zero. Condition (QC) reads $\mathbf{y} \in \mathrm{ri}(\mathbb{F}\,\mathrm{dom}\,H)$ is this case, and Condition $(QC^\star)$ is automatically satisfied. The reason is that the effective domain of $G_\star$ then is the whole of $\mathbb{R}^m$, and that $\mathrm{ri}\,\mathrm{dom}(H^\star \circ \mathbb{A}^\star)$ is nonempty, as the relative interior of a nonempty convex set.

Box constraints. The function V is defined in this case by

$$V(\boldsymbol{\eta}) = \delta_{\prod_{k=1}^n [-\beta_k, \beta_k]}(\boldsymbol{\eta}) = \sum_{k=1}^n \delta_{[-\beta_k, \beta_k]}(\eta_k).$$

Its convex conjugate is then given by

$$V^\star(\boldsymbol{\lambda}) = \sum_{k=1}^n \beta_k |\,\lambda_k\,| = \left\| \boldsymbol{\beta} \odot \boldsymbol{\lambda} \right\|_1,$$

in which $\odot$ denotes the pointwise product. In this case, Condition (QC) reads

$$\left(\prod_{k=1}^{m}(y_k-\beta_k,y_k+\beta_k)\right)\bigcap \mathrm{ri}(\mathbb{F}\,\mathrm{dom}\,H)\neq\emptyset.$$

Condition $(QC^\star)$ is, again, automatically satisfied. Since the corresponding function $\tilde{D}$ is a composite function, the maximization of $\tilde{D}$ will be efficiently performed by the proximal gradient algorithm.

Smoothed inner approximation of box constraints. Let $\psi_\beta\colon \mathbb{R}\to(-\infty,\infty]$ be defined by

$$\psi_\beta(\eta)=\begin{cases}\dfrac{\eta}{2}\ln\dfrac{\beta+\eta}{\beta-\eta}+\beta\ln\sqrt{1-\dfrac{\eta^2}{\beta^2}} & \text{if } \eta\in(-\beta,\beta),\\ \infty & \text{otherwise.}\end{cases}$$

The smooth inner approximation of box constraints with the same box as in the previous paragraph, and its convex conjugate, are respectively given by

$$V(\boldsymbol{\eta})=\sum_{k=1}^{m}\varepsilon_k\psi_{\beta_k}(\eta_k)\quad\text{and}\quad V^\star(\boldsymbol{\lambda})=\sum_{k=1}^{m}\beta_k\varepsilon_k\ln\mathrm{ch}\left(\frac{\lambda_k}{\varepsilon_k}\right),$$

in which ε_k controls the proximity of $V(\boldsymbol{\eta})$ to the target box constraint potential and, thereby, the proximity of the corresponding conjugate functions. Figure 5.1 gives an idea of the approximation of the box constraint potential and its conjugate. We note that the constraint qualification condition and its dual counterpart are exactly as in the box constraint case. The advantage here is that the dual function is differentiable, allowing the use of some powerful quasi-Newtonian optimization method, such as BFGS.

Least squares. Given a symmetric positive definite matrix Q, the weighted least square relaxation corresponds to the case where

$$V(\boldsymbol{\eta})=\frac{1}{2\alpha}\left\|\eta\right\|_{Q^{-1}}^2\quad\text{and}\quad V^\star(\boldsymbol{\lambda})=\frac{\alpha}{2}\left\|\boldsymbol{\eta}\right\|_{Q}^2.$$

The standard least square relaxation is obtained by letting $Q=I$, the identity matrix. If the covariance matrix Σ of the data vector, regarded as a

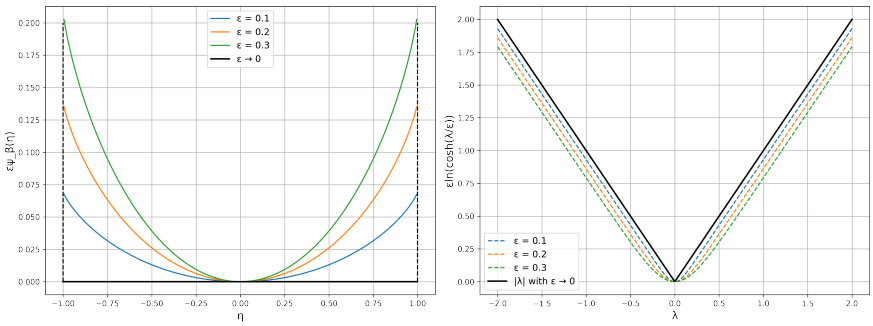


Figure 5.1: Plots of $\varepsilon\psi_\beta(\eta)$ and its conjugate for $\beta = 1$ and various values of ε.

random vector, is available or estimated, we may choose $Q = \Sigma$. The corresponding relaxation then entails the square Mahalanobis distance between the data $\mathbf{y}$ and the moment vector $\mathbb{F}p$. It is remarkable that, one more time, things look nicer on the dual side: no matrix inversion is required to evaluate $\tilde{D}(\boldsymbol{\lambda})$. Finally, the constraint qualification and its dual counterpart are automatically fulfilled here.

5.5 Conclusion

In this chapter, we studied the Maximum Entropy approach to the problem of reconstructing a spatial density from a finite statistical sample. The moment constraints associated with the choice of feature functions are relaxed using potential functions, as in [42].

The dualization of the problem allows the calculation of the solution density through the maximization of a dual problem in finite dimension, whose properties depend on the chosen potential function. Partially finite convex programming, as well as conjugacy of convex integral functionals, constitute the major ingredients of the proposed developments. In all cases, the dual problem is tractable, and the primal-dual relationship allows obtaining the solution.

The tools developed in this chapter will make it possible to solve the problem of reconstructing spatial densities without resorting to discretization of the problem. This is a definite advantage, since the choice of a

discretization introduces a certain bias, which is contrary to the spirit of the Maximum Entropy method.

The choice of a family of feature functions remains a crucial question in this field. The authors of this chapter are currently focusing on this issue, which will be the subject of a subsequent publication.

Chapter 6

Conclusion and further work

6.1 Conclusion

In the first part of this thesis (Chapter 2), we concentrated on applying Bayesian inference to improve flight trajectory predictions: Bayesian inference provides a powerful framework for incorporating prior knowledge and updating predictions based on new information. This allows for a more nuanced and probabilistic approach to trajectory prediction. By leveraging Bayesian methods, we can better estimate the uncertainties inherent in trajectory predictions and develop more effective strategies for managing air traffic in increasingly congested airspace. Our BayesCoordLSTM model, which integrates Bayesian optimization and coordinate transformation into a hybrid ConvLSTM framework, has demonstrated superior performance in predicting flight paths with higher accuracy and efficiency compared to existing methods.

In the second part (Chapters 3 to 5), we focused on analyzing air traffic complexity using concepts derived from information theory. We developed a method to construct air traffic complexity maps in two parts: first, at each point in the airspace, we estimate the angular probability density of the headings of aircraft entering a circular cell centered at the considered point, and then we evaluate the local complexity.

The first part, which is the subject of the article presented in Chapter 3, uses the Maximum Entropy principle combined with the method of moments. Fourier coefficients are estimated by empirical averages from trajectory data over a fixed time window, and an angular density maximizing the entropy while respecting the estimated moments is obtained via partially finite convex programming. The result is a GvM (Generalized von Mises) density.

The second part involves calculating a complexity index associated with the obtained GvM density and the intensity of the flow traversing the cell. The complexity associated with a given GvM is defined as a decreasing function of its geodesic distance to the uniform distribution. Calculating this distance is theoretically possible but quite complex, so we approximate the geodesic distance by the SKL (Symmetrized Kullback-Leibler) divergence. A complexity map can then be constructed by multiplying the previous index by the local traffic density.

Currently, the local traffic density is simply estimated by counting the number of aircraft crossing the considered cell. However, we questioned the notion of traffic density. A more instantaneous notion than the one used in the approach described here could make sense in air traffic management. This motivated the writing of the article presented in Chapter 5. In this article, we consider a spatial distribution of points (which could be the positions of aircraft at a given moment) as resulting from a random draw according to an unknown law. We then propose an estimation of this law using the Maximum Entropy principle combined with the method of moments. Again, partially finite convex programming forms the theoretical core of the study.

6.2 Further work

Building on the foundations laid by this research, several avenues for future work can be pursued to further enhance the capabilities and applications of our methodologies:

Take into account weather conditions for flight trajectory prediction: Future research should focus on developing techniques for dynamic adaptation of hyperparameters in the BayesCoordLSTM model. This could involve real-time adjustments based on changing flight scenarios and weather conditions, improving the model's adaptability and real-time predictive accuracy.

Consider discrepancies between heading distributions at entry and exit points of cells: Exploring different criteria for defining and measuring air traffic complexity could enhance our understanding and improve the accuracy of complexity maps. By comparing aircraft density at the entry and exit points of a region (cell S), we can identify if aircraft are changing their paths within that area. A significant difference in these densities suggests alter-

ations in trajectories, indicating unpredictability and increasing the overall complexity of the air traffic system.

Efficient geodesic distance computation: Future work could investigate on optimizing algorithms for faster geodesic distance computation, incorporating real-time data for dynamic updates, and leveraging parallel processing to handle larger datasets. Exploring these areas may enhance both accuracy and efficiency in practical applications.

This research has established a robust framework for predicting flight trajectories and assessing air traffic complexity. As air traffic volumes increase, the deployment of advanced analytical tools such as those developed in this study will become increasingly critical. By continuing to refine and expand upon our methodologies, we can contribute to the ongoing efforts to enhance the safety, efficiency, and effectiveness of air traffic management systems worldwide.

References

[1] Abbasimehr, Hossein and Paki, Reza. "Improving time series forecasting using LSTM and attention models". In: *Journal of Ambient Intelligence and Humanized Computing* 13.1 (2022), pp. 673–691. ISSN: 1868-5137, 1868-5145.

[2] Abdar, Moloud et al. "A review of uncertainty quantification in deep learning: Techniques, applications and challenges". In: *Information fusion* 76 (2021), pp. 243–297.

[3] Adrían, García and Daniel, Delahaye and Manuel, Soler. "Air traffic complexity map based on linear dynamical systems". In: *hal-02512103* (2019), pp. 1–35.

[4] Ahmed, Mohammed A and El-Sheimy, Naser and El-Bakry, Hosam and Moussa, Ahmed. "Machine learning-based prediction of air traffic complexity and delay". In: *Journal of Air Transport Management* 77 (2019), pp. 97–109.

[5] Alharbi, Abdulrahman and Petrunin, Ivan and Panagiotakopoulos, Dimitrios. "Deep learning architecture for UAV traffic-density prediction". In: *Drones* 7 (2023), pp. 1–31.

[6] Amari, Shun-ichi. *Information geometry and Its applications*. Applied Mathematical Sciences. Springer Japan, 2016. ISBN: 9784431559788.

[7] Amari, Shun-ichi and Karakida, Ryo and Oizumi, Masafumi. "Information geometry connecting Wasserstein distance and Kullback–Leibler divergence via the entropy-relaxed transportation problem". In: *Information Geometry* 1 (2018), pp. 13–37.

[8] Amersfoort, Joost and Smith, Lewis and Teh, Yee and Gal, Yarin. "Simple and scalable epistemic uncertainty estimation using a single deep deterministic neural network". In: *ICML*. 2020.

[9] Arribas, Ignacio and Rodas-Jordá, Alberto and Toro-Rueda, Jose M del and Parés, Ferran. "Evaluation of air traffic complexity with geometric metrics and machine learning techniques". In: *IEEE Access* 9 (2021), pp. 40894–40905.

[10] Atencia, Miguel and Stoean, Ruxandra and Joya, Gonzalo. "Uncertainty quantification through dropout in time series prediction by echo state networks". In: *Mathematics* 8.8 (2020). ISSN: 2227–7390.

[11] Banavar, Sridhar and Sheth, Kapil S. and Grabbe, Shon. "Airspace complexity and its application in air traffic management". In: *2nd USA/EUROPE Air traffic management R&D seminar* (1998).

[12] Beck, Amir. *First-order methods in optimization*. SIAM, 2017.

[13] Behmardi, Behrouz and Raich, Raviv and Hero, Alfred O. "Entropy estimation using the principle of maximum entropy". In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2011, pp. 2008–2011.

[14] Bender, Steven R and Smith, Andrew E. "A geometric airspace complexity metric for unmanned aerial systems". In: *Journal of Intelligent & Robotic Systems* 83.1 (2016), pp. 121–133.

[15] Bing, Dong. "Aircrafts monitoring and early warning based on air traffic complexity measure analysis". In: *The Open Automation and Control Systems Journal* 6 (2014), pp. 1563–1569.

[16] Bogert, Kenneth. *Notes on Generalizing the Maximum Entropy Principle to Uncertain Data*. 2022. arXiv: 2109.04530 [cs.IT].

[17] Boomsma, Wouter and Ferkinghoff-Borg, Jesper and Lindorff-Larsen, Kresten. "Combining experiments and simulations using the maximum entropy principle". In: *PLoS computational biology* 10.2 (2014), e1003406.

[18] Borwein, Jonathan and Lewis, Adrian. *Convex Analysis*. Springer, 2006.

[19] Borwein, Jonathan M and Lewis, Adrian S. "Partially finite convex programming, Part I: Quasi relative interiors and duality theory". In: *Mathematical Programming* 57.1 (1992), pp. 15–48.

[20] Borwein, Jonathan M and Lewis, Adrian S. "Partially finite convex programming, part II: Explicit lattice models". In: *Mathematical Programming* 57.1 (1992), pp. 49–83.

[21] Botea, Mariana and Ganesh, Arvind. "Prediction of air traffic complexity using machine learning algorithms". In: *Journal of Air Transport Management* 52 (2016), pp. 11–21.

[22] Broyden, Charles George. "The convergence of a class of double-rank minimization algorithms 1. general considerations". In: *IMA Journal of Applied Mathematics* 6.1 (1970), pp. 76–90.

[23] Burg, John Parker. "The relationship between maximum entropy spectra and maximum likelihood spectra". In: *Geophysics* 37.2 (1972), pp. 375–376.

[24] Cai, Guowei and Chen, Ben M and Lee, Tong Heng. "Coordinate systems and transformations". In: *Unmanned rotorcraft systems* (2011), pp. 23–34.

[25] Caticha, Ariel and Giffin, Adom. "Updating probabilities". In: *AIP Conference Proceedings*. Vol. 872. 1. American Institute of Physics. 2006, pp. 31–42.

[26] Caticha, Ariel and Mohammad-Djafari, Ali and Bercher, Jean-François and Bessière, Pierre. "Entropic Inference". In: *AIP Conference Proceedings*. AIP, 2011.

[27] Chen, Jeng-Shyang and Wu, Chen-Chien and Wang, Chien-Hua and Su, Yu-Chih and Lee, Cheng-Che. "Agent-based approach for assessing unmanned aircraft systems integration in the national airspace system". In: *IEEE Access* 5 (2017), pp. 11151–11160.

[28] Chen, Yunfei and Chen, Linlin and Yao, Wei and Sun, Jian. "A novel machine learning approach to air traffic complexity prediction and management". In: *Transportation Research Part C: Emerging Technologies* 102 (2019), pp. 251–266.

[29] Chen, Zhengmao and Guo, Dongyue and Lin, Yi. "A deep Gaussian process-based flight trajectory prediction approach and Its application on conflict detection". In: *Algorithms* 13.11 (2020), p. 293. ISSN: 1999-4893.

[30] Cheng, Cheng et al. "Machine learning-aided trajectory prediction and conflict detection for internet of aerial vehicles". In: *IEEE Internet of Things Journal* (2021). ISSN: 2327-4662, 2372-2541.

[31] Chung, Junyoung and Gülçehre, Çaglar and Cho, KyungHyun and Bengio, Yoshua. "Gated feedback recurrent neural networks". In: *CoRR* abs/1502.02367 (2015). arXiv: 1502.02367.

[32] Coppola, Gianluca and Mandolini, Marco and Pieroni, Alessio. "Airspace complexity assessment using topological methods". In: *IEEE Access* 8 (2020), pp. 75711–75720.

[33] Council, National Research and Engineering, Division on and Sciences, Physical and Aeronautics and Board, Space Engineering and Autonomy Research for Civil Aviation, Committee on. "Autonomy research for civil aviation: Toward a new era of flight". In: *National Academies Press* 2.3 (2014).

[34] De Leege, Arjen and Paassen, Marinus van and Mulder, Max. "A machine learning approach to trajectory prediction". In: *AIAA Guidance, Navigation, and Control (GNC) Conference*. 2013, pp. 1–10.

[35] Degas, Augustin et al. “A survey on artificial intelligence (AI) and explainable ai in air traffic management: Current trends and development with future research trajectory”. In: *Applied Sciences* 12.3 (2022), p. 1295. ISSN: 2076-3417.

[36] Delahaye, Daniel and García, Adrían and Lavandier, Julien and Chaimatanan, Supatcha and Soler, Manuel. “Air traffic complexity map based on linear dynamical systems”. In: *Aerospace* 9.5 (2022). ISSN: 2226-4310.

[37] Delahaye, Daniel and Puechmorel, Stéphane. “Air traffic complexity: Towards an intrinsic metric”. In: *Proceeding of the 3rd USA/Europe Air Traffic Management R and D Seminar*. 2000.

[38] Delahaye, Daniel and Puechmorel, Stéphane. *Modeling and optimization of air traffic*. John Wiley & Sons, 2013. ISBN: 9781848215955.

[39] Delahaye, Daniel and Puechmorel, Stephane. “Air traffic complexity based on dynamical systems”. In: *49th IEEE Conference on Decision and Control (CDC)*. Atlanta, GA, USA: IEEE, Dec. 2010, pp. 2069–2074. ISBN: 978-1-4244-7745-6.

[40] Dias, Fernando HC and Rey, David. “Robust aircraft conflict resolution under trajectory prediction uncertainty”. In: *Operations Research Letters* 50.5 (2022), pp. 503–508.

[41] Ding, Jingtao et al. *Artificial intelligence for complex network: Potential, methodology and application*. 2024. arXiv: 2402.16887 [cs.SI].

[42] Dudík, Miroslav and Phillips, Steven J and Schapire, Robert E. “Maximum entropy density estimation with generalized regularization and an application to species distribution modeling”. In: *Journal of Machine Learning Research* (2007), pp. 1217–1260.

[43] Facchinei, Francisco and Lucidi, Stefano and Palagi, Laura. “A truncated Newton algorithm for large scale box constrained optimization”. In: *SIAM Journal on Optimization* 12 (1999).

[44] Fang, Jianqiang and Li, Xiaoyan and Wei, Na and Lv, Xiaofei. “Agent-based simulation of collaborative decision making for air traffic management”. In: *IEEE Access* 7 (2019), pp. 135695–135707.

[45] Fu, Jun and Zhou, Wei and Chen, Zhibo. “Bayesian spatio-temporal graph convolutional network for traffic forecasting”. In: *arXiv:2010.07498 [cs]* (2020). arXiv: 2010.07498.

[46] Gao, Yi and Tang, Tao and Li, Xiaoyan and Liu, Fang. “Agent-based approach for modeling and simulation of air traffic management in en-route airspace”. In: *IEEE Access* 7 (2019), pp. 61689–61699.

[47] Gao, Zhenyu and Yu, Yue and Wei, Qinshuang and Topcu, Ufuk and Clarke, John-Paul. *Noise-aware and equitable urban air traffic management: An optimization approach*. 2024. arXiv: 2401.00806 [eess.SY].

[48] Gatto, Riccardo. "Some computational aspects of the generalized von Mises distribution". In: *Statistics and Computing* 18 (2008), pp. 321–331.

[49] Gatto, Riccardo and Jammalamadaka, Sreenivasa Rao. "The generalized von Mises distribution". In: *Statistical Methodology* 4.3 (2007), pp. 341–353. ISSN: 1572-3127.

[50] Girardin, Valerie. "Méthodes de réalisation de produit scalaire et de problème de moments avec maximisation d'entropie". In: *Studia Mathematica* 3.124 (1997), pp. 199–213.

[51] Gomes-Gonçalves, Erika and Gzyl, Henryk and Mayoral, Silvia. "Density reconstructions with errors in the data". In: *Entropy* 16.6 (2014), pp. 3257–3272.

[52] Gresele, Luigi and Marsili, Matteo. "On maximum entropy and inference". In: *Entropy* 19.12 (2017), pp. 1–16.

[53] Gzyl, Henryk. *Joint probabilities under expected value constraints, transportation problems, maximum entropy in the mean, and geometry in the space of probabilities*. 2021. arXiv: 2109.01166 [math.PR].

[54] Hestenes, Magnus R. "The conjugate-gradient method for solving linear systems[1]". In: *Numerical analysis* 6 (1956), pp. 83–102.

[55] Hiriart-Urruty, Jean-Baptiste and Lemaréchal, Claude. *Fundamentals of convex analysis*. Springer Science & Business Media, 2004.

[56] Hochreiter, Sepp and Schmidhuber, Jürgen. "Long short-term memory". In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[57] Hong, Youkyung and Kim, Youdan and Lee, Keumjin. "Application of complexity map to reduce air traffic complexity in a sector". In: *AIAA Guidance, Navigation, and Control Conference*. American Institute of Aeronautics and Astronautics, 2014. ISBN: 978-1-62410-317-9.

[58] Huang, Chiou-Jye and Kuo, Ping-Huan. "A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities". In: *Sensors* 18.7 (2018).

[59] Huang, Jin and Ding, Weijie. "Aircraft trajectory prediction based on Bayesian optimised temporal convolutional network–bidirectional gated recurrent unit hybrid neural nework". In: *International Journal of Aerospace Engineering* 2022 (2022). Ed. by Sobel, Kenneth M., pp. 1–19.

[60] Jaynes, Edwin T. "Information theory and statistical mechanics". In: *Physical review* 106.4 (1957), p. 620.

[61] Jaynes, Edwin T. *Probability theory: The logic of science*. Cambridge university press, 2003.

[62] Jones, Donald R. "A taxonomy of global optimization methods based on response surfaces". In: *Journal of global optimization* 21 (2001), pp. 345–383.

[63] Juntama, Paveen and Chaimatanan, Supatcha and Alam, Sameer and Delahaye, Daniel. "A distributed metaheuristic approach for complexity Rreduction in air traffic for strategic 4D trajectory optimization". In: *International Conference on Artificial Intelligence and Data Analytics for Air Transportation*. 2020, pp. 1–9.

[64] Kagan Abram, M. and Linnik Yuri, V. and Radhakrishna Rao, C. *Characterization problems in mathematical statistics*. 1973.

[65] Khayyam, Hamid and Yildirim, Süleyman and Yücesan, Enver. "Comparative evaluation of air traffic complexity metrics for en-route airspace". In: *Transportation Research Part C: Emerging Technologies* 119 (2020), p. 102725.

[66] Lakshmanan, Rajmadan and Pichler, Alois. "Soft Quantization Using Entropic Regularization". In: *Entropy* 25.10 (2023), p. 1435.

[67] Laudeman I.; Shelden, S. "Dynamic density: An air traffic management metric". In: *Technical Report, NASA/TM-1998-112226; NASA: Boulder, CO, USA* (1998).

[68] Lee, Keumjin and Feron, Eric and Pritchett, Amy. "Describing airspace complexity: Airspace response to disturbances". In: *Journal of Guidance, Control, and Dynamics* 32.1 (2009), pp. 210–222. ISSN: 0731-5090, 1533-3884.

[69] Li, Chengyu and Chen, Qian and Yan, Xiaoming and Wang, Jin and Lin, Xingwei. "Learning air traffic as images: A deep convolutional neural network for airspace operation complexity evaluation". In: *IEEE Access* 8 (2020), pp. 19732–19740.

[70] Li, Huanhuan and Jiao, Hang and Yang, Zaili. "AIS data-driven ship trajectory prediction modelling and analysis based on machine learning and deep learning methods". In: *Transportation Research Part E: Logistics and Transportation Review* 175 (July 2023), pp. 1–39. ISSN: 13665545.

[71] Li, Qiang and Guan, Xinjia and Liu, Jinpeng. "A CNN-LSTM framework for flight delay prediction". In: *Expert Systems with Applications* 227 (Oct. 2023), pp. 1–16. ISSN: 09574174.

[72] Li, Weigang and Alves, CJP and Omar, N. "Knowledge-based system for air traffic flow management: Timetable rescheduling and centralized flow control". In: *Transactions on Information and Communications Technologies* (1993), pp. 655–670.

[73] Liu, Hong and Lin, Yi and Chen, Zhengmao and Guo, Dongyue and Zhang, Jianwei and Jing, Hailong. "Research on the air traffic flow prediction using a deep learning approach". In: *IEEE Access* 7 (2019), pp. 148019–148030.

[74] Liu, Hua and Li, Xiaoyan and Wei, Na and Lv, Xiaofei. "Agent-Based Simulation of Aircraft Taxiing Operations". In: *IEEE Access* 5 (2017), pp. 18936–18945.

[75] Liu, Yan and Zhang, Jiazhong. "Predicting traffic flow in local area networks by the largest Lyapunov exponent". In: *Entropy* 18.1 (2016), p. 32.

[76] Ma, Lan and Tian, Shan. "A hybrid CNN-LSTM model for aircraft 4D trajectory Prediction". In: *IEEE Access* 8 (2020), pp. 134668–134680. ISSN: 2169-3536.

[77] Maksimov, Vyacheslav Mikhailovich. "Necessary and sufficient statistics for the family of shifts of probability distributions on continuous BiCompact groups". In: *Theory of Probability & Its Applications* 12.2 (1967), pp. 267–280.

[78] Maréchal, Pierre. "A note on entropy optimization". In: *Approximation, Optimization and Mathematical Economics*. Springer, 2001, pp. 205–211.

[79] Maréchal, Pierre. "On the Principle of Maximum Entropy on the Mean as a methodology for the regularization of inverse problems". In: *Probability Theory and Mathematical Statistics* (1998).

[80] Maréchal, Pierre and Navarrete, Yasmín and Davis, Sergio. "On the foundations of the maximum entropy principle using Fenchel duality for Shannon and Tsallis entropies". In: *Physica Scripta* 99.7 (2024), p. 075265.

[81] Marwala, Tshilidzi and Nelwamondo, Fulufhelo V. "A multi-agent system for air traffic flow management: A hybrid approach". In: *IEEE Transactions on Intelligent Transportation Systems* 22.7 (2020), pp. 4369–4381.

[82] Mirikitani, Derrick and Nikolaev, Nikolay. "Recursive Bayesian recurrent neural networks for time-series modeling". In: *Transactions on Neural Networks* 21.2 (2010), pp. 262–274.

[83] Mondoloni, Stephane and Liang, Diana. "Airspace Fractal Dimensions and Applications". In: *USAEurope ATM Seminar*. Dec. 2001.

[84] Moré, Jorge J. "Recent developments in algorithms and software for trust region methods". In: *Mathematical Programming The State of the Art: Bonn 1982* (1983), pp. 258–287.

[85] Nelder, John A and Mead, Roger. "A simplex method for function minimization". In: *The computer journal* 7.4 (1965), pp. 308–313.

[86] Nghiem, Thi-Lich and Le, Viet-Duc and Le, Thi-Lan and Delahaye, Daniel and Maréchal, Pierre. "Angular probability density reconstruction by Maximum Entropy". preprint. July 2024.

[87] Pang, Yutian and Xu, Nan and Liu, Yongming. "Aircraft Trajectory Prediction using LSTM neural network with embedded convolutional layer". In: *Annual Conference of the PHM Society* 11.1 (2019). ISSN: 2325-0178, 2325-0178.

[88] Pang, Yutian and Zhao, Xinyu and Hu, Jueming and Yan, Hao and Liu, Yongming. "Bayesian Spatio-Temporal grAph tRansformer network (B-STAR) for multi-aircraft trajectory prediction". In: *Knowledge-Based Systems* 249 (Aug. 2022), p. 108998. ISSN: 09507051.

[89] Pérez Moreno, Francisco and Gómez Comendador, Víctor Fernando and Delgado-Aguilera Jurado, Raquel and Zamarreño Suárez, María and Antulov-Fantulin, Bruno and Arnaldo Valdés, Rosa María. "How has the concept of air traffic complexity evolved? Review and analysis of the state of the art of air traffic complexity". In: *Applied Sciences* 14.9 (2024), pp. 1–21.

[90] Polpo, Adriano and Stern, Julio and Louzada, Francisco and Izbicki, Rafael and Takada, Hellinton. "Bayesian inference and maximum entropy methods in science and engineering". In: *Springer Proceedings in Mathematics & Statistics* (2018).

[91] Powell, Michael JD. *A direct search optimization method that models the objective and constraint functions by linear interpolation*. Springer, 1994.

[92] Powell, Michael JD. "An efficient method for finding the minimum of a function of several variables without calculating derivatives". In: *The computer journal* 7.2 (1964), pp. 155–162.

[93] Prandini, Maria and Piroddi, Luigi and Puechmorel, Stephane and Brazdilova, Silvie Luisa. "Toward air traffic complexity assessment in new generation air traffic management systems". In: *IEEE Transactions on Intelligent Transportation Systems* 12.3 (2011), pp. 809–818. ISSN: 1524-9050, 1558-0016.

[94] Qiao, Sun and Han, Nan and Zhu, Xin and Shu, Hong and Zheng, Jiao and Yuan, Chang. "A dynamic trajectory prediction algorithm based on Kalman Filter". In: *Acta Electronica Sinica* 46 (2018), pp. 418–423.

[95] Qin, Wanting and Tang, Jun and Lao, Songyang. "DeepFR: A trajectory prediction model based on deep feature representation". In: *Information Sciences* 604 (2022), pp. 226–248.

[96] Razvan, Bucuroiu and Stéphanie, Vincent. *European Network Operations Plan 2019 (2024) Rolling Seasonal Plan*. Tech. rep. EUROCONTROL, 2019-2024.

[97] Rényi, Alfréd. *Probability theory*. Courier Corporation, 2007.

[98] Rey, David and Rapine, Christophe and Fondacci, Remy and El Faouzi, Nour-Eddin. "Subliminal speed control in air traffic management: Optimization and simulation". In: *Transportation Science* 50 (2016), pp. 240–262.

[99] Rivero, Cristian Rodriguez et al. "Time series forecasting using Recurrent neural networks modified by Bayesian inference in the learning process". In: *2019 IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI)*. Barranquilla, Colombia, 2019, pp. 1–6.

[100] Rockafellar, R Tyrrell. *Convex analysis*. Vol. 18. Princeton university press, 1970.

[101] Rockafellar, R Tyrrell. "Convex integral functionals and duality". In: *Contributions to nonlinear functional analysis*. Elsevier, 1971, pp. 215–236.

[102] Rockafellar, R Tyrrell. "Integral functionals, normal integrands and measurable selections". In: *Nonlinear Operators and the Calculus of Variations*. Springer, 2006, pp. 157–207.

[103] Rockafellar, Ralph. "Integrals which are convex functionals". In: *Pacific journal of mathematics* 24.3 (1968), pp. 525–539.

[104] Rooij, Gijs de and Stienstra, Amber and Borst, Clark and Tisza, Adam B and Paassen, Marinus M van and Mulder, Max. "Contributing factors to flight-centric complexity in en-route air traffic control". In: *Proceedings of the 15th USA/Europe Air Traffic Management Research and Development Seminar*. 2023.

[105] Salinas, David and Flunkert, Valentin and Gasthaus, Jan and Januschowski, Tim. "DeepAR: Probabilistic forecasting with autoregressive recurrent networks". In: *International Journal of Forecasting* 36.3 (2020), pp. 1181–1191. ISSN: 0169-2070.

[106] Shafienya, Hesam and Regan, Amelia. "4D flight trajectory prediction based on ADS-B data: A comparison of CNN-GRU models". In: *Aerospace Conference (AERO)*. 2022, pp. 01–12.

[107] Shafienya, Hesam and Regan, Amelia C. "4D flight trajectory prediction using a hybrid deep learning prediction method based on ADS-B technology: A case study of Hartsfield–jackson Atlanta international airport (ATL)". In: *Transportation Research Part C: Emerging Technologies* 144 (2022), pp. 1–12.

[108] Shang, Tongfei and Xiao, Kunpeng and Han, Kun. "Target operator trajectory prediction method based on attention mechanism and LSTM". In: *Journal of Physics: Conference Series* 2037.1 (2021), pp. 1–6. ISSN: 1742-6588.

[109] Shi, Xingjian and Chen, Zhourong and Wang, Hao and Yeung, Dit-Yan and Wong, Wai-Kin and Woo, Wang-chun. "Convolutional LSTM network: A machine learning approach for precipitation Nowcasting". In: *CoRR* (2015).

[110] Shi, Xingjian and Chen, Zhourong and Wang, Hao and Yeung, Dit-Yan and Wong, Wai-kin and Woo, Wang-chun. "Convolutional LSTM network: A machine learning approach for precipitation nowcasting". In: *arXiv* (2015). eprint: 1506.04214 (cs.CV).

[111] Shi, Zhiyuan and Xu, Min and Pan, Quan. "4D flight trajectory prediction with constrained LSTM network". In: *IEEE Transactions on Intelligent Transportation Systems* 22.11 (2021), pp. 7242–7255. ISSN: 1524-9050.

[112] Siddique, Talha and Mahmud, Shaad and Keesee, Amy and Ngwira, Chigomezyo and Connor, Hyunju. "A survey of uncertainty quantification in machine learning for space weather prediction". In: *Geosciences* 12.1 (2022). ISSN: 2076–3263.

[113] Spencer, D.A. "Applying artificial intelligence techniques to air traffic control automation". In: *The Lincoln Laboratory Journal* 2.3 (1989), pp. 537–554.

[114] Sutskever, Ilya and Vinyals, Oriol and Le, Quoc V. "Sequence to sequence learning with neural networks". In: *arXiv* (2014). eprint: 1409.3215 (cs.CL).

[115] Teboulle, Marc. "A simplified view of first order methods for optimization". In: *Mathematical Programming* 170.1 (2018), pp. 67–96.

[116] Tran, Phu N. and Nguyen, Hoang Q. V. and Pham, Duc-Thinh and Alam, Sameer. "Aircraft trajectory prediction with enriched intent using Encoder-Decoder architecture". In: *IEEE Access* 10 (2022), pp. 17881–17896.

[117] Tsallis, Constantino and Mendes, Renio S and Plastino, Anel R. "The role of constraints within generalized nonextensive statistics". In: *Physica A: Statistical Mechanics and its Applications* 261.3-4 (1998), pp. 534–554.

[118] Varun S., Sudarsanan. "Deep learning framework for trajectory prediction and in-time prognostics in the terminal airspace". PhD thesis. Purdue University Graduate School, 2022.

[119] Vidosavljevic, A. and Delahaye, D. and Sunil, E. and Bussink, F. and Hoekstra, J. "Complexity analysis of the concepts of urban airspace design for METROPOLIS project". In: *ENRI Int. Workshop on ATM/CNS. Tokyo, Japan.* 2015.

[120] Wan, Guihong and Maung, Crystal and Schweitzer, Haim. "Improving the accuracy of principal component analysis by the maximum entropy method". In: *31st International Conference on Tools with Artificial Intelligence (ICTAI).* IEEE. 2019, pp. 808–815.

[121] Wang, Chengzhang and He, Yu and Xu, Jun and Zhou, Chunyang. "A machine learning approach for air traffic complexity prediction and management". In: *Journal of Advanced Transportation* 2019 (2019), pp. 1–15.

[122] Wang, Huai-zhi and Li, Gang-qiang and Wang, Gui-bin and Peng, Jian-chun and Jiang, Hui and Liu, Yi-tao. "Deep learning based ensemble approach for probabilistic wind power forecasting". In: *Applied Energy* 188 (2017), pp. 56–70. ISSN: 0306-2619.

[123] Wang, Xing and Jiang, Xinhua and Chen, Lifei and Wu, Yi. "KVLMM: A trajectory prediction method based on a variable-order Markov model with kernel smoothing". In: *IEEE Access* 6 (2018), pp. 25200–25208.

[124] Wang, Zhengyi and Delahaye, Daniel and Farges, Jean-Loup and Alam, Sameer. "Complexity optimal air traffic assignment in multi-layer transport network for urban air mobility operations". In: *Transportation Research Part C: Emerging Technologies* 142 (Sept. 2022), pp. 1–23. ISSN: 0968090X.

[125] Wang, Zhengyi and Liang, Man and Delahaye, Daniel. "Short-term 4D trajectory prediction using machine learning methods". In: *SID 2017, 7th SESAR Innovation Days.* 2017.

[126] Wei, Na and Li, Xiaoyan and Lv, Xiaofei and Yang, Xiaodong. "Agent-based simulation of air traffic flow management strategies". In: *IEEE Access* 6 (2018), pp. 32216–32228.

[127] West, Jon and Zhang, Yu and Yildirim, Tolga and Blanks, Brenden. "Agent-based simulation of air traffic control operations for small unmanned aerial systems". In: *IEEE Access* 7 (2019), pp. 56115–56125.

[128] Winter, Joost C de and Verhagen, Wim J and Ellerbroek, Joost. "Agent-based simulation of collaborative decision making in air traffic flow management". In: *IEEE Transactions on Intelligent Transportation Systems* 19.9 (2018), pp. 2771–2782.

[129] Wyndemere. "An evaluation of air traffic control complexity". In: *Technical Report, NASA 2-14284; NASA: Boulder, CO, USA* (1996).

[130] Xin, Liu and Hailong, Pei and Jianqiang, Li. "Trajectory prediction based on particle filter application in mobile robot system". In: *2008 27th Chinese Control Conference*. 2008, pp. 389–393.

[131] Xu, Zhengfeng and Zeng, Weili and Chu, Xiao and Cao, Puwen. "Multi-aircraft trajectory collaborative prediction based on social Long Short-Term Memory network". In: *Aerospace* 8.4 (2021). ISSN: 2226-4310.

[132] Yang, Hang and Liu, Zhonghua and Xiao, Yuguang. "Evaluation of air traffic complexity using entropy-based metrics". In: *Entropy* 21.10 (2019), p. 958.

[133] Yang, Yandong and Hong, Weijun and Li, Shufang. "Deep ensemble learning based probabilistic load forecasting in smart grids". In: *Energy* 189 (2019), pp. 1–12.

[134] Zalinescu, Constantin. *Convex analysis in general vector spaces*. World scientific, 2002.

[135] Zeng, Weili and Chu, Xiao and Xu, Zhengfeng and Liu, Yan and Quan, Zhibin. "Aircraft 4D trajectory prediction in civil aviation: A review". In: *Aerospace* 9.2 (2022). ISSN: 2226-4310.

[136] Zeng, Weili and Chu, Xiao and Xu, Zhengfeng and Liu, Yan and Quan, Zhibin. "Aircraft 4D trajectory prediction in civil aviation: A review". In: *Aerospace* 9.2 (Feb. 2022), pp. 1–19. ISSN: 2226-4310.

[137] Zhang, Rongjie and Wu, Wenjuan and Li, Liang and Li, Zhaowei. "Machine learning techniques for air traffic complexity evaluation". In: *Transportation Research Part C: Emerging Technologies* 111 (2020), pp. 463–479.

[138] Zhang, Tao and Gao, Yang and Zhang, Chengwei. "Short-term 4D trajectory prediction based on KF joint EKF parameter identification". In: *Journal of Civil Aviation University of China* 34.5 (2016), pp. 1–4.

[139] Zhang, Xiaoge and Mahadevan, Sankaran. "Bayesian neural networks for flight trajectory prediction and safety assessment". In: *Decision Support Systems* 131 (Apr. 2020), p. 113246. ISSN: 01679236.

Appendix

A Computing conjugates

In this appendix, we give computation details for the conjugate functions involved in the Maximum Entropy problem of Chapter 3. We start with the conjugate of

$$g_\circ(\eta_\circ, \boldsymbol{\eta}) = -\frac{\alpha}{2}\|\mathbf{z} - \boldsymbol{\eta}\|^2_{\Sigma^{-1}} - \delta_{\{1\}}(\eta_\circ).$$

We have:

$$\begin{aligned}
&(g_\circ)_\star(\lambda_\circ, \boldsymbol{\lambda}) \\
&\quad = \inf_{\eta_\circ, \boldsymbol{\eta}} \left\{ \langle (\lambda_\circ, \boldsymbol{\lambda}), (\eta_\circ, \boldsymbol{\eta}) \rangle + \frac{\alpha}{2}\|\mathbf{z} - \boldsymbol{\eta}\|^2_{\Sigma^{-1}} + \delta_{\{1\}}(\eta_\circ) \right\} \\
&\quad = \inf_{\eta_\circ} \left\{ \lambda_\circ \eta_\circ + \delta_{\{1\}}(\eta_\circ) \right\} + \inf_{\boldsymbol{\eta}} \left\{ \langle \boldsymbol{\lambda}, \boldsymbol{\eta} \rangle + \frac{\alpha}{2}\|\mathbf{z} - \boldsymbol{\eta}\|^2_{\Sigma^{-1}} \right\}.
\end{aligned}$$

Clearly, the first *infimum* is attained at $\bar{\eta}_\circ = 1$, yielding the value $\lambda_\circ$. The second *infimum* is attained at the point $\bar{\boldsymbol{\eta}}$ where the gradient of the function $\boldsymbol{\eta} \mapsto \langle \boldsymbol{\lambda}, \boldsymbol{\eta} \rangle + \frac{\alpha}{2}\|\mathbf{z} - \boldsymbol{\eta}\|^2_{\Sigma^{-1}}$. A straightforward computation shows that the gradient of the latter function is equal to $\boldsymbol{\lambda} - \alpha\Sigma^{-1}(\mathbf{z} - \boldsymbol{\eta})$, which implies that

$$\bar{\boldsymbol{\eta}} = \mathbf{z} - \frac{1}{\alpha}\Sigma\boldsymbol{\lambda}.$$

It follows that

$$\begin{aligned}
&\inf_{\boldsymbol{\eta}} \left\{ \langle \boldsymbol{\lambda}, \boldsymbol{\eta} \rangle + \frac{\alpha}{2}\|\mathbf{z} - \boldsymbol{\eta}\|^2_{\Sigma^{-1}} \right\} \\
&\quad = \left\langle \boldsymbol{\lambda}, \mathbf{z} - \frac{1}{\alpha}\Sigma\boldsymbol{\lambda} \right\rangle + \frac{\alpha}{2} \left\langle \frac{1}{\alpha}\Sigma\boldsymbol{\lambda}, \Sigma^{-1}\left(\frac{1}{\alpha}\Sigma\boldsymbol{\lambda}\right) \right\rangle \\
&\quad = \langle \boldsymbol{\lambda}, \mathbf{z} \rangle - \frac{1}{2\alpha} \langle \boldsymbol{\lambda}, \Sigma\boldsymbol{\lambda} \rangle .
\end{aligned}$$

Putting things together, we obtain:

$$(g_\circ)_\star(\lambda_\circ, \boldsymbol{\lambda}) = \lambda_\circ + \langle \boldsymbol{\lambda}, \mathbf{z} \rangle - \frac{1}{2\alpha} \langle \boldsymbol{\lambda}, \Sigma\boldsymbol{\lambda} \rangle .$$

Next, we address the case where the estimated covariance matrix is singular. In this case, the inverse covariance matrix does not exist and the Mahalanobis distance degenerates. We check here that the appropriate formula is that given in Equation (3.6).

Lemma 24. *Let Q be a symmetric positive semi-definite matrix and let*

$$\varphi(\mathbf{x}) := \frac{1}{2} \langle \mathbf{x}, Q\mathbf{x} \rangle$$

Then, the convex conjugate $\varphi^\star$ is given by

$$\varphi^\star(\boldsymbol{\xi}) = \begin{cases} \left\langle \boldsymbol{\xi}, Q^\dagger \boldsymbol{\xi} \right\rangle & \text{if } \boldsymbol{\xi} \in \operatorname{ran} Q, \\ \infty & \text{otherwise.} \end{cases}$$

Proof. By definition,

$$\varphi^\star(\boldsymbol{\xi}) = \sup_{\mathbf{x}} \left\{ \langle \boldsymbol{\xi}, \mathbf{x} \rangle - \frac{1}{2} \langle \boldsymbol{\xi}, Q\mathbf{x} \rangle \right\}.$$

The gradient of the function to be supremized is equal to $\boldsymbol{\xi} - Q\mathbf{x}$. If $\boldsymbol{\xi} \in \operatorname{ran} Q$, then the gradient vanishes at any $\bar{\mathbf{x}}$ such that $Q\bar{\mathbf{x}} = \boldsymbol{\xi}$, for example at $\bar{\mathbf{x}} = Q^\dagger \boldsymbol{\xi}$, since $QQ^\dagger \boldsymbol{\xi} = \boldsymbol{\xi}$ (recall that $QQ^\dagger$ is the orthonormal projection onto the range of Q. Then,

$$\varphi^\star(\boldsymbol{\xi}) = \left\langle \boldsymbol{\xi}, Q^\dagger \boldsymbol{\xi} \right\rangle - \frac{1}{2} \left\langle Q^\dagger \boldsymbol{\xi}, QQ^\dagger \boldsymbol{\xi} \right\rangle = \frac{1}{2} \left\langle \boldsymbol{\xi}, Q^\dagger \boldsymbol{\xi} \right\rangle.$$

If $\boldsymbol{\xi} \notin \operatorname{ran} Q$, then pick $\mathbf{x}_\circ \in \ker Q \setminus \{\mathbf{0}\}$. For every $t \in \mathbb{R}$,

$$\langle \boldsymbol{\xi}, t\mathbf{x}_\circ \rangle - \frac{1}{2} \langle \boldsymbol{\xi}, Q(t\mathbf{x}_\circ) \rangle = t \langle \boldsymbol{\xi}, \mathbf{x}_\circ \rangle. \tag{6.1}$$

Since $\mathbf{x}_\circ \in \ker Q = (\operatorname{ran} Q^\star)^\perp = (\operatorname{ran} Q)^\perp$ and we assumed that $\boldsymbol{\xi} \notin \operatorname{ran} Q$, the scalar product $\langle \boldsymbol{\xi}, \mathbf{x}_\circ \rangle$ is necessarily nonzero, which shows that the left hand side in Equation (6.1) can be made as large as desired. This implies that $\varphi^\star(\boldsymbol{\xi}) = \infty$ in this case.

A byproduct of the lemma is that the function in Equation (3.6) is a closed proper convex. Since such functions are equal to their biconjugates,

this justifies that $(g_\circ)_\star$ is unchanged, and subsequently so are the dual problems $(\mathscr{D})$ and $(\tilde{\mathscr{D}})$.

Finally, we deal with the computation of the convex conjugate of the function

$$h_\circ(t) = \begin{cases} t \ln t & \text{if } t > 0, \\ 0 & \text{if } t = 0, \\ \infty & t < 0. \end{cases}$$

By definition,

$$h^\star(\tau) = \sup_{t \in \mathbb{R}} \{\tau t - h(t)\}.$$

Suppose that the *supremum* is attained at $\bar{t} \in \mathbb{R}_+^*$. Then

$$\tau - h'(\bar{t}) = \tau - \ln \bar{t} - 1 = 0,$$

so that $\bar{t} = e^{\tau - 1}$. We see *a posteriori* that $\bar{t}$ maximizes the function $x \mapsto \tau t - h(t)$, since the maximum can't be attained on $\mathbb{R}_-$. Therefore,

$$h^\star(\tau) = \tau \bar{t} - h(\bar{t}) = \tau e^{\tau - 1} - e^{\tau - 1}(\tau - 1) = e^{\tau - 1}.$$