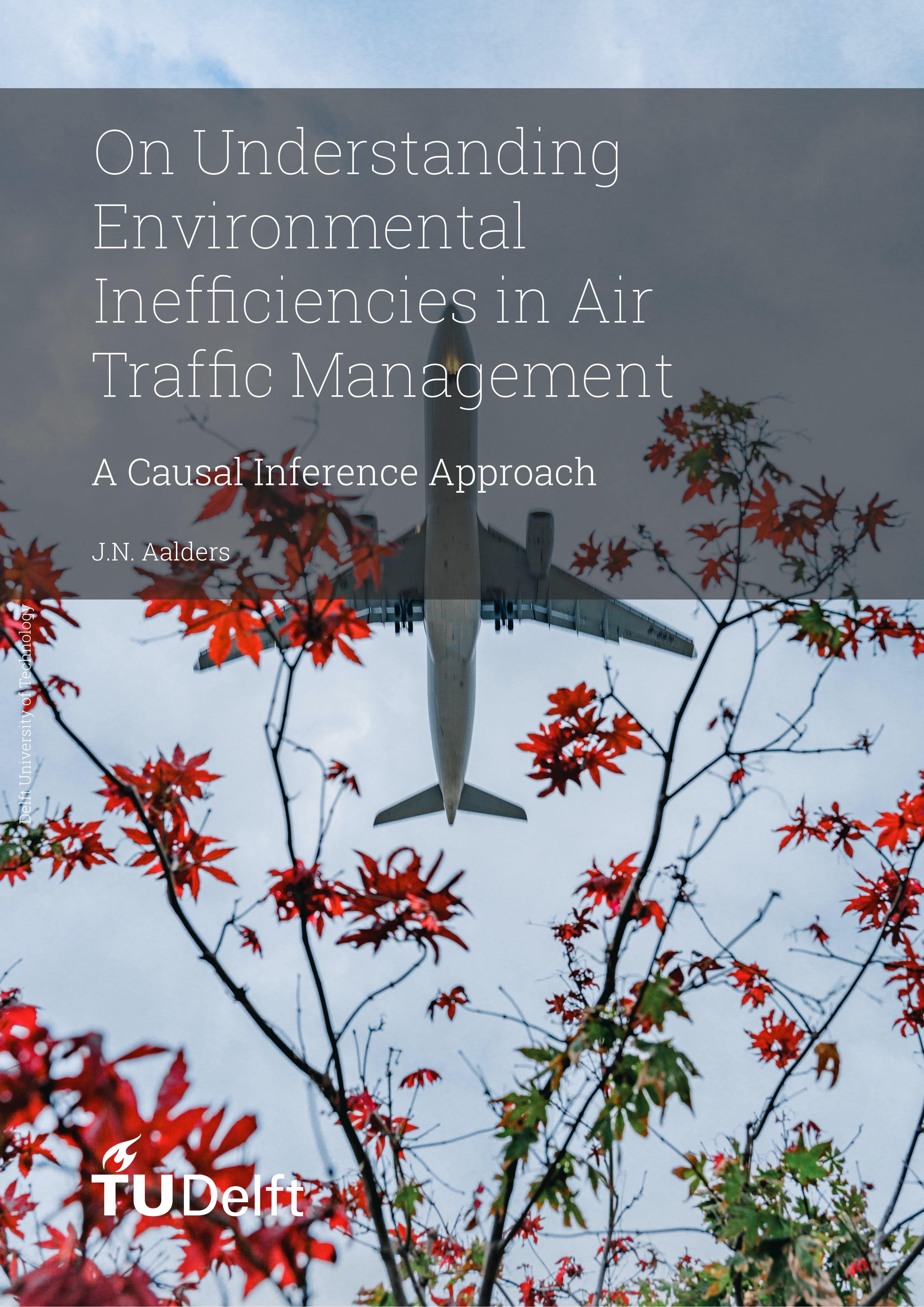
On Understanding Environmental Inefficiencies in Air Traffic Management

A Causal Inference Approach

J.N. Aalders

Delft University of Technology



TU Delft

This page is intentionally left blank.

On Understanding Environmental Inefficiencies in Air Traffic Management

A Causal Inference Approach

by

J.N. Aalders

to obtain the degree of
Master of Science at the Delft University of Technology,
to be defended on Tuesday March 5, 2024.

Student number:	4650166	
Thesis Committee:	Prof. dr. ir. M. Snellen	TU Delft, Chair
	Dr. I.C. Dedoussi	TU Delft, Supervisor
	Dr. J. Sun	TU Delft, Supervisor
	Dr. ir. F.D.A. Quadros	TU Delft, Supervisor
	Dr. M.J. Ribeiro	TU Delft, External Examiner

Cover: Low Angle View of an Airplane by Marstion

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

This page is intentionally left blank.

Preface

Thank you to everybody who, in any way, large or small, has contributed to my wonderful time at Delft University of Technology!

This thesis marks the end of my student time here, a journey I've enjoyed every step of the way. From the diverse people I've met to the unforgettable experiences I've had, each moment has contributed to shaping the person I am today. TU Delft and its student culture have not only provided high academic standards but also pushed me to grow in countless ways. Its support for extracurricular activities has provided me with opportunities to explore and develop various skills beyond the classroom. As I continue to the next adventure, I carry the name of a Delft Engineer with great pride!

The subject of this research, in large part, originated from an opportunity provided by Dr. Irene Dedoussi and Dr.ir. Flávio Quadros which, through a small project, allowed me to explore my motivation to contribute to a sustainable future of aviation. Building on this interest, and with the addition of Dr. Junzi Sun to the team, my perfect thesis was formed by combining air transport operations with their environmental impacts. With a larger group of supervisors than usual, we set off to try to explain the causes of environmental inefficiencies in European aviation!

Irene, Junzi, and Flávio, thank you very much for guiding me through this process as it would not have been possible without your structure and enthusiasm. My gratitude also extends to Prof.dr.ir. Mirjam Snellen for chairing this committee as well as providing feedback through critical questions during the milestone meetings.

Finally, I hope this novel application of causal inference methods will provide you, the reader, with new insights and ideas. Enjoy reading the work!

J.N. Aalders
Delft, February 2024

This page is intentionally left blank.

Contents

This page is intentionally left blank.

List of Figures

This page is intentionally left blank.

List of Tables

This page is intentionally left blank.

Nomenclature

List of Abbreviations

Abbreviation	Definition
ACC	Area Control Center
AEIC	Aviation Emissions Inventory Code
AMS	Amsterdam Airport
ATC	Air Traffic Control
ATCo	Air Traffic Controller
ATE	Average Treatment Effect
ATM	Air Traffic Management
BADA	Base of Aircraft Data
BGO	Bergen Airport
CATE	Conditional Treatment Effects
CO	Carbon Monoxide
CO_2	Carbon Dioxide
COVID	Coronavirus Disease
CFSP	Collaborative Flight Planning System
CTR	Tower Control
EU	European Union
FIR	Flight Information Regions
GWP	Global Warming Potential
GTP	Global Temperature Potential
HC	Unburned Hydrocarbons
HFE	Horizontal Flight Efficiency
ICAO	International Civil Aviation Organization
IPCC	International Panel on Climate Change
NN	Neural Network
NO_X	Nitrogen Oxides
O-D	Origin-Destination
OSL	Oslo Airport
PM	Particulate Matter
RCT	Randomized Control Trial
R&D	Research & Development
RF	Radiative Forcing
SAA	Special Activity Airspace
SO_2	Sulfur Dioxide
SID	Standard Instrumental Departure
SSR	Sum of Squared Residuals
SST	Total Sum of Squares
STAR	Standard Arrival Route
TMA	Traffic Maneuvering Area
TMI	Traffic Management Initiative
US	United States
USD	United-States Dollar
VOC	Volatile Organic Compound

This page is intentionally left blank.

1 Introduction

In line with the increasing urgency in the sustainable transition of aviation, recent years have seen an increase in publications related to measures reducing the environmental impact of flight [1]. While novel aircraft propulsion systems would allow for zero-emission flight to be achieved, the implementation time frame of such solutions risks extending beyond current sustainable development goals. On the other hand, innovations aimed at reducing aviation's impact through operational measures are able to be implemented on a larger scale more quickly than their hardware innovation counterparts [2]. This fact highlights the significance of such innovations in the mitigation of aviation's impact as these would apply to all types of operational aircraft in the sky [3].

While previous studies have examined the environmental inefficiencies of Air Traffic Management (ATM) operations [4], and other optimal flight paths [5], the actual steps needed to, on an operational level, reduce aviation's environmental footprint remain largely unclear. This study is motivated by the desire to understand ATM-related environmental inefficiencies by bridging the gap between the sources of environmental inefficiencies and their causal effects. It aims to provide a scientific foundation for flight operators and air traffic controllers to identify the primary factors contributing to excess emissions in air travel and assess their impact. This knowledge will empower them to make informed decisions to pave the way for a sustainable future in aviation.

Ultimately, the project's objective is to assess the potential of causal analysis methods in uncovering the most significant inefficiencies in ATM operations that contribute to the negative environmental impact of aviation. The report is structured such that, through a scientific paper, Part I presents the final findings of the performed research, and Part II presents the surrounding literature which forms the scientific foundation of this study.

This page is intentionally left blank.

Part I

Scientific Paper

This page is intentionally left blank.

A Causal Inference Approach to Understanding Environmental Inefficiencies in Air Traffic Management

J.N. Aalders
Supervisors: Dr. I.C. Dedoussi, Dr. J. Sun, Dr.ir. F.D.A. Quadros, and Prof.dr.ir. M. Snellen

Sustainable Air Transport, Faculty of Aerospace Engineering
Delft University of Technology, The Netherlands

***Abstract*—Addressing the increasingly urgent need for sustainable aviation solutions, this paper explores operational innovations as a quicker and more scalable addition to novel zero-emission propulsion systems. Through the use of regression-based causal inference methods, this study aims to understand the relationship between flight fuelburn inefficiency and the factors causing these inefficiencies. Such an approach allows for the attribution of inefficiencies to factors on an overall scale, requiring less specific domain knowledge for initial results. A case study, involving a sample of 100,000 flights, representative of European operations, reveals that airspace structure (3.2% increase in inefficiency) and turbulence along the flight plan (2.5% increase) are the leading causes, while variations in average airspeed, congestion, and crosswind contribute the least to flight inefficiency. A compilation of the results shows that the performed analysis leaves 61% of the observed flight inefficiency unaccounted for. Future work would include the exploration of different metrics even closer to actual climate and air quality effects, as well as detailed uncertainty quantification. The developed flight inefficiency prediction model allows experimentation with counterfactual scenarios, contributing to the global transition towards more sustainable air transport networks.**

***Index Terms*—causal inference, sustainable aviation, air traffic management, machine learning applications.**

I. Introduction

The urgent need for sustainable aviation systems has prompted an increasing number of publications exploring measures to mitigate the environmental impact of flights [1]. While novel zero-emission propulsion systems are envisioned, their implementation time frame risks extending beyond current sustainable development goals. In contrast, operational innovations offer a quicker and more scalable approach to address aviation's impact [2], with the potential to impact all currently operational aircraft [3].

While previous studies have delved into the generation of optimal flight trajectories [4–10], the current state-of-the-art, in large part, treats flight trajectories as an isolated problem rather than considering them in a network form by taking into account potential interactions between flights. This simplification overlooks a crucial aspect of real-world Air Traffic Management (ATM) dynamics, limiting the implementation of such optimal trajectories into actual air transport operations [11].

Bridging the gap between the analysis of individual flights and the one of multiple flights, former studies have assessed the performance of ATM networks as a whole [12–15]. Previous work in this domain has primarily relied on simple metrics, such as route extensions, time, or capacity, with recent efforts focusing on more fuel-centric metrics [15]. However, the limited use of metrics closely tied to environmental impacts, such as fuelburn, has resulted in an incomplete understanding, often relying on proxy indicators.

Advancing the understanding of flight inefficiency, causal inference methods have, to a limited extent, previously been applied. These aim to attribute a portion of flight inefficiency to measured causal factors. Often these have been confined to a narrow scope, typically centered around a small number of routes [16], a single Area Control Center (ACC) [17], or a specific phase of flight [18]. To the best of the authors' knowledge, none have studied inefficiency at a continental scale nor considered fuelburn as the reference inefficiency metric. This work aims to further understand ATM-related inefficiencies by considering Europe-wide fuelburn in a causal inference analysis, providing a scientific foundation for flight operators and air traffic controllers to identify the driving factors contributing to the environmental impact of air travel.

With the current state-of-the-art, improving flight efficiency could be envisaged through a "bottom-up" approach. This entails determining feasible optimal trajectories for implementation in reality. While this approach allows for specific restrictions to be considered, it necessitates significant computational resources and information, leading to results that may not be easily generalized to a larger scale.

In contrast, this paper proposes a "top-down" approach to expand the scope of ATM performance and improvement studies through a detailed use of causal inference methods. By attributing causal relations to inefficiency factors on an overall scale, this approach allows for subsequent detailed analysis. The advantage of such an approach lies in its requirement for less specific domain knowledge to obtain initial results, providing relevant directions for future research.

Thus far, this paper's introduction has provided the motivation behind the performed research. Section II details the methods used for the analysis. In Section III the selected case study is defined with the results presented in Section IV. The cumulated findings of the paper are presented Section V with a discussion, together with the conclusion and recommendations found in Section VI.

II. Methodology

A. Causal Inference by Regression Modeling

1. General Causal Inference Framework

Causal inference is the discipline that aims to determine the cause-and-effect relationships governing variable change [19]. Not to be confused with correlation, which simply describes how two variables tend to change together.

A fundamental issue of causal inference is the inability to observe the outcome variable both in the presence of a certain treatment variable and what would have happened without this intervention of this same treatment variable [20]. The difference between these two outcomes can be attributed to the causal effect of the treatment variable. Commonly, Randomized Control Trials (RCT) are considered the most reliable method for quantifying this effect.

However, in applications such as the one of this study, it would be practically infeasible to perform such a trial. Alternatively, regression-based causal inference methods are proposed where regression models can be used to compare actual situations with counterfactual situations. Stemming from the "potential outcome framework" [21], the causal impact of treatment can mathematically be formulated as described in Equation 1.

$$T = E[Y^a - Y^{a*}], \tag{1}$$

where a and $a*$ represent the different exposure levels of treatment A, Y represents the potential outcome under these conditions, and T the treatment effect.

The robustness of causal inference findings is commonly quantified using p-values that estimate the probability that a more extreme value than observed from the test, is found assuming that the null hypothesis is true, i.e., no causal relation exists. Mathematically this is described in Equation 2.

$$\text{p-value} = P\left(|X| \geq |x| \, | H_0\right), \tag{2}$$

where X is the test statistic, x is its observed value and H_0 the null hypothesis.

2. Flight Inefficiency Prediction Model

Within the causal inference framework, a prediction model is developed to predict a flight's efficiency from input factors of interest. While previous work has implemented linear regression [16] and Random Forest [17], this paper proposes the use of a gradient-boosting algorithm (XGBoost [22]). This is because boosting algorithms can be less prone to over-fitting when applied to a wider range of data sets [23]. Additionally, the faster training time of the XGBoost algorithm would allow for the training of multiple models, offering comparison possibilities to this study. Hyper-parameter tuning of the model is done using an open-source Bayesian optimization package, Hyperopt [24]. The final set of hyperparameters is described in Appendix E.

The performance of the prediction model is quantified by considering to what extent the variations within the dataset are explained by the trained model, using the adjusted R^2 metric, as applied in the work of Liu et al. [16] and Marcos et al. [17]. Mathematically this metric is described in Equation 3.

$$\begin{aligned} R^2 &= 1 - \frac{\text{SSR}}{\text{SST}}, \\ \text{Adjusted } R^2 &= 1 - \frac{(1-R^2)\cdot(n-1)}{n-k-1}, \end{aligned} \tag{3}$$

where SSR represents the Sum of Squared Residuals, SST is the Total Sum of Squares, n, is the number of observations, and k is the number of features. For legibility, this paper will refer to adjusted R^2 as R^2.

B. Flight Features

The following subsection presents the features describing possible causes of inefficiency for a given flight. These input factors are then used by the regression model to predict the efficiency factor of a given flight.

1. Feature Definition

The features used as input for the flight inefficiency prediction model are initially compiled and iterated from domain knowledge in part acquired from the literature described in Section I. Feature definition iteration was then performed using a correlation matrix, to redefine, combine, or remove features with a Spearman correlation index above 0.8. The final correlation matrix is detailed in Appendix C.

This step is performed to reduce the possible effects of multicollinearity. While this phenomenon does not per-see affect prediction performance with forest-based methods, it may cause unreliable feature importance results, which in turn could have undesired effects during causal inference by reducing the independence and interpretability of features.

To be able to implement categorical information in a regression model, the data need to be encoded to a numerical value. For this application, the categorical data was encoded using target encoding. This method simply replaces the categorical feature with the average outcome ("target") value of all instances of this category. The choice for encoding is made due to its ability to handle high cardinal categorical features while preserving the overview of features, facilitating the creation of counterfactual scenarios required within the causal inference framework.

2. Causal and Mediating Factors

Within the final selection of flight features, a distinction was made between causal and mediating factors. While the former is the subject for which this study attempts to quantify the causal effect on flight inefficiency, the latter represents factors that may indirectly affect flight inefficiency. As such mediating factors do not necessarily affect flight inefficiency themselves, but may do so in combination with a causal factor. It should be noted that causal factors may also show mediating effects. The allocation of features to the causal factor or mediating factor group was done using domain knowledge.

C. Flight Sampling

Due to computational restrictions, it would be practically infeasible to compute flight inefficiency requiring sampling to be applied to the entire population of available flights. While previous research [16, 17] have opted to limit the scope of study to a manageable number of flights, this paper proposes the use of a scaled test to determine the required sample size for accurate estimation of results over the entire population. For the quantification of similarity, the Jaccard similarity index is used, as described in Equation 4.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}, \tag{4}$$

where A is the sample of size n, and B is the entirety of the available data. The index ranges from 0 to 1, where 0 indicates no similarity, and 1 identical data sets. The factors considered for similarity are described in Appendix D.

1. Determining Required Sample Similarity

The performed analysis relies on the assumption that similar data sets will provide similar causal inference results. From this assumption scaled causal inference simulations are performed where the error between causal inference using the entire scaled population set, and multiple different-sized scaled sample sets are compared. The different scaled samples are then related to the population set through the Jaccard similarity index. The relation between the different sample and population sets is described in Figure 1, where the entire population represents the totality of the available dataset, and scaling refers to random sampling from this dataset. The sample size ranges from 100 to the scaled population size, increasing in increments of 500 flights.

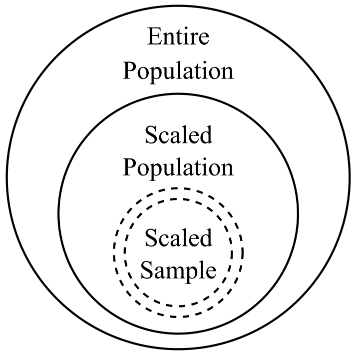


Fig. 1. Illustration of the relationship between the entire population, scaled population, and scaled sample sets.

For each scaled population set, the smallest possible similarity index, for which the causal effect error, of the scaled sample, is consistently within 5% of the causal effects computed using the scaled population. This is then repeated over multiple scaled population sizes. This analysis aims to gain insight into what the required Jaccard similarity index of a sample is, to return causal effects with an acceptable error margin if this was performed with the entire population. The size of the sample to population sizes is in this scaled test selected such that their ratio would be similar as one would expect with the entire dataset.

2. Determining Required Sample Size

To determine the required sample for this analysis, the Jaccard similarity index between the entire population of available data, and the different sample sizes using different sampling methods was determined. Two common sampling methods are considered: (1) random sampling, and (2) stratified sampling with the strata described in Appendix D. The minimal required sample size then corresponds to the sample size for which the required sample similarity, from the previous paragraph, is achieved.

III. Description of Case Studies

A. Considered Flights

1. Scope of Case Study

The case study was conducted for a year (2019) of pre-Covid European flights. This geographical selection was done by filtering for the flights for which the departure and arrival ICAO airport codes start with 'E' or 'L', while excluding airports on the French overseas territories, Portuguese Islands of Azores and Madeira, as well as the Norwegian island of Svalbard. Additional filtering of flights was applied due to limitations in data availability and aircraft performance models. In the end, only the four months available in the Eurocontrol R&D data [25] (March, June, September, and December) and flights with aircraft types compatible with the used aircraft performance model, OpenAP [26], were considered. Finally, military, business, general aviation, and flights with the same departure as arrival airport were excluded from this analysis.

Overall, 2,012,702 flights were considered before sampling. This resulted in an average of 17,000 daily flights considered in the analysis.

2. Sampling of Flights

Through the method described in Section II.C the final sample of flights was determined. Figure 2 plots the scaled population size, for the smallest sample's Jacquard similarity index required for the causal inference results of the sample, to be within 5% of the results returned using the entire population. The results show how the required Jaccard similarity index converges below 0.025 as the population size increases beyond 4,000 flights.

Figure 3 shows the Jaccard similarity index to the entirety of the dataset described in the previous paragraph, for a sample of a given size. It is found that stratified sampling and random sampling show similar performance for small sample sizes but that stratified sampling outperforms random when the sample size increases beyond 100,000 flights.

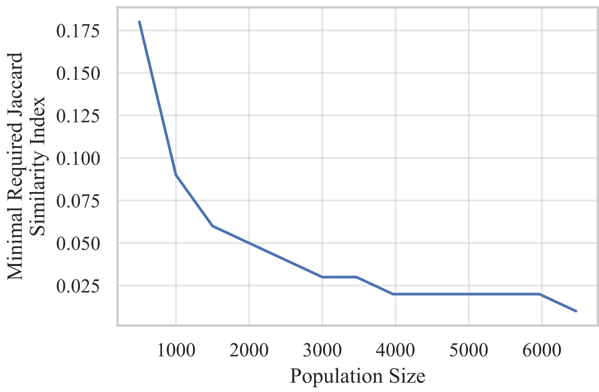


Fig. 2. Smallest required Jaccard similarity index for causal inference error to population size, to population size.

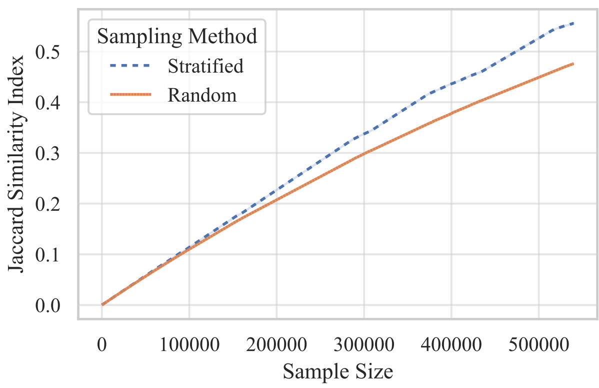


Fig. 3. Sample Jaccard similarity index to sample size.

The combination of both figures indicates that a sample size larger than 20,000 would lead to results similar to the ones retrieved from considering the entire population dataset. In addition, for this sampling size, the choice of sampling method would have a limited effect on the results. The final sample selected for causal inference is a random sample of 100,000 flights. This is because such a sample would be a manageable number of flights to run given the computational resources available for this research. This sample size would also provide more than enough data to account for the flight trajectory optimizer failing or other sources of error during the analysis.

B. Flight Inefficiency, Wind, and Airspace Data

1. Flight Inefficiency

Flight inefficiency at individual flight level was quantified by considering the ratio of actual over optimal fuelburn using OpenAP [26] and its corresponding trajectory optimizer [7]. This ratio-based flight inefficiency metric was selected due to it allowing for broader comparison between flights of different distances, as compared to excess fuelburn metrics. The actual flight trajectories were taken from the Eurocontrol R&D database [25] and up-sampled into a trajectory of 15-second resolution using the traffic library [27]. Flight inefficiency of the total sample of flights is defined by the sum of all actual fuelburn, over the sum of all optimal fuelburn. This is retrieved using the relation described in Equation 5,

$$y_{\text{sample}} = \frac{\sum_{i=1}^{N} y_{\text{flight}_i} \cdot FB_{\text{opt}_i}}{\sum FB_{\text{opt}}}, \tag{5}$$

where N represents all the flights in the sample, y_{sample} is the inefficiency ratio of the entire sample, y_{flight_i} the inefficiency ratio of flight i, and FB_{opt} is the optimal fuelburn for that flight. An overview of the number of flights considered and flight inefficiencies can be found in Table II, while the distribution of flight inefficiencies in the considered sample is attached to Appendix B. The discrepancies in the number of flights compared to the original sample size are a result of the flight trajectory optimizer, OpenAP.top [7], failing to converge.

A simple analysis of the flights, for which trajectory optimization has failed, indicates that the optimizer's performance was not disproportionately affected by extreme wind conditions. Additionally, no geographical biases in the failed set of flights were observed. When considering aircraft types, disparities in optimization success rates were found, however, the author deemed that attempting to correct for this bias may inadvertently introduce other biases. Moreover, the set of take-off mass ratios also shows different optimization success rates suggesting the potential for increased uncertainty in the causal upper (or lower) bounds of the causal relations. This impact appears negligible given the small observed variations in Section IV.A. Further details of the analysis on the failed set of flights are provided in Appendix F.

2. *Wind and Airspace Data*

Used both for determining fuelburn and for flight features, wind data is retrieved from the ERA5 hourly reanalysis model [28]. As the used optimizer only accepts windfields at a certain time, the considered windfield for each flight is the wind conditions for the hour of departure. This is to compare actual and optimal trajectories in the same windfield, together with limiting the computational load of the analysis.

Additional information used for flight features is obtained from the Eurocontrol R&D database [25] and includes the last filed flight plans, time spent in each Flight Information Region (FIR), aircraft (type) used, airline, as well as the previously mentioned actual trajectory flown by a given flight.

C. Selected Features

The final list of features utilized is presented in Table I while the final overview of feature definitions is presented in Appendix A. The final selection was based on domain knowledge and aims to provide a complete coverage of possible inefficiency factors while limiting possible overlap between causal factors. The details on how the final set of features is defined are presented in Appendix A.

TABLE I. Final selection of features.

Causal Factors	**Mediating Factors**
Wind	Departure Airport
Cross Wind	Arrival Airport
Turbulence	Aircraft Type
Airspace Structure	Airline
Departure Congestion	
Arrival Congestion	
En-Route Congestion	
Average Air Speed	

D. Final Flight Inefficiency Prediction Model

The achieved performance of flight inefficiency prediction models for each take-off mass fraction is detailed in Table II. The final R^2 of the model is in line with previous studies where the model's R^2 used in the study of Marcos et al. [17] ranges between 0.7 and 0.9, while the ones used in Liu et al. [16] range between 0.2 and 0.8.

The computed 18% flight fuelburn inefficiency is in line with the findings from Sun et al. [29] which applies the same trajectory data source and optimization algorithm to a smaller subset of flights between France and The Netherlands. Eurocontrol [15] estimates a fuelburn inefficiency of around 10% for which the difference can be attributed to the fact that for their estimate, the best-performing trajectory is considered as reference rather than the overall optimum, as is the case in the estimates presented in Table II.

E. Description of Counterfactual Scenarios

To quantify the causal effect of a given feature, the considered counterfactual scenario is the one where the influence of each inefficiency factor leads to minimal flight inefficiency. By utilizing the model detailed in Section III.D, one can analyze the scenario where a specific factor maintains a constant value across all flights. This analysis allows for the retrieval of the average flight inefficiency ratio for the entire set of flights, assuming this variable had remained constant throughout the entire dataset. Repeating this simulation over a wide range of values leads to the causal relationship between the influence of each factor, and the resultant flight inefficiency ratio.

TABLE II. Summary of flight inefficiency computations and prediction model results.

Take-Off Mass Fractions	*0.65*	*0.75*	*0.85*	*0.95*
Flights Considered				
Number of Flights	81,484	73,246	64,814	57,7476
Average Inefficiency	1.181	1.183	1.182	1.182
Model Performance				
R^2	0.501	0.520	0.536	0.550
Percentage Mean Error	0.09%	0.08%	0.08%	0.08%

IV. Results

A. Relation of Causal Factors to Flight Inefficiency

The following subsection provides insight into the relation between flight inefficiency and the different causal factors through their figures. Also, the distribution of the factor in the used data is plotted in the background. Each plot presented in this subsection is, assuming a normal distribution, cropped to show a range of two standard deviations (95% of values), of the counterfactual value range. Estimations are made for a range of take-off mass fractions between 0.65 and 0.95, denoted as m0.

1. *Wind*

Figure 4 provides the found relation between the along the Great Circle track wind and flight inefficiency. This relation suggests the minimum flight inefficiency to be around median tailwind ($-11.6\ m/s$) conditions with inefficiency steadily increasing upon deviation from this minimal value.

Figure 5 shows the relation between the cross Great Circle track wind, and flight inefficiency. The results indicate a slightly decreasing inefficiency between 0 and 13.2 m/s crosswind. Beyond this value, a small increase in inefficiency with increasing crosswind is found.

2. *Turbulence*

Figure 6 shows flight inefficiency to steadily increase with increased turbulence on the last filed flight path.

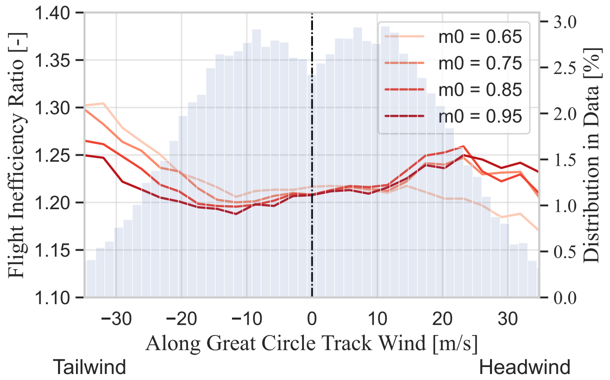


Fig. 4. Causal relation between along Great Circle track wind and flight inefficiency, accompanied by the distribution of along Great Circle track wind in the sample of flights.

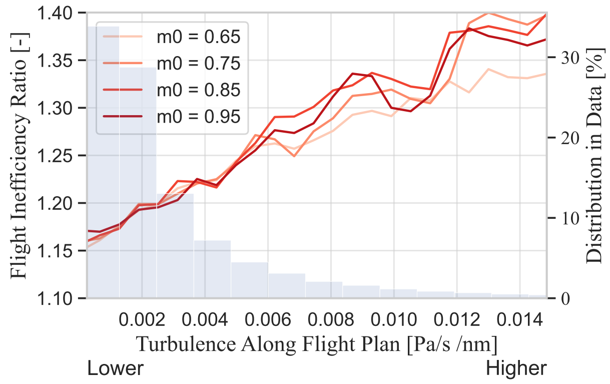


Fig. 6. Causal relation between turbulence along flight plan and flight inefficiency, accompanied by the distribution of turbulence along flight plan in the sample of flights.

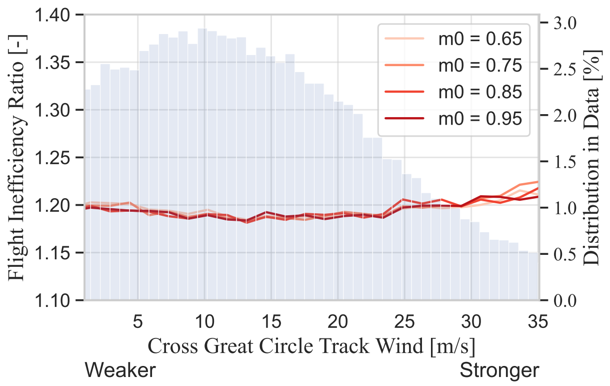


Fig. 5. Causal relation between cross Great Circle track wind and flight inefficiency, accompanied by the distribution of cross Great Circle track wind in the sample of flights.

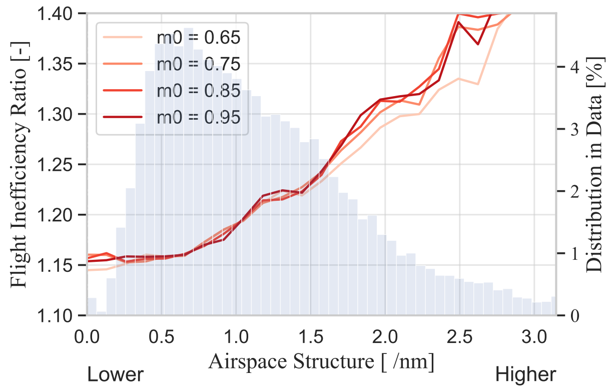


Fig. 7. Causal relation between airspace structure and flight inefficiency, accompanied by the distribution of airspace structure in the sample of flights.

Minimal flight inefficiency is found as the turbulence along the flight path reaches zero.

3. Airspace Structure

Figure 7 exposes a strong relation between airspace structure scores and flight inefficiency with the strongest increase in flight inefficiency after median airspace structure metric values. Minimal flight inefficiency is returned as the airspace structure metric reaches zero.

4. Excess Congestion

Figure 8, Figure 9, and Figure 10 respectively represent the departure, arrival, and en-route congestion. The former two relations indicate little causal relation between excess departure or arrival congestion on flight inefficiency. The latter plot shows minimal inefficiency to be at the median en-route congestion with operations becoming more inefficient when deviating from this median congestion. The increase in inefficiency is found to be larger for below-average than for above-average congestion.

5. Average Airpeed

Figure 11 shows the relation between the difference in the average airspeed per route and flight inefficiency. Here zero represents the average speed flown on that route. Average airspeed (zero on the plot) represents the least inefficient operations with deviations from this value leading to an increase in flight inefficiency. Surprisingly, a decrease in average airspeed leads to a sharper increase in flight inefficiency than an increase in average airspeed.

B. Causal Effect of Causal Factors

While Section IV.A provides insight into the relations between inefficiency factors and flight inefficiency, the following subsection attributes an inefficiency contribution to a given factor by setting its influence to the one where minimal flight inefficiency is found. This counterfactual value is retrieved by taking the minima, averaged over all four mass fractions, of the curves presented in Section IV.A. Table III presents a summary of the found causal effects for which the results are split between the average change of efficiency and the variance of the distribution of the considered flights.

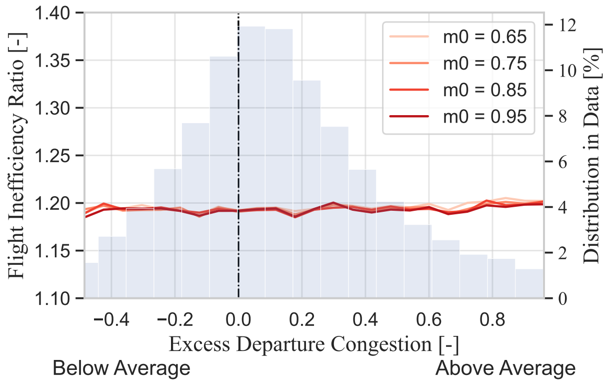


Fig. 8. Causal relation between departure congestion and flight inefficiency, accompanied by the distribution of departure congestion in the sample of flights.

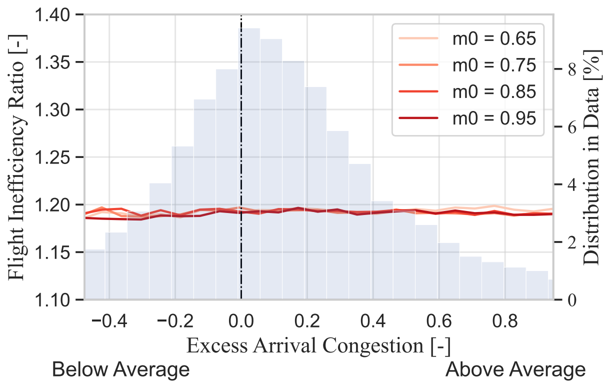


Fig. 9. Causal relation between arrival congestion and flight inefficiency, accompanied by the distribution of arrival congestion in the sample of flights.

When considering the effect of each factor, it was found that airspace structure, together with turbulence along the flight plan contributes the most to the inefficiency of pre-COVID flight operations. The low p-values associated with these computed effects confirm the presence of a causal relationship. On the other hand, the relatively high p-values associated with wind and average airspeed indicate that, from the performed analysis, no conclusive causal effects can be attributed to these features.

The change in outcome variance represents the extent the spread of flight inefficiencies, in the sample, is caused by a certain variable. The results show airspace structure and turbulence to be responsible for the most variation between flight inefficiencies, whereas congestion contributes the least. Interestingly, average airspeed is found to be responsible for a relatively large reduction in outcome variance considering its small causal effect. Furthermore, the uncertainties regarding the unknown take-off mass differ between factors where turbulence and along Great Circle track wind along the flight plan provide the largest uncertainty.

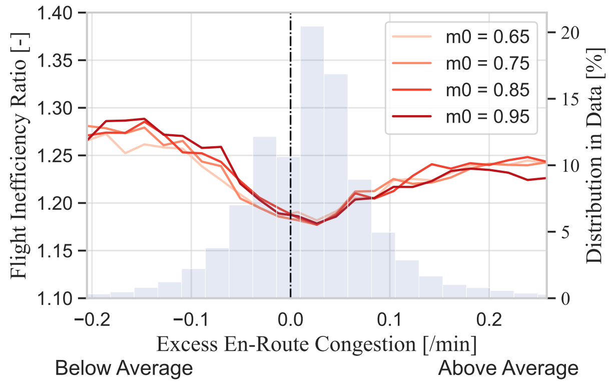


Fig. 10. Causal relation between en-route congestion and flight inefficiency, accompanied by the distribution of en-route congestion in the sample of flights.

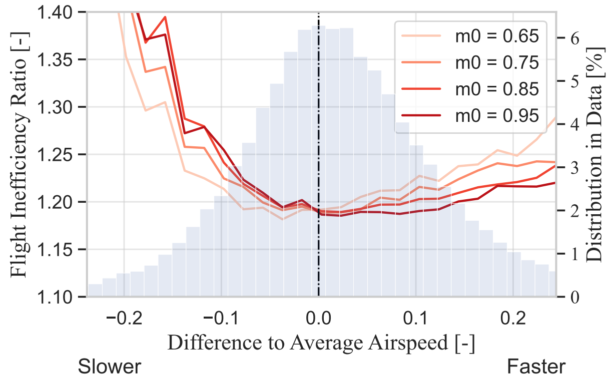


Fig. 11. Causal relation between difference to average airspeed and flight inefficiency, accompanied by the distribution of difference to average airspeed in the sample of flights.

C. Comparison of Total Flight Inefficiency Computations

The following subsection aims to provide insight into the completeness of the considered causal factors by comparing the total flight inefficiency using different methods. This comparison is shown in Table IV.

TABLE IV. Total flight inefficiency comparison including mass uncertainty.

Sum of Effects	*Minimal Effect*	**Reference Value**
-0.054 ± 0.026	-0.071 ± 0.020	-0.182 ± 0.001

Where the "sum of effects" represents the sum of each causal effect, the "minimal effect" value considers the counterfactual scenario where all the causal inefficiency factors are set to their minimum, and the "reference value" is retrieved by comparison of actual trajectories and optimal, as described in Section III.B.

By comparing the "sum of effects" and the "minimal

TABLE III. Summary of causal inference results.

	Causal Effect / Change in Outcome Value		Change in Outcome Variance
	incl. Mass Uncertainty	*p-value*	*incl. Mass Uncertainty*
Wind	0.012 ± 0.012	0.317	-0.016 ± 0.002
Crosswind	-0.003 ± 0.001	0.000	-0.014 ± 0.001
Turbulence	-0.025 ± 0.006	0.000	-0.041 ± 0.005
Airspace Structure	-0.032 ± 0.003	0.000	-0.043 ± 0.000
Departure Congestion	-0.002 ± 0.001	0.150	-0.010 ± 0.001
En-Route Congestion	-0.007 ± 0.001	0.000	-0.011 ± 0.001
Arrival Congestion	-0.000 ± 0.000	0.003	-0.011 ± 0.001
Average Airspeed	0.004 ± 0.002	0.264	-0.017 ± 0.001

effect", one can understand the level of overlap between factor effects, i.e., how much a portion of a factor's effect is also visible in another factor. When both these values are equal one would conclude that the set of features are completely independent. The case where the sum of effects is smaller than the minimal effect would indicate the presence of mediating relations between features while the opposite would indicate an overlap between factor definitions. The results, presented in Table III, show a 0.017 (24% of minimal effect) lower sum of effects than the minimal effect, suggesting the presence of mediating effects.

Similarly, one can compare "minimal effect" inefficiency to the "reference value" to understand how much of the observed inefficiency is covered by the selected set of causal factors. The results show a 0.111 (61% of the reference value) lower minimal effect inefficiency than the reference values for the same sets of flights, indicating that the selected features do not completely cover the seen inefficiency, through comparison of optimal and actual trajectories, in flight operations.

D. Sensitivity Analysis

Figure 12 shows a simple sensitivity analysis performed on the computed causal effects. Presented are the causal effects of the considered factors returned over multiple rounds of simulation, found when randomly splitting the available data in half, as well as the removal of one of the considered features. The findings from the sensitivity analysis suggest that the causal inference results are similarly sensitive to the data used as the inclusion of features.

Per feature, the variance in determined causal effects differs as well. For instance, the effects of airspace structure and turbulence show large variations while the other factors remain relatively constant. Additionally, certain effects, such as turbulence, show a larger sensitivity to feature removal than data split, possibly indicating an increased connection of this feature to other features.

When considering the average value of the causal effect of certain features, different causal effects are found from both sources of sensitivity. For instance, airspace structure shows on average a lower causal effect when subjected to data split as compared to feature removal, also indicating a possible increased connection between features. These hypotheses are further strengthened by the correlation matrix presented in Appendix C.

V. Discussion

A. Comparison with Literature

Summarized in Table III, this study found turbulence along flight plans and airspace structure to be the largest contributors to flight inefficiency in European airspace. While the reliability of such results is in large part dependent on the comparison to other publications, few papers are available on the subject.

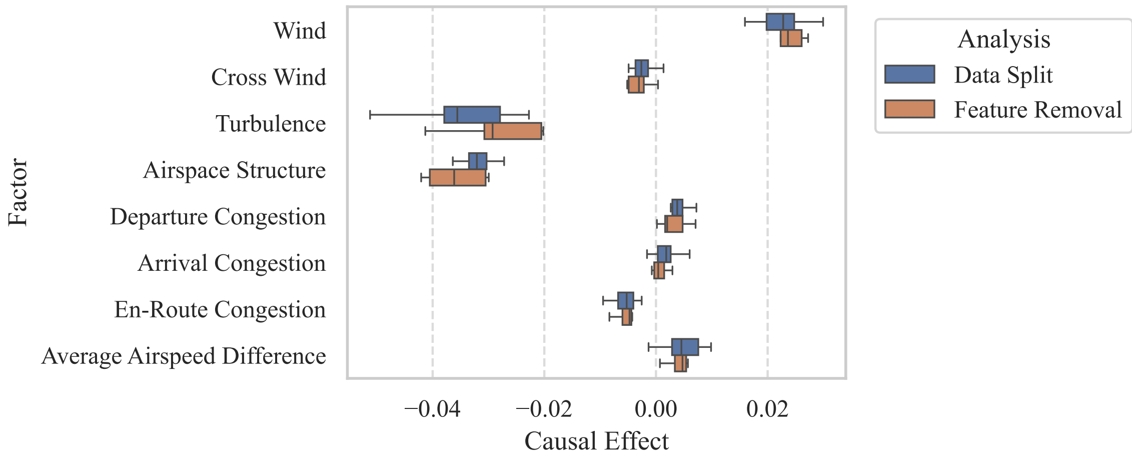


Fig. 12. Sensitivity analysis.

This discrepancy highlights the requirement for more research on the topic. The author is not aware of any papers that are directly comparable to this one, however, a pair of similar studies is available for limited comparison.

1. Weather Contribution

The work of Liu et al. [16] found over a limited selection of US-based origin-destination pairs that convective weather events and wind are the largest contributors to lateral flight inefficiency. While this research makes similar conclusions for turbulence, it can not attribute a similar causal effect to wind. This difference is not surprising as it can be explained by the fact that flights deviate from the optimal lateral routing to make use of favorable winds, as described by Vergnes et al. [6].

2. Airspace Structure and Air Traffic Control

Marcos et al. [17] performed a causal inference of cruise flight at the level of a single European Area Control Center. The results of that publication suggest that route structure is the most influential factor, and congestion a low influence. Which is in line with the results described in Table III.

B. Coverage of Selected Features

Section IV.C presents an initial insight into the coverage of the selected features. While the analysis, shown in Table IV, suggests that 61% of the observed inefficiency is not covered by the selection, it should be noted that the reverse statement, i.e., that 39% of the observed inefficiency is covered, does not hold. This is a consequence of the fact that the mediating effects of the not-included features, on the included features are unknown. It may be that through the addition of another feature, the percentage covered by the feature selection changes more than only the addition of the causal effect of this new feature. Therefore, the usage of this analysis can only be interpreted as a measure of completeness where no quantitative conclusions can be taken from.

Similarly, one is not able to extrapolate the extent of mediation between features, from the observed values in Section IV.C, due to unknown effects from the not included features. The current results underline the importance of considering these effects in future work.

C. Excess Congestion Definition

Figure 8, Figure 9, and Figure 10 consider the relation between congestion and flight inefficiency by considering the actual congestion of a flight in the hour of the flight. In reality, however, airspace capacity is managed by controlling flights in the hours before high congestion is expected to spread the load. Therefore, it may be that there is a misalignment between the used congestion values in this study, and the flights that were actually affected by congestion. This misalignment could be a potential explanation for the counter-intuitive increase in inefficiency seen at reduced en-route congestion observed in Figure 10.

D. Relation Between Airspeed and Fuelburn

Figure 11 at first glance presents somewhat surprising results for which the finding can be better understood when considering absolute fuelburn. At lower airspeed, the required thrust to overcome drag may not necessarily decrease as induced drag becomes more important. Combined with the longer flight times expected with a lower airspeed, the total fuelburn of the flight will increase. Alternatively, at higher airspeed, the induced drag will decrease while the parasite drag increases, causing a net increase in drag. Combined with a reduced flight time this could overall result in similar total fuelburn. The fact that the results show a change in the trend line around the average airspeed indicates that most flights operate at minimal drag conditions.

Another factor contributing to these results can be provided by the fact that flying at higher altitudes is often connected with flying at higher speeds and vice versa. Flights at higher altitudes will have to overcome less drag due to reduced air density, thereby requiring less fuel, and decreasing the inefficiency ratio.

E. Extension of Results to Excess Fuelburn

By directly comparing actual and optimal fuelburn, this work considers all flight control dimensions (lateral, vertical, speed, and time) thereby carrying over fewer of the limitations from previous studies which often utilize inefficiency metrics with a limited number of dimensions, such as lateral inefficiency. This in turn makes the results of Section IV present more relevant insights into the improvement potential of current operations.

That said, the fuelburn inefficiency ratio does present some limitations when considering absolute excess fuelburn. As such, inefficiencies are not evenly distributed through all flight lengths where shorter flights tend to have higher inefficiency values than longer ones. This however does not per se translate to large excess fuelburn quantities as shorter flights also tend to in total burn less fuel. Extending this point to the results of Section IV, means that in the current findings, shorter flights will be more represented than if one were to consider absolute excess fuelburn values.

F. Extension of Results to Environmental Effects

While this study provides an important step towards understanding the causes of fuelburn inefficiencies, the reader should be aware of the limited application of these results to actual climate and air quality impacts. The development of causal inference studies on these impacts will in large part be dependent on the development of rapid optimizers and estimators for these outcomes of interest. Due to the more complicated nature of these topics, including not only different emission species but also local atmospheric conditions, more variation in the outcome variable would be expected, requiring more data and features to provide accurate results.

G. Take-Off Mass Uncertainties

Most notable in Figure 4 and Figure 11, a flip in the "ranking" of causal relations returned from each mass fraction is observed. This particularity could be explained by the fact that on one side of the flip, the actual take-off mass may be underestimated (or overestimated) through the selected mass fraction range. As such for headwind conditions, the same mass range is likely to underestimate the actual take-off mass while for tailwind conditions, when take-off masses are more likely to be smaller, the mass range will be overestimating the actual mass.

In the same figures, areas of larger difference between the estimates from the four mass fractions can be noted. Conditions of larger differences could indicate areas of larger uncertainties related to the unknown take-off mass assumption.

H. Multitude of Sensitivities and Uncertainties

As this work combines many different models and data sources, the results are sensitive to a large part of uncertainties extending beyond the scope of the analysis. As such uncertainties regarding engine and aircraft-specific performance from OpenAP [26], or possible errors in the Eurocontrol [25] dataset are largely unaccounted for in this analysis. This, combined with a multitude of other uncertainties from the causal inference framework, presents a large limiting factor on the possible sensitivity analysis.

VI. Conclusions and Recommendations

Concluding this paper may lead to more questions than answers, however, that is exactly the added value of the proposed top-down approach to understanding flight inefficiency. What started as a simple comparison between actual and optimal flight trajectories has made evident several possible leads to explaining the reason behind these differences.

A causal inference analysis on a scaled set of 100,000 flights, representing all flights within a year of European flights, led to causal relations between flight fuelburn inefficiency and the causal factors of interest. Then, the causal effects of flight inefficiency in the European airspace were attributed to these factors. It is suggested that airspace structure (3.2% increase in inefficiency) and turbulence along the flight plan (2.5% increase) are the leading causes of flight inefficiency while changes in average airspeed, congestion, and crosswind contribute the least. Subsequent analysis of the results shows that in the current selection of inefficiency factor part of flight inefficiency can be attributed to mediating effects between factors. Additionally, it was found that 61% of the observed flight inefficiency remains unaccounted for.

Based on the results, recommendations for future research are provided. First, to take the next step in understanding the causes of environmental inefficiencies in air traffic management networks, it will be required to consider different inefficiency metrics even closer to actual climate and air quality effects.

Secondly, a detailed uncertainty quantification of the performed methods would be beneficial to provide a certainty range for the retrieved causal effects. By compiling many sources of data and various models, so have their errors and uncertainties been combined. This effect is currently in large part unknown.

Finally, different scopes of study should be considered to isolate specific areas of inefficiency together with their mediating effects. For instance, by only considering flights between one origin-destination pair, one can mitigate part of the observed effects of airspace structure in the final distribution of flight inefficiencies. Similarly, one could consider causal effects at individual flight levels to better understand the connection between retrieved causal effects and changes in spatiotemporal trajectories.

References

[1] T. Ryley, S. Baumeister, and L. Coulter, "Climate change influences on aviation: A literature review," *Transport Policy*, vol. 92, pp. 55–64, 6 2020.

[2] IATA, "Net zero 2050: operational and infrastructure improvements Fact sheet," Tech. Rep., 2022.

[3] IPCC, *Aviation and the Global Atmosphere.* Cambridge: Cambridge University Press, 1999.

[4] T. G. Reynolds, "Analysis of lateral flight inefficiency in global air traffic management," *8th AIAA Aviation Technology, Integration and Operations (ATIO) Conference*, 2008.

[5] C. Ramee, K. Junghyun, M. Deguignet, C. Justin, S. Briceno, and D. N. Mavris, "Aircraft Flight Plan Optimization with Dynamic Weather and Airspace Constraints," in *International Conference for Research in Air Transportation.* Georgia Institute of Technology, 2020.

[6] F. Vergnes, J. Bedouet, X. Olive, and J. Sun, "Environmental Impact Optimisation of Flight," in *10th International Conference for Research in Air Transportation.* FAA & EUROCONTROL, 2022.

[7] J. Sun, "OpenAP.top: Open Flight Trajectory Optimization for Air Transport and Sustainability Research," *Aerospace*, vol. 9, no. 7, 7 2022.

[8] Y. Lim, A. Gardi, and R. Sabatini, "Optimal Aircraft Trajectories to Minimize the Radiative Impact of Contrails and CO2," *Energy Procedia*, vol. 110, pp. 446–452, 3 2017.

[9] H. Yamashita, F. Yin, V. Grewe, P. Jöckel, S. Matthes, B. Kern, K. Dahlmann, and C. Frömming, "Analysis of Aircraft Routing Strategies for North Atlantic Flights by Using AirTraf 2.0," *Aerospace 2021, Vol. 8, Page 33*, vol. 8, no. 2, p. 33, 1 2021.

[10] E. Roosenbrand, J. Sun, I. Dedoussi, D. Stam, and J. Hoekstra, "Assessing and Modelling Climate Optimal Flights Using Open Surveillance and Remote Sensing Data," in *10th International Conference for Research in Air Transportation.* FAA & EUROCONTROL, 2022.

[11] A. Simorgh, M. Soler, D. González-Arribas, S. Matthes, V. Grewe, S. Dietmüller, S. Baumann, H. Yamashita, F. Yin, F. Castino, F. Linke, B. Lührs, and M. M. Meuser, "A Comprehensive Survey on Climate Optimal Aircraft Trajectory Planning," *Aerospace 2022, Vol. 9, Page 146*, vol. 9, no. 3, p. 146, 3 2022.

[12] T. G. Reynolds, "Air traffic management performance assessment using flight inefficiency metrics," *Transport Policy*, vol. 34, pp. 63–74, 7 2014.

[13] P. Pasutto, K. Zeghal, and D. Brain, "Initial analysis of vertical flight efficiency in cruise for European city pairs," in *AIAA Aviation and Aeronautics Forum and Exposition, AIAA AVIATION Forum 2021.* American Institute of Aeronautics and Astronautics Inc, AIAA, 2021.

[14] J. L. Leones, M. P. Morales, L. D'Alto, P. S. Escalonilla, D. F. Herrer, M. S. Bravo, F. C. Cámara, M. Mateo, B. MacNamee, S. Wang, A. Grover, and P. Plantholt, "Advanced flight efficiency key performance indicators to support air traffic analytics: Assessment of European flight efficiency using ADS-B data," *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, vol. 2018-September, 12 2018.

[15] EUROCONTROL, "Environmental Assessment: European ATM Network Fuel Inefficiency Study | EUROCONTROL," Tech. Rep., 2020.

[16] Y. Liu, M. Hansen, M. O. Ball, and D. J. Lovell, "Causal analysis of flight en route inefficiency," *Transportation Research Part B: Methodological*, vol. 151, pp. 91–115, 9 2021.

[17] R. Marcos, R. Herranz, R. Rodríguez Vázquez, P. García-Albertos, and O. G. Cantú Ros, "Application of Machine Learning for ATM Performance Assessment-Identification of Sources of En-Route Flight Inefficiency," in *Eighth SESAR Innovation Days*, 2018.

[18] J. B. T. Szenczuk, M. C. R. Murça, W. S. Sant'anna Souza, and R. A. Gomes, "Causal Analysis of Vertical Flight Inefficiency During Descents," *Transportes*, vol. 29, no. 4, p. 2486, 11 2021.

[19] X. Gejingting, J. Ruiqiong, W. Wei, J. Libao, and Y. Zhenjun, "Correlation analysis and causal analysis in the era of big data," in *IOP Conference Series: Materials Science and Engineering*, vol. 563, no. 4. Institute of Physics Publishing, 8 2019.

[20] A. Gelman, J. Hill, and A. Vehtari, "Causal Inference Using Regression on the Treatment Variable," in *Regression and Other Stories.* Cambridge University Press, 7 2020.

[21] D. B. Rubin, "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, vol. 66, no. 5, pp. 688–701, 1974.

[22] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

[23] W. Martinez and J. B. Gray, "The role of margins in boosting and ensemble performance," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 6, no. 2, pp. 124–131, 3 2014.

[24] J. Bergstra, D. Yamins, and D. D. Cox, "Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures," vol. 28, 2013.

[25] EUROCONTROL, "Structure and sample of the Aviation Data for Research Repository," 2020.

[26] J. Sun, J. M. Hoekstra, and J. Ellerbroek, "OpenAP: An Open-Source Aircraft Performance Model for Air Transportation Studies and Simulations," *Aerospace 2020, Vol. 7, Page 104*, vol. 7, no. 8, p. 104, 7 2020.

[27] X. Olive, "Traffic, a toolbox for processing and analysing air traffic data," vol. 4, no. 39, pp. 1518–1519, 2019.

[28] H. Hersbach, P. de Rosnay, B. Bell, D. Schepers, A. Simmons, C. Soci, S. Abdalla, M. Alonso-Balmaseda, G. Balsamo, P. Bechtold, P. Berrisford, J.-R. Bidlot, E. de Boisséson, M. Bonavita, P. Browne, R. Buizza, P. Dahlgren, D. Dee, R. Dragani, M. Diamantakis, J. Flemming, R. Forbes, A. Geer, T. Haiden, E. Hólm, L. Haimberger, R. Hogan, A. Horányi, M. Janiskova, P. Laloyaux, P. Lopez, J. Munoz-Sabater, C. Peubey, R. Radu, D. Richardson, J.-N. Thépaut, F. Vitart, X. Yang, E. Zsótér, and H. Zuo, "Operational global reanalysis: progress, future directions and synergies with NWP," Tech. Rep. 27, 3 2018.

[29] J. Sun, X. Olive, and D. Delahaye, "Evaluating Aviation Emission Inefficiencies and Reduction Challenges with Electric Flights Based on an analysis of flights from 2019 in the Dutch and French airspaces," in *12th SESAR Innovation Days*, 2022.

[30] J. Sun and I. Dedoussi, "Evaluation of Aviation Emissions and Environmental Costs in Europe Using OpenSky and OpenAP," *Engineering Proceedings 2021, Vol. 13, Page 5*, vol. 13, no. 1, p. 5, 12 2021. [Online]. Available: https://www.mdpi.com/2673-4591/13/1/5/htmhttps://www.mdpi.com/2673-4591/13/1/5

Appendix

A. Feature Definition

The following appendix details the definition of the features used in this paper. All metrics are shifted such that zero indicates no or average contribution depending on the features.

1. Wind

The along the Great Circle track wind indicates the contribution of wind to the required air distance to be flown by a given flight. This is computed by taking the component in line with the direct route between the origin and destination of the average wind component at these two locations. The other component of the average wind is the cross Great Circle track wind. Both are visually represented in Figure 13.

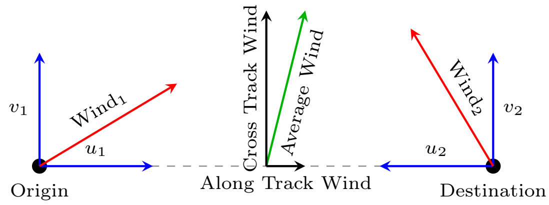


Fig. 13. Illustrative explanation of wind features.

2. Turbulence Along Flight Path

Turbulence along the flight path is defined as the sum of absolute vertical wind components along the last filed flight path. This can be expressed mathematically by Equation 6.

$$T_{\text{Flight Path}} = \sum_{i=1}^{n} |w_{\text{Vertical}_i}|\,, \tag{6}$$

where $T_{\text{Flight Path}}$ is the turbulence along the flight path, n is the number of data points along the last filed flight path, and w_{Vertical_i} represents the vertical wind component at the i-th data point.

3. Airspace Structure

The airspace structure metric aims to provide an estimate of how much a given flight's operations were affected by specific airspace-related restrictions. These include Standard Instrument Departures (SIDs), Standard Terminal Arrival Routes (STARs), entry/exit points between Flight Information Regions (FIRs), or areas of restricted airspace.

Air traffic constraint points are defined by, per origin-destination pair, counting the number of times a given coordinate is included in a flight trajectory. The points with frequencies beyond the 95th percentile are then taken forward as major air traffic constraint points, where the departure and arrival coordinates were excluded. The contribution of airspace structure is estimated by considering the number of major air traffic constraint points, from all origin-destination pairs, crossed by a flight per nautical mile flown, as described in Equation 7.

$$S_{\text{Airspace}} = \frac{\sum_{i=1}^{N} F_{\text{i}}}{d_{\text{flown}}}, \tag{7}$$

where S_{Airspace} is the airspace structure metric, N is the number of major air traffic constraint points, F_i the frequency of flights passing over that given constraint point, and d_{flown} is the actual distance flown by the aircraft of that flight. Figure 14 presents a visual example of flight routes with high and low constrained scores, ie., the airspace structure metric, on trajectories from Oslo (OSL) or Bergen (BGO), to Amsterdam (AMS).

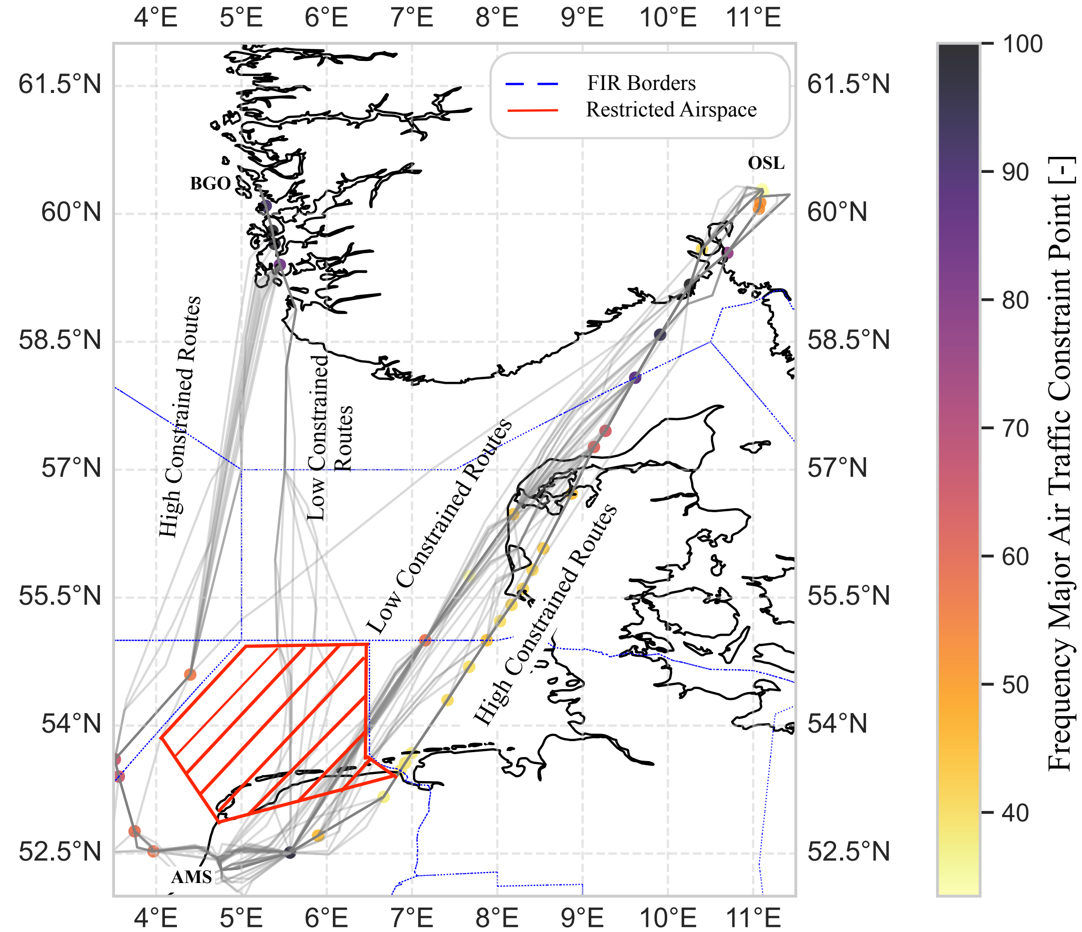


Fig. 14. Visual example of air traffic constraint points for flight departing from OSL and BGO towards AMS.

It should be noted that for legibility only the major constraint points computed for the origin-destination pairs OSL to AMS and BGO to AMS are plotted. In the actual analysis, all major constraint points from all included origin-destination pairs are considered. This is to be able to capture airspace-related restrictions from flights taking a vastly different route than the majority between these city pairs, providing a more complete picture of airspace structure. An example of this can be seen between OSL and AMS where a single trajectory passes westwards of the shown restricted airspace to join the last part of the highly constrained routes between BGO and AMS.

4. Departure and Arrival Congestion

The congestion values of a flight at departure or arrival are given by Equation 8.

$$C_{Dep/Arr} = \frac{N_{TO}|_h + N_L|_h}{\text{Average}(N_{TO} + N_L)|_h} - 1, \tag{8}$$

where $C_{Dep/Arr}$ is the congestion value, $N_{TO}|_h$ and $N_L|_h$ are, respectively, the number of take-off or landings at that given airport at a given hour h.

5. En-Route Congestion

Equation 9 gives the mathematical representation of the en-route congestion metric used in the analysis. It is defined as the sum of the percentual difference in congestion in a given Flight Information Region (FIR) multiplied by the time spent in that region along an entire flight,

$$C_{\text{ER}} = \left(\frac{\sum_{i=1}^{n} (N_{FIR_i}|_h \cdot T_{\text{FIR}_i})}{\sum_{i=1}^{n} (\text{Average } (N_{FIR_i})|_h \cdot T_{\text{FIR}_i})} \right) \cdot \frac{1}{\sum_{i=1}^{n} (T_{\text{FIR}_i})} - 1, \tag{9}$$

where C_{ER} is the congestion metric, $N_{FIR_i}|_h$ is the number of flights within that FIR at hour h, T_{FIR_i} is the time spent in that FIR of the flight of interest, and n represents the different FIRs crossed by the flight of interest.

6. Average Airspeed Difference

The metric used to describe the difference in flight speed is mathematically described in Equation 10.

$$S_{diff} = \frac{\frac{d_{act}}{t_{act}} - w_{\text{along track}}}{\text{Average}\left(\frac{d_{act}}{t_{act}} - w_{\text{along track}}\right)|_{O-D}} - 1, \quad (10)$$

where S_{diff} is the speed metric, d_{act} the actual distance flown, t_{act} the actual flight time from take-off to landing, $w_{\text{along track}}$ the along Great Circle track wind component, and $O-D$ represents a given origin-destination pair.

B. Distribution of Computed Flight Inefficiency

Figure 15 shows the right-skewed distribution of flight inefficiency ratios for the considered sample of flights. The ratios were computed for different mass fractions (m0) and correspond to the actual over optimal fuelburn returned from OpenAP [30].

The difference in peak height between mass fractions is due to the differences in the number of flights per mass fraction. Noteworthy is the fact that a small portion of flights (2.5 %) show flight inefficiency ratios below one, which irrationally would indicate the optimal operations are less efficient than the actual ones. This comes from modeling uncertainties related to the unknown specifics of a given flight such as, among others, the take-off mass, specific engine, or air-frame performance. These uncertainties can then lead the model to return flight inefficiency ratios below one.

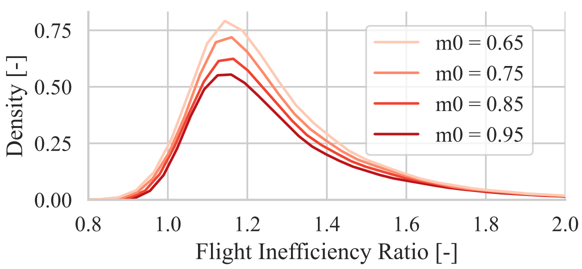


Fig. 15. Distribution of computed flight inefficiency for the considered sample of flights.

C. Feature Correlation

Figure 16 presents the final correlation matrix used to select and define the final set of features to be used in the flight inefficiency prediction model. Important is to note that no correlations above 0.8 are found.

D. Similarity Strata

Table V presents the strata used to determine sample similarity with each other. The selection of these strata was made using domain knowledge to, with readily available data, cover the main driving areas leading to flight inefficiency.

TABLE V. Strata used to determine Jaccard similarity.

Similarity Strata
Peak/Off-Peak Hours
Departure Airport
Arrival Airport
Month
Weekday
Hour of Day
Aircraft Type
Airline

E. Hyperparameter Tuning

The final hyperparameter values used in the developed prediction, as well as the range for which the parameter was tuned, are described in Table VI.

TABLE VI. Hyperparameter tuning results.

Hyperparameter	Range	Final Value
Learning Rate	1×10^{-10} to 1	1.90×10^{-1}
Max Depth	2 to 24	20
N Estimators	10 to 999	809
Subsample	0.5 to 0.8	0.79
Colsample Bytree	0.5 to 0.8	0.60
Scale Pos Weight	0.01 to 2	1.34
Reg Lambda	0 to 299	58
Gamma	0 to 299	0

F. Failed Flights

A. Wind and Geographical Location

Figure 17 presents the distribution of failed to successful flights against the min-max normalized wind metrics, plotted on the y-axis, including all mass fractions. The distribution of flights is normalized such that the areas of failed and successful flights are the same. In the plot, it can be observed that the distribution of flights does not vary significantly between failed and successful flights indicating that, most likely, no bias from extreme winds has inadvertently been induced by the optimizer failing. Furthermore, when plotting the routes of the failed flights on a map, no concentration was found along a geographical location which could indicate the presence of geographical biases.

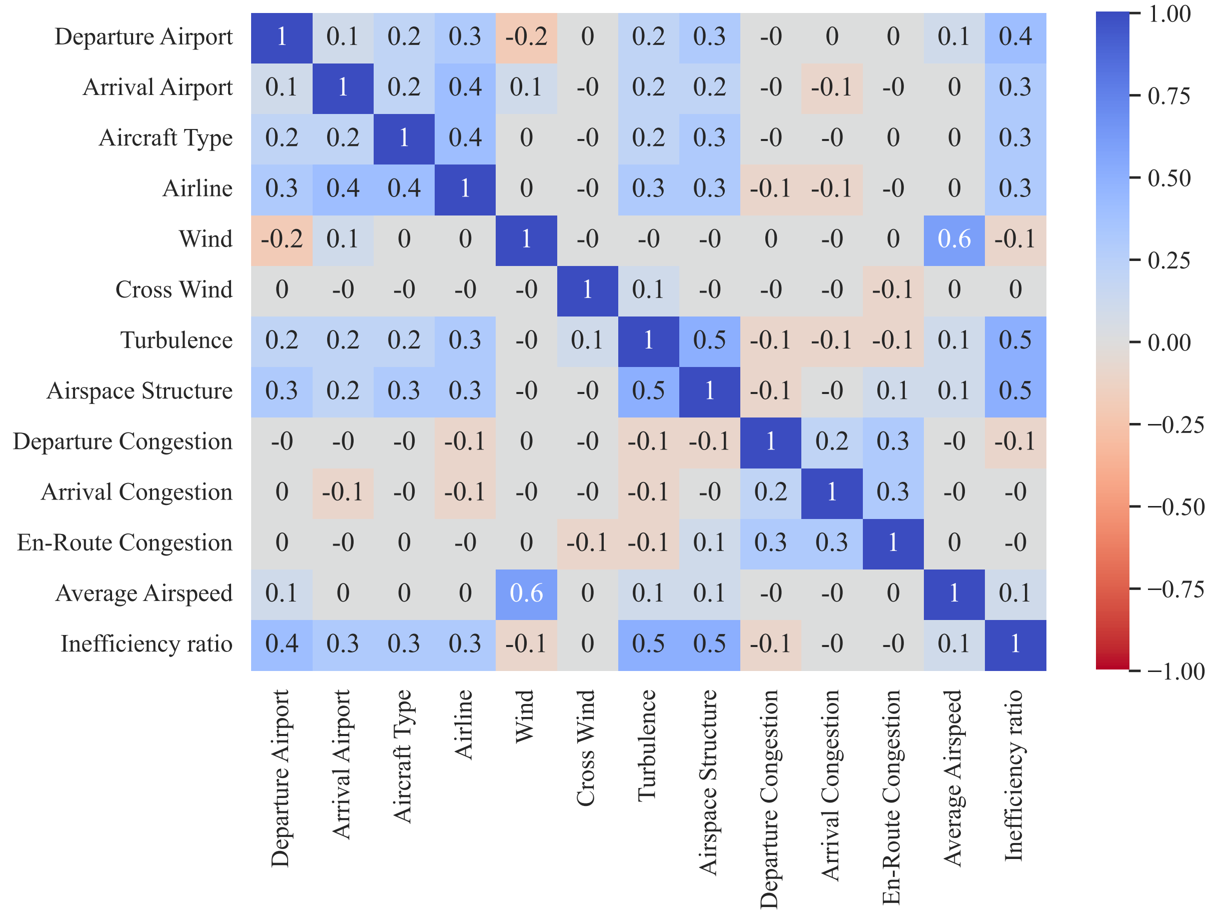


Fig. 16. Spearman correlation matrix of causal features included in the model.

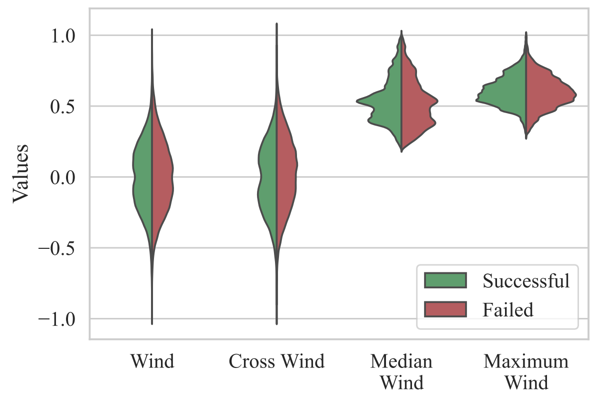


Fig. 17. Comparison of normalized wind metric distribution in flights where optimizer has failed or succeeded for all mass fractions combined.

B. Aircraft Type and Take-Off Mass Fraction

Figure 18 presents the distribution of failed to successful flights to aircraft type codes including all mass fractions. It can be noted that certain aircraft types are over- or under-represented in the set of successfully optimized flights. As such, in the final results, A320 operations are over-represented compared to other aircraft type operations. While the argument could be made that this bias should be corrected by readjusting the distribution of aircraft types in the final set of selected flights, doing so may inadvertently cause other biases unknown to the author. For this reason, this bias was accepted in the results.

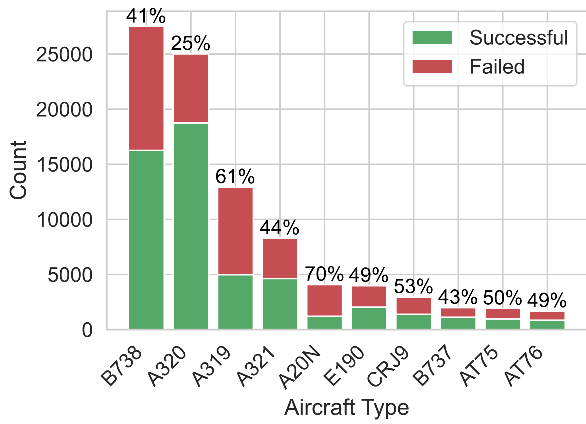


Fig. 18. Comparison of aircraft type distribution in flights where optimizer has failed or succeeded for all mass fractions combined, with annotated the failed percentage.

This page is intentionally left blank.

Part II

Supplementing Literature Review

This page is intentionally left blank.

2
Environmental Impact of Aviation

Understanding the environmental inefficiencies of ATM networks requires the understanding of what elements of flight negatively impact the environment. It is then important to consider how these effects are quantified as well as what the theoretical environmental optima would look like. The following chapter provides the relevant state-of-the-art for the proposed research topic of this study.

2.1. Sources of Environmental Impact

When considering the environmental impact of aviation, this can be split into two main areas: the emission of pollutants and the production of excess noise [3]. The former impacts the Earth's air by inducing climate change as well as adverse health effects related to the worsening of air quality, while the latter also causes negative health effects next to annoyance for surrounding communities [6]. Within the scientific community, research in this field is often split into three subjects: climate impact, air quality assessment, and excess noise studies.

In this study, a focus will be placed on the emissions side of aviation's impact, measured by the excess fuel burned due to flight operations. The considerations supporting the use of fuel burn instead of more direct climate or Air Quality (AQ) metrics rely on the added complexity required by having to combine an atmospheric model with emission data, which is required to incorporate detailed environmental impact metrics at the level of the current state-of-the-art. A significant limitation of this choice is that it will not be possible to take into account the effects of contrail-cirrus cloud formation. Noise considerations were also not taken into account as within the scope of focusing on the cruise portion of the flight, the effects of noise disturbance will be negligible.

A review of scientific literature [7] has highlighted the sector's focus on the climate impact of aviation, and in particular the resultant CO_2 emissions of aero-engine combustion during flight. However, as of the 1999 Special International Panel on Climate Change (IPCC) report [3] it has been known that non-CO_2 effects from aviation fuel burn emissions represent a significant portion of aviation climate impact. This discrepancy emphasizes the need to further advance the understanding, and awareness, of aviation's environmental impact beyond CO_2 emissions.

More recent studies, confirm the large effect of non-CO_2 effects, such as the one of Lee et al. [8] for which the current estimate of aviation is that it is responsible for approximately 5% of human-induced Radiative Forcing (RF). Figure 2.1 provides an overview of the relative elements contributing to the climate impact of aviation globally. An important takeaway from this figure is to consider the large confidence intervals for non-CO_2 effects in the current knowledge which can make studies relying on accurate estimation of aviation impact on the environment, including non-CO_2 effects, difficult. For the sector to reduce its contribution to global warming, it will be required to operate at net-zero CO_2 emissions while also being able to reduce the warming effects from non-CO_2 emissions. This advance in the understanding of non-CO_2 can also be identified as a major trend within the scientific community.

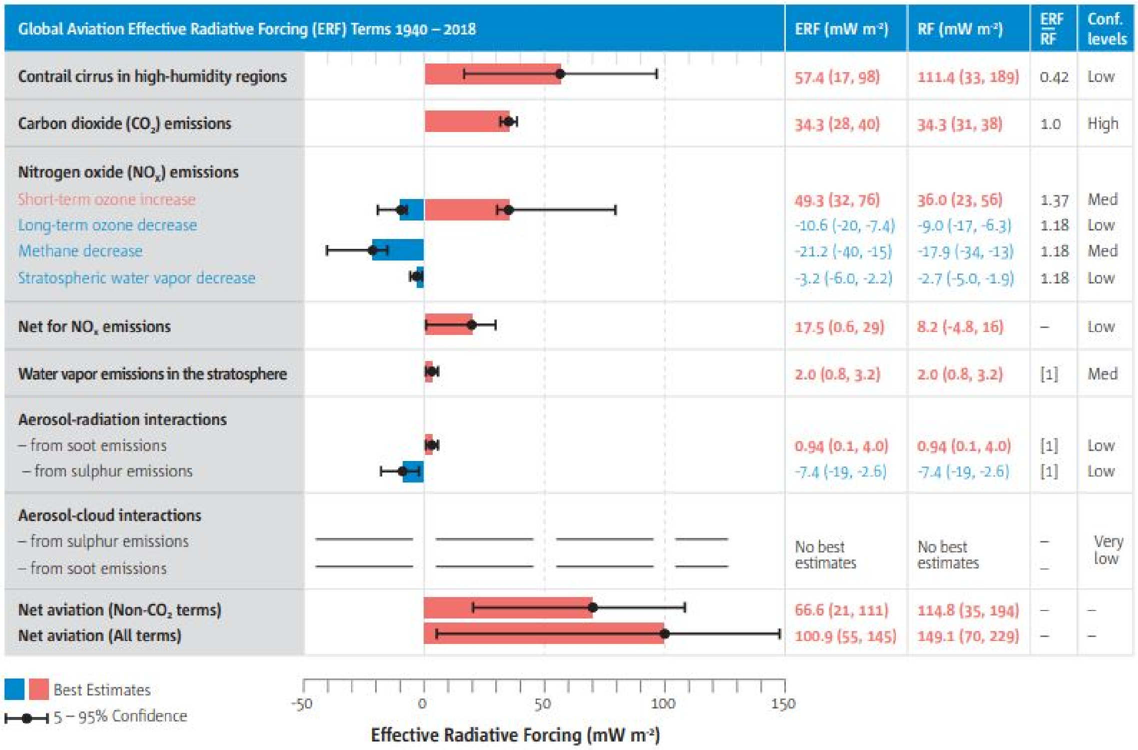


Figure 2.1: Best estimates for climate forcing terms for global aviation from 1940 to 2018 [9] [8].

Regarding air quality degradation considerations, in addition to carbon dioxide, aircraft engines emit a range of pollutants, such as nitrogen oxides (NO_X), particulate matter (PM), sulfur dioxide (SO_2), volatile organic compounds (VOCs), carbon monoxide (CO), and unburnt hydrocarbons (HC), which upon breathing have been linked to adverse health effects. These adverse health effects include impaired immune, cardiovascular, and respiratory functions, next to an increased response to allergens. Additionally, the emission of these species can lead to environmental damage through the formation of smog [10] [9].

That said, it is important when assessing the results of such studies to consider that a reduction in fuel burn will, due to the different effects of emissions species, not necessarily result in a reduced detrimental climate or air quality impact [3]. Additionally, a reduction in air quality impact does not always entail a reduced climate impact and vice versa. This particularity comes from the climate impact of contrail formation which may in some cases overrun the benefits achieved from air quality improvements from total emission reductions.

In section 2.2, an overview of the identified quantification methods in literature will be provided. Together with what sources of impact for the envisaged scope of this study a decision can be made to environmental impact metrics.

2.2. Quantification of Excess Environmental Impact

In the context of quantitatively assessing the surplus environmental impact of aviation, it is required to understand the quantification of the impact associated with a specific (set of) flight(s), as well as delineate the theoretical minimum. This section outlines the different methodologies identified in the existing literature.

2.2.1. Determining Flight Emission

Various methods exist to translate flight activities into emissions with or without the inclusion of positional information as input. The first type of methods employed aim at simulating actual aircraft trajectories, and from there determining the engine emissions using accurate aircraft performance models such as the Base of Aircraft Data (BADA) [11], or an open-source alternative such as the one developed by Sun et al. [12]. An advantage of such models is that they are able to take into account the different engine settings, particularly relevant for the estimation of CO, UHC, NO_X, and soot emissions. This combined with actual flight profiles increases the accuracy of the resulting estimations.

Other methods use important simplifications to extract emission estimates while using fewer computational resources, such as the one employed by Simone et al. [13], which utilized historical airport departure and arrival information to gauge aviation emissions. More recent examples of such a method being applied are found in the studies of Quadros et al. [14] and Seymour et al. [15]. Alternatively, as highlighted in a Eurocontrol report [16], a flight phase-focused approach that approximates an entire flight to the sum of different flight phases, in combination with time or distance performance indicators, could also be used for rapid emission estimation.

Depending on the performed analysis a trade-off would therefore need to be made to determine the required level of detail. While trajectory-simulation-based techniques provide accurate spatiotemporal and magnitude information on engine emissions, they require significantly more computational resources and are often restricted by data availability. A consequence of this limitation is that only a limited amount of flights can be taken into account, as opposed to methods simplifying flight movements.

2.2.2. Impact Quantification

With emission information, a global impact estimation could be made. As aviation emissions alter Earth's chemical concentrations in the atmosphere, the spatial-temporal emission information can then be combined with atmospheric data and models (or assumptions) to determine the global climate impact [17]. As mentioned in section 2.1, combining emission information with atmospheric data extends beyond the scope of this study due to its complexity and required computation resources which will be required to achieve this level of accuracy.

Fundamentally, however, the climate and air quality impact of aviation is not simply proportional to the amount of emissions as also briefly explained in section 2.1 [8] [18]. With the aim of advancing the body of knowledge in this field, it would be desirable to nonetheless consider the effects on climate and air quality degradation due to aviation operations. A suitable method found in the literature would be to consider the societal climate and air quality impact as derived in the work of Grobler et al. [19], and employed in the study of Sun et al. [20].

This metric assigns a price (USD) for the societal costs induced by the increased Radiative Forcing (RF) for the climate impact cost, and mortality for the air quality cost, both a consequence of the pollution induced by the aero-engine emissions. Combined with the ability of aircraft performance models such as OpenAP [12] to return emission information this would provide a rapid, and initial, insight towards non-CO_2 effects at play, allowing for broader generalization of this study's results. For cruise-level emissions in the European domain, the costs correspond to the ones presented in Table 2.1 [20] [19].

Table 2.1: Aggregated marginal climate cost and air quality cost due for cruise-level emissions based on Monte Carlo simulations for the EU domain from [19] used in [20].

Emissions Species	Unit	Climate Cost	Air Quality Cost (EU)
CO_2	USD/tonne CO_2	45	-
H_2O	USD/tonne H_2O	2.8	-
SO_x	USD/tonne S	-20 000	42 000
NO_X	USD/tonne NO_X as NO_2	-940	31 000

Table 2.1 continued from previous page

Emissions Species	Unit	Climate Cost	**Air Quality Cost (EU)**
CO	USD/tonne CO	-	270

2.2.3. Optimal Reference Trajectory

To be able to quantify a given "inefficiency" one will be required to define an "optimal" such that one can compare actual observations. In literature, three different methods of doing this have been identified (1) flown distance reduction, (2) fuel burn reduction, and (3) climate impact reduction.

The first form of inefficiency quantification involves comparing the flight's "ground distance", that is, the total distance on the ground traced directly below the aircraft. In theory, the shortest route would then be the greater-circle route between the origin and the destination airport. While this is a simple and efficient method to estimate the most optimum route, study results may be misleading when efficiency gains due to wind are not taken into account [21].

The second method builds on the shortcomings of the first and involves reducing fuel burn emissions by minimizing air distance instead of ground distance. The advantage of this form of optimization is that it accounts for all control dimensions (lateral, vertical, speed, and time), and through such better links to environmental consequences [21]. Additionally, this method is able to include weather phenomena as is the case in the optimizer developed by Ramee et al. [22]. Current airline practices indicate a preference for flight plans that align with optimized fuel burn routes, as it reduces operational costs [23]. Additionally, comparing actual flights to these optimal operations would provide a better insight into environmental consequences compared to the first described method. To make these studies possible, open-source optimizers such as OpenAP.top [24] have shown promising results in providing optimal routing comparable to closed-source alternatives. An additional step that could be taken in sustainability studies consists of utilizing emission species to infer climate or air quality metrics without the inclusion of detailed atmospheric models or data, such as described in subsection 2.2.2 and Table 2.1. This principle is also applied by OpenAP.top [24] to perform trajectory optimizations on various outlooks (20, 50, and 100 years) of Global Warming Potential (GWP) and Global Temperature Potential (GTP). Methods that do include detailed atmospheric models or data are summarized below.

The third method, and the most recent development in this field, consists of routing for reduced climate impact. While routing for reduced fuel burn will reduce direct emissions from aviation, it may be so that through this method flights are routed through climate-sensitive regions, where warming contrails could be formed. For this reason, various studies, such as the one from Lim et al. [25], Yamashita et al. [26], or Roosenbrand et al. [27] have looked into developing free routing models that consider these effects by combining satellite remote sensing, atmospheric science, and aircraft surveillance data. One main advantage of these types of studies is that it is able to include the climate effects of contrails, however, the resulting estimations do show a considerable amount of variability originating from variations in weather conditions [28]. As a result, it may be difficult to include these effects in the quantification of an inefficiency factor, as obtaining consistent results would be more challenging.

A point to underscore regarding the current state-of-the-art in optimal flight trajectories is the focus on addressing flight inefficiency as an isolated optimization problem. Existing studies have often not accounted for the potential interactions that may arise when multiple flights operate in a network system, thereby simplifying a crucial aspect of real-world air traffic management dynamics. This scientific gap in the literature limits the implementation of current optimal trajectories into actual air transport operations [29].

2.2.4. Quantification Uncertainties and Challenges

Exact answers to inefficiency quantification questions are however hard to provide the inherent uncertainties regarding the results. The reviewed literature highlights the following considerations.

Most apparent in analyses of multiple short flights, variations in trajectories lead to large variations in the estimated inefficiency results making accurate conclusions more challenging [20] [30]. When considering excess fuel burn metrics, it should additionally be noted that variations in vertical flight profiles will be additional sources of variation, as denoted in the study of Pasutto et al. [31].

Secondly, the lack of access to essential data regarding take-off mass and cost index settings, which is necessary for calculating fuel burn inefficiency, can pose challenges for the study. This data is often restricted due to commercial sensitivity concerns. Sun et al. [32] for this reason accounted for the unknown take-off mass by considering a masses fraction between 0.65 and 1 times relative to the maximum take-off mass. Finally, for full flight trajectory simulations such as the ones utilizing BADA [11] or OpenAP [12] significant computational resources are required potentially limiting the scope of possible analysis.

This page is intentionally left blank.

3

Air Traffic Management Inefficiencies

While chapter 2 dives into the specific environmental impact of a flight, the following subsection aims to bridge the gap between studying one aircraft and the operations required for a network of multiple flights. This is done by assessing where additional sources of (environmental) inefficiencies come into play in an Air Traffic Management (ATM) network.

The study of Reynolds [4] presents an overview of the various sources of flight inefficiencies in the different flight phases, simplified in three areas (1) Departure terminal area, (2) en-route, and (3) arrival terminal area, as seen in Figure 3.1. Although this classification offers valuable information about potential causes of inefficiency in a flight's time distribution, linking these sources to stakeholders who potentially could act on these identified inefficiencies becomes more challenging. For this reason, in this study, it is proposed to classify ATM system inefficiencies into four categories (1) Weather-related, (2) Airspace-related, (3) Air Traffic Control-related, and (4) Airline-related.

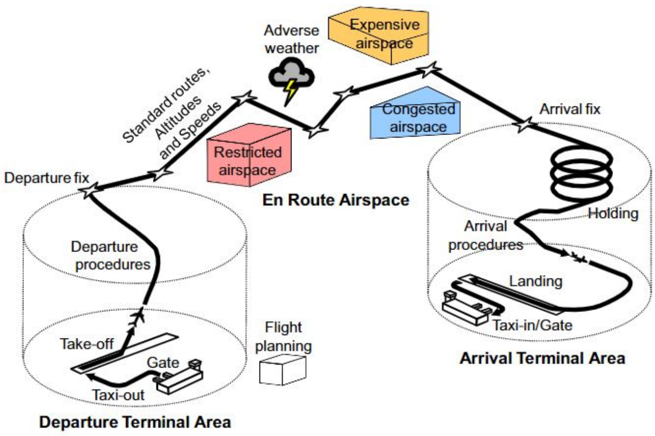


Figure 3.1: Overview of potential causes of flight inefficiency [4].

The following section will present how each of the four used inefficiency categories affects flight efficiency, as well as how these factors have been quantified in scientific literature.

3.1. Weather-Related Inefficiencies

The contributing factors to this type of inefficiency are the effects of convective weather phenomena and wind. Convective weather can force aircraft to deviate from their intended flight path to avoid turbulence, while wind conditions can result in longer air distances being flown due to sections of the flight being flown into a headwind.

Liu et al. [33] employed for their analysis US-based data sets consisting of both wind field information as well as an hourly summary of convective events. The effect of wind on a given flight was then quantified by considering the sum over the entire flight of the individual wind field components in line with the nominal direction of flight. Through this, an estimate of the additional or reduced “air-distance” contribution to the flight can be made. In the same study, the effect of convective weather is quantified by an exposure metric proportionally weighted to the distance between the convective phenomenon and the aircraft flight track. For EU-based studies considering weather effects, the ERA5 data repository [34] is commonly used.

3.2. Airspace Related Inefficiencies

This type of inefficiency encompasses various kinds of flight restrictions that arise from the organization of the airspace, such as predefined routes, restricted or special-use airspace, and route deviations due to the partitioning of airspace into different Flight Information Regions (FIRs). The author proposes to classify airspace-related inefficiency factors into two main groups (1) structures for workload distribution, and (2) fixed routes for traffic simplification. The latter can be both during en-route phases and during departure or arrival procedures through Standard Instrumental Departures (SIDs) and Standard Terminal Arrival Routes (STARs). The former refers to the separation of airspace and tasks into different (upper) FIRs, Traffic Maneuvering Areas (TMA), and Tower Control (CTR).

Previous studies have looked at different factors to understand specific sources of airspace-related inefficiencies better. For instance, the study of Pasutto et al. [31], considered the inefficiency contribution of the airspace through comparison of the “best-performer”, defined as the best achievable profile considering all constraints between O-D pairs, and the “ideal performer” defined as the theoretical most optimum profile. The best-performer route was determined by the routing employed by the 10% of best-performing flights per season.

Another approach is one implemented in the study of Marcos et al. [35] which considered, among others, a Horizontal Flight Efficiency (HFE) metric that could be split into two components (1) local extension, i.e. the extra distance flown compared to the shortest possible distance between entry/exit points within an FIR, and (2) the interface contribution, i.e. the distance between the ideal entry/exit point for a flight, and the planned entry/exit point. It should be noted that for the classification of inefficiency sources, the local components can directly be attributed to the sector in question while another sector could also influence the interface component.

3.3. Air Traffic Control-Related Inefficiencies

Air Traffic Control inefficiencies, in essence, arise from two different reasons, the first originates from the various measures implemented to ensure separation between aircraft, ranging from vectoring to holding patterns. The other reason could be requests made from the flight deck for airline-specific efficiency reasons such as a change in cruise altitude or a direct-to-way point. Commands given by an Air Traffic Controllers (ATCo) are typically one of three options: heading, speed, or altitude changes.

Studies such as one of Olive et al. [36] present both rule-based and statistical methods to detect ATC commands such as en-route flight plans and arrival holding patterns. While these methods would provide the insight required for a detailed analysis of the role of ATC during a flight, other studies have employed proxy indicators to estimate the overall influence of ATC commands during a flight. For instance, Marcos et al. [35] approximated the impact of ATC by considering congestion in a given FIR or the time of departure (peak/off-peak hours).

3.4. Airline Related Inefficiencies

The choices made by airlines before and during the flight, such as the selected route, velocity, and requested flight level are influential inefficiency factors. In essence, operators seek to reduce their direct operating costs, which encompass various factors such as fuel costs, flying time costs, and route charges [16]. This may result in the operation of flights over longer distances, leading to increased fuel consumption, in order to save costs on route charges (related to Air Traffic Management), or flying at lower altitudes to reduce flying time costs (related to airline internal factors). For this reason, flight plans submitted by airlines may not necessarily represent the most fuel-efficient routes due to factors such as the cost index, speed, fuel and flight planning policies, and the use of Collaborative Flight Planning System (CFSP) software [16]. Specific to European studies it should be noted that the effect of ATC charges is more significant due to large differences happening over a small area [4].

In the available literature, the work of Leones et al. [37] proposes revised ATM Key Performance Indicators (KPIs) that introduce the airspace user's viewpoint. This analysis made use of equity indicators to capture how inefficiency indicators are spread between different airlines. This is done by looking at the mean inefficiency for all airlines combined, and then the difference of each airline's specific means to highlight different underlying strategies.

A factor interesting to highlight from this study is the different strategies employed by airlines with regard to the treatment and execution of their flight plans. While some may almost always reach their requested flight level, some may only do so a limited number of times. Additionally, by applying this method "company routes" could potentially be identified highlighting the preferred routing of an airline [37].

While these factors are generally not included or related to ATM's performance, they do have a significant impact on the fuel consumption between O-D pairs. Future work to improve the environmental efficiency of air travel will have to consider the collaboration between the various airspace stakeholders [16].

3.5. Summary of Inefficiency Factors

Summarizing the above-described categorization of inefficiencies the following Table 3.1 proposes a complete overview of ATM network inefficiency (sub)categories, an aspect not explicitly found in the studied state-of-the-art.

Table 3.1: Categorized summary of air traffic management inefficiency factors.

Categories	Sub-Categories	Description
Weather	Wind	Extra "air distance" flown due to the difference between air and ground speeds.
	Convective Weather	Deviation from optimal flight path due to adverse weather conditions.
Airspace	Workload Distribution	Procedures linked to the division of airspace into different regions (FIR, TMA, CTA).
	Fixed Routing	Set points in a route which should be followed.
Air Traffic Control	Separation	Intervention of Air Traffic Controllers to ensure a safe distance between aircraft.
	Flight-Deck Request	Pilot's initiative to deviation from the flight plan.
Airline	-	(Economic) Choices made by an airline at all stages of operation.

Combining the reviewed literature on this subject, two areas of study have been found to be underrepresented in the state-of-the-art. The first area is within the airline-related inefficiencies where, most likely due to the commercially sensitive nature of this topic, relatively few published studies are available. This leaves an area of uncertainty regarding the reason behind variations in flight inefficiency. With the current openly-available knowledge the extent of airline-related factors remain in large part invisible, effectively excluding them from most performed analysis.

For the second area of knowledge, the complexity of global airspace structures and challenges in separating airspace from ATC-related inefficiencies, together with limited openly available data, make the generalization of results to large airspace difficult. This is more so the case for European flights compared to US flights, where an immense amount of variables have to be taken into account for a complete representation to be made. This has limited the extent to which these inefficiencies have been studied in large international airspaces, to simpler metrics such as the comparison of flights to a “best-performer”.

Furthermore, while there is substantial literature on air traffic management inefficiencies, most studies rely on simple metrics like route extensions, time, or capacity, with recent efforts focusing on more fuel-centric metrics [16]. This means little understanding of air traffic management performance on metrics close to the environmental impact is available leaving only speculations to be made based on the available “proxy” indicators as for example route extension and time. Most commonly flight inefficiency has been defined as the excess distance flown, leaving the use of excess fuel burn, air quality, or climate impact metric for future research.

4

Causal Inference of Flight Trajectories

While in general, it was found that only a few papers were available that look into utilizing data to understand causality in the context of aviation operations, some examples of use cases were found. The following presents these findings by first looking at what specific causal inference methods were used, before discussing how the available data was processed and used in this case.

4.1. Introductory Principles of Causal Inference

It is important to distinguish between correlation and causation. Correlation shows how two variables tend to change together, while causation uncovers the cause-and-effect relationship, revealing the laws governing variable changes [38]. Presented in the foundational “potential outcome framework” proposed by Rubin [39], the causal impact of an exposure on an outcome may be mathematically formulated as follows:

$$T = E[Y^a - Y^{a*}] \tag{4.1}$$

Where a and $a*$ represent the different exposure levels of treatment A, Y represents the potential outcome under these conditions, and T the treatment effect. The causal effect of a variable is then defined as the change in outcome resulting from a specific treatment while holding all other variables constant.

Central to estimating causal effects is to account for the effect of so-called confounders, which are variables that affect both the treatment variable and the outcome. A common example confounder in medicine studies is age as it will have an effect on whether a patient is assigned a given treatment, as well as the rate of recovery of a patient. Causal inference methods then rely on methods that are able to account for the effect of confounders [40].

Both the treatment and confounder effects may influence the outcome directly and indirectly, via a mediator which itself has no direct effect on the outcome. Both types of effects may differ. Figure 4.1 graphically summarizes the different ways factors can affect an outcome of interest.

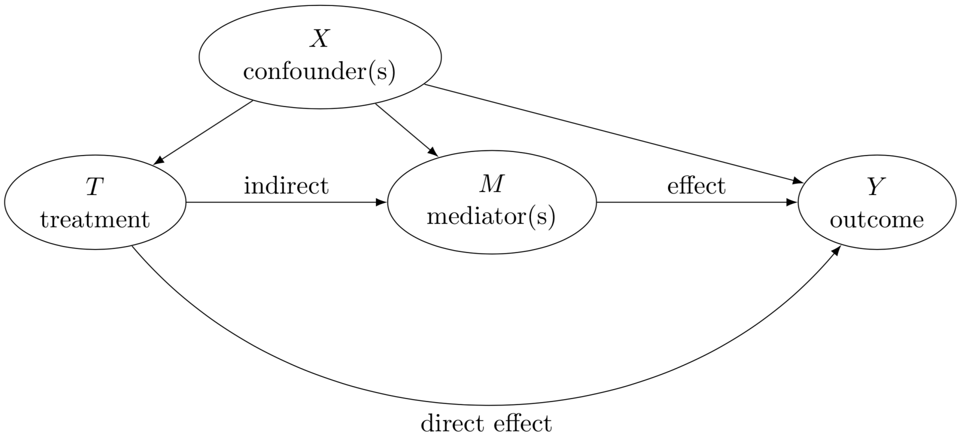


Figure 4.1: Graphical summary of factor influence on outcome variable within the causal inference framework [41].

Generally, two principles have been identified in the literature: (1) propensity-score-based methods, and (2) regression-based methods.

Propensity-score-based methods, as the name suggests, rely on a propensity score, i.e. "probability of treatment" to account for the effect of confounding on the outcome. The propensity score can then be used to group or weigh observed outcomes with different treatments, but similar propensity scores. Visually this process would be depicted as in Figure 4.2.

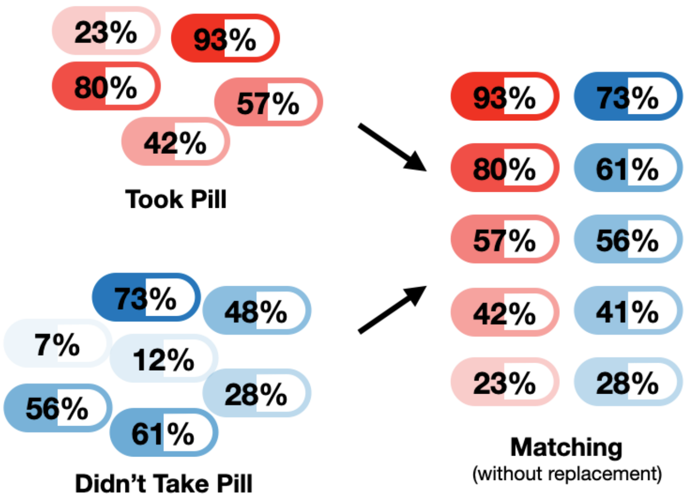


Figure 4.2: Visual representation of propensity-score based methods of causal inference [1].

Where then the average difference of the propensity score-matched pairs, called Individual Treatment Effect (ITE), results in the treatment's Average Treatment Effect (ATE).

Regression-based methods rely on applying machine learning techniques to observational data to predict an outcome based on covariate and treatment inputs. The resultant model can then be used to simulate counterfactual outcomes from which the causal effects of the input variables can be determined. Mathematically this can be represented as described in Equation 4.2 [42].

[1]Visuals taken from : https://towardsdatascience.com/propensity-score-5c29c480130c

$$
\begin{aligned}
ITE_i &= \hat{f}\left(X_i = 1, Z_i\right) - \hat{f}\left(X_i = 0, Z_i\right) \\
ATE &= \frac{1}{N}\sum_i ITE_i
\end{aligned}
\tag{4.2}
$$

Where $\hat{f}$ represents the selected regression model, X_i the treatment status, Z_i the covariates, and N the number of observations.

From its historical origins, causal inference is often described using medical jargon. While this offers a clear explanation of the method's functioning, it may lead to confusion regarding how this can be applied to an aviation use case. To address this, definitions of the used jargon for this aviation use case are provided:

- **Outcome:** The degree to which excess fuel burn is induced by a given situation. This could for example be quantified as an inefficiency percentage or a quantity of excess fuel burn.
- **Treatment:** An ATM inefficiency factor for which the effect on the outcome is to be analyzed. These will correspond to the elaborated factors identified in chapter 3.
- **Confounders:** All other factors that may influence the outcome and treatment effects of interest, which are not currently being seen as the treatment variable.
- **Mediators:** Factors that may indirectly affect the treatment outcome through a treatment and/or confounder variable. This would correspond to factors not necessarily discussed in chapter 3, but which may have an effect.

An advantage offered by these data-driven causal inference methods is that they allow for causal analysis to be performed in fields where it may not (practicably) be possible to perform a Randomized Control Trial (RCT), such as is the case in large-scale Air Traffic Management studies. Additionally, the above-mentioned framework allows for more robust conclusions to be made as compared to correlation-based analysis traditionally used in low-risk fields of study [38].

A significant limitation of this method however is the requirement of having sufficient knowledge of the subject at hand to define causal relations between factors. This is due to the fact that current causal inference methods are not able to account for hidden confounders meaning that all should be defined beforehand. In the future, this limitation could potentially be overcome when methods around causal discovery, i.e. the determination of causal relations, have matured further [43].

4.2. Methods Used in Literature

In the available literature, various aviation applications were found for causal inference. The following section summarizes the found literature by first presenting the studies related to flight trajectories, before developing studies applied to other domains but for which the methods are still of interest.

Flight Trajectory Studies The following studies aim to understand what the cause is of certain variations in flight routes. Marcos et al. [35] used a random forest regression model to understand inefficiency sources at the Area Control Center (ACC) level. Random forests were chosen due to their ability to capture non-linear dependencies and their resistance to overfitting compared to single decision trees or Neural Networks (NN). The relative importance of features was obtained by evaluating the rate of misclassification when excluding one feature from the out-of-bag dataset. Using the method developed by Wong et al. [44], the training data was augmented by adding samples with random noise, resulting in reduced error between validation and testing.

Liu et al. [33] took a different approach in their study on lateral flight efficiency in the US. They used two subsequent regression models, logistic regression and linear regression, to model the strategic route choice before en-route deviations. This approach aimed to mitigate overfitting risks when considering all factors in a single model. The study built upon the observation of significant variations in inefficiency for similar flight conditions in a previous study [45]. It suggested that route inefficiency is influenced by a two-step decision process involving strategic routing and tactical rerouting.

Szenczuk et al. [46] employed a single linear regression model to study vertical flight efficiency during descent at two major airports. The reduced issue of overfitting was attributed to the narrower focus of the study. Linear regression models, including the one used in this study, have a limitation in that they can only consider linearly related factors.

Other Causal Inference Studies Besides studies on flight trajectories, within aviation other applications of causal inference have been found. Interesting methods used in those studies are presented in the following paragraphs.

Shah [47] applied causal inference to evaluate the impact of ground delay absorption on overall airborne delay in the ATM community. The study advocated for the use of propensity score methods, as it was argued that it would provide more robust results than regression-based models. The main reason is that propensity score methods are grounded in formal statistical frameworks, and avoid extrapolating beyond the observational data, as could be the case with regression-based methods.

Baker et al. [48] used the Granger causality framework to examine the causal relationship between regional aviation and economic growth. However, this framework's link to causal relations is debated as it relies heavily on temporal relationships rather than causal ones [49]. This doubt relates to the fallacy of "Post hoc ergo proter hoc", where conclusions are based solely on the order of events [50].

Wilke et al. [51] investigated the causes of airport surface safety occurrences and used correlation-based statistical methods to gain insights into potential factors contributing to these events.

4.3. Feature Selection and Data Preparation

To apply the methods described in section 4.2, the model's input features must be determined. In the causal inference framework, this refers to selecting the causal factors for analysis. The strategies commonly used are extensions of each other where one first relies on supporting literature and data availability, potentially supplemented by the author's expertise. Then secondly, one can opt to employ an analytical approach. Both these strategies are presented in this section.

Literature, Available Data, and Own Insight Approach In the available literature, the majority of the found studies performed feature selection through literature review and own insights of the author on the topic. Examples of papers where this method of feature selection has been applied are the ones from Liu et al. [33], and Szenczuk et al. [46].

It should be noted that as this study is performed in the US, data availability will differ for European studies. As such, the implementation of Traffic Management Initiatives (TMIs) and Special Activity Airspace (SAA), both used in the study of Liu et al. [33], indicators are not in the same way applicable to studies in European airspace.

What is interesting to note in the feature selection of these studies, is the different chosen levels of detail for the features. For example, convective weather phenomena were split into five variables in the study of Lui et al. [33], while in the study of Szenczuk et al. [46] a single indicator was used to cover all aspects. This is most likely the result of a trade-off between analysis complexity, and the data availability in combination with the scope of the study, as the former considers the cruise segment of flight, and the latter only the descent phase.

Although differences remain in the feature selection between studies, the following groups of considered causal features could be identified:

1. Nominal route variations
2. Wind contribution
3. Convective weather
4. Flight path restrictions
5. Congestion

Analytical Approach In the study of Marcos et al. [35], analytical methods for data exploration and selection for input features are presented. The method involved visually analyzing an initial selection of features from expert consultation to determine their suitability for the prediction of the chosen flight efficiency indicator. In parallel, to support the visual justifications, the following four statistical correlation indicators were computed between the efficiency indicator and the different features to help support the selection: Peasron's, Spearman's, Distance, and Mutual Information correlation factors.

A feature heavily correlated to flight efficiency can hint towards a possible causal relation between these two factors. Next to this, the outputs of these correlation metrics can then also be used to assess what methods could possibly be used for causal inference. For instance, if a feature scores high in Pearson's correlation, a linear regression model could be suitable whereas if it were to score high in Spearman's correlation and low in Pearson's correlation a non-linear relationship is likely. Then a linear regression model would not be suitable.

4.4. Findings of Previous Studies

Summarizing the results of previous studies will not only allow for the comparison of findings but for an idea to be formed of what results one could expect when performing similar studies. The following section presents the found causal influence as well as an introduction to future work discussions.

Feature Influence Results While it should be noted that the considered studies did not explicitly address fuel burn. Therefore, the conclusions drawn regarding inefficiency causes cannot be directly generalized to excess fuel burn, as discussed in chapter 3. Nonetheless, the findings and insights from these previous studies serve as valuable starting points for the proposed research.

Regarding a limited selection of origin-destination (O-D) pairs in the US analyzed by Liu et al. [33], significant variations were observed among different O-D pairs. Overall, convective weather and wind effects were identified as having the greatest impact on flight inefficiency, although it should be noted that the study also emphasized the fact that no single identifiable cause can be noted.

At the Area Control Center (ACC) level of analysis, as demonstrated in the European study by Marcos et al. [35], the route structure and flight direction were suggested as the most influential factors. It is important to note that this study did not consider weather effects. Contrary to expert expectations, congestion, and daily variability were found to have a low influence, potentially due to the study being conducted in February, a period with lower flight intensity.

Regarding flight trajectory performance during arrival at two major Brazilian airports, the study conducted by Szenczuk et al. [46] highlighted the significant role of airspace structure in vertical performance variations. Also, it was observed that certain trajectory patterns consistently showed lower efficiency. In terms of causal effects, this study provided evidence of the impact of arrival demand, weather conditions, and traffic flow management on vertical performance. Among the observed factors, convective weather phenomena were identified as the most significant contributors to inefficiency.

Limitations and Recommendations for Future Work Just as important to consider are the limitations and recommendations for future work from the previously performed research.

In the work of Liu et al. [33] difficulties were encountered when attempting to pinpoint the various causes of flight inefficiency due to all effects being found to be relatively similar. A possible reason given for was given to be the largely homogeneous distribution of the attributes across the various flight path alternatives. This would mean that not enough variation of the features would be available in the training model, thereby making it harder to estimate the effect on flight efficiency due to this factor when it is varied. Alternatively, it could be possible that for the flights considered, the identified causes average each other out. The study of Marcos et al. [35], highlighted the lack of data as a source of error due to (1) exclusion of both wind and convective weather conditions, and (2) ATC shortcuts which currently can't be deterministically predicted.

Furthermore, in the course of this literature review, an observation was made that previous research predominantly focused on the primary causal factors of interest, often overlooking the potential influence of mediators. This omission raises the likelihood of confounding in the obtained results.

5 Conclusion of Literature Review

The performed literature review has highlighted a research gap that could contribute to a better understanding of the necessary measures for reducing the environmental impact of aviation. While previous studies have focused on the quantification of aviation's environmental impact and others on optimizing flight trajectories for reduced impact, this study aims to understand the underlying dynamics leading to the disparity between current and optimal operations. Although some studies have explored the causality of Air Traffic Management (ATM) inefficiencies, to the best of the author's knowledge, none have evaluated inefficiency causality with metrics closely tied to environmental improvements, such as excess fuel consumption.

To employ metrics such as excess fuel burn, the developed model will need to consider all control dimensions (lateral, vertical, speed, and time) simultaneously, a particularity not previously required in existing studies. In the current state-of-the-art, researchers have previously made use of excess fuel burn metrics by comparing actual trajectories with optimal routes computed by flight path optimizer. However, this approach has not been extended to the scale of the entire European airspace with the same level of detail. The inclusion of additional control dimensions also requires the consideration of a broader spectrum of ATM inefficiency factors, all of which impact fuel consumption. While an overall comprehension of factors that may contribute to a change in flight efficiency is present, only a few studies have considered the entirety of inefficiency factors together. To fill this research gap, both of these scientific innovations must be addressed through well-founded assumptions and decisions regarding flight subset sampling and the selection of causal features.

The aim is to establish a scientific foundation that flight operators and air traffic managers can utilize to identify the key factors responsible for excessive emissions in air travel and to determine their level of influence. By doing so, these stakeholders will be equipped with the necessary tools to make informed decisions regarding the path toward a sustainable future for aviation. Furthermore, the novel framework proposed in this study may facilitate more precise estimates of inefficiencies in aviation emission inventories. Ultimately, the objective of this project is to create and assess a framework for the large-scale use of causal inference methods in uncovering the most significant inefficiencies in Air Traffic Management (ATM) that contribute to the adverse environmental effects of aviation. Further confirming this research gap, the use of causal inference methods in the context of aviation environmental impact analysis has also been hypothesized in the state-of-the-art compiled by Gao et al. [52].

References

[1] T. Ryley, S. Baumeister, and L. Coulter, “Climate change influences on aviation: A literature review,” *Transport Policy*, vol. 92, pp. 55–64, Jun. 2020, ISSN: 0967-070X. DOI: 10.1016/J.TRANPOL.2020.04.010.

[2] International Air Transport Association, “Net zero 2050: operational and infrastructure improvements Fact sheet,” Tech. Rep., 2022. [Online]. Available: www.iata.org/flynetzero.

[3] International Panel on Climate Change, *Aviation and the Global Atmosphere*, J. E. Penner, D. H. Lister, D. J. Griggs, D. J. Dokken, and M. McFarland, Eds. Cambridge: Cambridge University Press, 1999. [Online]. Available: https://www.ipcc.ch/report/aviation-and-the-global-atmosphere-2/.

[4] T. G. Reynolds, “Air traffic management performance assessment using flight inefficiency metrics,” *Transport Policy*, vol. 34, pp. 63–74, Jul. 2014, ISSN: 0967-070X. DOI: 10.1016/J.TRANPOL.2014.02.019.

[5] J. Rosenow, D. Strunck, and H. Fricke, “Trajectory optimization in daily operations,” *CEAS Aeronautical Journal*, vol. 11, no. 2, pp. 333–343, Jun. 2020, ISSN: 18695590. DOI: 10.1007/S13272-019-00429-7/FIGURES/8. [Online]. Available: https://link.springer.com/article/10.1007/s13272-019-00429-7.

[6] M. Basner, C. Clark, A. Hansell, *et al.*, “Aviation Noise Impacts: State of the Science,” *Noise & Health*, vol. 19, no. 87, p. 41, Mar. 2017, ISSN: 19984030. DOI: 10.4103/NAH.NAH{_}104{_}16. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5437751/.

[7] L. Rupcic, E. Pierrat, K. Saavedra-Rubio, N. Thonemann, C. Ogugua, and A. Laurent, “Environmental impacts in the civil aviation sector: Current state and guidance,” *Transportation Research Part D: Transport and Environment*, vol. 119, p. 103 717, Jun. 2023, ISSN: 1361-9209. DOI: 10.1016/J.TRD.2023.103717.

[8] D. S. Lee, D. W. Fahey, A. Skowron, *et al.*, “The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018,” *Atmospheric Environment*, vol. 244, p. 117 834, Jan. 2021, ISSN: 1352-2310. DOI: 10.1016/J.ATMOSENV.2020.117834.

[9] “European Aviation Environmental Report,” European Union Aviation Safety Agency, Tech. Rep., 2022.

[10] S. H. Yim, G. L. Lee, I. H. Lee, *et al.*, “Global, regional and local health impacts of civil aviation emissions,” *Environmental Research Letters*, vol. 10, no. 3, p. 034 001, Feb. 2015, ISSN: 1748-9326. DOI: 10.1088/1748-9326/10/3/034001. [Online]. Available: https://iopscience.iop.org/article/10.1088/1748-9326/10/3/034001%20https://iopscience.iop.org/article/10.1088/1748-9326/10/3/034001/meta.

[11] A. Nuic, D. Poles, and V. Mouillet, “BADA: An advanced aircraft performance model for present and future ATM systems,” *International Journal of Adaptive Control and Signal Processing*, vol. 24, no. 10, pp. 850–866, Oct. 2010, ISSN: 1099-1115. DOI: 10.1002/ACS.1176. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/acs.1176%20https://onlinelibrary.wiley.com/doi/abs/10.1002/acs.1176%20https://onlinelibrary.wiley.com/doi/10.1002/acs.1176.

[12] J. Sun, J. M. Hoekstra, and J. Ellerbroek, “OpenAP: An Open-Source Aircraft Performance Model for Air Transportation Studies and Simulations,” *Aerospace 2020, Vol. 7, Page 104*, vol. 7, no. 8, p. 104, Jul. 2020, ISSN: 2226-4310. DOI: 10.3390/AEROSPACE7080104. [Online]. Available: https://www.mdpi.com/2226-4310/7/8/104/htm%20https://www.mdpi.com/2226-4310/7/8/104.

[13] N. W. Simone, M. E. Stettler, and S. R. Barrett, “Rapid estimation of global civil aviation emissions with uncertainty quantification,” *Transportation Research Part D: Transport and Environment*, vol. 25, pp. 33–41, Dec. 2013, ISSN: 1361-9209. DOI: 10.1016/J.TRD.2013.07.001.

[14] F. D. A. Quadros, M. Snellen, J. Sun, and I. C. Dedoussi, “Global Civil Aviation Emissions Estimates for 2017–2020 Using ADS-B Data,” *Journal of Aircraft*, pp. 1–11, Nov. 2022, ISSN: 15333868. DOI: 10.2514/1.c036763.

[15] K. Seymour, M. Held, G. Georges, and K. Boulouchos, “Fuel Estimation in Air Transportation: Modeling global fuel consumption for commercial aviation,” *Transportation Research Part D: Transport and Environment*, vol. 88, p. 102 528, Nov. 2020, ISSN: 1361-9209. DOI: 10.1016/J.TRD.2020.102528.

[16] Eurocontrol, “Environmental Assessment: European ATM Network Fuel Inefficiency Study | EUROCONTROL,” EUROCONTROL, Tech. Rep., 2020. [Online]. Available: https://www.eurocontrol.int/publication/environmental-assessment-european-atm-network-fuel-inefficiency-study.

[17] S. Matthes, V. Grewe, K. Dahlmann, *et al.*, “A Concept for Multi-Criteria Environmental Assessment of Aircraft Trajectories,” *Aerospace 2017, Vol. 4, Page 42*, vol. 4, no. 3, p. 42, Aug. 2017, ISSN: 2226-4310. DOI: 10.3390/AEROSPACE4030042. [Online]. Available: https://www.mdpi.com/2226-4310/4/3/42/htm%20https://www.mdpi.com/2226-4310/4/3/42.

[18] F. D. Quadros, M. Snellen, and I. C. Dedoussi, “Regional sensitivities of air quality and human health impacts to aviation emissions,” *Environmental Research Letters*, vol. 15, no. 10, p. 105 013, Oct. 2020, ISSN: 1748-9326. DOI: 10.1088/1748-9326/ABB2C5. [Online]. Available: https://iopscience.iop.org/article/10.1088/1748-9326/abb2c5%20https://iopscience.iop.org/article/10.1088/1748-9326/abb2c5/meta.

[19] C. Grobler, P. J. Wolfe, K. Dasadhikari, *et al.*, “Marginal climate and air quality costs of aviation emissions,” *Environmental Research Letters*, vol. 14, no. 11, p. 114 031, Nov. 2019, ISSN: 1748-9326. DOI: 10.1088/1748-9326/AB4942. [Online]. Available: https://iopscience.iop.org/article/10.1088/1748-9326/ab4942%20https://iopscience.iop.org/article/10.1088/1748-9326/ab4942/meta.

[20] J. Sun and I. Dedoussi, “Evaluation of Aviation Emissions and Environmental Costs in Europe Using OpenSky and OpenAP,” *Engineering Proceedings 2021, Vol. 13, Page 5*, vol. 13, no. 1, p. 5, Dec. 2021, ISSN: 2673-4591. DOI: 10.3390/ENGPROC2021013005. [Online]. Available: https://www.mdpi.com/2673-4591/13/1/5/htm%20https://www.mdpi.com/2673-4591/13/1/5.

[21] T. G. Reynolds, “Analysis of lateral flight inefficiency in global air traffic management,” *8th AIAA Aviation Technology, Integration and Operations (ATIO) Conference*, 2008. DOI: 10.2514/6.2008-8865. [Online]. Available: https://arc.aiaa.org/doi/10.2514/6.2008-8865.

[22] C. Ramee, K. Junghyun, M. Deguignet, C. Justin, S. Briceno, and D. N. Mavris, “Aircraft Flight Plan Optimization with Dynamic Weather and Airspace Constraints,” in *International Conference for Research in Air Transportation*, Georgia Institute of Technology, 2020. [Online]. Available: https://smartech.gatech.edu/handle/1853/63783.

[23] F. Vergnes, J. Bedouet, X. Olive, and J. Sun, “Environmental Impact Optimisation of Flight,” in *10th International Conference for Research in Air Transportation*, FAA & EUROCONTROL, 2022.

[24] J. Sun, “OpenAP.top: Open Flight Trajectory Optimization for Air Transport and Sustainability Research,” *Aerospace*, vol. 9, no. 7, Jul. 2022, ISSN: 22264310. DOI: 10.3390/aerospace9070383.

[25] Y. Lim, A. Gardi, and R. Sabatini, “Optimal Aircraft Trajectories to Minimize the Radiative Impact of Contrails and CO2,” *Energy Procedia*, vol. 110, pp. 446–452, Mar. 2017, ISSN: 1876-6102. DOI: 10.1016/J.EGYPRO.2017.03.167.

[26] H. Yamashita, F. Yin, V. Grewe, *et al.*, “Analysis of Aircraft Routing Strategies for North Atlantic Flights by Using AirTraf 2.0,” *Aerospace 2021, Vol. 8, Page 33*, vol. 8, no. 2, p. 33, Jan. 2021, ISSN: 2226-4310. DOI: 10.3390/AEROSPACE8020033. [Online]. Available: https://www.mdpi.com/2226-4310/8/2/33/htm%20https://www.mdpi.com/2226-4310/8/2/33.

[27] E. Roosenbrand, J. Sun, I. Dedoussi, D. Stam, and J. Hoekstra, “Assessing and Modelling Climate Optimal Flights Using Open Surveillance and Remote Sensing Data,” in *10th International Conference for Research in Air Transportation*, FAA & EUROCONTROL, 2022.

[28] V. Williams and R. B. Noland, "Variability of contrail formation conditions and the implications for policies to reduce the climate impacts of aviation," *Transportation Research Part D: Transport and Environment*, vol. 10, no. 4, pp. 269–280, Jul. 2005, ISSN: 1361-9209. DOI: 10.1016/J.TRD.2005.04.003.

[29] A. Simorgh, M. Soler, D. González-Arribas, *et al.*, "A Comprehensive Survey on Climate Optimal Aircraft Trajectory Planning," *Aerospace 2022, Vol. 9, Page 146*, vol. 9, no. 3, p. 146, Mar. 2022, ISSN: 2226-4310. DOI: 10.3390/AEROSPACE9030146. [Online]. Available: https://www.mdpi.com/2226-4310/9/3/146/htm%20https://www.mdpi.com/2226-4310/9/3/146.

[30] A. Filippone and B. Parkes, "Evaluation of commuter airplane emissions: A European case study," *Transportation Research Part D: Transport and Environment*, vol. 98, p. 102 979, Sep. 2021, ISSN: 1361-9209. DOI: 10.1016/J.TRD.2021.102979.

[31] P. Pasutto, K. Zeghal, and D. Brain, "Initial analysis of vertical flight efficiency in cruise for European city pairs," in *AIAA Aviation and Aeronautics Forum and Exposition, AIAA AVIATION Forum 2021*, American Institute of Aeronautics and Astronautics Inc, AIAA, 2021, ISBN: 9781624106101. DOI: 10.2514/6.2021-2346.

[32] J. Sun, X. Olive, and D. Delahaye, "Evaluating Aviation Emission Inefficiencies and Reduction Challenges with Electric Flights Based on an analysis of flights from 2019 in the Dutch and French airspaces," in *12th SESAR Innovation Days*, 2022.

[33] Y. Liu, M. Hansen, M. O. Ball, and D. J. Lovell, "Causal analysis of flight en route inefficiency," *Transportation Research Part B: Methodological*, vol. 151, pp. 91–115, Sep. 2021, ISSN: 0191-2615. DOI: 10.1016/J.TRB.2021.07.003.

[34] H. Hersbach, P. de Rosnay, B. Bell, *et al.*, "Operational global reanalysis: progress, future directions and synergies with NWP," eng, Tech. Rep. 27, Mar. 2018. DOI: 10.21957/tkic6g3wm. [Online]. Available: https://www.ecmwf.int/node/18765.

[35] R. Marcos, R. Herranz, R. Rodríguez Vázquez, P. García-Albertos, and O. G. Cantú Ros, "Application of Machine Learning for ATM Performance Assessment-Identification of Sources of En-Route Flight Inefficiency," in *Eighth SESAR Innovation Days*, 2018.

[36] X. Olive, J. Sun, A. Lafage, and L. Basora, "Detecting Events in Aircraft Trajectories: Rule-Based and Data-Driven Approaches," *Proceedings 2020, Vol. 59, Page 8*, vol. 59, no. 1, p. 8, Dec. 2020, ISSN: 2504-3900. DOI: 10.3390/PROCEEDINGS2020059008. [Online]. Available: https://www.mdpi.com/2504-3900/59/1/8/htm%20https://www.mdpi.com/2504-3900/59/1/8.

[37] J. L. Leones, M. P. Morales, L. D'Alto, *et al.*, "Advanced flight efficiency key performance indicators to support air traffic analytics: Assessment of European flight efficiency using ADS-B data," *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, vol. 2018-September, Dec. 2018, ISSN: 21557209. DOI: 10.1109/DASC.2018.8569584.

[38] X. Gejingting, J. Ruiqiong, W. Wei, J. Libao, and Y. Zhenjun, "Correlation analysis and causal analysis in the era of big data," in *IOP Conference Series: Materials Science and Engineering*, vol. 563, Institute of Physics Publishing, Aug. 2019. DOI: 10.1088/1757-899X/563/4/042032.

[39] D. B. Rubin, "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, vol. 66, no. 5, pp. 688–701, 1974.

[40] L. Yao and Y. Li, "A Survey on Causal Inference," vol. 1, no. 1, 2020.

[41] *Causal inference with mediation using machine learning in high-dimensional settings – Parietal*. [Online]. Available: https://team.inria.fr/parietal/research/statistical-and-machine-learning-methods-for-large-scale-data/causal-inference-using-machine-learning-in-high-dimensional-settings/.

[42] A. Gelman, J. Hill, and A. Vehtari, "Causal Inference Using Regression on the Treatment Variable," in *Regression and Other Stories*, Cambridge University Press, Jul. 2020, ISBN: 9781139161879. DOI: 10.1017/9781139161879.

[43] D. Kaltenpoth and J. Vreeken, "Causal Discovery with Hidden Confounders using the Algorithmic Markov Condition," in *39th Conference on Uncertainty in Artificial Intelligence*, 2023. [Online]. Available: https://eda.rg.cispa.io/prj/pepsi/.

[44] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: when to warp?" *2016 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2016*, Sep. 2016. DOI: 10.1109/DICTA.2016.7797091. [Online]. Available: https://arxiv.org/abs/1609.08764v2.

[45] M. Hansen, Y. Liu, C. Chuang, D. Lovell, and M. O. Ball, "En Route Performance in the National Airspace System," 2017.

[46] J. B. T. Szenczuk, M. C. R. Murça, W. S. Sant'anna Souza, and R. d. A. Gomes, "Causal Analysis of Vertical Flight Inefficiency During Descents," *Transportes*, vol. 29, no. 4, p. 2486, Nov. 2021. DOI: 10.14295/transportes.v29i4.2486.

[47] A. Shah, "Causal Inference for ATM Counterfactual Estimation," *Internal RAND Corporation report. Unpublished.*,

[48] D. Baker, R. Merkert, and M. Kamruzzaman, "Regional aviation and economic growth: cointegration and causality analysis in Australia," *Journal of Transport Geography*, vol. 43, pp. 140–150, Feb. 2015, ISSN: 0966-6923. DOI: 10.1016/J.JTRANGEO.2015.02.001.

[49] A. Shojaie and E. B. Fox, "Granger Causality: A Review and Recent Advances," *Annual Review of Statistics and Its Application*, vol. 9, pp. 289–319, Mar. 2022, ISSN: 2326831X. DOI: 10.1146/ANNUREV-STATISTICS-040120-010930. [Online]. Available: https://www.annualreviews.org/doi/abs/10.1146/annurev-statistics-040120-010930.

[50] J. Tobin, "Money and Income: Post Hoc Ergo Propter Hoc?" *The Quarterly Journal of Economics*, vol. 84, no. 2, pp. 301–317, 1970. [Online]. Available: https://www.jstor.org/stable/1883016.

[51] S. Wilke and A. Majumdar, "A Multi-national Causal Analysis of Airport Surface Safety Occurrences," 2011. DOI: 10.2514/6.2011-7005. [Online]. Available: http://arc.aiaa.org.

[52] Z. Gao and D. N. Mavris, "Statistics and Machine Learning in Aviation Environmental Impact Analysis: A Survey of Recent Progress," *Aerospace 2022, Vol. 9, Page 750*, vol. 9, no. 12, p. 750, Nov. 2022, ISSN: 2226-4310. DOI: 10.3390/AEROSPACE9120750. [Online]. Available: https://www.mdpi.com/2226-4310/9/12/750/htm%20https://www.mdpi.com/2226-4310/9/12/750.