Prediction of Aircraft Trip Fuel Deviations for Fuel Loading Decisions with a Deep Time Series Approach

Msc. Thesis report

Reinoud Lampe

Delft University of Technology

TU Delft

Prediction of Aircraft Trip Fuel Deviations for Fuel Loading Decisions with a Deep Time Series Approach

Msc. Thesis report

by

Reinoud Lampe

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on December 19^{th} 2022.

Student number:	
Project duration:	15 Nov 2021 – 19 Dec 2022
Thesis committee:	Dr. B.F. Lopes Dos Santos
	Dr. M. Baptista
	Dr. C. Varriale
	Dr. L. Li
	Ir. D.A. Bos

TU Delft

Acknowledgements

Dear reader,

Before you, you have the final report of my MSc Thesis that marks the end of my time as a student at the Delft University of Technology. I am grateful for this experience and have enjoyed the journey. Starting my bachelor's in Aerospace Engineering with a lot of scepticism on the aircraft enthusiasts around me, it is interesting to see I became an aviation enthusiast myself.

I would like to thank Lishuai Li, for all your support. Even though most of our meetings were online, you helped me to stay motivated. High-quality research is hard to achieve, but I have learned more about this from you. I also want to thank Bruno Santos, to assist me in the project, and providing me with great feedback on the research setup. Furthermore, I want to thank Xingting Zhu, for all the help on doing research in the area of fuel loading.

I was very thrilled to hear I could visit SunExpress Airlines and see the operations of an airline from the inside. Many thanks to Diederik-Jan Bos, for showing me around and making sure I had a good experience. During the research, it was great to hear about the practicalities of the airlines from you. Thereby I want to thank Emre Murphy and all the others I've met at SunExpress, for the support and conversations about the operations of an airline.

Last but not least, I would like to thank all of you who supported me during the writing of my Thesis. Susan, for always being there, both the ups and downs, to help me to stay focused. My family, for being there to listen to everything I have done during the thesis. My roommates, who were there to discuss my progress of the research, and my friends with whom I could enjoy study sessions and coffee breaks.

Reinoud Lampe,
Delft, December, 2022

Contents

List of Figures

List of Tables

List of Abbreviations

2TS	2-layered Time Series model
AF	Alternate Fuel
APU	Auxiliary Power Unit
ATC	Air Traffic Control
ATM	Air Traffic Management
CTC	Cost-To-Carry
EASA	European Union Aviation Safety Agency
ELF	Estimated Landing Fuel
FDM	Flight Data Recorder (also FDR)
FPS	Flight Planning System
FRF	Final Reserve Fuel
FTTS	Fixed Time interval Time Series
GB	Gradient Boosting
GRN	Gated Residual Network
ICAO	International Civil Aviation Organization
IFR	Instrumental Flight Rules
IFTS	Individual Flight Time Series
LSTM	Long-Short Term Memory Recurrent Neural Network
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
METAR	Meteorological Aerodrome Report
OD-pair	Origin-Destination pair
OFP	Operational Flight Plan
RF	Random Forests
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SARIMAX	Seasonal Auto Regressive Integrated Moving Average
SID	Standard Instrument Departure Route
SPI	Safety Key Performance Indicator
STAR	Standard Arrival Route

SVM	Support Vector Machine
TAF	Terminal Area Forecast
TF	Trip Fuel
ZFW	Zero Fuel Weight

Introduction

Decreasing fuel consumption is an increasingly important goal in the aviation industry. The cost of fuel accounts for roughly 25% of the total costs of an airline, which makes it an important contributor to the profitability of airlines [77]. Besides, growing concerns about greenhouse gas emissions make reducing fuel consumption a high priority within aviation. From an Air Traffic Management (ATM) perspective, controllers try to minimize delays by maximizing the airspace capacity and opt for more freedom in the trajectory selection of aircraft in order to reduce fuel burn. Other methods are performed by airlines, which renew their fleet to more modern and fuel-efficient aircraft, aim for other fuel-saving strategies like trajectory and taxi procedure optimization and reduce the unnecessary weight of utilities on board.

One way to limit fuel consumption is by reducing the aircraft's weight. The largest variable weight of the aircraft is the loaded fuel. Regulations from ATM and procedures of airlines make sure buffers of fuel are loaded to avoid fuel shortage. Ryerson, [73], found that 2.21% - 4.48% of aircraft fuel is burned due to carrying unused fuel and 0.70% - 1.04% comes from loading fuel beyond what is needed to ensure a safe flight. Although there is potential for fuel savings, parties argue that the fuel loaded on aircraft should actually be increased in order to meet the safety target of the European Union Aviation Safety Agency (EASA) ([28]). Since safety concerns arise when one opts to reduce fuel on flights, changing fuel loading procedures is a complex topic, where trust in the models is of high importance.

Currently, airline flight dispatchers and pilots load extra fuel, to overcome the experienced uncertainties in trip fuel consumption. Based on traffic, weather forecast and experience, decisions are made on the fuel to be loaded. Airlines often present mean deviations per flight leg or statistical contingency fuels at multiple levels, to indicate the expected deviations of the flight. However, data-driven methods have been proven to predict these deviations well, resulting in more accurate fuel-loading decisions [4, 49, 53, 96]. In research, general parameters are used to describe the air networks system predictability, based on descriptive metrics of historical flights as the mean and percentiles of the fuel and flight time deviation. These metrics are usually defined per flight leg and not updated in the time set. Therefore, the current research will contribute by introducing an analysis on the relationship between mission parameter deviations and trip fuel deviations. Thereby, a time series model will be created, which updates the state of the system's predictability throughout the year. These results will be evaluated and used for fuel loading advice. The main research objective is as follows:

> *To more accurately predict trip fuel burn by creating a data-driven model with detailed historical flight information such that loading discretionary fuel could be minimized while maintaining the same standards of reserve fuel usage.*

The trade-off of fuel savings on the one side versus the risk of fuel depletion on the other, make the subject a delicate topic. With data-driven insights, more knowledge is available on both aspects, which may result in better decisions. With the goal of reducing greenhouse gasses in mind, this research will be one of the many steps towards greener aviation.

The research is performed in close collaboration with SunExpress Airlines. They ensured data availability of historical flights, required for the research. Thereby, employees have given valuable insights, important to the current topic.

This thesis report consists of three components. Part I presents the paper of the research. The paper is a stand-alone document, containing the important aspects of the research. Part II presents the literature study performed prior to writing the thesis. This part is written in February 2022, so, as a result, the content is not fully aligned with the content of the rest of the report. The literature study contains information on fuel loading practices and prediction models. Finally, Part III presents additional chapters to the research performed in the paper. First, the preprocessing of the data will be discussed, and thereafter, extra analysis on the results is presented.

I

Scientific Paper

Prediction of Aircraft Trip Fuel Deviations for Fuel Loading Decisions with a Deep Time Series Approach

Reinoud Lampe,*

Delft University of Technology, Delft, The Netherlands

Abstract

Reducing fuel consumption is an increasingly important topic within aviation. One approach to accomplish this goal is reducing excess fuel weight being loaded on aircraft. Flight dispatchers and pilots load extra fuel to account for uncertainties present in trip fuel consumption, which is currently computed by the flight planning system (FPS). In this paper, a time-series-based model is proposed that predicts deviations in trip fuel consumption of commercial flights. The aim is to assist dispatchers on fuel loading decision, using the proposed model. A 2-layered time series is proposed, able to capture temporal patterns present in deviations of trip fuel consumption. The first layer is a fixed time interval time series, grouping flights per time interval, to estimate average trip fuel deviation for the coming time intervals. The output of this layer is used for the second layer, which is a sequential time series model, modelling each flight individually, able to capture patterns present in the individual flight information. To estimate the trip fuel deviations for coming flights, input features from the operational flight flan, weather descriptive features from terminal area forecasts and historical flight descriptive input features from the flight data recorder are used. The new prediction model is able to reduce the root mean squared error (RMSE) of trip fuel predictions of the FPS by 26% and reduces the RMSE compared to the baseline gradient boosting model by 5.4%. Using a fixed-buffer loading strategy, 0.12 - 0.39% of fuel consumption could be reduced, depending on the desired safety key performance indicators, which leads to yearly savings of up to $1.5 million and 4,792 tonnes of CO_2.

1 Introduction

Decreasing fuel consumption is an increasingly important goal in the aviation industry. The cost of fuel accounts for roughly 25% of the total costs of an airline, which makes it an important contributor to the profitability of airlines [Statista, 2021]. Besides this, the increasing concern about the emission greenhouse gasses makes fuel burn reduction a top priority within the aviation industry. In order to reduce aircraft fuel consumption, Air Traffic Control (ATC) tries to minimize delays by maximizing the airspace capacity and opt for more freedom in the trajectory selection of aircraft. Airlines renew their fleet to more modern and fuel-efficient aircraft, aim for other fuel-saving strategies like trajectory and taxi procedure optimization, and reduce the unnecessary weight of utilities on board.

One way to limit fuel consumption is by reducing the aircraft's weight, with the largest source of excess weight being excess fuel [Irrgang, 2011], since an increased aircraft weight leads to a higher required thrust setting of the engines for a given speed and altitude. [Ryerson et al., 2015] found that 2.21% - 4.48% of aircraft fuel is burned due to carrying unused fuel and 0.70% - 1.04% comes from loading more fuel than necessary to ensure a safe flight. Pilots and flight dispatchers decide on extra fuel to be loaded for each flight, to account for uncertainties in trip fuel consumption, in order to avoid fuel depletion or flight diversion. The trip fuel consumption is estimated by the Flight Planning System (FPS), and model imperfections, weather uncertainties and traffic congestion lead to deviations from these estimates. These uncertainties result in more fuel being loaded by dispatchers [Kang, 2017]. So, more accurate predictions could result in less extra fuel being loaded on aircraft. On the other side, parties argue that the loaded fuel on aircraft should actually be increased in order to meet the safety target of the European Union Aviation Safety Agency (EASA) [Drees et al., 2017]. Again, improved trip fuel predictions may lead to less risk on fuel shortage for flights, knowing which flights are expected to deviate from the FPS estimate. Since safety concerns arise when one opts to reduce fuel weight on flights, changing fuel loading procedures is a complex topic, where trust in fuel predictions is of high importance.

In the end, the pilot-in-command is responsible to ensure that enough fuel is loaded on each flight, being assisted by the flight dispatcher. Regulations on fuel loading are provided by the International Civil Aviation

*Msc Student, Air Transport and Operations, Faculty of Aerospace Engineering, Delft University of Technology

Organization (ICAO), to make sure there are fuel buffers loaded on each flight. In addition, fuel is categorized, where each category has its intended purpose. The fuel categories are listed below, retrieved from [ICAO, 2018]. Not present in the list is tankering fuel, which is fuel loaded when it is economically beneficial to bring fuel to the destination airport. The regulated fuel categories are as follows:

- **Taxi fuel** The amount of fuel expected to be consumed before take-off, including auxiliary power unit fuel consumption.
- **Trip fuel (TF)** The amount of fuel needed to fly from take-off to landing at the intended destination airport, taking into account operating conditions of the flight.
- **Contingency fuel** The amount of fuel required to compensate for unforeseen factors such as weather and traffic. It is required to be 5% of the planned trip fuel, but not lower than the amount required to be able to fly for 5 minutes at holding speed at 1500ft at the destination airport. Exemptions can be made for airlines such that the contingency fuel is required to be 3% of the fuel.
- **Destination alternate fuel (AF)** The amount of fuel required to be able to fly to an alternate airport. This includes fuel for a missed approach and the flight to the alternate airport including climb and descent to the alternate airport. An alternate airport is only required when the flight is operating in Instrumental Flight Rules (IFR). In case there is no alternate airport, an extra fuel buffer should be present, equal to the amount to be able to fly for 15 minutes at holding speed at 1500 ft altitude.
- **Final Reserve Fuel (FRF)** The amount of fuel to be used in emergency scenarios. The amount of fuel shall be enough to be able to fly for 30 minutes at holding speed at 1500 ft above the destination airport, using the estimated mass on arrival at the (alternate) airport. The amount of FRF required depends on the geographical location, as the Federal Aviation Administration (FAA) sets the FRF to 45 minutes, while the EASA sets the FRF to 30 minutes.
- **Discretionary fuel** The extra amount of fuel that may be loaded by flight dispatchers or the pilot-in-command. Non-regulated alternate fuel is also considered discretionary fuel.

Since there are different types of fuel on board of an aircraft, the risk of fuel depletion increases when each buffer is being used. Pilots are required to frequently monitor the remaining fuel and the Estimated Landing Fuel (ELF). A MAYDAY emergency is present when the ELF is less than the FRF, which is the highest level of emergency. The second level is a declaration of MINIMUM FUEL, which should be made when a change in the current clearance would immediately result in a MAYDAY emergency. The third level is that pilots are required to formally ask for delay information of the destination when the ELF is less than FRF and AF. These levels of emergencies should be used to assess the risk on emergencies on flights caused by the fuel loading procedure.

In order to achieve flights with a higher fuel efficiency and lower risk on fuel shortage, the accuracy of the predictions of fuel consumption needs to be increased. The goal of this paper is to create a model that assists dispatchers on fuel loading decisions, using TF deviation predictions. Deviations in TF are estimated, such that the values can be directly transmitted in expected extra fuel that will be used for a specific flight. The TF deviations analysed in this paper are defined by Equation 1, where TF_{dev}, TF_{plan}, AF_m, and TF_{used} represent the TF deviation, the planned TF, the mandatory AF and the actual TF. b_{div} represents a binary variable which is equal to 1 for diverted flights and equal to 0 for non-diverted flights. The outcomes of the study may be compared to other studies predicting total trip fuel directly, by adding the planned fuel to the estimated deviations. The new predictions could be used to support on fuel loading decisions, by informing flight dispatchers and pilots on expected TF_{dev}. Furthermore, pilots may gain more trust in the system using the new predictions leading to less conservative fuel loading. This way, a more accurate decision on the risk on fuel shortage of each flight could be made, potentially further reducing conservative fuel loading. Besides assisting on fuel loading decisions, the trip fuel prediction improvements could also help with operational problems such as determining the weight balance and altitude during the flight, optimization of tankering fuel and the aircraft assignment problem.

$$TF_{dev} = TF_{plan} + b_{div} * AF_m - TF_{used} \tag{1}$$

Based on the research gaps found in literature present in section 2, the paper enhances knowledge on TF_{dev} prediction models for fuel loading advice through the following contributions:

- A 2-layered Time Series (2TS) model is proposed that enhances TF_{dev} consumption predictions by using available information of preceding flights, to make the most use of temporal and spatial information available.
- The 2TS model and simpler variations of the model and the gradient boosting model are evaluated in a fixed buffer loading strategy, such that the benefits of advancements in TF_{dev} prediction model can be put in perspective.

- Descriptive mission parameters of flights are computed and tested against TF_{dev}, to investigate the causes of TF_{dev}.

The paper is structured as follows. First, the literature review and is present in section 2. Thereafter, the prediction model is proposed in section 3 alongside a fuel loading strategy. The results of the proposed prediction model are presented in section 5 and compared to the predictions of the FPS and a baseline model. The paper is concluded in section 6, with recommendations on further research.

2 Literature Review

[Trujillo, 1996] investigated reasons for excess fuel loading and found that pilots load extra fuel mainly based on past personal experience and traffic and weather forecasts. The problem has not been addressed often until [Ryerson et al., 2015] performed statistical analysis on fuel data from a single airline, and used these results to generate potential fuel savings for all the major airlines in the US. They found that 2.21% - 4.48% of aircraft fuel is burned due to carrying unused fuel and 0.70% - 1.04% comes from loading more fuel than necessary to ensure a safe flight. Thereby, they introduced a cost-to-carry calculation method, to estimate fuel savings when carrying less fuel. Fuel burn because of excess discretionary fuel loading is also studied by [Hao et al., 2016], to measure the unpredictability and costs of an air transportation system where they found that 6-11 minutes of fuel would be loaded less if there would be no unpredictability in the aviation system. [Gomes et al., 2020] studied fuel loading regulations in Brazil and they found that 0,21% of fuel consumption could be reduced, without increased risk on fuel shortage, if the regulations would be changed. [Kang et al., 2018] introduced machine learning methods for fuel loading decisions to create confidence bounds for fuel usage on individual flights. The study showed that weak base learners, with Gradient Boosting (GB) being the best performing, reduced the Root Mean Squared Error (RMSE) by up to 50% and ensembles of base learners improved accuracy by 2-5%, which leads to fuel savings of up to 2%. Important predictors were proven to be the trip fuel and flight time distributions of the prior year, based on the same month, Origin-Destination (OD)-pair and hour block. Additionally, multiple machine learning algorithms were tested by [Achenbach, 2018], to predict flight time and fuel burn for international flights, where GB was the best-performing model among short-haul flights, with an RMSE reduction of 28%. The flight time of prior flights was found to be the most important parameter for flight time predictions. [Kang and Hansen, 2021] used machine learning models, with the random forest algorithm as the best performing base learner, to estimate the statistical contingency fuel for airlines. Using scaled predictions, savings of 0.16-0.55% were found. The paper addresses the difference in the distribution of TF_{dev} for different flight legs and the use of traffic-based input parameters of the prior year. [Zhu and Li, 2021] incorporated delay states at the origin, destination and other airports in the network by using a spatial weighted long-short-term recurrent neural network model to estimate flight times in order to improve fuel loading decisions. They found that outliers in flight time, which are flights with a high risk of fuel depletion, benefit the most from the newly proposed methods with an RMSE reduction of 50%, while non-outlier flights benefited less, with a 22% RMSE reduction. [Khan et al., 2019], created a self-constructing neural network to estimate TF_{used} for separate flight legs, with an average RMSE reduction from 4% to 58%, heavily dependent on the bias present in estimations. Other flight parameters including the flight altitude, mach number, and wind direction were estimated to predict the TF_{used} by [Khan et al., 2021].

To summarize, TF predictions have been studied and improved compared to FPS TF predictions, with varying results, mostly based on the experimental set-up. Geographical differences and airline practices affect the results of the model, and the procedure of splitting training and test data chronologically or at random affects the model performance. Thereby, flights of each OD pair are sometimes modelled separately, because of the large differences in flight behaviour. Furthermore, TF_{dev} from prior flights helps to improve predictions for coming flights. One limitation of literature is that diverted flights are omitted in most studies. Diverting flights could be critical for safety analysis, and is therefore evaluated on extra fuel usage in the current study. Another limitation is that although TF_{used} predictions can often be improved with more advanced models, the fuel savings resulting from the models are often compared with the airline practice, and not with simpler versions of the model. It is, however, important to compare prediction models for their performance on costs and risk on fuel shortage to assess the importance of improving TF_{used} predictions.

Related research to TF_{dev} predictions are prediction models for specific components of the flight. In the current research, they are referred to as mission parameters. Fuel flow of flights has been studied for different phases of the flight [Uzun et al., 2021, Trani et al., 2004, Baklacioglu, 2016, Sen, 2009, Baumann and Klingauf, 2020], with RMSE reductions of up to 55% compared to models used by airlines, using operational and environmental settings of a flight. Especially estimations of the fuel flow for the climb and descent are improved in the studies. Flight track wind uncertainty is studied by [Vazquez et al., 2017] using probabilistic models of wind. They found that the standard deviation of the trip consumption scales linearly with wind uncertainty. Moreover, uncertainty in fuel consumption is larger for headwinds than it is for tailwinds. The prediction of the flight trajectory, being the lateral and vertical profile of the flight in combination with the velocity, has been studied by for instance [Khan et al., 2021, Soler et al., 2012, Luke L. Jensen and Reynolds, 2023, Alligier and Gianazza, 2018, Gallego et al., 2019]. The RMSE of altitude estimations reduced by roughly 48% and for velocity estimations by 25%. Thereby, when better predictions of the flight trajectory are present, less in-flight re-optimization is required which makes changes to the intended flight path. Also, flight delays are modelled, which is a major source of extra flight time and fuel consumption. [Sternberg et al., 2017] analysed more than 200 papers relevant on flight delays, with a focus on terminal area delays, taxi delays, airline-specific delays and delay propagation. Despite the wide coverage in literature, [Zhu and Li, 2021] argue that most delay studies are intended for planning purposes, and not to model deviating flight time and so, TF_{dev}.

The improved prediction for mission parameters may be used to increase the accuracy of flight plan parameters, optimize operational strategies, communicate more efficiently with ATC, or model emissions more precisely. Some studies mention that improved mission parameter predictions can help improve TF_{used} predictions. Despite this, little is known about the importance of mission parameters on TF_{dev} encountered in practice. Therefore, they will be computed in this paper, and compared with TF_{dev}, to identify the most important parameters to improve on TF_{dev} predictions.

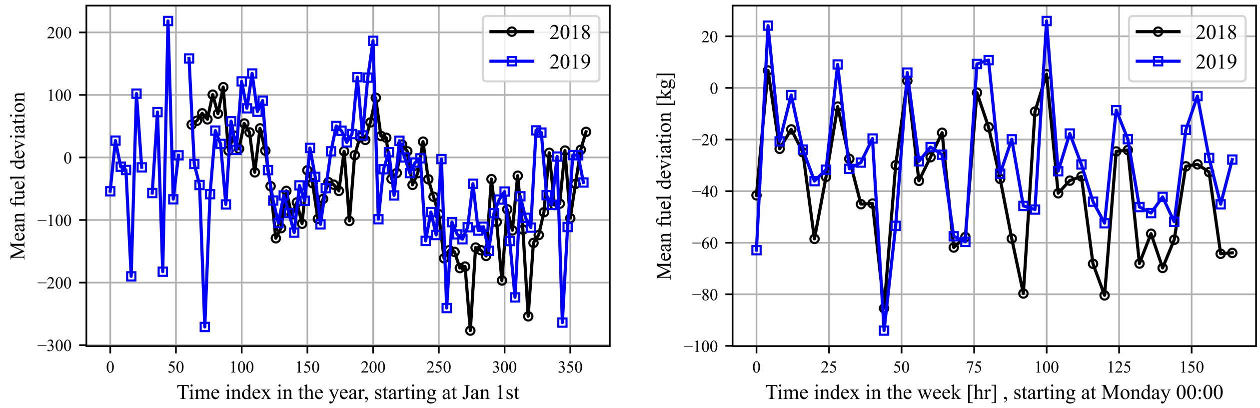
In this paper, the problem of predicting TF_{dev} will be modelled as a time series. The first reason that time series are used is that inputs based on mission parameter deviations of prior flights have shown to be important predictors, and the second that the distribution of TF_{dev} predictions varies greatly between OD-pairs and other defined clusters. Time series models can be used to extract critical temporal information on mission parameters to predict TF_{dev}. Thereby, sequences of data may be generated per OD-pair or other cluster, while the dependencies withing a sequence may be the same across the entire model. The implementation of time series problems is often done by using statistical models [De Gooijer and Hyndman, 2006]. Recent developments in time series are based on neural networks, with its time series variant, the recurrent neural network [Zhang et al., 2020]. Since its introduction, improvements to the model are made, each tackling issues within the model. A recently developed model is the Temporal Fusion Transformer (TFT) introduced by [Lim et al., 2021], which beats the prediction performance of other benchmark models. TFT is a novel transformer-based architecture, showing promising results in terms of prediction performance on various problems including traffic speed prediction ([Zhang et al., 2022]), wind speed prediction [Wu et al., 2022], and power demand prediction ([Li et al., 2023]). The TFT model will be explained in more detail in section 3.1.2. In order to identify the full potential of time-series models for TF_{dev} predictions, the state-of-the-art TFT model will be used.

3 Methodology

Improved fuel loading decisions may be enabled by an improved TF_{dev} prediction model. Fuel loading advice for flight dispatchers will be created using two steps. The first step is the creation of a prediction model for TF_{dev}. A 2-layered Time Series (2TS) model is proposed to demonstrate the performance of deep time series for TF_{dev} predictions. More information on the model is presented in section 3.1, and the input features of the model are described in section 3.2. Using the newly created predictions, a fixed-buffer fuel loading strategy is used to investigate the benefits in terms of fuel saving and Safety Key Performance Indicators (SPI) associated with the created prediction model, further explained in section 3.3.

3.1 Prediction Model

The 2TS model is a time series prediction model using deep learning, to demonstrate the performance of deep time series models for TF_{dev} predictions. The TFT model is used to make predictions within the 2TS model. Values of TF_{dev}, change over time, as presented in Figure 1, for the average value of TF_{dev} per day of the year and time of the week. The figures are created using the dataset presented in section 4. In literature, temporal information is often included using month, day of week and hour block input features. Hour blocks are used since a discretization of 1 hour would lead to results with a lot of noise. In practice, however, the average TF_{dev} per day or hour block is not constant throughout the year and between consecutive years. Figure 1 shows that the average TF_{dev} per hour block is higher towards the end of the week for the year 2019 than for 2018, showing a difference in the temporal pattern of TF_{dev} in the week. These changes may be caused by changes in the operations of an airline, changes in traffic congestion in the air network, differences in ATC allowances and more. Thereby, there are also temporary changes. Visible in the yearly time series in Figure 1, is that a roughly similar pattern is present for TF_{dev} in two consecutive years. Despite this, there are differences present, for example at day index 272. In this period, the average TF_{dev} for 2018 is more than 100 kg lower than it is for 2019. These temporary changes could be caused by environmetal effects, such as a period of extreme weather, or changes in the passenger mix, caused by for example school holidays. Time series enhance the use of available information when planning coming flights using data from prior flights. Furthermore, flights behave comparably within groups based on OD-pairs while they behave differently between groups, found in the literature study. For this, time series work well as they can be modelled for each specific group of flights, such that shared information within this group is used for coming flights. This way, information on deviations from the flight plan from recent flights for a certain OD-pair could be used to determine expected TF_{dev} for coming flights. In the current study, spatial levels are defined to represent groups or clusters of flights based on the OD pair. The spatial levels defined are: the entire network, the planned arrival airport, the planned departure airport and the planned flight leg. Each flight belongs to a group in each of the spatial levels.



(a) Yearly time series, with average TF_{dev} per day of the year, grouped per four days

(b) Weekly time series, with average TF_{dev} per hour block of the week, grouped per four hours

Figure 1: Average values of TF_{dev} throughout the year and week

There is no trivial solution when modelling flights in an airlines network as a time series. Flights originate and arrive at various airports, have different flight durations and have different instances for the creation of the Operational Flight Plan (OFP). Introducing time series to the TF_{dev} prediction problem, two separate methods are created and thereafter combined. Hence, the model is called the 2-layered time series. A Fixed Time Interval Time Series (FTTS) is proposed to model seasonality effects present in the data, and to consider time series for different spatial levels, defining them as sequences that contain the groups of datapoints. As flights do not arrive on fixed time intervals in reality, flight data needs to be grouped and transformed per time intervals, which is set as target to predict coming time intervals. The predicted targets are then collected for each flight and serve as inputs for a second model, that is able to predict individual flight target data, instead of grouped values. A Individual Flight Time Series (IFTS) is proposed to allow individual flight data, including flight plan and weather information, to be used from preceding flights with the same OD-pair. This way, the relationships between flights within a OD-pair may be directly used for predictions of coming flights.

3.1.1 Model architecture

The model architecture, with its inputs and and relationships will be described in the current section. Figure 2 shows a schematic representation of the proposed 2TS model, with its sub-models to estimate TF_{dev} for a flight. The FTTS model uses time interval features as inputs, to predict features for the following time intervals. The IFTS model uses features of the to-be-predicted and preceding flights, to predict for multiple coming flights.

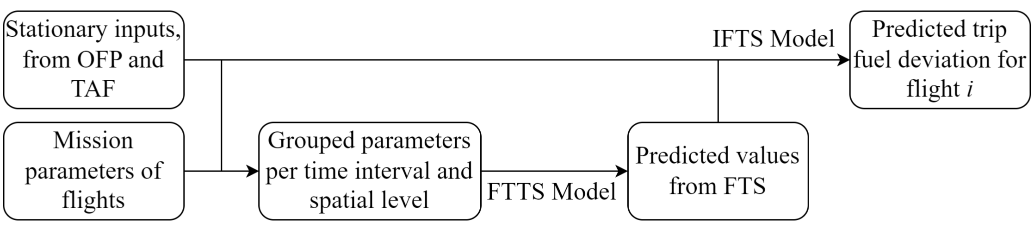


Figure 2: Prediction model for trip fuel consumption

A summary of the used inputs and parameters in the model is present in Table 4. A distinction is made between flight specific parameters, time interval parameters and model parameters. For the TF_{dev} denoted as y_i, predictions, denoted by $\hat{y}_i$, are created for each flight $i \in [0, N_f]$, where N_f represents the number of flights used in the dataset. Each flight has stationary parameters from the OFP and Terminal Aerodrome Forecasts (TAF) used in the model and are defined as vectors $\mathbf{o}_i \in \mathbb{R}^{N_f * N_O}$ and $\mathbf{t}_i \in \mathbb{R}^{N_f * N_T}$ respectively, where N_O indicates the number of OFP input features and N_T the number of TAF input features. Thereby, for each flight, the mission parameters $\mathbf{m}_i \in \mathbb{R}^{N_f * N_M}$ are computed. So, each flight has a combined vector of features $\mathbf{p}_i \in \mathbb{R}^{N_f*(N_O+N_T+N_M)}$. An instant of time is denoted by t and each flight has an arrival time t_i^{arr} and a planning time t_i^{plan}. Inputs $\mathbf{o}_i$ and $\mathbf{t}_i$ are considered to be known for all flights, also for coming flights. This information is only known during the planning phase in reality. However, as time series predictions would only be used at the planning time, the assumption can be made that all future inputs are known, without the model behaving differently in reality. The inputs $\mathbf{m}_i$ are considered known once an aircraft has landed, so when $t_i^{arr} < t$.

Within the dataset, groups of flights are created, based on the OD-pair of the flight. Thereafter, the flights are sorted on their respective arrival time. This way, data from preceding and succeeding flights can be selected easily using indexing. For example, $\mathbf{o}_{i-k_m:i+h_m}$ represents the set of OFP input data from flight i, the k_m flights that arrived before flight i with the same OD-pair and the h_m flights that arrived after flight i with the same OD-pair.

Table 1: Summary of model parameters

Parameter, vector or matrix symbol	**Description**
	Per flight i
$\mathbf{p}_i \in \mathbb{R}^{N_f*(N_O+N_T+N_M)}$	Combined features of a flight
$\mathbf{o}_i \in \mathbb{R}^{N_f*N_O}$	OFP input features of a flight
$\mathbf{t}_i \in \mathbb{R}^{N_f*N_T}$	TAF input features of a flight
$\mathbf{m}_i \in \mathbb{R}^{N_f*N_M}$	Mission parameter features of a flight
$\mathbf{g}_i \in \mathbb{R}^{N_f*N_{sg}}$	Spatial groups of flight
$t_i^{arr} \in \mathbb{R}^{N_f}$	Arrival time index of flight
$t_i^{plan} \in \mathbb{R}^{N_f}$	Planning time of flight
$TI_{g,i}^{arr} \in \mathbb{R}^{N_f*N_{sg}}$	Arrival time interval of flight i and spatial group $g \in \mathbf{g}_i$
$TI_{g,i}^{plan} \in \mathbb{R}^{N_f*N_{sg}}$	Planning time interval of flight i and spatial group $g \in \mathbf{g}_i$
$h_{g,i}^{FTTS} \in \mathbb{R}^{N_f*N_{sg}}$	Required forecast horizon for FTTS for flight i and spatial group $g \in \mathbf{g}_i$
$h_i^{IFTS} \in \mathbb{R}^{N_f}$	Required forecast horizon for IFTS for flight i
$\mathbf{y}_i^{IFTS} \in \mathbb{R}^{N_f*h_m}$	Predictions from the IFTS model for flights within the forecast horizon, starting with flight i
$\hat{\mathbf{y}}_{g,i}^{FTS} \in \mathbb{R}^{N_f*N_{sg}}$	Predicted FTTS parameters for a flight and spatial group $g \in \mathbf{g}_i$
$\hat{y}_i \in \mathbb{R}^{N_f}$	Predictions of TF_{dev} per flight
	Per time interval TI
$\mathbf{fp}_{g,TI} \in \mathbb{R}^{(\sum_{g\in\mathcal{G}} N_{TI_g})*(N_{fm}+N_{ft})}$	Combined features per time interval
$\mathbf{fm}_{g,TI} \in \mathbb{R}^{(\sum_{g\in\mathcal{G}} N_{TI_g})*N_{fm}}$	Mission parameter features per time interval and group
$\mathbf{ft}_{g,TI} \in \mathbb{R}^{(\sum_{g\in\mathcal{G}} N_{TI_g})*N_{ft}}$	Temporal features per time interval and group
$\mathbf{Y}_{g,ti}^{FTTS} \in \mathbb{R}^{(\sum_{g\in\mathcal{G}} N_{TI_g})*N_{ft}*h_m}$	Predictions from the FTTS model, being the predicted parameters for each time interval within the forecast horizon
	Per time series model m
h_m	Forecast horizon
k_m	Look-back window

Besides parameters for each flight i, there are parameters defined per time interval TI. Time intervals are created for the FTTS model, which represent periods of time, in which data of flights is grouped. The FTTS model will be used for different spatial levels, indicated with $\mathbf{g}_i \in \mathbb{R}^{N_f*N_{sg}}$, where N_{sg} represents the spatial levels, being the OD-pair, arrival airport, departure airport of flight i and the entire network. So, time intervals are created for each group within a spatial level. Each flight arrives at time intervals $TI_{g,i}^{arr} \in \mathbb{R}^{N_f*N_{sg}}$ and is planned at time intervals $TI_{g,i}^{plan} \in \mathbb{R}^{N_f*N_{sg}}$. Then, all combined parameters per time interval of a specific group are defined as $\mathbf{fp}_{g,TI} \in \mathbb{R}^{(\sum_{g\in\mathcal{G}} N_{TI_g})*(N_{ft}+N_{fm})}$, where $\mathcal{G}$ represents the set of all groups used in the model, N_{TI_g} the number of time intervals for group g, N_{ft} the number of temporal features, and N_{fm} the number of mission parameter features used per time interval. This vector consists of the mission parameter features

$\mathbf{fm}_{g,TI} \in \mathbb{R}^{(\sum_{g\in\mathcal{G}} N_{TI_g})*N_{fm}}$ and temporal parameters, $\mathbf{ft}_{g,TI} \in \mathbb{R}^{(\sum_{g\in\mathcal{G}} N_{TI_g})*N_{ft}}$. The temporal parameters are known for all time intervals, while the mission parameters are only known for arrived flights.

Using the inputs, $\hat{y}_i$ is computed using the set of equations shown in Equation 2. First, a matrix of parameters $\mathbf{Y}_{g,ti}^{FTTS} \in \mathbb{R}^{(\sum_{g\in\mathbb{G}} N_{TI_g})*N_{ft}*h_m}$ is computed using the FTTS model. $f_1()$ represents the function used in the FPS model. Each model has a forecast horizon h_m and look-back window k_m. Using parameters of all preceding intervals within the look-back window and all temporal parameters for the to-be-predicted time intervals, $\mathbf{Y}_{g,ti}^{FTTS}$ is generated, which is a matrix consisting of the vectors of predicted parameters of a time interval, for all time intervals in the forecast horizon. By subtracting TI_i^{plan} from TI_i^{arr}, the required forecast horizon $h_{g,i}^{FTTS}$ for the FTTS model is computed. Depending on this value, the appropriate forecasted vector is selected from all prediction matrices $\mathbf{Y}_{g,ti}^{FTTS}$. This way, each flight is assigned predicted FTTS values $\hat{\mathbf{y}}_{g,i}^{FTTS}$ for all spatial groups the flight belongs to. Once the predicted FTTS values are calculated, prediction vectors are created for the IFTS model, using the OFP, TAF, mission parameter and FTTS features. Now, the IFTS model function is referred as $f_2()$. The prediction vector contains a prediction for a succeeding flights within forecast horizon h_m after flight i. Again, the correct prediction should be selected from all the prediction vectors. Now, the required forecast horizon, h_i^{IFTS} is based on the amount of flights with the same OD-pair, that landed between t_i^{plan} and t_i^{arr}. Using this value, the correct predicted value $\hat{y}_i$ may be selected.

$$\begin{aligned}
\mathbf{Y}_{g,ti}^{FTTS} &= f_1(\mathbf{fp}_{g,ti-k_m:ti-1}, \mathbf{ft}_{g,ti:ti+h_m}) \\
\hat{\mathbf{y}}_{g,i}^{FtTS} &= \mathbf{Y}_{g,TI_{g,i}^{plan}}^{FTTS}[h_{g,i}^{FTTS} - 1] \\
\mathbf{y}_i^{IFTS} &= f_2(\mathbf{o}_{i-k_m:i+h_m}, \mathbf{t}_{i-k_m:i+h_m}, \mathbf{p}_{i-k_m:i-1}, \hat{\mathbf{y}}_{g\in\mathbf{g}_i,i-k_m:i+h_m}^{FTTS}) \\
\hat{y}_i &= \mathbf{y}_{i-(h_i^{IFTS}-1)}^{IFTS}[h_i^{IFTS} - 1]
\end{aligned} \tag{2}$$

Fixed Time Interval Time Series Model The FTTS model is created to generate inputs for preceding models for TF_{dev} estimations. The outputs of the FTTS will be used for the IFTS model. The FTTS is designed such that it enables the use of multiple spatial levels, that will be treated as sequences, and the use of repeating patterns apparent in the fixed time blocks.

The two driving parameters for the design of the FTTS are the length of the time intervals, $L_g^{TI} \in \mathbb{R}^{N_G}$, and the transformation of the grouped data within each time interval, $\phi_g(), \forall g \in \mathbb{G}$. The length of the time interval determines the generalisation performance versus the precision of time-dependent data. A large time interval results in time blocks with many flights in one block such that each block has a representative data distribution, while a low time interval allows for detailed allocation of data to specific time intervals. This accurate temporal information would be lost for large time interval series. The transformation of the data within each time interval determines what values to predict for each time interval.

Each group, or entity, in a spatial level, has different dataset sizes. L_g^{TI} is chosen such that the time-space is discretized in as small steps as possible, while there should be a low number of time intervals with zero flights within a time interval. This means that within each spatial level, TI_L changes, depending on the size of the group. The possible time intervals considered in this research are: 3 hours, 6 hours, 1 day, 1 week and 1 month. A group is assigned to the L_g^{TI} where the dataset size is more than twice the amount of time intervals ($\frac{TS}{L_g^{TI}}$), such that a fair balance between discretization and empty data sets are found. TS denotes the total time span of the dataset. A TFT model is trained on every set of groups within a spatial level with the same L_g^{TI}, as the model is able to train on multiple sequences at once. Model parameters that change per TI_L are the look-back window k_m and the forecast-horizon h_m. The values per L_g^{TI} are shown in Table 2.

Table 2: Look-back window k_m and forecast horizon h_m for each time interval length L_g^{TI}

$\mathbf{L}_g^{TI}$	$\mathbf{k}_m$	$\mathbf{h}_m$
1 month	1	2
1 week	4	2
1 day	14	2
6 hours	12	3
1 hour	48	8

In the current research, the mean and standard deviation of a parameter within the vector of parameters in a time interval are used as transformations $\phi_g()$. The mean may be computed with only one data point in a time interval and the standard deviation using only two data points. Other transformations such as quantiles of the dataset, require more data for a good representation. If there are not enough data points for a transformation, the average value of the training set is used. The set of target variables $\hat{\mathbf{y}}^{FTTS}$ is not equal to the complete set of mission parameters $\mathbf{m}$, as only the trip fuel deviation, flight time deviation, fuel flow deviation, descent fuel burn deviation, and cruise fuel burn deviation are used, with their corresponding mean and standard deviation. The computation of the input parameters is described in section 3.2.

Individual Flight Time Series Model The IFTS model is used to make predictions of target data for individual flights, combining inputs $\mathbf{o}_i$, $\mathbf{t}_i$, $\mathbf{m}_i$ and $\hat{\mathbf{y}}_{g,i}^{FTS}$. As flights need to be arrived at the destination in order to generate $\mathbf{m}_i$, not all flights can use the previously planned flight as input upon the planning time. Therefore, a multi-horizon forecast is created, such that the required horizon may be selected, depending on the number of preceding flights that are planned, but have not landed yet. For most flight legs, a 1-step ahead forecast will suffice (in fact, for 90% of the flights in the current dataset). Since the accuracy of the model may decrease for multi-horizon forecasts, a 1-step ahead IFFT and a τ-step ahead IFFT is created, where τ represents the maximum time-step ahead required in the dataset. Moreover, the time between the planned arrival time of a flight and the following flight for a specific leg is used as input parameter, such that the model uses information on irregular time intervals.

3.1.2 Temporal Fusion Transformer

The TFT is used as the time series model in the 2TS architecture to predict TF_{dev} for coming flights, which is denoted by $f_1()$ and $f_2()$ in Equation 2. The TFT has beaten the prediction performance of other advanced time-series models, because of its ability to distinguish long-term relationships by the attention mechanism as well as short-term dependencies by the LSTM layers. Besides top-level accuracy, the method has more advantages. First of all, the model is able to predict values for multiple time series sequences in a single model, even with unequal time series lengths. This way, similar behaviour between sequences may be used to train the model, while only preceding inputs of the respective group are used. Second, the TFT has two selection mechanisms, deciding which data to process in the model and whether to use non-linear or linear processing. This way, only relevant data is used and the model is not overly complex. Third, the TFT incorporates static input data in multiple places with embeddings of static categorical features, being able to model differences between static groups. Fourth, the attention mechanism in the metric allows the user to output the feature importance of the model variable. Especially for human experience-related predictions, such as the current fuel loading problem, the interpretability of the model is of utmost importance.

The TFT distinguishes 3 types on input data for each entity i, previously defined as groups within a spatial level. Each entity, being different separate flight legs, has time steps $t \in [0, T_i]$ with static covariates $\mathbf{s}_i \in \mathbb{R}^{m_s}$. Thereby it has inputs $\chi_{i,t} = [\mathbf{z}_{i,t}^T, \mathbf{x}_{i,t}]^T$, which is divided in unknown observed inputs, which is only available for previous time-steps, $\mathbf{z}_{i,t} \in \mathbb{R}^{m_z}$ and known inputs $\mathbf{x}_{i,t} \in \mathbb{R}^{m_x}$. Each input parameter could be categorical or numerical, where categorical features are represented as vector embeddings. Equation 3 presents the form of each forecast, where $\hat{y}_i(t, \tau)$ is a point prediction of the τ-step-ahead forecast at time t, and $f()$ is a prediction model. k represents the look-back window for past information.

$$\hat{y}_i(t, \tau) = f(\tau, y_{i,t-k:t}, \mathbf{z}_{i,t-k:t}, \mathbf{x}_{i,t-k:t}, \mathbf{s}_i) \tag{3}$$

The architecture of TFT is presented in Figure 3, which will be expanded upon. The full explanation of the model is presented by [Lim et al., 2021]. The TFT is designed to combine multi-horizon forecasting with interpretable insights. Its architecture has elements specially designed to deal with different types of time series. The key components present in the model's architecture are:

1. **Static covariate encoders.** Static input data may have relevant context for the relationship of input parameters. Therefore, static features are integrated in multiple places in the network, with the encoding of context vectors to condition temporal dynamics. The static covariates are transformed into four different context vectors, which are used into separate locations in the temporal fusion decoder, being the variable selection, the local processing of temporal features and the enriching of temporal features with static information. The integration of static encoders leads to a distinction between time series of different entities, while similar patterns are shared. In the current study, OD-pairs are used as static variables.
2. **Gating mechanisms and variable selection.** The precise relationship between exogenous inputs and targets is often unknown in advance, and it is unknown if non-linear processing is required, or that simpler linear models will suffice. Therefore, there are Gated Residual Network (GRN) blocks present

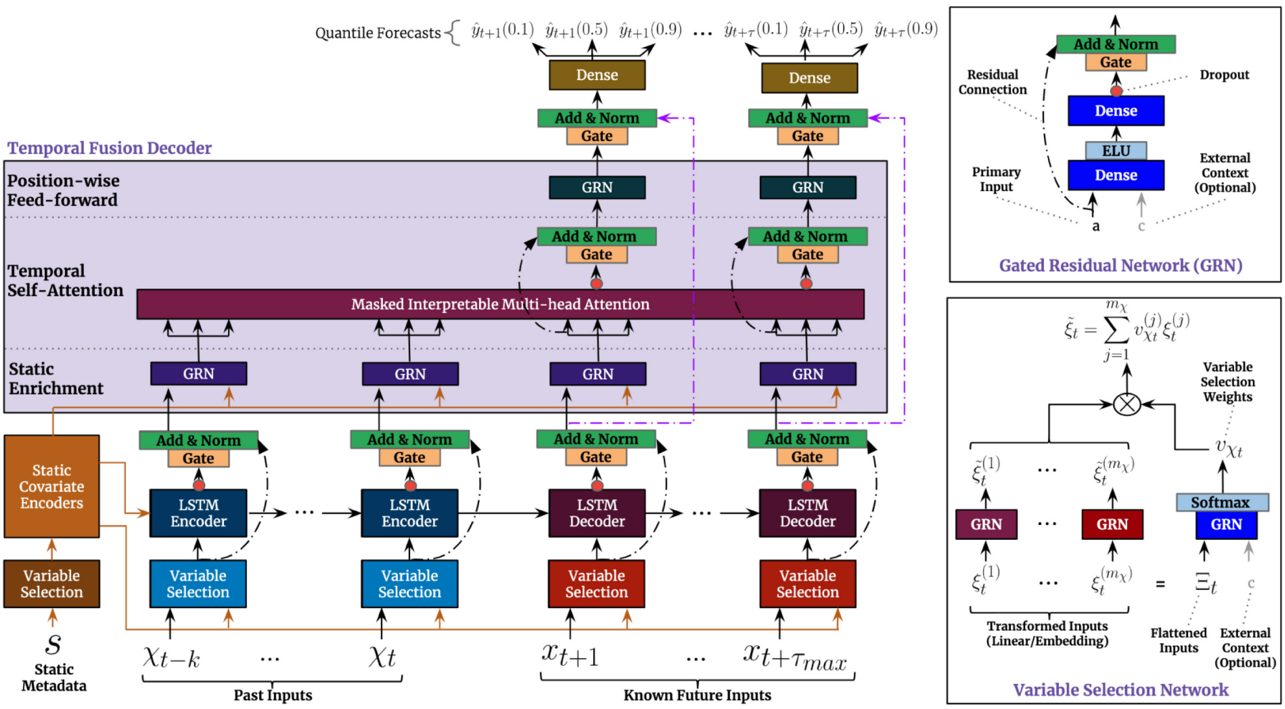


Figure 3: The TFT architecture, from [Lim et al., 2021]

in multiple places in the model architecture. The primary input **a** and an optional context vector **c**, transformed from the static covariates **s**, are combined in an architecture with a residual connection and an activation function. This way, inputs can be not transformed at all, transformed using a simple linear function, or using a more complex non-linear function. Each input variable is first transformed using an embedding or linear transformation for a categorical or numerical input variable, respectively. Thereafter, each input variable has its GRN, which determines how the information should go through. Irrelevant components of the architecture and irrelevant input variables can be filtered out this way, to achieve a better representation of the problem.

3. **A sequence to sequence layer for local processing.** States of past data affects the data of future data points. Therefore, a LSTM-based sequence-to-sequence layer is created. Using only this layer, the model would be equal to an LSTM model [Hochreiter and Schmidhuber, 1997].

4. **A temporal self-attention decoder** The outputs of the sequence-to-sequence layer are, after being transformed by a GRN, decoded in a temporal self-attention layer. This layer interprets the temporal relation between the inputs and selects the important information to extract. Long-term dependencies are especially well extracted in this layer since the output of each time step does not need to be transformed by a sequence-to-sequence layer.

[Lim et al., 2021] specifies that the model is able to train on multiple loss functions simultaneously. Despite this, a single loss function is used, in order to compare the results to other models. A point prediction is created, for which the MSE loss is used, presented in Equation 4.

$$MSE(y, \hat{y}) = (\hat{y}_i - y_i)^2 \tag{4}$$

3.1.3 Baseline methods

The proposed 2TS model will be compared to other models, to analyse its behaviour and assess its prediction performance. The other models exist of components of the 2TS model and high-performing models in literature. Using the available input data, the GB model, introduced by [Friedman, 2001], is found to have a high prediction performance for trip fuel predictions [Kang, 2017, Achenbach, 2018]. The GB regression algorithm combines multiple base learners to create a strong, generalized combined model. Often, the base learners are decision trees. In a decision tree, binary choices are made based on the input data, where the predicted outcome is based on those decisions. Many of those decision trees can be created, using different parts of the data and different input features. Combining those decision trees is called bagging [Breiman, 1996]. GB aims to minimize the residual errors from previous trees, such that it may converge to a better predicting model. When the GB model is used, input parameters of prior flights are not used in the model.

Moreover, the performance of the model is assessed by creating simpler variations of the model. An IFTS model using only individual flight information will be tested, without FTTS inputs. Second, a GB model will be chosen using known covariates and fixed time interval predictions. A summary of the tested models is presented in Table 3, with each model name, and the used model for the IFTS and FTTS components of the 2TS model.

Table 3: Prediction models that are evaluated, built out of two sub-models

Model Name	**Fixed Time Interval model**	**Individual Flight model**
2TS	FTTS	IFTS
FTTS	FTTS	GB
IFTS	-	IFTS
GB	-	GB

3.2 Feature engineering

The creation of the input features is described in the following section. The input features are used for a TF_{dev} prediction model during the planning phase of a flight. All the available input data used in the prediction should be available at this point in time. The planning time of each flight and availability of the data is rounded per hour, for easier processing of the data.

The input data is divided into three types, which are visualized in section 3.1. The first is the flight static data known during the planning phase of the flight, consisting of the flight plan information and weather information available from TAF. The second group is data from preceding flights including the same static information as well as mission profile parameters of the flight. The third group is spatial time block predictions, built out of time series predicting the mean over/underburn at a given spatial level for a given time duration of the arrival time of the flight.

3.2.1 OFP features

The OFP has a detailed description of the intended flight path of an aircraft, as well as important parameters describing the to-be-performed flight. These parameters may influence the average value or chance on large values of TF_{dev}. Table 4 shows the parameters available in the OFP used for the prediction model. The OD-pair, Arrival airport, and Departure airport all influence the deviations, because terminal efficiencies deviate per airport. The parameters are categorical variables in the TFT model, the OD-pair is used as static input. Temporal parameters are the Year, Month, Day of Week and Hour Block, all used as categorical variables. The Hour Block is divided into blocks of three hours. The other parameters describe the general characteristics of the flights. Combined, the inputs created the input vector $\mathbf{o}_i$, used in the IFTS model. In the FTTS model, temporal features are used, which are the month, day of week and hour block, which make the input vector $\mathbf{fm}_{g,TI}$. Dependent on L_g^{TI}, the Hour Block and Day of Week features are omitted. the TFT model is able to transform categorical features. For the GB model, one-hot encoding is used for categorical features.

Table 4: Summary of OFP input data used in the model for STS

Flight Identification			
OD-pair	Arrival airport	Departure airport	Year
Month	Day of week	Hour Block	
Flight description			
Planned trip fuel	Planned flight time	Planned distance	Great circle distance
SID + STAR distance	Max Altitude	Take-Off Weight	Max wind speed
Avg. head wind	Aircraft performance factor		

3.2.2 TAF features

TAFs are weather forecasts available for most airports. The TAFs contain weather predictions for the area around the aerodrome indented for aviation. Using these forecasts, weather-related input features are created. Depending on the weather predictions, terminal efficiencies change, making weather forecasts important predictors for large arrival delays. Important elements in the forecast are: horizontal visibility, vertical visibility, precipitation and runway cross-wind. According to regulations, AF is only required when there is IFR expected at the arrival time at the arrival airport plus or minus one hour.

A bad weather, W_b and very bad weather, W_{vb} indicator are used as input features for the model. W_{vb} is equal to one when there is either: low IFR, snow or thunderstorm predictions, and zero otherwise. W_b is equal to one when there is either: any precipitation, IFR or crosswind above 20 knots on the runway. The two features are combined into the input vecotr $\mathbf{t}_i$, used for the IFTS model.

3.2.3 Historical flight mission parameters

Flight data of preceding flights are used in 2TS. There are many possible mission parameters that describe the flight, besides the TF itself. Some parameters may be estimated more accurately than others compared to the FPS [Khan et al., 2021]. By splitting a flight in separate elements, some information may be estimated accurately, to use for coming predictions. The computation of descriptive parameters will be described in the current section. All of the descriptive parameters will serve as input for the 2TS model, while only some parameters will serve as input for the FTTS model. The flight descriptive parameters of interest are: distance, fuel flow, velocity, flight time, flight altitude, effective wind, and aircraft weight. Each parameter is defined as deviation from the flight plan, consistent with TF_{dev} as the target variable. Thereby, special occurrences are computed as well. The special occurrences serve as input for the IFTS model and may be used for further analysis on causes of trip fuel consumption. The combination of mission parameters per flight is presented as $\mathbf{m}_i$, used for the IFTS model. In this section, a description of waypoints will be given first, and the computation of the mission parameters thereafter.

Some of the descriptive parameters require en-route flight information, to compare the OFP data to the Flight Data Recorder (FDM) data. Therefore, the flight will be segmented, to be able to compute en-route values of desired parameters. The flight is segmented as follows. The flight plan consists of waypoints along the route with lateral straight flight legs in between. These waypoints usually correspond to entry/exit points of countries or airways. For each segment, either a cumulative value or an average value of the desired mission parameter is extracted. This value may then be compared to the FDM data. An example of a segment of a flight is shown in Figure 4. The large, blue points indicate waypoints of the OFP, while the small, green points indicate FDM measurement points. The average or accumulated value of all the FDM points may then be compared to the description of the waypoint in the OFP. The set of OFP waypoints is defined as $\mathbf{OFP}_i$, which contains mission parameter data per waypoint $\mathbf{ofp}_{w,i}$ and the set of FDM points as $\mathbf{FDM}_i$ with parameters per point $\mathbf{fdm}_{w,i}$. Each OFP waypoint is linked to a single FDM measurement point, by measuring the smallest lateral distance. Then, all FDM measurement points in between two waypoints are assigned to the subsequent waypoint. The first FDM point assigned to an OFP waypoint $\mathbf{ofp}$ is presented as $\mathbf{fdm}_{o_1}$ and the last point as $\mathbf{fdm}_{o_2}$. Thereafter, the different mission profile parameters may be computed per segment. FDM time steps are taken for every 20 seconds, which is deemed a small enough time step to lead to accurate results. A flight start, t_i^{start} is defined to start at the beginning of the runway. Since, there is no direct measurement when that instance is, the first moment in the air, t_i^{air} is used, subtracting a fixed time to include the take-off time, t^{run}. The flight ending, t_i^{end}, is defined to be a fixed time instance, t^{lan} after the first second on the ground after the flight, t_i^{ground}. t_{run} is set to three minutes and t_{lan} to two minutes, the reasoning is present at the computation description of the trip fuel.

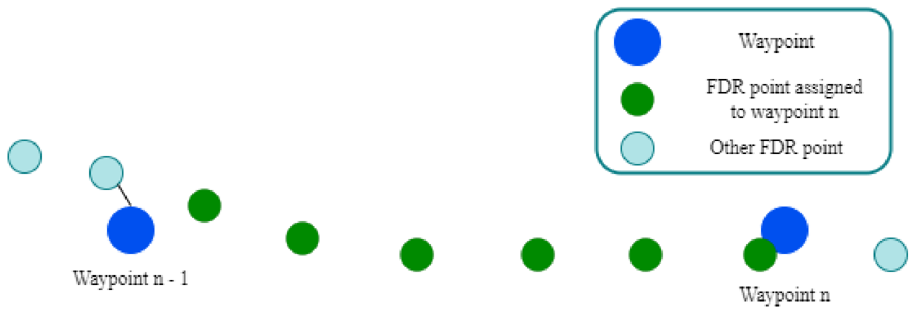


Figure 4: Link of FDM data and OFP data

Trip fuel The trip fuel of each flight, $TF_i \in \mathbb{R}^{N_f}$ is the target for the prediction models. TF is measured by comparing the fuel weight at the start of the flight and at the end of the flight, as described in Equation 5, where W_t^f indicates the fuel weight at time t. The weight sensors are not fully accurate, as accelerations cause the weight in the tank to increase. Therefore, any weight measurement during take-off is unreliable. So, a proper length t_{run} of three minutes is used and two minutes for t_{lan} to overcome these issues. The trip fuel is computed for the entire flight. When the trip fuel is desired between two waypoints, TF_i^{way}, the fuel flow is integrated over time to get the resulting value.

$$TF_i = W^f_{t=t_{end}} - W^f_{t=t_{start}} \tag{5}$$

Distance A distance input feature $DIS^i \in \mathbb{R}^{N_f}$ is computed to compare the actual versus the planned flight distance, shown in Equation 7. The flight distance of a segment is defined as the distance between two waypoints, calculated using Equation 6, where R_E is the radius of the earth, ϕ_i, the latitude of point i and λ_i the longitude of point i. Actual distances are computed using the FDM datapoints and the planned distance using the OFP data.

$$d^{i,j}_{line} = R_E * 2arcsin\sqrt{sin^2(\frac{\phi_i - \phi_j}{2}) + cos\phi_i cos\phi_j sin^2(\frac{\lambda_i - \lambda_j}{2})} \tag{6}$$

$$DIS^i = d_{plan} - \sum_{f \in \mathbf{FDM}} d^{f,f+1}_{line} \tag{7}$$

Flight time The difference in flight time, $FT^i \in \mathbb{R}^{N_f}$ is defined as the duration spent in a certain segment, shown in Equation 8. The flight time between two waypoints is described in the OFP. The flight time from the FDM is defined as the number of FDM data points, multiplied by 20 seconds. Flight time deviates when another distance is flown than intended, or when the velocity changes during the segment.

$$FT^i = FT^i_{plan} - (t_{start} - t_{end}) \tag{8}$$

Fuel Flow The fuel flow of each datapoint, FF^i of the FDM is averaged, to compare the deviation in fuel flow to the flight time and computed in Equation 9. N_{FDM} represents the number of FDM points in a flight. The fuel flow may change because of aircraft performance or weight differences compared to the OFP, or because of a different trajectory than planned.

$$FF^i = FF^i_{plan} - \frac{\sum_{f \in \mathbf{FDM}} f_{FF}}{N_{FDM}} \tag{9}$$

Velocity The velocity, $V^i \in \mathbb{R}^{N_f}$ along a segment is defined as the mach number and computed as shown in Equation 10. N_{OFP} represents the number of OFP points in a flight. Since the altitude may deviate from the OFP and the FDM, the mach number resembles a better comparison of velocity then air or ground speed. Flights change their speed from planned speed either because of a new optimization given the current environmental settings, or new clearances on speed given by ATC.

$$V^i = \frac{\sum_{o \in \mathbf{OFP}} (o_{mach} - \frac{\sum_{f=\mathbf{fdm}_{o_1}}^{f=\mathbf{fdm}_{o_2}} f_{mach}}{\mathbf{fdm}_{o_2} - \mathbf{fdm}_{o_1}})}{N_{OFP}} \tag{10}$$

Altitude For the altitude of the mission profile, the maximum altitude is measured. The maximum altitude correlates less with other mission profile parameters than the average altitude. The maximum altitude of the OFP is compared to the FDM to get the altitude difference of each flight, $ALT^i \in \mathbb{R}^{N_f}$.Only a single value of the altitude is present in the OFP, at the location of a waypoint. Therefore, the altitude of the FDM datapoint closest to the waypoint is used. Aircraft operate more efficiently at higher altitudes, but environmental, ATC and internal factors including weight changes affect the maximum achievable altitude.

$$ALT^i = \frac{\sum_{o \in \mathbf{OFP}} (o_{alt} - \mathbf{fdm}_{o_2,alt}}{N_{OFP}} \tag{11}$$

Wind component The wind component, or effective wind, is defined as the wind vector along the flight direction of the aircraft and denoted by $Windc \in \mathbb{R}^{N_f}$. For each segment between waypoints, the average effective wind is stated in the OFP. The effective wind is computed by measuring the vector parallel to the flight direction. A negative sign indicates tailwind. The effective wind may deviate because the wind speed differs from the prediction, or a different altitude or route is flown, with different wind speeds and flight directions.

$$Windc^i = \frac{\sum_{o \in \mathbf{OFP}} (o_{windc} - \frac{\sum_{f=\mathbf{fdm}_{o_1}}^{f=\mathbf{fdm}_{o_2}} f_{windc}}{\mathbf{fdm}_{o_2} - \mathbf{fdm}_{o_1}})}{N_{OFP}} \tag{12}$$

Weight The weight used for the current paper is the Zero Fuel Weight (ZFW), to avoid induced effects by extra fuel loaded on the aircraft. The ZFW is predicted using the aircraft weight, crew and utilities weight and passenger weight and denoted by $ZFW^i \in \mathbb{R}^{N_f}$. Mainly the passenger weight remains uncertain since no-show passengers may be there, or the weight of clothing or luggage of passengers varies. Upon departure, a load sheet indicates the actual number of passengers on board. The resulting change in ZFW is no direct mission profile parameter but affects other mission parameters.

$$ZFW = ZFW^{OFP} - ZFW^{FDM} \tag{13}$$

Rerouting Rerouting is defined as a significant change in lateral position from the intended flight path. The value for a significant deviation is up for interpretation and is set to 50nm for the current study. The minimum lateral deviation is computed per OFP waypoint, and from these waypoints, the maximum value is selected. Then, a flight can have a different route as a shortcut or a longer route. Only negative reroutings are of interest, which are flights with a lateral deviation of more than 50nm and a negative total distance difference.

$$lat_{dev} = max(min(d_{line}^{o,f})), f \in FDM, o \in OFP \tag{14}$$

Holding There is a clear definition for a holding pattern, but it is more complicated to identify holding patterns automatically. Often, pilots manually control the aircraft, leading to non-perfect circles. Thereby, holding patterns may be entered and left in multiple locations. Therefore, a holding pattern is identified as an aircraft being on a the same position as it had done previously, with a one-minute minimum difference between the two datapoints. For being on the same location, a margin of 0.5nm is used.

Missed approach When a pilot is not confident that is can make a safe landing, or decelerate quick enough before the aircraft reaches the end of the runway, the pilot may decide to take-off again and redo the landing procedure. This manoeuvre leads to a large increase of fuel consumption. A missed approach in computed when the aircraft reaches its final altitude plus 1000ft and is on the same location as the final destination with a 1 mile accuracy.

Diversions Diverted flights are flights landing at another airport than is prescribed in the OFP. Diversions have been excluded in literature, while they are critical for fuel loading analysis. As was the case for rerouting, there is a distinction between positive and negative diversions. Early descents have very low fuel burn since the entire flight time is drastically lowered. Therefore, only diversions with a negative fuel burn are considered in the current analysis, to avoid fuel loading decisions based on early stopped flights.

Diversions are special cases in fuel deviation analysis. As soon as a diversion takes place, alternate fuel, which was previously reserved, is now considered normally usable by the pilot. So, when modelling TF_{dev}, the deviation from the actual fuel consumption compared to the planned trip fuel and alternate fuel is considered if a flight is diverted. This way, the buffer that should be taken to avoid final reserve fuel is modelled.

Flight phase fuel consumption The flight is divided into flight phases, in order to assess the fuel consumption in the different flight phases. Per flight phase, the fuel used is computed by integrating the fuel flow over the duration of the flight phase. The actual fuel burn is compared to the planned fuel burn. Three phases are being defined, which are the climb, cruise and descent phase, visualized in Figure 5. The cruise phase starts when the first waypoint reaches the initial altitude and ends at the last waypoint above the initial altitude. Since the phases are calibrated to the flight plan, it may be that the actual flight is not in cruise altitude yet, while crossing the horizontal location of a waypoint. By separating the flight in different phases, critical elements may be identified causing fuel consumption differences per flight phase.

$$F_{phase} = OFP_{F_{phase}} - \sum_{f \in} FDR_{phase} f_{FF} * f_t \tag{15}$$

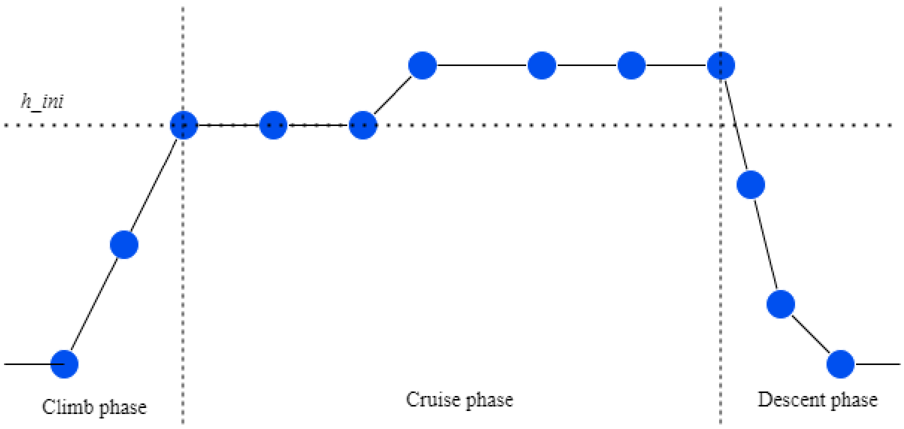


Figure 5: Different flight phases during a typical flight

3.3 Fuel loading strategy

The fuel loading predictions made in section 3.1 are used for a fuel loading strategy which is presented in the current section. The aim of the new fuel loading strategy is to reduce excess fuel loading weight, without compromising on flight safety. First, benefits and SPIs are discussed in section 3.3.1. Thereafter, a fuel loading strategy is explained in section 3.3.2.

3.3.1 Benefit assesment

The benefits in terms of less carried fuel are computed using the new fuel loading strategies. First, the reduced weight is computed by comparing the airline fuel loading to the new fuel loading strategy. Then, by using a Cost-to-Carry (CTC) factor, the extra consumed fuel is calculated for a given flight [Ryerson et al., 2015]. For each individual flight, the CTC factor is computed using Equation 16, where β_1 and β_2 are used from [Ryerson et al., 2015], with values of 0.02 and $4.852 * 10^{-5}$, respectively. The reduced fuel burn F_{red}, is computed using Equation 17.

$$\gamma_i = \frac{\beta_1}{dis_i} + \beta_2 \tag{16}$$

$$F_{red} = \gamma_i * (F_{load}^{2TS} - F_{load}^{FPS}) * dis_i \tag{17}$$

For an airline-wide benefit assessment, the fuel savings per flight are combined. Using reduced fuel usage, the reduced emissions and reduced monetary savings could be computed, which are both for airlines. The reduction in CO_2 emissions are computed using the value from the U.S. Environmental Protection Agency on jet fuel, which is 9.75kg per gallon of jet fuel [EPA, 2018]. Then, monetary savings are computed using a price of 3$/gallon for the jet fuel price [IATA, 2022].

Studies on fuel loading often use FRF usage as risk indicator [Zhu and Li, 2021, Achenbach, 2018, Kang, 2017], since FRF usage indicates emergency scenarios. To gain trust from airline dispatchers and pilots however, more metrics are required, such that models may be evaluated across different standards. EASA introduced SPI's for fuel loading, intended to be used to allow airlines to change their contingency fuel procedure if the standards on the SPI's can be adhered [EASA, 2022]. Not all SPI's are applicable for the current research. Those SPI's can either not be measured using the current dataset or would not change based on the changes in fuel loading in the current research. The list of used SPI's are:

- SPI_{cont}: Flights using more than 100% of the contingency fuel.
- SPI_{FRF}: Landings with less than the FRF remaining, which is equal to 30 flight minutes.
- $SPI_{45,60}$: Landings with less than a fuel equivalent of a specific number of minutes remaining, which are set to 45 min and 60 min for the current study.

These SPIs are based on the fuel remaining in the aircraft. However, pilots and dispatcher currently load discretionary AF. To allow a fair comparison, SPIs based on AF usage should be definted as well. Thereby, pilots should monitor the remaining fuel during the flight, and act if the landing is estimated to use alternate fuel, which is a risk on fuel shortage. Therefore, SPIs regarding alternate fuel usage are created. As there is a distinction between mandatory AF and discretionary AF, two SPI's are made:

- SPI_{alt_m}: Flights landing with less than FRF and mandatory AF combined
- SPI_{alt_d}: Flights landing with less than FRF and discretionary AF combined

3.3.2 Fuel Loading Strategy

Two types of fuel loading strategies are tested, with multiple variations. To allow for a relatively simple and comparable model, fixed flight time buffer strategies are created. The fuel buffer is defined in flight time, to comply with the FRF definition. One minute of flight time is computed as $\frac{1}{30}$ of the FRF, which is equal to a flight time of 30 minutes. The first loading strategy is designed such that no discretionary alternate fuel is used, while the second uses the same amount of discretionary fuel as is currently used by the case study airline. This way, the benefits of a new prediction model and benefits of limiting the amount of alternate fuel can be evaluated separately. A fixed buffer is loaded for each flight, to deal with large trip fuel consumption deviations, not predicted by the prediction model. The fuel loading strategies are defined in Equation 18 and Equation 19. TF is based on the newly created trip fuel predictions. FB, which represents the fixed buffer, is set, such that on the validation set, the value of SPI_{45} is the same as is the case for the airline fuel loading practice. SPI_{45} is used instead of SPI_{FRF}, as low number of flights use FRF in the dataset. FBA, which represents the fixed buffer for flights with discretionary AF, is set such that the same value of SPI_{alt} is found. For interpretable results, the fixed buffer may only be equal to multiplications of 5 minutes. SPI_{cont} is computed as if the same contingency fuel as the airline is currently loading would be loaded using the new strategies. A loading strategy receives a name with the model used, the FB and the FBA. For example, $2TS_{15}$ refers to a strategy using 2TS predictions and a value of 15 minutes for FB and $FPS_0 + alt{-}10$ refers to a strategy using the FPS predictions, with no FB, all discretionary fuel being loaded and a FBA of -10 minutes.

$$\text{Fixed buffer strategy: } LF = TF + AF_m + FRF + FB \tag{18}$$

$$\text{Fixed buffer and alternate fuel strategy: } LF = TF + AF + FRF + FB + FBA \tag{19}$$

4 Case study description

Data is collected from a partner airline to test the proposed model. The airline is a medium-sized European airline, operating domestic and international flights within Europe. Data is collected from the 4^{th} of March 2018 until the 31^{st} of December 2019, such that the Covid pandemic does not affect the results. Flight data prior to the 4^{th} of March is scarce and therefore excluded. The dataset is described in section 4.1. An exploratory analysis on the importance of mission parameters is present in section 4.2

4.1 Data description

Data is obtained from both the OFP data per flight and the FDR data per flight. For each flight, the most recent OFP is used. This leads to varying planning times before departure in reality. For simplicity, a fixed planning time of two hours is used, needed for the connection of TAFs and required prediction horizons. By merging the two datasets, a total of 75,315 available flights are found. When the flight leg, arrival airport or departure airport has less then 50 flights, the respective input parameter is set to 'other'. This way, all groups of spatial levels are represented in the training, validation and test set. Flights belonging to a group 'other' are often irregular flights after for example a diversion to an airport not used in normal operations.

TAFs are typically issued every 6 hours by airports, which occasionally updates in between these time steps. All forecasts are collected for the used arrival airports from *navlost.eu*, [s.r.o., 2022]. For smaller airports, there is missing data in the TAFs. This causes 11,650 flights to not have TAF forecast younger than 18 hours prior to departure. For these flights, no information on wind or percipitation is present, but information on the horizontal and vertical visibility is present, as it is present in the OFP.

The data is split chronologically in a training, validation and test set, dividing all months in the dataset into one of the sets. The division is shown in Figure 6. Months need to be considered complete, as this is the largest time interval in the FTTS model. As a result, the training, validation and test set sizes are unequal. The dataset is split chronologically, as look-ahead biases would be present otherwise, with unrealistic model performance as result. The sizes of the training, validation and test set were aimed to be 60/20/20. The dataset sizes are shown in Table 5.

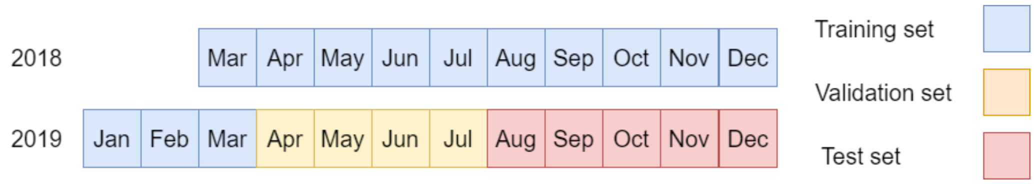


Figure 6: The training, validation, test and casestudy sets

Table 5: Descriptive parameters for the training, validation and test set

Set	**Size**	**Mean TF_{dev} [kg]**	**Std. TF_{dev}**	**Major outliers [%]**	**Medium outliers [%]**	**Minor outliers [%]**
Training	39706	-32	234	0.12	1.14	5.4
Validation	15903	10	222	0.05	0.6	3.2
Test	17846	-68	222	0.13	1.26	6.49

Outlier sets are defined at the 5%, 1% and 0.1% percentile of the dataset, defined as minor, medium and major outliers. This way, the performance of prediction models for different types of outliers may be evaluated. For reference, in the current dataset, the boundaries of the outlier sets are approximately -400, -600 and -1200 kgs of TF_{dev}. A summary of the different datasets is present in Table 5. The percentage of outliers changes per dataset. A changing average TF_{dev} is the reason for this, as the standard deviation is relatively similar.

Dependent on the required inputs of the proposed model, data from prior sets can be used as input data for preceding sets. This is the case for the validation and test set. In these sets, past inputs can be treated as inputs for time series, even though they are not present in the set themselves. This way, data loss is avoided and consistent input data for each prediction is used. When there is no prior data within the training set, the datapoint is not considered in the model. Depending on the look-back length, the training dataset size will change. Otherwise, inconsistent input data is used which may lead to inaccurate training. However, since the training data set size changes, the results will also change, depending on the look-back length of the model. For the fixed interval time series, the month of February 2018 is also considered as input, which is filled with the average values of the respective group.

4.2 Mission parameter analysis

Exploratory results on the historical mission profile parameters are described in the current section. The mean and standard deviation of each of the mission parameters is presented in Table 6. The mean values and standard deviations for each parameter are presented in both the applicable unit as percentages. Percentages are based on the planned values of the respective parameter. The wind difference is compared to the planned airspeed as a percentage. Thereby, the R-squared (R^2) and the is presented, which is the measure of the explained variance of the dependent variable, the TF_{dev}, by the independent variable. Finally, a linear regression is created for each special flight indicator as independent variable and TF_{dev} as dependent variable. The regression coefficient is presented.

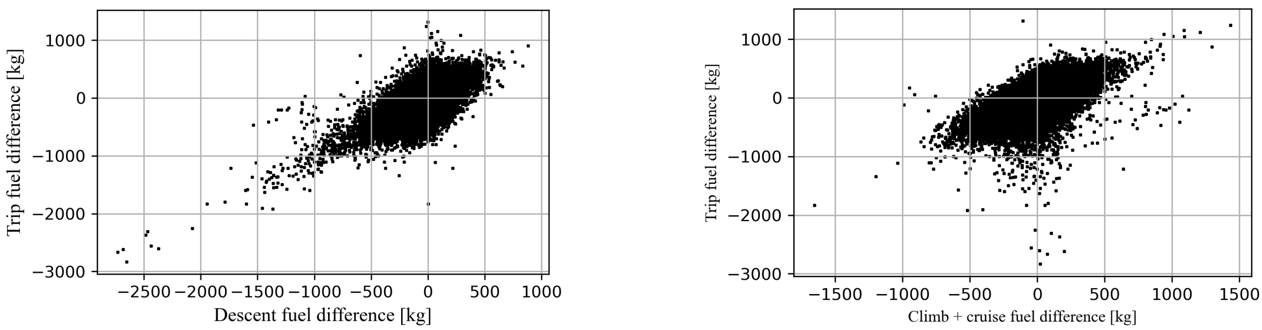
First of all, some parameters have a mean other than zero, which is the case for the distance, time, fuel flow and weight difference for a flight. To have a consistent pairing between the planned and actual values, a mean close to zero is expected. This could be explained by the fact that usually, a flight route is selected that is allowed by ATC certainly, while often shortcuts are present in practice. Thereby, the ZFW may change because of numerous reasons, with an important reason being the number of no-show passengers, inducing large variations in ZFW. Interestingly, the fuel flow is underestimated, probably caused by the ZFW differences. Although the current predictions result in an average error of zero for fuel consumption, the flight plan is not correct for determining other variables, meaning that calculations in the OFP are miss-aligned. So, to get more accurate calculations of the trip fuel, improving accuracy of other variables will help. Secondly, the standard deviation of most descriptive parameters is around or below 5% for parameters describing the entire flight, while a larger spread is observed for specific phases of the flight, with the descent phase with the highest standard deviation. Even though the cruise phase typically has a higher total fuel burn, the standard deviation of the descent phase is higher, indicating a higher uncertainty.

To improve prediction of TF_{dev}, it is important to improve the accuracy of mission parameters with a high R^2. The highly explained variance of the fuel by the distance difference and the descent fuel difference shows that

Table 6: Means and standard deviations of mission parameters, expressed in the relevant unit and in percentage compared to the flight plan.

Parameter	**Mean**	**Std**	**Mean [%]**	**Std [%]**	R^2	**Reg. coefficient [kg]**
Trip fuel difference [kg]	-19.7	215	0.108	4.45	-	-
Distance difference [nm]	35.8	25.9	4.50	4.51	0.32	-
Time difference [hrs]	0.106	0.089	5.01	5.12	0.24	-
Fuel flow difference [kg/hr]	-40.2	91.1	-1.77	3.84	0.12	-
Mach difference [-]	0.059	0.043	7.53	5.47	0.08	-
Altitude difference [100ft]	4.55	11.8	1.23	3.2	0.03	-
Wind difference [kts]	2.73	3.1	0.61	0.70	0.02	-
Weight difference [kg]	320	956	0.54	1.68	0.06	-
Climb fuel difference [kg]	-2.5	119.8	0.39	13.6	0.12	-
Cruise fuel difference [kg]	0.6	139.9	0.6	7.3	0.19	-
Descent fuel difference [kg]	-20.8	143	0.12	25.8	0.38	-
Holding [-]	0.013	-	1.3	-	0.05	-456
Approach [-]	0.005	-	0.5	-	0.03	-498
Reroute [-]	0.002	-	0.2	-	0.01	-360
Diversion [-]	0.0009	-	0.09	-	0.002	-368

these two parameters have the most effect on TF_{dev} of a flight. Wind and altitude differences in a flight have a large influence on the trip fuel consumption as was found in literature, but the deviation from the flight plan does not have a large effect on TF_{dev}. So, more exact planning of those variables may lead in more fuel efficient flight, but will lead to marginal improvements in predictions of TF_{dev}. Figure 7 shows the relationship of fuel deviations in flight phases versus the total fuel deviation. Two insights are found in these figures. First, minor and medium outliers in TF_{dev} are caused by both the cruise+climb phase and the descent phase. So, for an analysis of fuel loading, all phases need to be considered, each with its own important parameters. Secondly, it is visible that most major outliers are caused by descent fuel deviations, which implies that arrival delays cause the most extreme fuel deviations.



(a) Descent fuel

(b) Climb + cruise fuel

Figure 7: Scatter plots of the trip fuel deviation versus the deviation in flight phases for non-diverted flights

In Table 6, regression coefficients for special flights are shown. This value represents the average difference in TF_{dev} if the flight belongs to the special flight scenario group. For the diverted flights, the deviation is on top of the mandatory AF. The average values in TF_{dev} for a special case results in a minor outlier on TF_{dev}. So, it is not enough to predict each of the flight scenario's as a binary value, as the differences within these groups are large, and the cause of major outliers remains unclear.

Another interesting topic in fuel loading is discretionary AF. Regulations do not force dispatchers or pilots to load AF, but still, this is done in many cases. Even though chances may be little to divert, because of destination aerodrome terminal inefficiencies, it may happen. In the used dataset, 19% of the negatively diverted flights had no IFR and so, did not need AF. Thereby, the percentage of flights using discretionary AF that diverted, is seven times lower then the flights using mandatory AF. So, the discretionary AF loaded is considerably less effective than mandatory AF being loaded. It is impossible to predict all diverted flights. But, one can estimate the diverted flights as overburn cases. If the extra fuel would be enough for such an alternate flight, the diversion would be possible.

4.3 Hyperparameters

The hyperparameters for the 2TS model and GB model are presented in the current section. The explanation of each hyperparameter if explained by [Lim et al., 2021] and [Friedman, 2001], for the 2TS and GB model, respectively. As there are many hyperparameters, a grid search method is computationally expensive. Therefore, a baseline hyperparameter combination is set. Then, for each hyperparameter, a list of variations of that parameter, multiplied and divided by factors of two, is tested. If the accuracy of the validation set improves, this updated parameter value is selected. This way, all hyperparameters are updated from their baseline value. The resulting values are presented in Table 7 and Table 8.

Table 7: Hyperparameters for the 2TFT and STFT models

Hyperparameter	**IFTS value**	**FTTS value**
Encoder length	25	model dependent
Max epochs	50	50
Batch size	128	32
Learning rate	0.003	0.03
Hidden size	32	16
Hidden continuous size	16	8
Attention head size	1	1
LSTM layers	1	1
drop out	0.2	0.1

Table 8: Hyperparameters for the GB model

Hyperparameter	**GB Value**
Nr of estimators	500
Sub sample	0.8
Max features	$0.8 * n_{features}$
max depth	4

5 Computational results

The 2TS model proposed in section 3, is evaluated in the current section. 2TS is compared to the other baseline models in order to assess their performance. The baseline models are FTTS, IFTS, and GB, which are all discussed in section 3. Also, the FPS error is presented, to measure the increase in performance compared to current airline practice. A fuel loading strategy, based on the 2TS predictions is also assessed in section 5.2.

5.1 Prediction model results

To compare the different models, the RMSE, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the standard deviation of the error (Std) are computed. The results of the test are shown in Table 9 and the distribution of the prediction error is shown in Figure 8. The 2TS model outperforms all baseline models, which indicates a better prediction performance. The performance in increase is 1.2%, 1.9%, 5.4% and 26.4% compared to the FTFT, STFT, GB and FPS models, respectively. The improvement to the FPS model indicates that deviations in trip fuel consumption can be estimated, resulting in a lower error, and potentially help for fuel loading decisions. Compared to the GB model, which is a strong performing prediction model in other studies [Kang, 2017, Achenbach, 2018], the prediction performance increased marginally. Still, small improvements may have a large impact on fuel loading decisions. Although the prediction model using time series performs better than benchmarks, a significant amount of uncertainty remains present in TF_{pred}. Sensor errors of the trip fuel consumption and unpredictable events using the current inputs are causes for this. When aiming to reduce the RMSE of trip fuel, it is questionable how much further the error could potentially be reduced. Thereby, both the STFT and FTFT model have comparable errors, with a difference in RMSE of only 1%. So, the architectures of the models are different, while the benefits of creating a time series are almost the same. When combining the models, the RMSE decreases by 1.2-1.9% and the MAE by 1.8-2.7%. Using the combined model, the model is able to extract extra information, describing the behaviour of TF_{dev} for planned flights, to increase the prediction performance.

Table 9: Test set results for the proposed models and baseline models.

Model	**RMSE**	**MAE**	**MAPE**	**Std**
FPS	228.9	174.4	3.32	222.5
GB	177.5	132.7	2.64	177.0
STFT	170.5	125.8	2.51	170.1
FTFT	172.1	128.2	2.54	172.1
2TFT	168.0	123.5	2.49	167.7

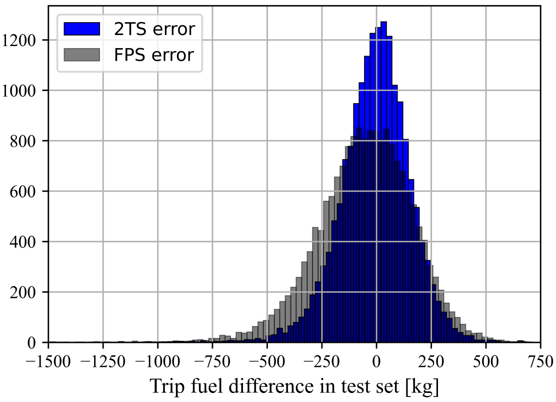


Figure 8: Predictions errors for the 2TS model and FPS

The maximum RMSE decrease in literature found using the GB model is 50% [Kang, 2017]. One reason for this difference is that the average TF_{dev} is close to zero in the current dataset, which is not the case for the other study. Comparing the standard deviations, the decrease is comparable, being 24-31% versus 20.4% in the current study. There are multiple possible explanations for this difference in prediction performance improvement. First, there are dataset differences, because of geography and average flight duration. Secondly, diverted flights and flights with tankering are included in the current dataset which is not the case for other studies. Thrid, a different division of the training, validation and test set is present, being chronologically ordered in the current study. The latter is more representative of reality but leads to a lower test set performance. So, by comparing the outcomes of models to the GB model, similar evaluations may be expected using other different datasets. Moreover, accuracy improvements of other studies may be deemed too optimistic, compared to current results because of the possible explanations of the differences.

The best performing model in literature is found to be stacking, with RMSE reductions of 2-5% compared to GB, [Kang, 2017]. The current improvements beat that range slightly. As deep learning and time series are introduced to the problem, an increase in performance was expected. Improving other model performances just slightly, the benefit of the time series introduction is unclear. To demonstrate the effectiveness of a time series model, a sensitivity analysis on the 2TS model is performed in Figure 9. The encoder length and inclusion of mission profile parameters in the model are tested. Up to an encoder length of 25, the model decreases the RMSE of the predictions. This indicates that incorporating past flights into a prediction model helps to predict the following flights. Thereby, when no mission parameters expect for the trip fuel would be used, the prediction performance decreases, for nearly all encoder lengths.

The variable importance of the 2TS model is shown in Figure 10, including the 10 most important decoder variables, the 5 most important encoder variables and the attention of the encoder, for an encoder length of 20. Among the most important variables of the 2TS model are predicted values of the FTTS model, which is expected as it already has a strong predictive performance of TF_{dev}. Interestingly, the cost index is among the top 10 important variables. This value is selected by the airline, to match the desired operations or aircraft performance. Apparently, the values consistently lead to deviations in trip fuel consumption. The exact relationship is unknown. The most important encoder variable is the fuel difference, which is expected, as it is the target variable of the model, and should have the highest correlation between consecutive values of the same parameter for flights with the same OD-pair. Furthermore, there is only a slight decrease in attention visible over the time indexes of preceding time steps for a prediction. So, there is a slight, but not strong decay in the correlation between TF_{dev} of preceding flights.

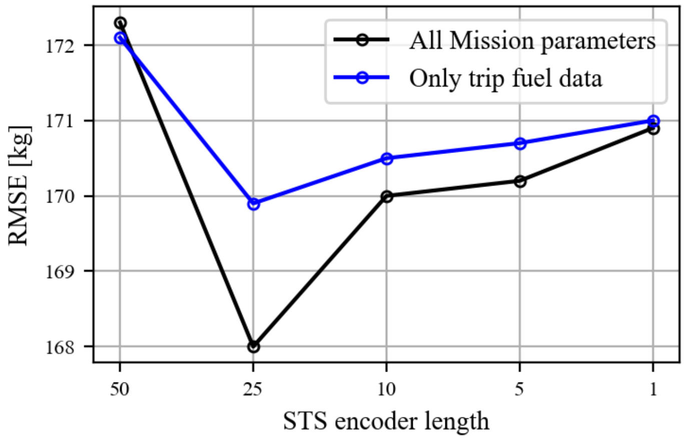


Figure 9: Sensitivity analysis on the encoder length of 2TS

5.2 Benefit assesment

The newly created predictions are tested using the fuel loading strategy described in section 3.3.2, in order to determine the benefits of using the 2TS model. Complete adherence to the fuel loading strategy is assumed, without any extra fuel loaded by pilots or dispatchers, to demonstrate the effectiveness of the model. Table 10 shows the resulting average fuel weight reduction using the new strategies and each of the SPIs, where the SPIs are measured in the percentage of flights change compared to the airline practice. Using the validation set, a 35-minute buffer is found for the 2TS model that results in the same SPI_{45} value as the airline has. Thereby, an economical strategy loading 10 minutes less is also presented. The same strategies are tested using the GB predictions and using the FPS predictions, in order to assess the benefits of improving prediction models. For the second strategy type, with the same discretionary AF loading, a 30-minute fixed buffer is found, reducing the buffer with 15 minutes for the flights with discretionary AF.

Depending on the desired SPIs, a fuel weight reduction of 1.05-5.99% may be achieved. The first result is that using the 2TS model, less fuel is loaded than using the GB model using the same time buffer, indicating a different average prediction value for TF_{dev}. The SPIs may be compared between the models. Using a conservative fixed buffer fuel loading strategy, the same or lower values for SPI_{FRF}, SPI_{45} and SPI_{60}, SPI_{alt_m} is found compared to the practice of the airline, which indicates less risk of fuel depletion. However the value of SPI_{alt_d} is worse for the fixed buffer strategy. Less flights have the option to divert to another airport this way. So, depending on the desires of flight dispatchers and pilots, one procedure may be experienced better than the other.

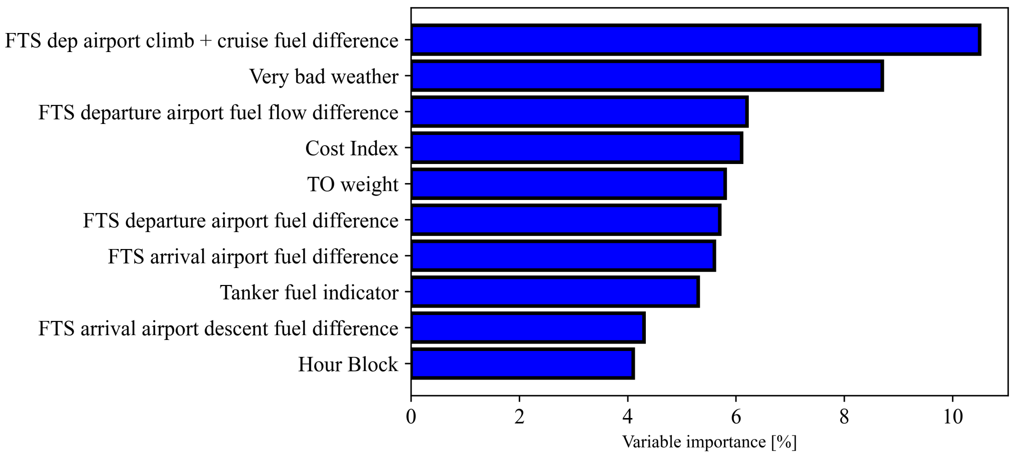
When a more aggressive fuel loading strategy is used, the values of SPI_{45} and SPI_{60} deteriorates, while the value of SPI_{FRF} remains similar. The percentage of less fuel being loaded increases to 6%, which might be worth the little increased risk on FRF usage.

A similar fuel loading strategy is presented for the GB model predictions and the FPS predictions. A result is that the combination of SPIs is worse for the other prediction models

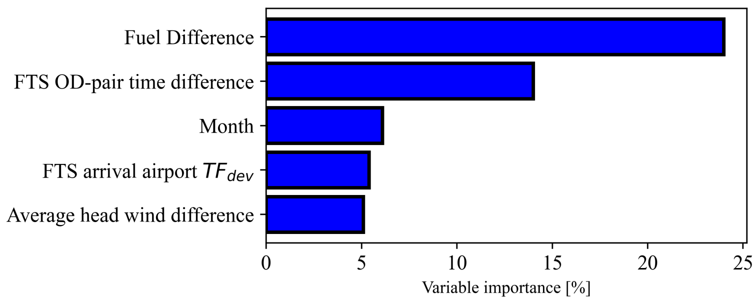
Using a fixed buffer and alternate fuel strategy, loading the same amount of discretionary AF, all SPIs are equal or better than current practice. A downside is that this strategy results in only 1% of fuel being loaded less. So, if airlines would follow one of the loading strategies, a fixed buffer strategy is considered the best solution, combining weight saving and remaining fuel. Another result is that for all the new models, SPI_{cont} is reduced the most. So, using the new predictions, higher trust of pilots and flight dispatchers in models may result. Furthermore, all strategies reduce the value for SPI_{alt_m}, being weather impacted flights that land using alternate fuel. Apparently, the fixed buffer on top of mandatory AF results in a more conservative fuel loading strategy. By reducing the fuel buffer for weather impacted flights, less fuel may be loaded potentially, if the values of the other SPIs would not change too much.

Figure 8 shows the distributions of the remaining fuel for the new two strategies and the airline practice. A wider spread for the airline practice is found, indicating less accurate fuel loading decisions. For the new strategies, two peaks are found, one for the regular flights and the other for weather-impacted flights, with mandatory AF loaded. The distribution of the fixed buffer strategy has a high peak, with remaining fuel for the vast majority of the flights.

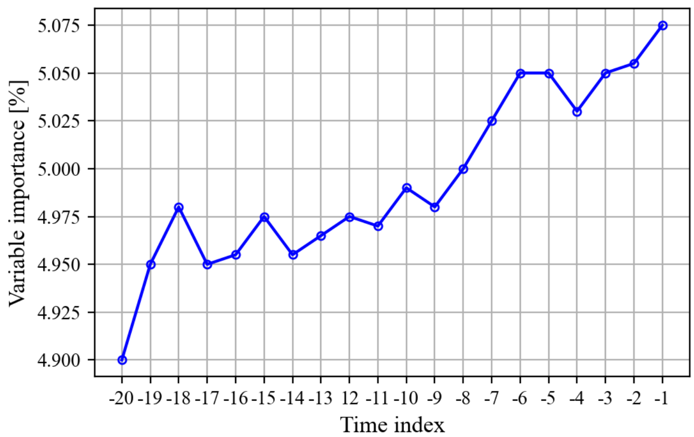
Then, the aggregated savings are presented in Table 11. Depending on the desired SPIs, costs may be reduced by 0.24 - 1.47 million \$ and CO_2 emissions by 747 - 4,792 tonnes. Using the fixed buffer and alternate strategy, fuel



(a) Decoder variables



(b) Encoder variables



(c) Encoder Attention

Figure 10: Variable importance of three different components of the 2TS model

Table 10: SPI analysis of different loading strategies. Values are presented in [%] fuel or [%] of flights. For confidentiality purposes of the case study airline, all values are presented as difference compared to the airline. A negative value indicates an improvement, whereas a positive value indicates deterioration.

Strategy	**Fuel Loaded**	**SPI**$_{FRF}$	**SPI**$_{45}$	**SPI**$_{60}$	**SPI**$_{alt_m}$	**SPI**$_{alt_d}$	**SPI**$_{cont}$
$2TS_{35}$	-2.01	0	-0.006	-0.61	-0.151	5.48	-9.90
$2TS_{25}$	-5.99	0	0.025	0.69	-0.091	25.94	-9.90
$2TS_{30} + alt_{-15}$	-1.05	0	0	-0.33	-0.126	-0.523	-9.90
GB_{35}	-1.89	0	0.006	0.11	-0.162	4.30	-9.43
GB_{25}	-5.54	0.006	0.044	0.72	-0.71	24.94	-9.43
FPS_{35}	- 2.25	0	0.025	0.194	-0.124	8.09	0
FPS_{25}	-6.23	0.025	0.065	4.61	-0.016	29.83	0

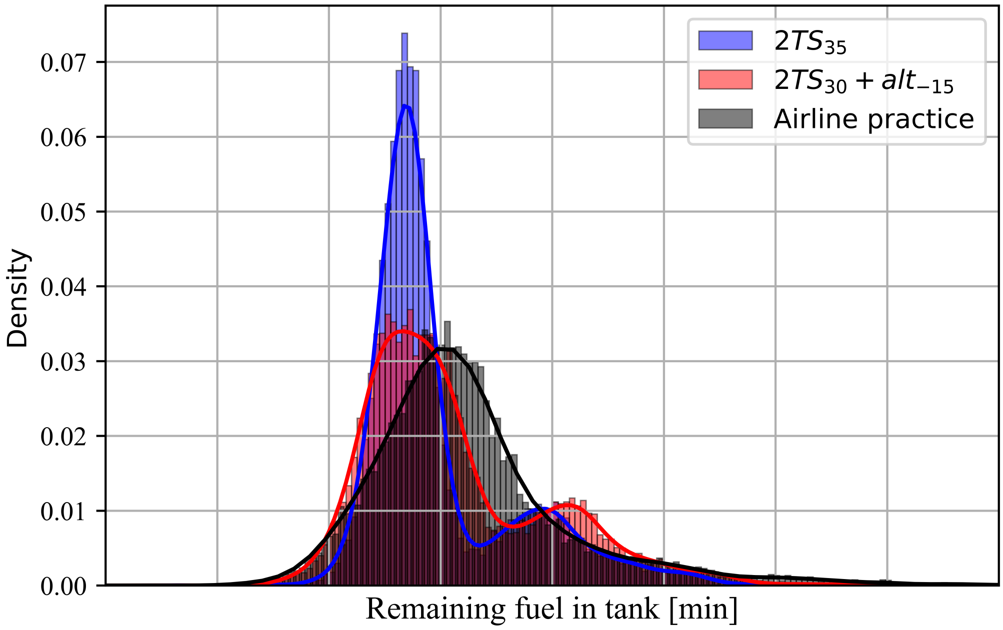


Figure 11: Minutes of fuel remaining upon landing for different loading strategies. The exact values are not presented, for confidentiality purposes

savings are less compared to the fixed buffer strategy. A fixed buffer strategy is recommended, as it works well for the most critical SPIs. Using a fixed buffer strategy, fuel savings of 0.12-0.39% are found, resulting in 0.48 - 1.47 million $ cost reduction and 747 - 4,792 tonnes CO_2 emissions reduction. There is no clear goal for the SPIs on each flight. Each pilot may decide on the amount to load extra individually, making exact determination of benefits unclear. Despite this, SPI_{cont} will always be improved using new predictions, potentially resulting in more trust in TF predictions. The fuel savings of 0.38% is lower than comparable studies found, with fuel savings up to 2%. The most probable reason is that the fuel loading practice is less conservative for the current case study airline compared to other studies. Hub and spoke airline are inclined relatively more to take less risk on diversions of flights, leading to a more conservative fuel loading strategy. So, trip fuel predictions have the highest potential on fuel consumption reductions for hub and spoke airlines. Thereby, the accuracy of the trip fuel predictions may be too optimistic in other studies.

Table 11: Airline wide savings

Strategy	**Fuel Savings [%]**	**Fuel savings [tonne]**	**Cost reduction [million $]**	**CO_2 reduction [tonne]**
$2TS_{35}$	0.117	481	0.466	1,515
$2TS_{25}$	0.385	1,525	1.474	4,792
$2TS_{30} + alt_{-15}$	0.058	238	0.230	747

6 Conclusions & Recommendations

In this paper, a model for improved trip fuel deviation predictions is presented, to assist dispatchers on fuel loading decisions. A 2 layered Time Series (2TS) prediction model is proposed, being able to capture critical temporal patterns present in trip fuel deviations. Using a fixed time interval layer, the model is able to identify temporal patterns and changes in them, for different groups to which a flight belongs. The individual flight layer is able to connect specific flight mission parameter data from prior flights to succeeding flights. Compared to the FPS trip fuel predictions, the RMSE of 2TS predictions is decreased by 26 %, and compared to the baseline GB model, the RMSE is decreased by 5.4%. The addition of mission parameters of flights in the model have shown to reduce the RMSE of predictions by 1.2% compared to a model without the mission parameters.

Using the newly created predictions for fuel loading decisions, fuel savings of 0.12 - 0.39% could be achieved using a fixed buffer loading strategy. Depending on the SPIs, airlines may decide on the most applicable strategy. Using the new predictions, a considerable amount of flight use less than 100% of the contingency fuel, while they do without the new predictions. So, when changing only the predictions and not the fuel loading strategy, the predictions could be advantageous for airlines.

The current paper only focused on fuel burn parameters of the case study of a single airline. To advance fuel loading decisions, a dataset of a larger scale is recommended. As a very low number of flights use Final Reserve Fuel, an even larger dataset is required to obtain a statistically significant dataset size, to analyse major deviations. As fuel consumption is sensitive data, flight times could be used for such a study. Descent trip fuel deviations are found to be the most important for major fuel deviations, so improving predictions on this would be beneficial for fuel loading decisions.

Thereby, the desired values for SPIs dictate the fuel consumption that can be saved for flights. There are no targets present for each of the SPIs. So, interviews with pilots and dispatchers, but also with regulators should be conducted to determine the desired targets for SPIs, along with is lower and upper bound. What is clear, however, is that using a similar loading strategy, better values for SPIs can be achieved.

Besides predictions for fuel loading decisions, accurate trip fuel predictions have the potential to assist with more problems in aviation. For example tanker fuel, flight trajectory, or tactical aircraft assignment decisions may be improved with updated trip fuel estimations. A benefit analysis on each of those problems versus the accuracy of trip fuel consumption is recommended, to potentially improve solutions to those problems.

One of the advantages of time series models is the ability to update predictions during irregular events, with the Covid-19 pandemic being a recent example. Robustness to changes in the system network could be studied to assess the effectiveness of time series model during irregular time periods.

References

[Sen, 2009] (2009). Estimation of aircraft fuel consumption. *Journal of Aircraft*, 46(4):1089–1093.

[EPA, 2018] (2018). Emission factors for greenhouse gas inventories.

[Achenbach, 2018] Achenbach, A. (2018). *Predictive analytics in airline operations : application of machine learning for arrival time and fuel consumption prediction.* WHU / Chair of Logistics Management.

[Alligier and Gianazza, 2018] Alligier, R. and Gianazza, D. (2018). Learning aircraft operational factors to improve aircraft climb prediction: A large scale multi-airport study. *Transportation research part C: emerging technologies*, 96:72–95.

[Baklacioglu, 2016] Baklacioglu, T. (2016). Modeling the fuel flow-rate of transport aircraft during flight phases using genetic algorithm ptimized neural networks. *Aerospace Science and Technology*, 49:52–62.

[Baumann and Klingauf, 2020] Baumann, S. and Klingauf, U. (2020). Modeling of aircraft fuel consumption using machine learning algorithms. *CEAS Aeronautical Journal*, 11:277–287.

[Breiman, 1996] Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

[De Gooijer and Hyndman, 2006] De Gooijer, J. G. and Hyndman, R. J. (2006). 25 years of time series forecasting. *International journal of forecasting*, 22(3):443–473.

[Drees et al., 2017] Drees, L., Mueller, M., Schmidt-Moll, C., Gontar, P., Zwirglmaier, K., Wang, C., Bengler, K., Holzapfel, F., and Straub, D. (2017). Risk analysis of the easa minimum fuel requirements considering the acare-defined safety target. *Journal of Air Transport Management*, 65:1–10.

[EASA, 2022] EASA (2022). Amc and gm to annex iv (part-cat) to commission regulation (eu) no 965/2012 - issue 2, amendment 20. *European Aviation Safety Agency.*

[Friedman, 2001] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

[Gallego et al., 2019] Gallego, C. E. V., Comendador, V. F. G., Carmona, M. A. A., Valdés, R. M. A., Nieto, F. J. S., and Martínez, M. G. (2019). A machine learning approach to air traffic interdependency modelling and its application to trajectory prediction. *Transportation Research Part C: Emerging Technologies*, 107:356–386.

[Gomes et al., 2020] Gomes, D., Oliveira, F., Centeno, J., Figueiredo, L., and Halawi, L. (2020). Issue 3 article 9 2020 part of the air and space law commons, legislation commons, and the other law commons scholarly commons citation scholarly commons citation youssef. *International Journal of Aviation, Aeronautics, and Aerospace*, 7.

[Hao et al., 2016] Hao, L., Hansen, M., and Ryerson, M. S. (2016). Fueling for contingencies: The hidden cost of unpredictability in the air transportation system. *Transportation Research Part D: Transport and Environment*, 44:199–210.

[Hochreiter and Schmidhuber, 1997] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

[IATA, 2022] IATA (2022). Jet fuel price monitor.

[ICAO, 2018] ICAO (2018). Annex 6, part i international commercial air transport aeroplanes, eleventh edition.

[Irrgang, 2011] Irrgang, M. (2011). A look at the state of airline fuel conservation.

[Kang, 2017] Kang, L. (2017). *Changing Fuel Loading Behavior to Improve Airline Fuel Efficiency*. University of California, Berkeley.

[Kang and Hansen, 2021] Kang, L. and Hansen, M. (2021). Quantile regressionbased estimation of dynamic statistical contingency fuel. *Transportation Science*, 55:257–273.

[Kang et al., 2018] Kang, L., Hansen, M., and Ryerson, M. S. (2018). Evaluating predictability based on gate-in fuel prediction and cost-to-carry estimation. *Journal of Air Transport Management*, 67:146–152.

[Khan et al., 2019] Khan, W. A., Chung, S.-H., Ma, H.-L., Liu, S. Q., and Chan, C. Y. (2019). A novel self-organizing constructive neural network for estimating aircraft trip fuel consumption. *Transportation Research Part E: Logistics and Transportation Review*, 132:72–96.

[Khan et al., 2021] Khan, W. A., Ma, H. L., Ouyang, X., and Mo, D. Y. (2021). Prediction of aircraft trajectory and the associated fuel consumption using covariance bidirectional extreme learning machines. *Transportation Research Part E: Logistics and Transportation Review*, 145:102189.

[Li et al., 2023] Li, D., Tan, Y., Zhang, Y., Miao, S., and He, S. (2023). Probabilistic forecasting method for mid-term hourly load time series based on an improved temporal fusion transformer model. *International Journal of Electrical Power Energy Systems*, 146:108743.

[Lim et al., 2021] Lim, B., Arık, S. Ö., Loeff, N., and Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764.

[Luke L. Jensen and Reynolds, 2023] Luke L. Jensen, R. John Hansman, J. V. and Reynolds, T. G. (2023). Commercial airline speed optimization strategies for reduced cruise fuel consumption. *AIAA AVIATION Forum*.

[Ryerson et al., 2015] Ryerson, M. S., Hansen, M., Hao, L., and Seelhorst, M. (2015). Landing on empty: Estimating the benefits from reducing fuel uplift in us civil aviation. *Environmental Research Letters*, 10.

[Soler et al., 2012] Soler, M., Olivares, A., Staffetti, E., and Zapata, D. (2012). Framework for aircraft trajectory planning toward an efficient air traffic management. *Journal of Aircraft*, 49(1):341–348.

[s.r.o., 2022] s.r.o., A. (2022). Navlost metar/taf archive. `https://www.navlost.eu/metar/request/`. Accessed: 2022-10-11T13:37:48.000Z.

[Statista, 2021] Statista (2021). Fuel costs of airlines worldwide from 2011 to 2022, as percentage of expenditure.

[Sternberg et al., 2017] Sternberg, A., Soares, J., Carvalho, D., and Ogasawara, E. (2017). A review on flight delay prediction. *Transport Reviews*, 41:499–528.

[Trani et al., 2004] Trani, A. A., Wing-Ho, F. C., Schilling, G., Baik, H., and Seshadri, A. (2004). A neural network model to estimate aircraft fuel consumption.

[Trujillo, 1996] Trujillo, A. C. (1996). *Uncertainties that flight crews and dispatchers must consider when calculating the fuel needed for a flight*. National Aeronautics and Space Administration, Langley Research Center.

[Uzun et al., 2021] Uzun, M., Demirezen, M. U., and Inalhan, G. (2021). Physics guided deep learning for data-driven aircraft fuel consumption modeling. *Aerospace 2021, Vol. 8, Page 44*, 8:44.

[Vazquez et al., 2017] Vazquez, R., Rivas, D., and Franco, A. (2017). Stochastic analysis of fuel consumption in aircraft cruise subject to along-track wind uncertainty. *Aerospace Science and Technology*, 66:304–314.

[Wu et al., 2022] Wu, B., Wang, L., and Zeng, Y.-R. (2022). Interpretable wind speed prediction with multivariate time series and temporal fusion transformers. *Energy*, 252:123990.

[Zhang et al., 2020] Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J. (2020). *Dive into Deep Learning.* `https://d2l.ai`.

[Zhang et al., 2022] Zhang, H., Zou, Y., Yang, X., and Yang, H. (2022). A temporal fusion transformer for short-term freeway traffic speed multistep prediction. *Neurocomputing*, 500:329–340.

[Zhu and Li, 2021] Zhu, X. and Li, L. (2021). Flight time prediction for fuel loading decisions with a deep learning approach. *Transportation Research Part C: Emerging Technologies*, 128:103179.

II

Literature Study
previously graded under AE4020

1
Introduction

Decreasing fuel consumption is an unifying goal in the aviation industry. Over the past decade, the cost of fuel accounts approximately for 25% of the total costs of an airline, which makes it an important contributor to the profitability of airlines [77]. This effect has been verified in research, where an 1% increase in fuel price, lead to 40% increase in operating costs in 2012 in the US [74]. The profit margins in the airline industry are razor thin due to the strong competitiveness, which makes it that fuel efficient airlines have a significant advantage over others. Besides, the increase in concern of emitting green house gasses results in fuel burn reduction being a high priority within aviation. Although fuel burn per passenger has dropped by 24% between 2004 and 2017 in Europe, global aviation still accounts for 3.8% of global CO_2 emissions [31]. To achieve the European Union goal of carbon neutrality by 2050, fuel burn reduction by aviation will be a key component. Currently, aviation fuel is tax free in many countries. If a carbon fee was added to airline ticket prices in order to compensate the environmental costs, air travel demand would decrease by 2.4% to 21% depending on the carbon fee price [66]. Because of the large impacts of fuel consumption, many old and current studies are focused on an aspect of reducing fuel consumption. From an Air Traffic Management perspective, controllers try to minimize delays by maximizing the airspace capacity and opt for more freedom in trajectory selection of aircraft in order to reduce fuel burn [18]. Airlines renew their fleet to more modern and fuel efficient aircraft, investigate the use of biofuels [47], aim for other fuel saving strategies like trajectory and taxi optimization, [53], [90], and reduce unnecessary weight of utilities on board [11].

Despite the well-known fact that reducing weight reduces fuel consumption, little discussion is present on reducing the amount of fuel on aircraft. The increase in weight of the aircraft causes an higher thrust required to fly at the same speed. Thereby, an increased weight also increases the maintenance costs of engines and landing gear [3]. Recent research found that 2.21% - 4.48% of aircraft fuel is burned due to carrying unused fuel and 0.70% - 1.04% comes from loading fuel beyond what is needed to ensure a safe flight [73]. Airlines choose to load extra fuel in order to reduce the risk of diversion to another airport and to avoid fuel depletion. In terms of direct operating costs, a diversion to another airport costs roughly 22,500 euro, which is excluding the costs of passenger dissatisfaction and network delay costs [13]. Especially for flag carriers, operating according to a hub and spoke network, delays may result in significant operating costs [8]. Besides the economical effects, diversion to another airport also affects the safety of the passengers. Since the fuel reserves are low in such a scenario, pilot decisions are key. Wrong or too late decisions may result in fuel depletion [28]. Before departure, the pilot in command is responsible for the amount of discretionary fuel loaded on board. So even if the Flight Planning System (FPS) presents an objective discretionary fuel advice, pilots tend to load extra fuel relatively subjectively. Research has shown that pilots tend to mistrust the fuel predictions of the FPS and increase the amount of fuel depending on past experience, weather and traffic forecasts [84].

In order to achieve the goal of reducing fuel on board, several approaches are found in literature. First of all, the airline policy on fuel loading could be analysed ([82], [48], [84]). For this, studies investigate pilot behaviour on fuel loading decisions and recommend improvements to limit the loading of excess discretionary fuel. Changing the regulations of local aviation authorities could also lead to a reduction of contingency fuel as is investigated for Brazil [36] and for China [80]. From an airline perspective, better fuel burn predictions over flights could lead to lower discretionary fuel loading, while ensuring the same level of reserve fuel usage, which is first investigated by Ryerson (2015) [73]. Robust optimization on fuel loading according to the fuel burn predictions could also lead to better fuel usage [13]. All approaches are related since the factors combined result in the actual fuel loading. Different policies could only be adapted when better predictions are present to ensure a safe flight. And better predictions are not useful if regulations require to load more fuel anyway. Due to the quantitative nature of the fuel burn prediction and fuel loading optimization, only these approaches are considered applicable for the current research. Furthermore, the current research will focus on European flights. Differences with other regions will be mentioned in chapter 2.

The report is organized as follows. First, the regulations on fuel loading, the current practice by airlines and existing research will be discussed in chapter 2. In this chapter, research gaps will be identified. Supervised learning methods useful for fuel burn prediction will be presented in chapter 3. Statistical models, machine learning models and neural networks will be described. Feature extraction methods and its application on the current research will be presented in chapter 4. Thereafter, fuel loading strategies and cost to carry analyses will be mentioned in chapter 5. The report will be concluded in chapter 6. Here, the research proposal will be stated.

2 Fuel loading operations

Existing information on fuel loading will be discussed in the current chapter. From this information, research gaps can be identified. Regulations on fuel loading are described in section 2.1. The operations for airlines regarding fuel loading are assessed in section 2.2. Thereafter, a literature study on existing research on fuel consumption modelling is performed in section 2.3. Uncertainties in fuel consumption and possible input data will be shown in section 2.4 and section 2.5. section 2.6 will conclude the chapter, by identifying research gaps and further goals for the literature study.

2.1. Regulations

The aviation sector is heavily regulated to ensure a safe and efficient flights. The International Civil Aviation Organisation (ICAO) defines guidelines for National Aviation Authorities (NAA) for aviation regulations in annexes. NAA's like the European Union Aviation Safety Agency (EASA) often copy the guidelines with minor deviations. Fuel regulations are introduced to decrease the chance of fuel depletion. Flight dispatchers and pilots need to live up to these regulations. The regulations present a minimum required amount of fuel to bring, which dispatchers and pilots can only enlarge before departure. The guidelines presented by ICAO will be described first, and the applicable regulations for this research of the EASA will be described afterwards.

2.1.1. ICAO

The ICAO and the International Air Transport Association (IATA) were the first organizations to provide guidelines on regulations for fuel loading. As stated by the ICAO, the regulations are traceable to the 1950's, when meteorological forecasts, aircraft fuel consumption predictions and Air Traffic Management (ATM) was worse then today [44]. Thereby, the criteria were outdated which lead to the fact that described provisions were insufficient for the modern flight planning and efficiency maximization. This is why operators often carried excess fuel compared to the regulations.

Advances in the flight planning and flight management systems (FMS) causes increased accuracy and predictability in fuel planning. These systems could predict fuel and contingency fuel using statistics on previous flights. Also alternate airport selection methodologies have improved over time. Moreover, improved ATM system reduce numerous safety risks, introducing opportunities for operational efficiency [44]. If NAA's opt for prescriptive regulations, operational flexibility is limited which doesn't allow for full advantage of modern tools. Performance based regulations however, enables operators to optimize their flight planning capabilities. Although efficiency could be increased and environmental effect could be decreased over the years, safety should be considered to central theme as is described by Figure 2.1.

The current regulations are stated in Annex 6 - Operation of Aircraft, Part 1 - International Commercial Air Transport Aeroplanes, 11^{th} edition, which, among others, gives guidelines on flight preparation and fuel loading. Paragraph 4.3.6 describes the fuel loading regulations and starts with the following requirement:

> An aeroplane shall carry a sufficient amount of usable fuel to complete the planned flight safely and to allow for deviations from the planned operation

Figure 2.1: The relationship between safety, efficiency and the environment as stated by ICAO, from [44]

This should be determined based on aircraft specific data on fuel consumption, aircraft mass, Notices to Airmen (NOTAM), current meteorological conditions and forecasts, ATM procedures and restrictions and effects on deferred maintenance items.

Thereafter, requirements are set on the pre-flight calculation of usable fuel required, which is build from several types of fuel. These are shown in Figure 2.2 and listed below:

- **Taxi fuel** The amount of fuel expected to be consumed before take-off, including auxiliary power unit (APU) fuel consumption.
- **Trip fuel** The fuel needed to fly from take-off to landing at intended destination airport, taking into account operating conditions.
- **Contingency fuel** The amount of fuel required to compensate for unforeseen factors. It is required to be 5% of the planned trip fuel, but not lower then the amount required to be able to fly for 5 minutes at holding speed at 1500ft at the destination airport.
- **Destination alternate fuel** The fuel required to be able to fly to an alternate airport. This includes fuel for a missed approach and the flight to the alternate airport including climb and descend to the alternate airport. In case there is no alternate airport, fuel equal to the amount to be able to fly for 15 minutes at holding speed at 1500 ft altitude is required. If two alternate airports are required, the fuel should be enough to fly to the airport which will cost the most fuel. If the destination airport is isolated, the fuel should be enough to fly for 2 hours at nominal cruise above the destination airport, including final reserve fuel.
- **Final reserve fuel** The amount of fuel to be used in emergency scenario's. The fuel shall be enough to be able to fly for 30 minutes at holding speed at 1500 ft above the destination airport, using the estimated mass on arrival at the (alternate) airport.
- **Additional fuel** Which is supplementary fuel, if the minimum fuel calculated for the other requirements is not enough to fly to an alternate airport at the most critical point along the the route, fly for 15 minutes at holing speed and make an approach and landing.
- **Discretionary fuel** The extra amount of fuel that may be loaded by the pilot-in-command.

The State of the Operator, or NAA, may allow variations to the pre-flight fuel calculations on fuel by the operator. A specific risk assessment performed by the operator needs to be conducted which demonstrates an equivalent level of safety.

The ICAO also states requirements on in flight fuel management. The pilot-in-command is responsible for ensuring that enough fuel is available to land the aircraft, with more than the final reserve fuel remaining. When less then the final reserve fuel is remaining, a situation of fuel emergency needs to be declared by the pilot. If the amount of fuel is not enough without using the final reserve fuel and alternate fuel, the pilot

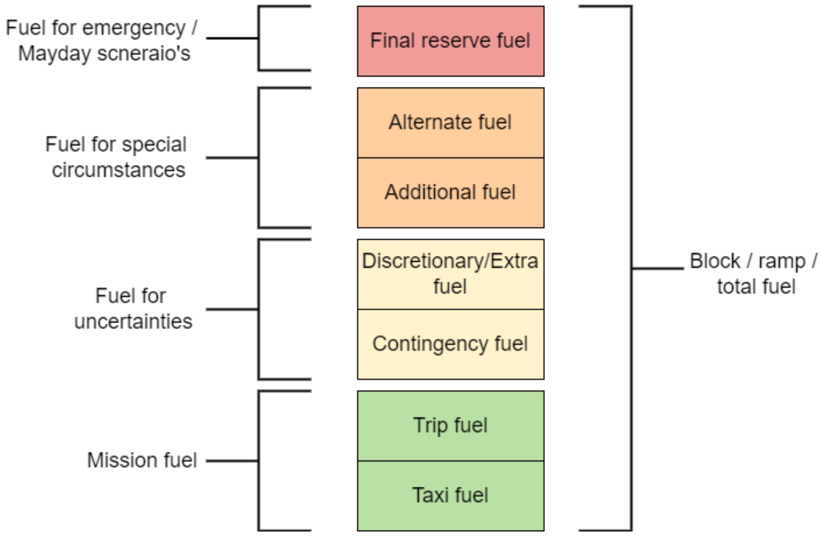


Figure 2.2: Types of fuel defined by ICAO

needs to request delay information from Air Traffic Control (ATC). If an aircraft is committed to land at a specific airport, and adjusting this results in landing with using final reserve fuel, a declaration of minimum fuel is needed. This is not an emergency situation, but an indication that an emergency situation is possible. So, when all the alternate and final reserve fuel is remaining upon landing, no extra communication to ATC is required, and the flight can be considered a normal flight. Once alternate fuel is not fully remaining upon landing, an emergency situation is possible with an increasing chance, the less alternate fuel is available.

2.1.2. EASA

The EASA published the Acceptable Means of Compliance (AMC) and Guidance Material (GM) to Annex IV Commercial air transport operations [Part-CAT] in which requirements on fuel loading are presented. The requirements are similar to the requirements set by ICAO with some deviations, mainly on contingency fuel. These deviations are listed per item below:

- **Trip fuel** The intended departure routing and expected arrival procedure should be included in the amount of fuel as well.

- **Contingency fuel** The amount of fuel should be not less than 3% of the planned trip fuel if an en-route alternate aerodrome (ERA) is available. This ERA should be within a radius of 20% of the trip distance, measured at 75% of the total flight plan distance. The amount of fuel could also be based on a statistical method that ensures an appropriate statistical coverage of the deviation from the planned to the actual trip fuel. For a flight less then 2 hours, 3% trip fuel is required. For a flight more then 2 hours, with an weather permissible ERA aerodrome, a 99% coverage is sufficient. If the destination airport has two separate runways available, a 90% coverage is sufficient. The amount has to be more than the amount needed to fly for 5 minutes at holding speed at 1500ft above the destination airport on top of the trip fuel. For the operator, there is also a possibility to opt for a reduced contingency fuel procedure, allowing the operator to chose a decision point along the route, from which the operator should have enough trip fuel to fly to destination airport and an other airport, including 5% contingency fuel from that point onwards.

- **Alternate fuel** For the alternate fuel, the intended departure routing and arrival procedure should be included in the fuel determination.

The extra regulations on contingency fuel are more complicated compared to the regulations of the ICAO. They allow airlines to reduce fuel loading further, in case safe flights can be ensured. Airline policies should state the percentage of flights that are allowed to land using alternate or final reserve fuel, which indicates the possibility of an emergency situation.

2.1.3. Other regions

Other regions have other NAA's which enforce different regulations on aircraft flying in the airspace. The regulations per region will not be discussed intensively. Important is that other research has been performed in other regions, where different regulations apply. Research has been performed in the US, China and Brazil, where the required contingency fuel is 0%, 10% and 10%, respectively [36]. Without changing the regulations, the contingency fuel determines the lower bound of the fuel loaded to ensure enough trip fuel. So, the potential of optimizing discretionary fuel loading is the highest in the US, whereas it is the lowest in China and Brazil.

2.2. Operations

Adhering to the regulation set by NAA's, airlines are free to their way of operations. Differences between airlines exist, but most use a similar structure. The structure of fuel loading and the related flight planning phase will be described in the current section, identifying critical points for discretionary fuel loading. The exact responsibilities of a flight dispatcher and pilot-in-command is different per airline. The pilot-in-command, now called pilot, is end-responsible for the amount of fuel on board. In the current section, the term 'flight dispatcher' is used for the entire crew that is responsible for the flight planning phase of aircraft, potentially including the pilot.

2.2.1. Airline practice

Prior to the departure of each flight, flight dispatchers decide on the operational settings of the aircraft, the to be flown flight trajectory and the amount of fuel that will be loaded. A flowchart of a typical flight planning procedure is shown in Figure 2.3. The flight planning phase takes place a couple of hours before departure. It is required to submit the flight plan to ATC. So postponing the flight planning phase increases the accuracy of weather and traffic forecasts, but brings more pressure on the flight dispatchers and ATC. The flight plan shall included a 4D flight plan and specifics on the aircraft, as described by Eurocontrol [1]. In Europe, the flight plan may be submitted up to 120 hours before departure. If a flight is delayed, the flight plan has to be altered in communication with ATC.

A Flight Planning System (FPS) is used to assist in the flight planning phase. This system optimizes the flight trajectory and computes the amount of fuel required for the trip. The system allows flight dispatchers to input constraints on the flight plan such as closed routes. It is also possible that the FPS presents multiple routes with its corresponding fuel consumption such that flight dispatchers are able to choose the desired trajectory. Current practice of airlines usually depends on the Simmod simulation program, based on the Advanced Fuel Burn Model (AFBM) developed by the FAA and the Base of Aircraft Data (BADA) developed by Eurocontrol. The simulation optimizes the energy balance of drag and thrust through fuel consumption for the different flight phases as shown in Figure 2.4. The program will be further discussed in subsection 2.3.1. Using a cost index, flight time vs fuel costs are weighted to reach the best result. The industry is constantly improving the flight planning systems in order to maximize the safety standards and operational efficiency. One example is the LIDO system designed by Lufthansa Systems [79]. Making advantage of the big data available in aviation [23], the system is able to improve pre-flight flight planning and in-flight trajectory planning. Historical data is available to test the influence of operational settings on fuel consumption. New improvements include en-route weather integration and in flight route alteration, based on information available in-flight. Based on changing weather forecasts, or opening of closing of airspaces, the system presents a new route in real-time. Advanced flight planning systems also present the Statistical Contingency Fuel (SCF), based on historical data [50]. This calculated SCF however, is inaccurate in some cases due to the small amount of input data.

The flight dispatcher selects the best route from the presented routes based on flight time, fuel consumption and personal experience. Then, the flight dispatcher will give an advice on the amount of discretionary fuel. Policies of airlines may result in a standard amount of discretionary fuel advice on certain flight legs, based on historical evidence [82]. Thereby, the flight dispatcher uses weather forecasts, traffic information and the statistical contingency fuel provided by the FPS to form an advice on discretionary fuel loading.

The pilot in command makes the final decision on discretionary fuel loading based on the information of the FPS, flight dispatcher advice, weather forecasts, traffic conditions and personal experience. It is unethical for

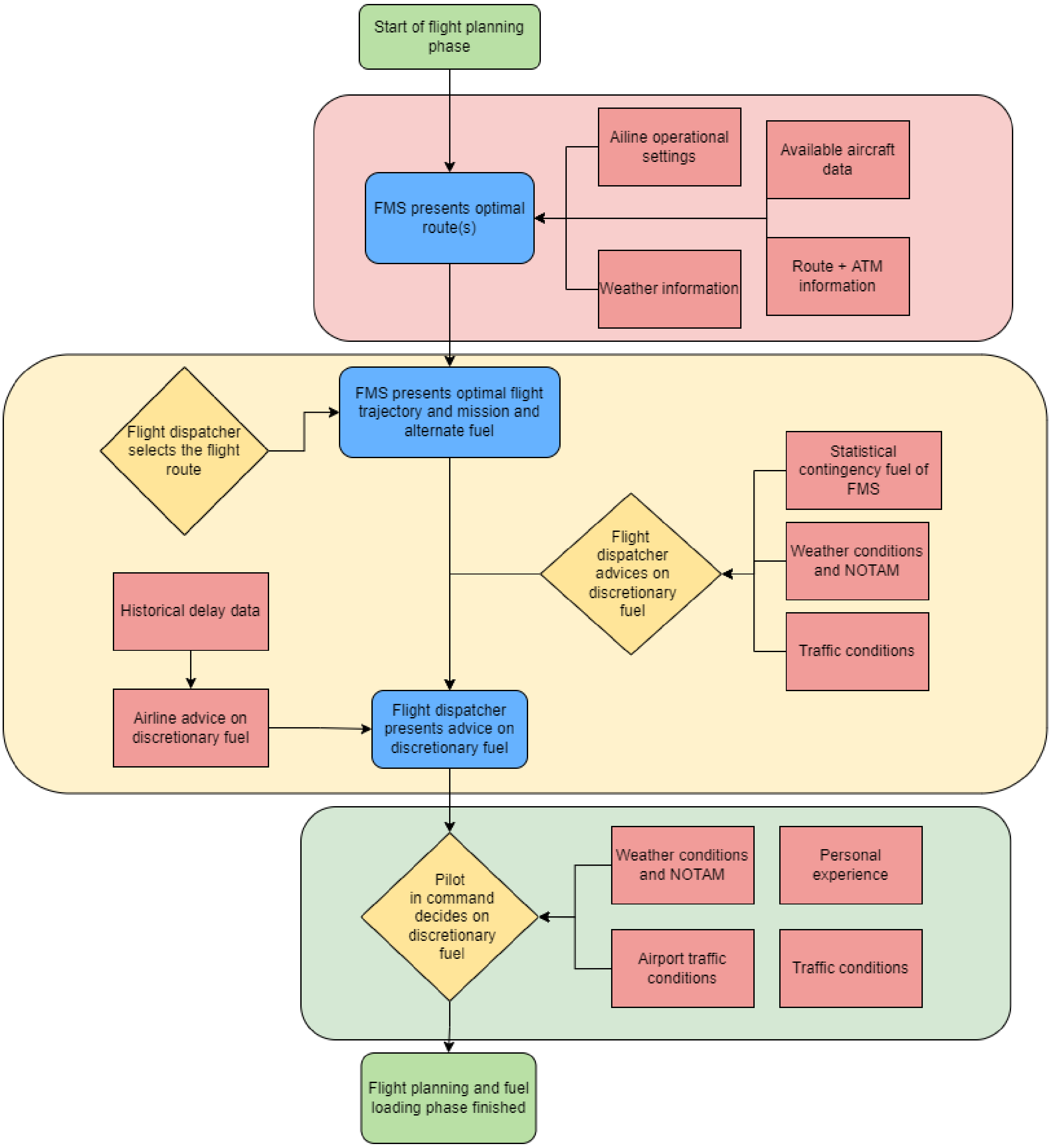


Figure 2.3: Flowchart of the flight planning phase

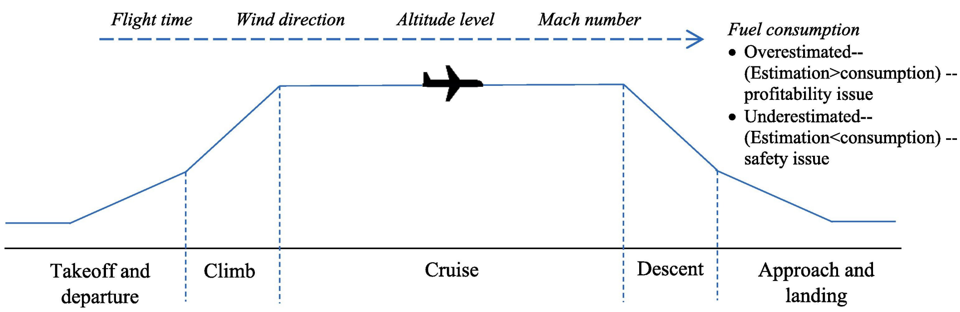


Figure 2.4: Flight phases with trajectory parameters and fuel consumption, from [53]

airlines to punish pilots for loading too much fuel and therefore, they may only advice on the amount to be loaded. This causes the discretionary fuel loading to be relatively subjective. Pilots may be risk-adverse or not, which leads to different values of discretionary fuel loading ([84], [48], [82]). The discretionary fuel must cover both taxi as trip fuel consumption. So, pilots may also load extra fuel for de-icing or long taxiing, since it is uncertain how much fuel that is going to cost. Pilots are a critical factor in reducing discretionary fuel. A single bad experience, such as a mayday situation, might affect future decisions made by the pilot. Therefore it is of upmost importance that the discretionary fuel advice should not lead to extra reserve fuel usage, since this will directly lead to mistrust of pilots.

Airlines do not only consider the amount of fuel required for the current flight but will also consider following flights in fuel loading. When loading fuel is cheaper at the current airport then at the next airport, it can economically wise to bring the cheaper fuel to the destination for the next flight. This fuel is called tankering fuel or ferrying fuel in literature. Bringing this fuel, however, causes extra fuel burn in return. Also maintenance costs are increased because of the increase in weight. Models are created to optimize the amount of tankering fuel on board [3]. If the amount of tankering fuel is larger than the amount that the discretionary fuel would be, the need for discretionary fuel is taken away. So, in these cases, there is no room for optimizing the amount of discretionary fuel, since it will be zero in almost all cases.

2.2.2. Safety considerations

Despite the benefits of minimizing discretionary fuel on commercial aircraft, risks are attached as well. Parties argue that the amount of discretionary fuel or contingency fuel should be increased instead of decreased. For the coming years, airspace congestion is expected to grow every year which increases the chance of delays [32]. Drees et al. [28] found that the current regulations on fuel loading are not in line with the required safety standards set by the EASA, which is an accident probability of $p_{target} = 1 \cdot 10^{-7}$. Although the fuel reserve may be sufficient to safely land the aircraft in theory, experiments on flight simulators have shown that pilot behaviour under stress-full circumstances may result in too little fuel for a safe landing. In the study, flights were simulated that started using 60 minutes of fuel left, 5 minutes before landing. Because of a landing gear malfunction, a missed approach had to be performed. The fuel consumption was increased afterwards due to the increased aerodynamic drag. A realistic environment was created, including miscommunication with ATC.

Moreover, the focus on fuel savings by mainly low cost carriers (LCCs) caused concern by pilots in the past [46]. Increasing the amount of fuel on each flight reduces the chance of fuel depletion. However, using more accurate fuel burn predictions, extra fuel only needs to be loaded on flights with a high uncertainty of fuel consumption. This way, potentially, both the average fuel consumption as the amount of dangerous fuel related situations could be decreased.

2.3. Research on fuel consumption

Improvements on fuel consumption modelling may be found in literature. Research can be found on taxi fuel and trip fuel consumption. Taxi fuel research will be discussed first. Fuel loading strategies will be mentioned shortly but discussed more extensively in chapter 5.

Nikoleris et al. [63] estimated fuel consumption during taxi operations and Khadilkar et al. [51] estimated taxi fuel consumption based on flight data recorder archives. Both studies focused on taxi fuel only. The thrust settings of the aircraft on ground are different then the settings during cruise. A detailed analysis on the number of times the aircraft stopped, accelerated and turned is performed. Including environmental conditions and traffic, an improved fuel consumption prediction could be made.

The taxi and flight phase are sequential. The weight left at the end of taxiing influences the trip fuel consumption slightly. Thereby, there are shared input parameters for fuel consumption predictions such as the weather conditions at the airport, which makes the fuel consumption of the two phases slightly correlated. However, since the trip fuel consumption is affected by many other parameters, it is assumed that the taxi fuel and trip fuel can be estimated separately and combined to a total fuel consumption.

The issue is however, that pilots load discretionary fuel based on taxiing conditions and flight conditions. So, in order to study discretionary fuel, one can not fully rely on the estimated trip fuel consumption. This way, fuel loading reductions can be estimated too positively since discretionary fuel may be loaded because of taxiing conditions. The current research will focus on trip fuel consumption only, because the largest portion of fuel is consumed in flights and because of the separable nature of taxiing and trip fuel consumption prediction. For a proper analysis, a state of the art taxi fuel prediction method should be used in combination with the currently used research. Achenbach (2021) [5], did use taxi fuel in fuel consumption predictions, which will be described in more detail in subsection 2.3.2

For trip fuel consumption estimation, two streams can be typically found. The first being mathematical, engineering or iterative approaches like the Simmod program currently used by airlines. Using models of aircraft on instantaneous fuel consumption, a route is created iteratively, thereby outputting the fuel consumption. The second are data-driven approaches using historical trip fuel consumption data to obtain a trip fuel estimate. In the current section, those streams will be analysed and compared. From this comparison, the best method for the current research will be chosen and will be divided further. The first methods excels in predicting the trip fuel consumption under normal conditions, whereas the second approach excels at incorporating special conditions such as holding before the destination airport. Both approaches will be investigated since both approaches may lead to improved trip fuel consumption predictions and reduce the need for discretionary fuel loading. The uncertainties found in both research area's can be used to create an improved model. Research from both methods can be concluded to introduce the research question for the current research. For clarity, when total trip fuel consumption is meant, the term 'trip fuel consumption' is used, whereas when in-flight fuel consumption per time instant is used, the term 'instantaneous fuel consumption' is used.

2.3.1. Iterative fuel consumption methods

New models are created to improve the accuracy of trip fuel consumption estimations. Studies often improve a single component of existing models. The most used model by airlines is the Simmod model [52]. This model is based on the Advanced fuel burn model (AFBM) developed by the FAA in 1982 [24] and the Base of Aircraft Data developed by Eurocontrol [65]. The AFBM provides physical formulas that lead to instantaneous fuel consumption. BADA contains empirical data on instantaneous fuel consumption of different aircraft under changing operating conditions. The Simmod model relies on an energy balance model throughout the entire flight. By balancing the thrust and drag of the aircraft under the estimated operational settings, the instantaneous fuel consumption at a given moment could be calculated. The energy used by the frictional forces of drag is compensated by the energy generated by the thrust of the engines. Flight altitude, speed and weather conditions are used as inputs for the model. Using operational flight data, parameters for mathematical models are gathered such as engine constants, airspeed, aircraft mass, atmospheric density, extra thrust needed for ascending, and wing area. The model calculates the trip fuel consumption by integrating the flight time over the instantaneous fuel consumption.

In order to create the optimal flight trajectory, the model generates a flight backwards, starting at the end of the flight, at the alternate airport. From this airport, with only the final reserve fuel left in the aircraft, a flight to the destination airport is constructed. Possible trajectories are discretized and the trajectory with the least amount of fuel usage is chosen. Then, from the destination airport, a flight is constructed towards the departure airport ending with the fuel weight that is just computed. The model iterates using the new values for final reserve fuel.

The speed is computed using the Cost Index (C_I) shown in Equation 2.1. A higher speed of the aircraft usually results in a higher instantaneous fuel consumption. The time costs is composed out of the crew costs, leasing costs and maintenance costs per time instant. The ratio compared to the fuel costs determines the optimal speed in that scenario.

$$C_I = \frac{TimeCost}{FuelCost} \tag{2.1}$$

By including wind and weather forecasts, the best trajectory and aircraft speed could be determined. Studies have shown that the BADA models works well in cruise conditions with deviations of 3% compared to actual instantaneous fuel consumption [61]. Pagoni et al [67] highlighted, however, that the information needed to determine coefficients for energy balance methods is not always available in a real scenario and it needs to be generated from other sources, what may limit the use of these models. Thereby, linear interpolation is used to determine values for operational setting in between test points, which reduces the accuracy. The introduction of big data in aviation offers potential to incorporate fuel consumption data from newer aircraft types without using the BADA database [22]. Inaccuracies for terminal phases remain an issue because of data scarcity and larger differences in operating conditions with an resulting accuracy of 22% [2].

Studies in the iterative approach focus on increasing the accuracy of instantaneous fuel consumption at given operational settings or on better procedures to determine the optimal flight trajectory. For the increase in accuracy of instantaneous fuel consumption estimations, data driven approaches such as neural networks are used to get better estimates of the parameters used in the mathematical models. Trani et al. [83] trained a neural network architecture in order to estimate instantaneous fuel consumption during climb, cruise and descent phases of flights, using experimental flight data of Fokker 100 and SAAB 2000 aircraft. They concluded that neural networks achieve high accuracy and can be used in fast-time simulations. Baklacioglu [14] designed a genetic algorithm-optimized neural network topology to predict the fuel flow-rate of a transport aircraft using real flight data focussing purely on the climb phase. Flight data from Boeing 737-800 transport aircraft was used to train the model. Flight altitude, mach number, engine speeds, flight time and fuel flow rates were used a raw inputs. Senzig at al [2] modelled fuel consumption of aircraft in the terminal area using data from the Boeing 777-300ER and Boeing 757-200. The new model has a accuracy of 5% or better up to 10,000 ft which is an impressive improvement compared to a accuracy of 22% by the BADA model. Uzun et al [85] developed a physics guided deep learning model for fuel consumption modelling. Using neural networks, parameters are trained to fit the physical models. The advantage is that physical models can be applied more easily to untested operational settings. Bauman et al [16] developed a machine learning model to estimate instantaneous fuel consumption in a recent study. New is that data from the flight data recorder could be directly transmitted to the ground station. This increases the potential of instantaneous fuel consumption modelling, since in flight data could be directly used in order to estimate instantaneous fuel consumption in the rest of the flight.

The improved instantaneous fuel consumption estimation could be implemented in the trajectory optimization algorithm to get improved trip fuel consumption predictions. Concluding from the studies, attention is put on modelling climb and descent phases, using real-time flight data and using physical formula's for understanding.

New models are created to generate the trajectory of the aircraft more efficiently. The goal is not only to generate a more efficient flight, but also a more reliable flight plan and so fuel consumption modelling. Khan [53] found that aircraft typically do not follow their intended flight plan, since other trajectories result in a lower fuel consumption. Soler [76] worked on optimal control for trajectory optimization, taking into account nonlinear effects. Alligier [10] worked on mass and speed estimations for aircraft during the climb phase in order to get a better estimation on the trajectory during the climb phase. The study compared the flown trajectories with the actual trajectories obtained from ADS-B data. Gallego [35] used traffic data to estimate the reliability of trajectory predictions. Khan, [53], integrated improved trajectory prediction with trip fuel consumption. Jensen, [60], optimized speed and trajectories based on improved instantaneous fuel consumption estimations. Vazques, [86] included wind effects in trajectory optimization.

Concluding from the trajectory research, studies improve the complexity of the trajectory optimization model, focus on a specific phase or introduce new effects as wind and traffic.

Essentially, the iterative approach is split into different problems, of instantaneous fuel consumption modelling, trajectory prediction and trajectory optimization, that result in the actual trip fuel consumption. This causes that studies are usually not focused on fuel loading advice, but mention it as a side effect. Khan (2021) [53], introduced one of the few iterative models for trip fuel consumption estimation.

So, instantaneous fuel consumption modelling is useful to integrate with trajectory optimization. The effects of operational changes or engine age could be tested using flight recorder data. However, total flight fuel burn is calculated iteratively. An error during the descent phase may have large influence on the total fuel burn. The uncertainty of wind or results of congestion like holding are difficult to incorporate. Due to differences in flight phases, new studies focus on only a single flight phase in order to beat the current state of the art. This limits the practicality by combining different models. Current used models rely on old data from just a set of aircraft, but the application of the flight data recorder may result in more updated values. The advantage of new research in instantaneous fuel consumption modelling is the potential of integrating it in current systems. When airlines already use the Simmod program, improvements could more easily by introduced.

2.3.2. Data-driven fuel consumption methods

The second stream, the data-driven approaches, predict the trip fuel as a whole, without defining the flown trajectory. This method is based purely on historical data of trip fuel consumption. A model is created to estimate the fuel consumption of previous flights using available information. This information should indicate uncertainties in the fuel consumption such that more accurate predictions could be made. The information may relate to trip fuel consumption causally, or is merely correlated without a direct causal relationship. More on input data will be described in section 2.5. An elaborate description of the research is presented, since the information is necessary for later analysis in the report.

The potential of reducing contingency and discretionary fuel loading has been introduced by Ryerson in 2015 [73], where he found that up to 1.04% of fuel is consumed due to carrying fuel beyond a reasonable buffer. The reasonable buffer in this research is enough for 99% of the flights, which is on the low side as will be discussed in chapter 5. He performed statistical analysis on the fuel remaining upon landing from a single airline and used these results to generate fuel savings for all the major airlines in the US. Thereby, a cost to carry calculation method for carrying extra fuel is introduced in this study to compute the fuel savings by fuel weight saving. The goal of the research is to determine the discretionary fuel for flights of six major airlines and so does not give predictions for individual flight fuel consumption.

In contrary, Ryerson also modelled trip fuel consumption in 2014 [72] in order to identify the extra fuel burn caused by delays and airport terminal inefficiencies. Here, fuel consumption per individual flight was calculated. Equation 2.2 was used to estimate the impact of delay and terminal inefficiencies on fuel consumption. c is the baseline fuel consumption value, d a vector of delay information, q a value representing assumed and actual TO weights, y a vector containing origin and destination information and w a vector of airport weather variables. The delay information is divided in departure delay, which can only take negative values and airborne delay, which can take positive and negative delays. Using data of 4000 aircraft in the US, he found that departure delay, airborne delay, planned flight time, and take off weight difference are all correlated at the 1% significant level. In this study, the relationship per origin and destination airport on fuel consumption is also shown. Different origins and destinations show different effects on fuel consumption because of terminal inefficiencies. *Concluding from the delays, delays account for a range of 0.75-1.75% of the fuel consumption, the 90th percentile is around 3%, whereas the largest values are in the order of 20%.* The resulting fuel consumption is shown in Figure 2.5. Excess flight time is also considered as a delay, which does not account for more then 1% of the fuel consumed for all flights. Although data is used that is only known after the flight takes place, causal relationships of delays on fuel consumption are found, which is harder to incorporate in the iterative approach. There is however, a baseline fuel consumption used, which makes advantage of improved instantaneous fuel consumption estimations.

$$f = g_n(c, d, g, y, w) + \epsilon \tag{2.2}$$

Statistical analysis and econometric methods are also used by Hao et al. [39] to model discretionary fuel for individual flights. The predictions are presented as measure for the unpredictability of the ATS. In this study, only input data available 2 hours before departure is used. historical traffic effects are included in the regression by incorporating averages and standard deviations of the airborne time per OD pair. It was found that a 1 min increase in the standard deviation of flight time increases contingency fuel loading by 0.88 min. Moreover, weather indicators at the destination and origin airport such as a low ceiling indicator and a indicator for low visibility are used on top of the average quarter hourly delay at the destination airport. This way, correlations are present to estimate the effect of unforeseen scenario's in trip fuel consumption modelling. An increasing uncertainty leads to an increasing contingency fuel loading, so a reduction in uncertainty may lead to a reduction in contingency fuel loading.

Gomes et al. [36] uses historical flight data to evaluate contingency fuel loading regulations in Brazil. A distribution of fuel reserves upon landing is created using almost 300,000 flights. The flights are grouped depending on their range, but other then that, no input data is used for the trip fuel consumption estimation. Using the determined distribution of fuel consumption, Monte Carlo simulations were ran. The goal was to compare the fuel left for 10% contingency fuel and 5% fuel loading. Only one flight of the 300,000 flights landed using reserve fuel, leading to an emergency scenario. This was case for both the loading regulations.

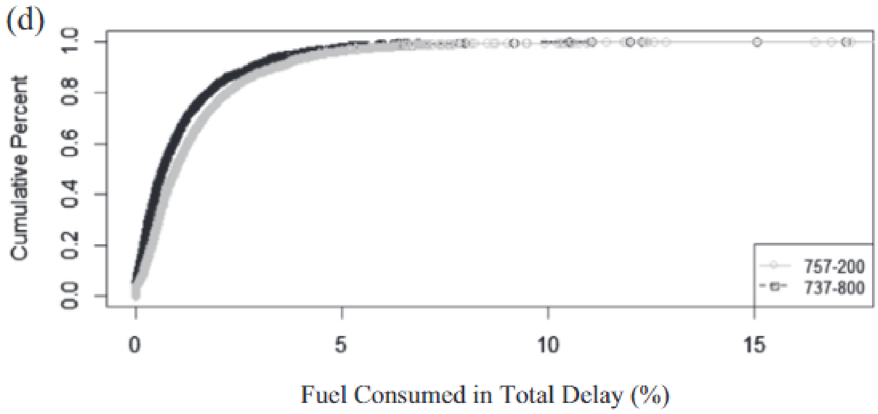


Figure 2.5: Percentage of fuel consumed in the delay of a flight, from [72]

Flights using alternate fuel are not explicitly mentioned. Undesired situations of using alternate fuel may increase when reducing the contingency fuel, but are not mentioned. Monte Carlo simulations allow users to properly investigate extreme scenario's in fuel consumption, but are costly to make for models containing many inputs, since distribution for input-output relations need to be generated.

Kang, [49], introduces machine learning methods for fuel loading decisions in 2018 by creating an ensemble of multiple machine learning algorithms in order to create confidence bounds for fuel usage on an individual flights. Machine learning methods have the potential to capture non-linear effects and thereby increase accuracy [81]. In this study, more information on the distribution of historical fuel consumption is used, namely different percentiles of historical airborne time and fuel consumption. Weather forecasts at origin and destination airport are used as well. Thereby, OD pairs are clustered in three groups based on flight direction and distance. Lasso, Ridge regression, Random Forests, Gradient boosting machines, k-nearest neighbors (KNN), multivariate adaptive polynomial spline regression (MARS) and two types of stacking are used as models. Detailed explanation is presented in chapter 3. Base learners already improved the RMSE of the FPS by 50%. The ensemble performed 2-5% better then base learners on RMSE. So, relatively simple models as Lasso and Ridge regression perform considerably better then FPS, but only slightly worse then more complex estimators. From the regression trees, the relative importance of input features can be found. Planned trip fuel and trip time were amongst the most important features and airborne percentiles thereafter. Using clustering techniques on the dataset, similar datapoints were grouped and 95% and 99% confidence bounds for those groups are constructed. The values are linearly extrapolated such that the amount of reserve fuel usage was similar too airline practice. No minimum was set on the discretionary fuel. The resulting benefits are in the range of 169-422 lbs of fuel savings per flight. Kang [48] also investigated fuel loading behaviour by pilots. He concluded that pilots value 1 kg of reserve fuel 1,200 times more than 1kg of extra fuel consumption. Resulting from the data set used, Figure 2.6 was found, showing the variability in discretionary fuel loading.

Kang also used machine learning methods in 2021, [50], this time predicting the statistical contingency fuel (SCF). The FPS outputs a SCF prediction, which is often inaccurate. Weather forecasts are not considered, which has a high correlation to SCF. Using a quantile regression as loss function, the SCF is modelled using quantile regression, gradient boosting, random forests and stacking. Input data included historical airborne distributions and weather forecasts at the origin and destination airport. Resulting from the predictions, a SCF lower then zero was found in some cases. This means that the estimated trip fuel is already high enough for 95% of the flights. Despite that, the airline had a policy of 10 minutes of minimum contingency fuel, which is also used in the benefit estimation. In total, 19 million $ could be saved using the new SCF estimations, whereas 61.5 million $ was found using no minimum for discretionary fuel.

Machine learning methods are also used by Achenbach, [5], where she used different algorithms to predict flight time and fuel burn for 20,000 international flights. Lasso, random forests and gradient boosting machines are used to estimate trip fuel consumption. The accuracy increased by 30% for the MAE. In this model, taxi fuel is included. Mission features, weather forecasts and (the error in) fuel consumption of previous flights is used. Also gate and runway information is used, which is correlated to taxi fuel. The gradient

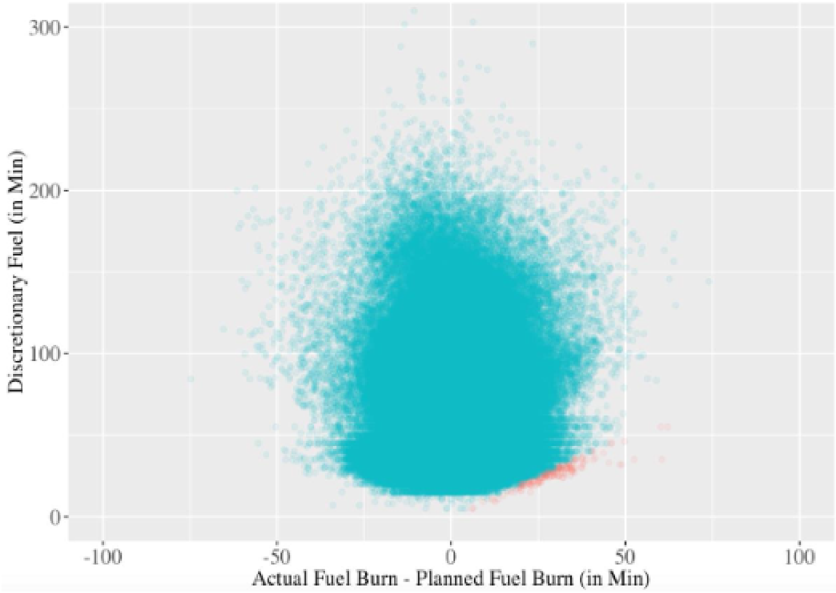


Figure 4.1 Discretionary fuel loading overview

Figure 2.6: Discretionary fuel loading on flight data, from [48]

boosting method allows for presentation of important input parameters, using this the fuel consumption of previous flights was found to be the most important.

Interestingly, Achenbach also modelled arrival times for different phases in the flight, including block off time, i.e. departure time. For this model, en-route weather forecast are taking into account. Regression, random forests and gradient boosting are used as estimators. The FPS uses weather forecasts for each waypoint in the flight planning phase. Achenbach compared the accuracy of the forecast with weather data during the flight, which includes temperature, humidity, wind speed and wind direction. For the in flight weather, the AMDAR data set is used. She found that the accuracy of temperature and wind direction had a MAE of 4.4% and 2.8% respectively. Even though the wind direction has a high influence on fuel consumption, the variations were not considered relevant and therefore, these data points were not considered as input point. Wind speed variations were higher, namely 17%. This may be explained by the fact that wind speed can be close to zero, resulting in high percentage differences. The target feature is remaining flight time for critical phases in the flight, including taxi time. Besides the en-route weather information, similar input features are used as in the fuel consumption model. Trial and error was used to get the best correlated features. Only weather forecasts at cruise level are used since the TAF covers the climb and descent. From the experiments, regression was found to perform best for the remaining flight time at departure without en-route weather data, while random forests performed better with weather data. The method improved the accuracy of estimation by 20%, while it improved by 25% when the en-route effects are considered. Again, the most important variable is flight time prediction error of previous flights.

The promising results of using en-route weather effect on arrival time are not considered in trip fuel consumption estimation. There is potential to improve the fuel consumption estimations this way.

Zhu [96] was the first to incorporate neural networks for trip fuel consumption. She did this by estimating the flight time of aircraft, and estimating the fuel consumption based on this flight time. The neural network is based on flight mission data, weather forecasts at the origin and destination airport and traffic delay states from ADS-B data. For the traffic features, a spatial weighted long-short term recurrent neural network model is used, such that delay states in the network can be captured as well. The spatial weighted term refers to training on the delay states of OD pairs. In the complete set the RMSE was reduced by 6% compared to Lasso stacking, whereas the RMSE reduced by 26% for the outlier set. This introduces the importance of treating the outlier set differently, since fuel loading decisions are based on misjudged past outliers. This study presented the effect of incorporating traffic information by using delay states in the network. Flight time increases due to holding can be predicted. The model predicts flight time in order to convert it to fuel consumption. From the previous studies, we found that flight time was one of the correlated parameters, but dot not contain all information on fuel consumption.

The main advantage of data-driven models is that uncertainties in fuel consumption over the entire trip will be considered. Even if the direct fuel consumption is accurate, the mathematical approach does not take into account delays or other events that affect fuel consumption. Using input data per month or OD pair, systematic effects and correlations can be identified.

There is, however, a pitfall in the approach which may let the resulting accuracy be too high. By introducing more specific parameters in order to identify correlations, chances on overfitting increases. Temporary effects such a month with bad weather, or maintenance on a runway of a specific airport, is included in the train set as well as in the test set. Estimations should only be able to be made using data that is available a few hours before departure in order to advice on discretionary fuel loading. A test set should be used in a different time frame then the model is trained in for realistic results. Furthermore, models created is studies use trip fuel consumption for the FPS as input in the model. So, a more accurate trip fuel consumption in the FPS automatically leads to a more accurate trip fuel consumption in the new created model. So, the complexity of the FPS should be described as well, in order to assess the performance of the new model.

2.3.3. Possible outputs

The current research focuses on the uncertainties for fuel loading and the advice on discretionary fuel. Before possible methods will be investigated, the desired output of the model should be stated. As discussed in section 2.1, fuel loaded on an aircraft is build for different types of fuel. Although the desired output is discretionary fuel, actually modelling the discretionary fuel is not useful. In some scenario's, discretionary fuel should be zero, since the contingency fuel is already enough to ensure a safe flight [49]. In order to capture all the effects of uncertainties and model flaws, the trip fuel should be estimated. This is also more useful to evaluate the accuracy of the FPS to compare vs other models.

Then, focusing on trip fuel, different values may be outputted. An actual fuel consumption estimation is useful in order to compare the model results with actual data. A confidence bound however, is more useful for discretionary fuel loading. Kang (2018) [49] trained its model on confidence bounds. In his reasoning, a 99% confidence bound should be enough to ensure a safe flight. Current practice of airlines is however, loading fuel beyond what is needed in 99% of the cases. The higher the percentage gets however, the higher the chance on overfitting of the model will be. For a 99.9% confidence bound, only 1 out of a 1000 flights is above this bound. In total, a low amount of flights will be left. Kang found that extrapolation on the 95% was in some cases better than extrapolating the 99% case. So, different levels of confidence bounds will be modelled. This will be further discussed in chapter 5.

2.3.4. Reflection on trip fuel consumption approaches

The trip fuel consumption approaches both have its advantages and disadvantages. They are summarized in Table 2.1. An increasing accuracy in iterative approaches leads to an increasing accuracy in data-driven methods as discussed in subsection 2.3.2. Still, the data-driven methods excel in capturing uncertainties encountered in flight. The goal of the current research is to gain trust of flight dispatchers in fuel loading models, such that discretionary fuel loading can be reduced. Because of the ability to incorporate uncertainties in fuel consumption well, data-driven methods will be used. It remains important that iterative approaches should be improved as well, to provide better estimations of the trajectory. This approach can also be used to model effects on specific parts of the route such as wind effects, which is also useful for identifying uncertainties using the data-driven method. Moreover, keeping track of operational setting such as the cost-index or aircraft trim settings affects the accuracy of the data-driven method, since these changes affect the correlations of other parameters. Using date information, most operational changes should be covered within the date correlations.

2.4. Uncertainties in trip fuel consumption estimation challenges

The factors that influence trip fuel consumption will be discussed in the following section. Using these factors, the completeness of existing research can be investigated. If certain factors can be included in an estimation model, the accuracy may be improved. The factors could be grouped in direct uncertainties, which affect instantaneous fuel consumption and indirect uncertainties, which influence the flight trajectory, speed

Table 2.1: Summary of fuel consumption methods in literature

Approach	Adavantages	Disadvantages
Iterative	Trip fuel accuracy Trajectory inclusion Makes use of available in-flight data	Complex Insensitive to unforeseen factors
Data-driven	Discretionary fuel accuracy Simple Robust, when including uncertainties	Insensitive to new operational settings Overfitting on correlations

or flight time and therefore influence the trip fuel consumption.

2.4.1. Direct fuel consumption uncertainties

Direct uncertainties affect the physical state of the aircraft and therefore influence the fuel consumption. Only factors that are unknown a couple of hours prior to departure will be considered in the current section. These factors could be integrated in the mathematical trip fuel consumption estimation approaches. In fact, the advances in the new developed systems such as the LIDO system are able to incorporate increasing amounts of data and so also incorporate variations in direct fuel consumption. Still, uncertainties remain, which explain the deviations in fuel consumption.

New operational settings such as a changed cost index or different airport arrival slots are known before the flight and don't influence the uncertainty in fuel consumption estimation. Ofcourse, if a new operational setting is unknown, it causes an increase in uncertainty of fuel consumption. When assuming no changes are present in operational settings, this assumption need to be verified in order to obtain realistic results.

Aircraft specific variations Although an aircraft type is designed in a single way, minor differences in the production always will be present. More importantly, over the life time of an aircraft, the aircraft degrades and gets different forms of maintenance. Aerodynamic properties changes because of surface issues and the engine performance changes because of degradation [47]. Studies assuming the same fuel consumption constant for same aircraft types, resulted in a root mean squared error of less then 1%, which suggests aircraft specific variations have limited effect on the direct fuel consumption [91].

To incorporate the effect of aircraft specific variations, one should use general input features. Aircraft and engine age, or the last time since an A-check or C-check might indicate aircraft performance degradation. Using each individual aircraft as input results in a high risk of overfitting, since correlations may be found that are caused by coincidence.

Weight variations If aircraft weight increases, thrust of the engines to maintain the same speed has to be increased which increases fuel consumption directly. The Dry Operating Weight (DOW) of an aircraft typically increases by 0.1/0.2 % per year due to maintenance additions up to a total of about 1% [47]. If possible, weighing the aircraft regularly, reduces this factor of uncertainty.

Besides DOW, there is an higher uncertainty in passenger weight. The weight per person and the number no-show passengers fluctuates. The flight planning and fuel loading phase takes place before boarding of the aircraft which makes it not possible to estimate the number of passengers exactly. Possible solutions are estimations on the number of no-show passengers on historical data and pushing the fuel loading phase as close to the boarding as possible. Weight per passenger differences per passenger and throughout the year. There are correlations found between months and passenger weight, [39], because of winter clothing. Using a assumptions of 3% fuel per payload kg per hour [47], a basic calculation results in a 5% passenger difference leading to approximately 70 kg difference in fuel consumption for a 3 hour flight in a B737-8.

Upon take-off, the fuel left from the taxiing phase also varies. Traffic at the airport may result in waiting on other aircraft, while the engines are running. The amount of fuel to ensure enough for the taxiing phase is not considered in the current research, but the weight left over will affect the trip fuel consumption. Moreover, the weight uncertainty also affects the weight balance of the aircraft. If the weight balance varies, trim settings of the aircraft result in drag changes.

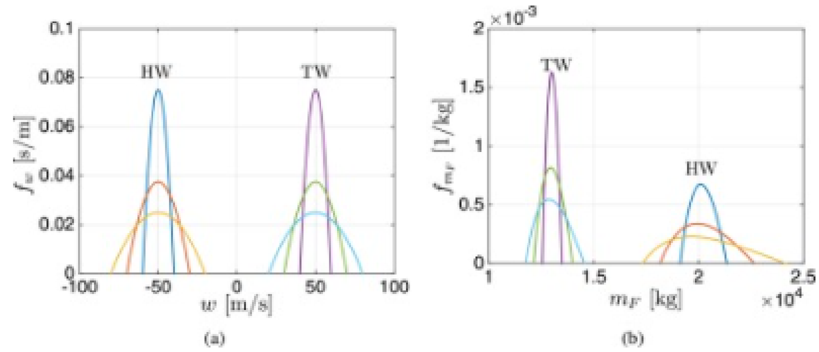


Figure 2.7: Probability density functions of wind speed (a) following a symmetric beta-distribution and its resulting probability denstiy functions of fuel consumption (b), from [86].

Weather uncertainties The conditions of the atmosphere directly influences fuel consumption. Wind speed affects the relative velocity of the aircraft in the air. For the same ground speed, another amount of thrust has to be exerted. The temperature and air humidity influences the drag coefficient of the aircraft and so, different values of thrust are needed. Weather forecast are uncertain due to the complexity of weather forecasting. Vazquez et al. [86] investigated the influence of wind uncertainty on fuel consumption. Different distributions of wind uncertainty are used to identify the relationship to fuel consumption by using mathematical models. One example is shown in Figure 2.7. In the study, he found that the standard deviation of fuel consumption scales almost linearly with wind uncertainty. A standard deviation of 10 m/s results in a standard deviation of 110-650 kg for a B767-400, flying for 3000km at cruise, changing per assumed wind distribution and head or tail wind. The mean of the fuel consumption was not effected however. From this paper, one can conclude that the probability distribution of predicted wind is important for fuel planning. Thereby, they concluded that head wind uncertainty has a higher impact on fuel consumption uncertainty then tail wind.

Weather uncertainties are difficult to assess because of inaccuracies in weather forecasts. Including the variance of the forecast in the predictions would entail information on the variance of the final result. Incorporating percentiles or other distribution information is useful to use as well. Indications for bad weather in weather forecasts or NOTAMs could be used to estimate the probability of severe weather occurring.

2.4.2. Indirect uncertainties

Indirect uncertainties make changes to the intended flight plan. The flight plan is created during the flight planning phase. Once the flight planning phase has past, the environment may have been changed or new information is available. Typically, this information is only present in-flight. The changes in flight plan influences the total trip fuel consumption. Only some of the flights will have changes in the flight plan, so these uncertainties can best be incorporated using data-driven approaches.

Trajectory alterations Trajectory alterations are intended alterations to adapt the flight plan to a more fuel efficient approach. In order to change the flight plan, admission by the ATC shall be requested. When no conflicts are present, the alterations may be implemented. With trajectory alterations, minor deviations from the intended flight plan are meant, namely altitude variations, speed variations and minor horizontal deviations caused by conflict detection. Atmospheric conditions will be different from the forecasts and lead to different a different optimal flight path. Mainly with different altitudes, wind speed and temperature varies. Without flying extra horizontal distance, better operating conditions can be found. Pagoni et al. [66] found that flight altitude of flights vary on the same day. Lovegren et al. [59] found that speed and altitude optimization in flights may result in 2% fuel consumption reduction.

Other differences in trajectories are changes in speed and minor deviations in the horizontal flight plan. Flight speed may change to plan the time of arrival at the destination airport. In Europe, Eurocontrol could suggest to increase or decrease the speed of a certain flight, in order to avoid congestion at the destination airport. Minor deviations take place due to differences in atmospheric conditions or due to conflict detection. Also for the conflict detection, ATC will request to alter the intended flight plan of users in the airspace. Within the trajectory on aircraft, speed alterations are considered as well. If the departure of a flight is delayed, airline may change their flight plan to a faster, less fuel-efficient, speed to recover time. Ryerson, [72], found

Figure 2.8: A situation where the flown trajectory (blue) is different from the planned trajectory (yellow), from [33]. In this scenario, a Belgian military airspace opened for commercial flights which allows the short-cut to be performed.

that 1 minute of departure delay results in approximately 1 kg of extra fuel burned for a B738. Also the excess planned flight time, the time above what the FMS predicts, is correlated to the fuel consumption due to longer airborne time, with 3 kg per extra minute.

Rerouting The European airspace is divided into airways, which aircraft are allowed to use. These airways provide an overview for Air Traffic Flow Management (ATFM) to help with congestion handling and conflict resolution. Airways could be closed or airspace could be congested. In those cases, ATFM suggest another route to the aircraft in flight. The maneuvering to the other routes and the difference in distance affects the fuel consumption of the trip. Still, the closing of airways in flight is very rare, since information on this is usually available earlier and so the flight plan is already adapted at the flight planning phase [64]. Airspace congestion will be elaborated upon in the next paragraph. Besides the closing of airways, it is also possible that airspaces could actually be opened. Military airspaces are considered inadmissible by the flight planning system, but these airspaces open up when the airspace is not used by the military. An example is shown in Figure 2.8. Depending on authorization of ATC, flights may enter the airspace. This leads to shortcuts and fuel savings.

So, routes closing will result in extra fuel burn while airspaces opening results in lower fuel burn. ATC like Eurocontrol will have the most data accessible on information of these event, which will be unavailable by airlines. Historical analysis on airspace closing or openings will be the best way to estimate the uncertainties of ATM induced rerouting.

Severe weather may also cause necessary reroutings to ensure a safe flight. For example thunderstorms are better to avoid reducing the risk of aircraft damage or losing control of the aircraft. When such a decision is made in flight, a new route needs to be generated. The increase in distance typically leads to an increase in fuel consumption. Studies have been performed on optimizing route alternative in flight [62]. Furthermore, a runway and a corresponding Standard Arrival Route (STAR) is selected in the flight planning phase. Different operating conditions may result in a change to another runway. Depending on the orientation of the runways, this results in a positive or negative difference in fuel consumption.

Holding When an airport or an airspace is congested, ATFM requests approaching aircraft to hold before entering the airport or airspace. This extra flight time causes extra fuel burn without covering distance towards the destination. Ryerson, [72], found that extra airborne time was correlated with fuel consumption by 25 kg per minute. A holding pattern is defined by ICAO and is shown in Figure 2.9. Aircraft may enter the holding pattern from four different entry points and fly complete laps thereafter. The time typically takes four minutes of flying. Airports have a limited number of entry points, defined by the open runways, which determines the chance of a queue building to land the aircraft. Also in this case, ATFM manages the incoming traffic, but holding aircraft temporary in the air is sometimes needed to allow for safe operations [7]. En-route

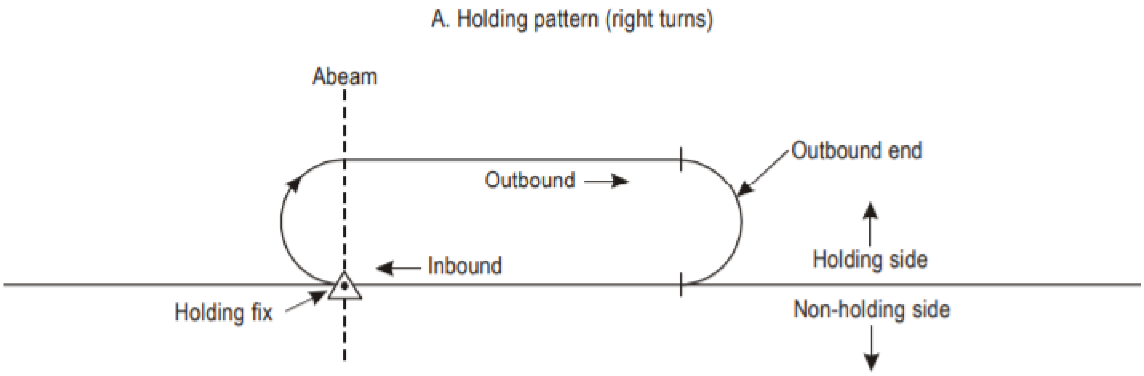


Figure 2.9: Holding pattern for aircraft, from [43].

holding does not occur regularly. Airspaces have multiple entry points and ATC can anticipate and manage the incoming flow by speed alterations. Airlines have to submit their flight plans prior to departure to ATC. Using this information, ATC can estimate the level of congestion for the coming period of time. When the airspace is congested, aircraft may be asked to extend their departure.
Airlines have historical data on the number of times aircraft are required to hold at the destination airport. The chance of holding at an airport can be deduced from this historical data. Especially busy airports such a higher chance for holding. The effect of sever weather on arrival delay appears often in research. Severe weather limits the airport runway capacity, which increase the chance of queues building. Among others, Reitmannand [9] investigated the correlation of METAR weather forecasts on arrival delays. When the airport is operating close to its maximum capacity, weather has a strong influence on the delays at the airport.

2.4.3. Visual presentation of the uncertainties

The uncertainties previously listed are visually presented in Figure 2.10. The factors are divided in severity, ranging from minor to major and predictability, ranging from high too low. The direct factors are indicated in red and the indirect factors are indicated in yellow. The indirect factors do not always affect the fuel consumption. However, the severity is taken when such a scenario takes place.

Uncertainties influence discretionary fuel the most when the predictability is low and the severity uncertain. This way, many flights should bring extra fuel to reduce the risk of reserve fuel usage. Studies have investigated the effect of predictability of holding on trip fuel consumption, discussed in subsection 2.3.2. Weight variations and trajectory alterations have been included in instantaneous fuel consumption and trip fuel consumption as well. Rerouting of aircraft and direct weather uncertainties influence the trip fuel consumption severely, but have not been studies regarding fuel planning. Estimating the effects on fuel consumption and predicting the chance of effect prior to departure may increase trip fuel consumption accuracy. From the figure, on can conclude that relatively certain factors influence the fuel consumption less then uncertain factors, in general. Changes in the small factors as weight an aircraft specific variations, won't cause the aircraft to use more then the available contingency fuel. However, those factors make the total fuel consumption in extreme cases as holding more uncertain. Increasing the accuracy of regular flights will also help to increase the accuracy of extreme flights.

2.5. Input data for fuel consumption

Using the uncertain factors discussed in section 2.4, relevant input data that affect one of the uncertain factors can be investigated. The input data can be used in predictive models described in chapter 3. Correlations can be found between the input data and trip fuel consumption and used for estimations in a similar scenario. Input data can be casually related, or merely correlated. Wind is stronger in the winter months, which may cause the month January to have a higher variation in fuel consumption, while en-route wind forecasts casually influences fuel consumption, which coincidentally might be higher in certain months. Input data may be correlated because of multiple reason. In winter months, passengers are heavier, wind is stronger and other operational policies of the airline may apply. To obtain an accurate estimation, these correlations should be

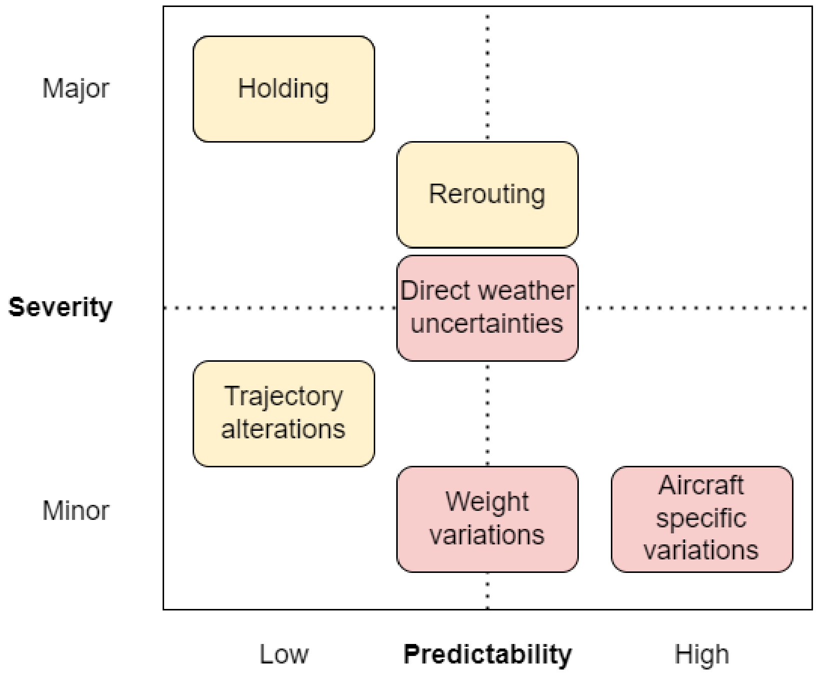


Figure 2.10: Assumed categorization of uncertainties in fuel consumption. Direct factors are shown in red, indirect factors in yellow.

split up as much as possible, Collectible data is discussed in the current section. Feature creation from the input data will be discussed in chapter 4.

A summary of possible input data is shown in Table 2.2. Not every possible single input feature is mentioned, since each paper has its own list of 50+ features ([48],[5], [96]). The input data is grouped such that new input data may be placed in one of the groups. Interesting to note is that increasing the types of inputs often increases the accuracy of prediction models. Achenbach [5] introduced en-route weather data on flight arrival modelling, which increased the accuracy by 5%. Zhu [96] introduced delay states of aircraft at airports, resulting in an improved prediction accuracy of 26% for the outlier set. In this study, a new model type is introduced as well, which is designed for the extra input features.

Interesting is the result that no research has considered en-route effects for fuel consumption yet. It is known that weather conditions influence the fuel consumption, and that the conditions are uncertain. An argument posed by Kang (2018) [49] is that trip fuel consumption estimation already takes into account weather forecasts, such that trip fuel consumption data may be used as proxy. However, trip fuel consumption is also correlated to other factors as the cost index and weight of the aircraft. Thereby, one would rely on the estimated trip fuel consumption purely. Factors important for discretionary fuel estimation would not be used such as the variance in predicted weather conditions.

The total number of flights is roughly similar per day due to the weekly schedules of airlines. Differences however exist in the actual positions of aircraft. This leads to conflict avoidance enforced by ATFM. The progress of aircraft in their flights will influence congestion at the destination airport which causes holding patterns. Thereby, ATFM allows trajectory variations if the airspace is not crowded. If this is the case, flight plan alteration have a higher chance to conflict with other aircraft. So beneficial altitude or trajectory alterations are not possible in congested airspaces. The time of day indirectly predicts congestion well because of repetitive patterns in the airspace. This might reduce the benefit of considering en-route traffic information as input.

The route to be flown is not used in input data either. Within an OD-pair, different routes are possible. The difference in distance is covered when taking the mission data as planned trip fuel consumption into account. However, the different routes might show different historical fuel consumption data. Thereby, shortcuts or closings of routes may appear more frequently in certain routes. Taking into account different route segments

Table 2.2: Summary of possible input data used for trip fuel consumption modelling

Input data	Relates to	Causal	Covered in literature
Flight mission data	All uncertainties	Partially	Multiple
Airport specific information	Rerouting Holding	Yes	Multiple
Date-time data	All uncertainties	No	Multiple
Historical mission data	All uncertainties	No	Multiple
Airport weather data	Weather effects Rerouting Holding	Yes	Multiple
En-route weather data	Weather effects	Yes	Achenbach (2021) [5], on arrival time
Traffic delay states	Holding Trajectory alteration Rerouting	Partially	Zhu (2021) [96]
En-route traffic	Trajectory alteration Holding Rerouting	Partially	Gallego et al. (2019) [35], on trajectory prediction
Route information	Rerouting	No	Multiple ([6], [62]), on in-flight rerouting

results allows information of other flights sharing similar parts of the route to be used for the current fuel consumption estimation. To determine the number of times a shortcut is used or a route is closed, one can cluster flown trajectories within an OD pair and compare the trajectories with the intended flight path. By detecting outliers in the flown route, information on historical rerouting is available.

2.6. Conclusion on fuel loading operations

A literature study on the regulations, operations, research and the uncertainties in fuel consumption has been performed. From the analysis, it is clear that studies focus on improving instantaneous fuel consumption predictions, trajectory optimization methods and trip fuel consumption predictions to increase the accuracy in fuel consumption estimations. In this research, data-driven methods will be used, since uncertainties as delays can be best incorporated this way. Studies improve state-of-the-art by improving the models [53] or by including new data inputs ([5], [96]). From the existing research, the following research gaps can be identified:

1. En-route weather forecasts are not included.
2. En-route traffic congestion information is not included.
3. Route information is not included.
4. Taxi fuel estimations are not included in discretionary fuel advice.
5. A fuel loading strategy including a risk analysis in not included.

The research gap of the fuel loading strategy will be further discussed in chapter 5. Not all research gaps can be filled in a single research. Taxi fuel estimations and a detailed fuel loading strategy is not considered in this report as is discussed in subsection 2.3.2 and will be discussed in chapter 5. The other research gaps are related and be combined well in a single research. Therefore, this research will fill the first three research gaps, using en-route data for trip fuel consumption estimations.

The two related research area's of supervised learning and feature engineering will be discussed in chapter 3 and chapter 4, respectively. Since the benefits of the trip fuel consumption is dependent on the used strategy, fuel loading strategy optimization methods will be discussed in chapter 5:

- **Supervised learning** Prediction on trip fuel consumption estimated using models trained on historical data. Using the historical target variable, the estimations fall under supervised learning. Different methods will be discussed in order to obtain the best method for the current research. Ways to reduce overfitting will be discussed as well.
- **Feature Engineering** Features are required to be generated from existing available data prior to departure. It is unknown what features are desired. Therefore, feature engineering needs to be investigated. Thereby, the high dimensional input data of air traffic data needs to be organized in order to use it as input features for the model.
- **Fuel Loading Optimization and Cost to Carry analysis** Once a trip fuel consumption estimation is present, a fuel loading strategy needs to be determined. Optimization methods may be used for this. The cost to carry analysis is used to determine the benefits of the new used approach.

3
Supervised Learning

When a model is desired that predicts a target variable based on input data, supervised learning methods can be used. Supervised learning is a broad term and research groups the models according to specific characteristics. To start with, supervised learning is used for classification and regression problems. When the target variable is categorized, classification methods is used while for continuous variables, regression methods are used. Still, most supervised learning models can be used for either classification or regression, with slight alterations. Thereby, studies group supervised learning methods in other categories as forecasting, regressions or machine learning methods [23]. In this chapter, the following distinction is made: Statistical (or econometric) models use regression techniques to identify correlations between input and output variables, described in section 2.1. Regularized regressions are also considered under statistical methods, even though that is not always done in studies. Machine learning methods are more advanced parametric or non parametric models including RF, SVM and KNN, discussed in **??**. Neural networks and deep learning methods are separated from other machine learning methods and described in section 3.3. All methods are trained on a loss function, which influences the resulting model. The loss function is in particular interesting for the current research, since asymmetric loss functions might be well suited for trip fuel consumption predictions [49], [50]. Furthermore, regularization and hyperparameters of each model is discussed at the description of the model.

3.1. Statistical models

Statistical methods identify correlations between input features and the target variable. This way, estimations on new data could be made using the trained parameters. Regression models and time series will be considered in the current section.

3.1.1. Regression models

The most simple and widely studied form of forecasting is a linear regression. A regression describes the relationship between a dependent variable and a set of independent variables. In the most general form, the regression model can be written as in Equation 3.1 [56]. Where $f(x_t;\beta)$ is a mathematical function of p independent variables $x_t = (x_{t_1},...,x_{t_p})$,unknown parameters $\beta_t = (\beta_{t_1},...,\beta_{t_m})$ and the noise of the model, ϵ_t. The noise of the model can be assumed to follow various distributions. The most used assumption on the noise is the Gaussian distribution, meaning a normal distribution with a to be determined mean and variance. A linear regression depends on a set of independent variables that are linearly added, which means that the derivatives of $f(x_t;\beta_t)$ with respect to the parameters β do not depend on β. For a non-linear regression, derivatives of $f(x_t;\beta_t)$ with respect to the parameters β do depend on β. The values for the parameters of β are trained by minimizing a loss $L(\hat{y}, y)$. For a linear regression, the most used loss is the squared error $E_\theta(\theta-\hat{\theta})^2$. A linear regression is fast in computing while a non-linear regression is not. Non-linear regressions however, are able to capture more complex structures in the data set.

$$y_t = f(x_t;\beta)+\epsilon_t \tag{3.1}$$

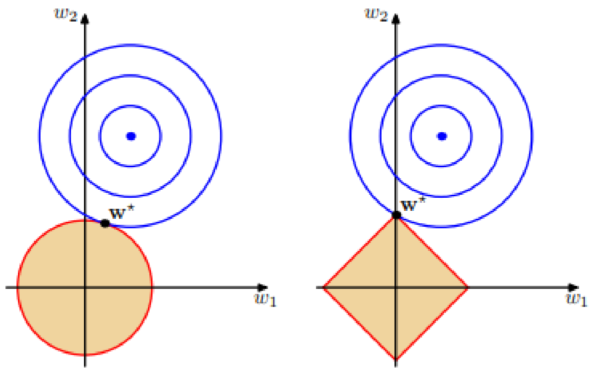


Figure 3.1: Visualization of regularization [12]. Contours of the unregularized error are shown in blue along with a constraint region on the weight, based on L_2 (left) and L_1 (right) regularization.

Regression types Linear regressions assuming Gaussian distributed errors have been discussed. More types of regressions exists, all with their (dis)advantages and specific use. Logistic regression is used when the target variable is binary. Using a sigmoid function, the probability of the value of the target variable can be computed. Multivariate regression is used when there are multiple inputs and outputs. For example, flight time can also be estimated alongside fuel consumption. Non-linear regressions are used when an input variable is non-linearly dependent on the target variable. The non-linear functions have to be stated before performing the regression. When many non-linear functions are tested, the input feature space becomes very large, increasing the chance of overfitting [81]. So a few non-linear functions should be tested or an optimization methods should be used as will be discussed in chapter 4.

[49] used multivariate adaptive polynomial spline regression (MARS) for fuel consumption estimation which is an algorithm for complex non-linear functions. The MARS is able to implement hinge functions in the regression, such that non-linear effects can be introduced. Due to the large number of functions possible, an optimization algorithm is used to determine a solution.

Regularization When using a high dimension of input variables, while only a few input-output data points exists, a high chance of overfitting is present. In this case, it appears that there are always correlations between the independent variables and the training data set. To overcome this issue, regularizes are used to reduce the overall weight size [12]. In the error function, an extra regularizer error is introduced, which penalizes high values of weights, shown in Equation 3.2. Here, E_{tot} is the total model error, E_D the regression error and E_w the regularizer error. This way, extremely large weights are reduces, which generalizes the resulting model. The most used regularizers are L_1 and L_2 regularizer using a extra loss $E_W(w) = |w|$ and $E_W(w) = |w|^2$, respectively. They are also called the Lasso regression and the Ridge Regression. Lasso makes it that many features will have a weight of zero and a few features will have a weight above one, which is demonstrated in Figure 3.1. The highest correlated features will have a non-zero weight, while others have a weight of zero. Using a Ridge Regression, all weights are lowered, but the method generally results in all weights being slightly above 0. High weight are penalized heavier then in the Lasso regression. The Lasso and Ridge regression technique's are useful for high dimensional data spaces, since unimportant input features will be discarded.

$$E_{tot} = E_D(\mathbf{w}) + \lambda E_W(\mathbf{w}) \tag{3.2}$$

Hyperparameters When performing a regression, one has to decide on the input features to use in the model and the functions to apply on these models. Technically, these are not hyperparameters, but will change the resulting model for different input combinations. For regularization, there is hyperparameter introduced, namely the parameter λ. However, the techniques use a hyperparameter, λ, which influences the result and so the accuracy of the prediction. When using all the available data to train and also test the model, a λ value of zero will usually results in the optimal value. Therefore, a test set needs to be introduced in order to test different values of λ.

3.1.2. Time series

If the dataset is sequential, time series may be used. Here, the target variable is defined at a time instant. For fuel consumption, one can state the fuel consumed by an aircraft upon landing. This value may be used as input data to predict the fuel consumption of a following flight. More generic, time series can be written as (First-Order) Difference Equations [38]. A difference equation is shown in Equation 3.3, where y_t is the target variable at time t, y_{t-1} the target variable at time $t-1$ and w_t an variable based on another function at time t. w_t may be the result from a linear regression. This way, data of previous time steps can be used in combination with a linear regression as shown in Equation 3.4. The advantage of using time series is that one can apply a function ϕ on the previous data points. This way more complex relationships can be trained.

$$y_t = \phi y_{t-1} + w_t \tag{3.3}$$

$$y_t = \phi y_{t-1} + f(x_t;\beta) + \epsilon_t \tag{3.4}$$

Two well-known time series are the Moving average (MA) model, the AutoRegressive model (AR) and a combination of the two, the AutoRegressive Integrated Moving Average (ARIMA) [27]. When using the MA model, one uses Equation 3.5 in order to incorporate the errors of previous estimations in the new estimation. Using an Exponential MA, the weights of the errors degrade exponentially. When using an AR model, Equation 3.6 is used. This time, the resulting values of previous time steps are used. The AR is similar to the difference equation shown in Equation 3.3. When only the previous time step is used, the model is described as ARIMA(1). Seasonal effects may be included using the value a fixed amount of time before the current time step. For fuel consumption, one can take 12 hours to measure daily effects and 7 days to measure weekly effects.

$$Y_t = \mu + \epsilon_t + \theta\epsilon_{t-1} + \theta_2\epsilon_{t-2} + ... \tag{3.5}$$

$$Y_t = c + \phi X_{t-1} + \phi_2 X_{t-2} + ... + \epsilon_t \tag{3.6}$$

Hyperparameters and regularization One can regularize time series similar to regularizing regression models. This way, the temporal dependency between data points is reduced. Choosing the amount of previous data-points can be seen as choosing a hyperparameter. The choice influences if seasonal effect are taken into account.

3.1.3. Reflection statistical methods

The first advantage of regressions are its simplicity in use and understanding. The method has been studied for a long time and widely. This makes it that many programs exist to perform a regression quick. When performing a regression, one needs to know its desired output, the number of parameters and input data. Furthermore, one can identify correlations between the input parameters and the output variable. This way, an analysis may be performed on the the effect of different input variables. Using the significance of input parameters, one can determine if an input is correlated to the output or not. Hao et al. [39] uses a correlation analysis to identify the relationship between different parameters and trip fuel consumption. Moreover, regressions can be made using few data points already. Using the *statsmodels* package in *phyton*, one is able to estimate parameters for the regressions and time series [78].

The limitation is that higher polynomial functions of a parameter or combinations of parameters are not used unless a user explicitly uses this as an input variable. Non-linear functions may be trained, but still one needs to define to possible functions. This way, possible correlations between (a combination of) input variables might not be found, which reduces the accuracy of the model. It is impossible to state all possible polynomials or combinations of parameters. Thereby, the computational speed will decrease significantly.

Regressions are used in trip fuel consumption research to indentify relationships, ([72], [73], [39]) or used as comparison or verification with other models ([48], [5]). Achenbach [5] used fuel consumption and flight time values of previous flights to estimate new fuel consumption. Both the value of one historical data point as the accumulative value of three flights is used as input parameter for the model. Using a time series however, more elegant solutions can be found. For this method one needs to define the time steps of data points. Precedent flights of a flight in the morning are further away then precedent flight of a flight in the evening. A proposed solution is to include the time past since the previous flight, either linearly or using another function, shown in Equation 3.7.

$$Y_t = c + \phi\frac{1}{(t-p)} * \phi y_p + f(x_t;\beta) + \epsilon_t + \theta\frac{1}{(t-p)}\epsilon_p \tag{3.7}$$

3.2. Machine learning techniques

Core machine learning techniques are discussed in the current section. Many alterations exist for specific applications. The core method will be described and if applicable, some extensions will be described as well.

3.2.1. Random forests

Classification and Regression Trees (CART) are introduced by Breiman in 1984 [20]. The methods divides the input space into regions. An example is shown in Figure 3.2. The decision tree splits the feature space in two, such that data points are separated by the decision boundary. The algorithm iteratively splits the rest of the input space and thereby creates decisions on the input variables. In the three, each node, or decision, is a numerical or categorical comparison on the input feature. The creation of nodes stops when no more splits could be made, or the maximum depth is reached in each branch. So in the end, the input feature space is split into subspaces. Each subspace may contain multiple datapoints.

Once the decision tree is created, new estimations on datapoints could be made using the decision tree. The decision rules will be applied on the input features to end up with a final subspace. If there are multiple other datapoints in this subspace, the average value or any other metric will be taken as final result. There are many possible ways to create a decision tree for a set of input parameters. So, when an algorithm is applied, many different decision trees may result.

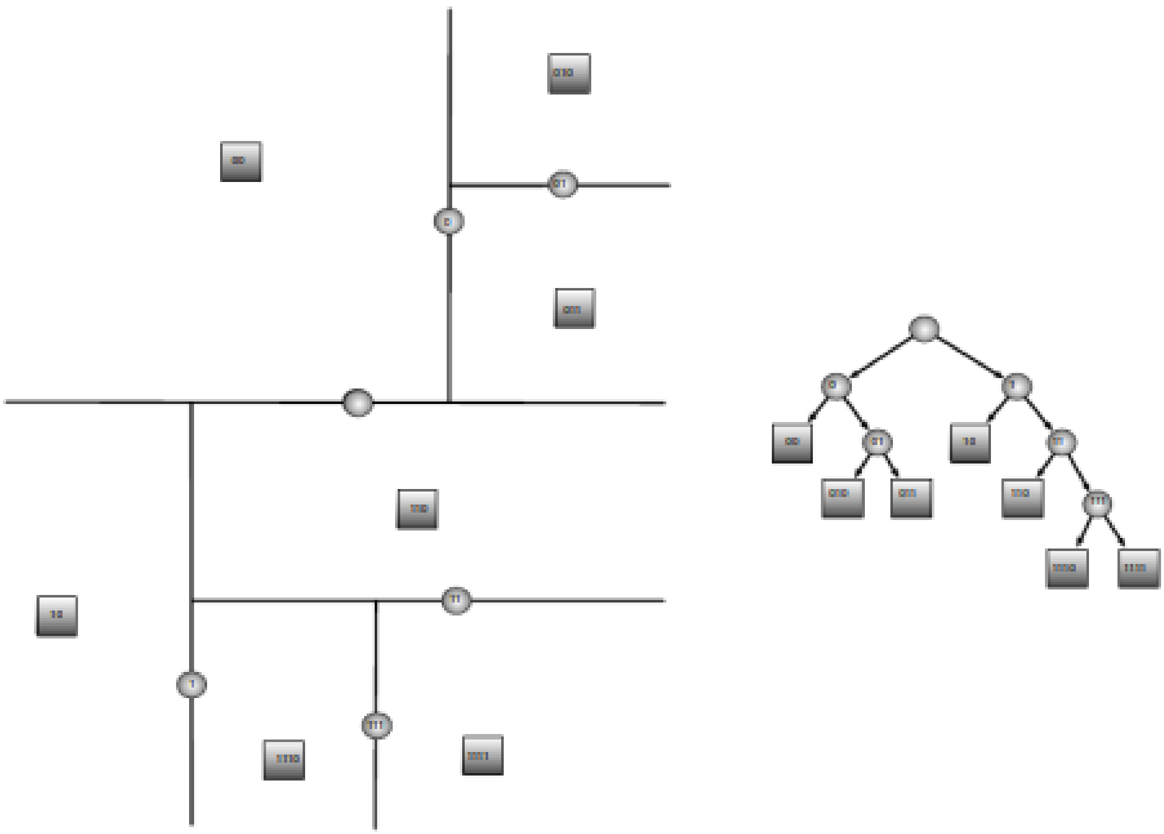


Figure 3.2: Decision tree visualization, from [17]

One can combine multiple decision trees into one model. This method is called random forests. It is essentially an ensemble learning method further discussed in **??**. Many trees are possible, so a large variation in the outcome of each tree is present. Using many different created trees, one can determine the outcome value based on a selection of the outputs of the decision trees.

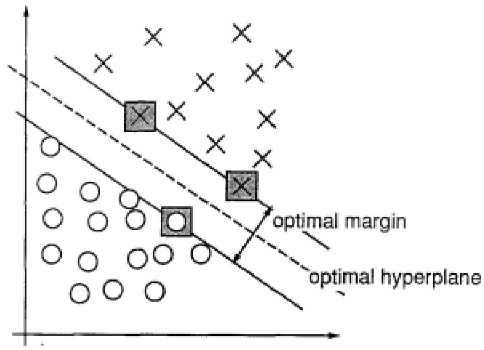


Figure 3.3: Support Vector Machine method, from [26].

Hyperparameters The hyperparameters determine the complexity of the model and the chance on overfitting. The max depth of the three determines the maximum amount of sequential decision to make. By limiting this value, the model generalizes better but limits the ability to model complex distributions. The minimum datapoint left in each subspace per split determines that outliers will not be split that easily. For random forests, the loss function determines the placement of decision boundaries. Gini loss and entropy loss are a often used for this [81].

Reflection Random forests have the ability to capture non-linear effects very well. Still, the implementation of the algorithm is relatively simple. Thereby, decision trees have the option the identify feature importance, which has a practical use in operations, similar to correlation values. The model is easy to interpret and explain. However, in high dimensional spaces, there is a high chance of overfitting. Thereby, parameter tuning will be required to get the best result. Using *scikit*, one is able to create regression trees in phyton.

Random forests have been used in trip fuel consumption estimation ([48], [5]). In both studies, the method resulted comparable to other machine learning methods.

3.2.2. Support Vector Regression

Boser, Guyon and Vapnik ,[19], [26], introduced support vector machines (SVM's) in 1992. The algorithm is designed for classification problems and works well in high dimensional data. The algorithm selects a hyper plane in the input space that maximizes the separation between two classes. The SVM is visualized in Figure 3.3. A characteristic of SVM's is that only a few datapoints are present on the decision boundary. The advantage of this is that if new datapoints are added, one only needs to check if the new datapoint is in between the other support vectors, to see if the model needs to be updated. The method works by minimizing the weights, shown in Equation 3.8, while subjecting to the constraints in shown in Equation 3.9.

$$\Phi = w \cdot w \tag{3.8}$$

$$y_i(x_i \cdot w + b) \geq 1,\ i = 1, ..., l \tag{3.9}$$

When the classes are not linearly separable, a kernel function may be used to transform the input feature space. This is shown in Figure 3.4. Various kernels exists, but the kernel to specify polynomials of the input features is shown in Equation 3.10. Other often used kernels are the Gaussian Kernel and the Sigmoid Kernel.

$$K(u, v) = \phi(u) \cdot \phi(v) = (u \cdot v + 1)^d \tag{3.10}$$

If the classes are not separable, slack is introduced. Wrong classified data points will have a penalty. The new objective function is shown in Equation 3.11 and the constraints are shown in Equation 3.12 and Equation 3.13

$$\Phi = w \cdot w + C(\sum_{i=1}^{l} \xi_i)^k,\ k > 1 \tag{3.11}$$

$$y_i(x_i \cdot w + b) \geq 1 - \xi_i,\ i = 1, ..., l \tag{3.12}$$

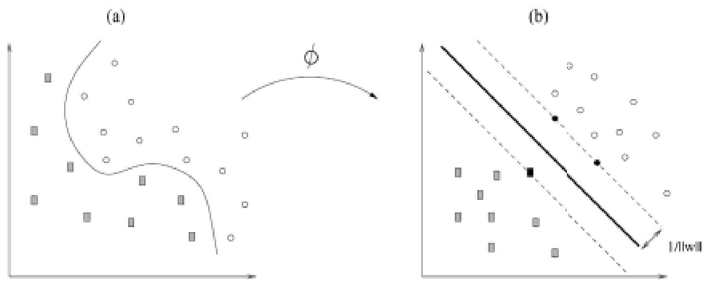


Figure 3.4: Kernel trick for support vector machines, from [75].

$$\xi_i \geq 0,\ i = 1, ..., l \tag{3.13}$$

Support vectors can also be used for regression. An error allowance, ϵ, on the target variable is introduced. Then the algorithm selects the hyperplane containing the most data points within the error range. The datapoints at the decision boundaries are still the support vectors. Thereby, the weights are minimized. The objective function results in Equation 3.14, subject to Equation 3.15.

$$\Phi = w \cdot w + C(\sum_{i=1}^{l} |\xi_i|)^k,\ k > 1 \tag{3.14}$$

vi

$$y_i(x_i \cdot w + b) \geq \epsilon + |\xi_i|,,\ i = 1, ..., l \tag{3.15}$$

Hyperparameters For SVR, the hyperparameters to tune are the kernels used, ϵ and C. The kernel determines the complexity of decision boundaries the SVR method can handle. By increasing ϵ, less data points around the mean affect the hyperplane orientation. Increasing C increases the effect of outliers. If C approaches infinity, a normal regression is found, neglecting some datapoints centered around the mean.

Reflection The SVR method is robust to outliers. This is actually undesired, since outliers in fuel consumption actually affect the safety of the flight. However, one is also able to detect outliers in the dataset. This might be useful to indicate risky scenario's. SVM's can be a useful model to detect outliers, but will be less useful to use as model to predict fuel consumption. An advantage is that confidence bounds can be produce relatively easily. The SVR method works well in high dimensional input spaces with little training data. The input dimension will be large in the current research, which shows potential. SVR's or SVM's have not been used in studies on data-driven trip fuel consumption estimations. SVM's however, they have been used in instantaneous fuel consumption ([88], [37], and traffic congestion prediction [94]. In the study on traffic congestion prediction, classes of congestion were first created, in order to separate them using SVM's.

3.2.3. K-nearest neighbor

The K-nearest neighbor (KNN) is a widely used non-parametric estimation method for classification and regression [17]. For regression, one determines the closest input vectors to the current input vector. The definition of close can be changed according to the desire of the user. Common used measures are the Euclidean distance or the city block/Manhattan distance, shown in Equation 3.16 and Equation 3.17, respectively. By scaling of input features, the distance between points in that feature direction will be reduced. This way, the importance of important features can be enlarged. The estimated output of the new datapoint is the average of the outputs of the closest input data vectors.

$$d_{x,y} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2)} \tag{3.16}$$

$$d_{x,y} = \sum_{i=1}^{n} (x_i - y_i) \tag{3.17}$$

Reflection The method is a simple model which is very easy to implement. However, KNN is sensitive to scaling of input features. In high dimensional input spaces this is undesired. Unimportant features may influence the outcome significantly. One should first perform solid feature extraction before the method can be useful. KNN has been used in trip fuel consumption research [48]. In this research, it performed the worst out of all other models, including the FPS. Probably, unimportant factors affected the model too much.

3.2.4. Model ensembling

The previously described statistical and machine learning methods all have their advantages and disadvantages. One can combine estimators in order to get a better final estimators. When combining estimators, the estimators are called 'weak learners'. Three often used methods are bagging, stacking and boosting. Introduced by Breiman (1996) [21], bootstrap aggregating (Bagging) was intended to reduce generalization error by individual algorithms. By averaging the outcomes of different estimators, which usually have different errors on the test set, the beneficial factors of different estimators are used. Besides averaging, the maximum, the median or any other heuristic for combining the estimators may be applied. Typically, different samples of the data set are used to train a model, and the models are then combined in an ensemble [17]. An example is shown in Figure 3.5, where different classifiers for the number eight are trained on different data sets. By combining the two models, the ensemble recognizes that two circles are needed to identify an eight. Random forests is an example of bagging. When applying a bagging algorithm, different strategies for selecting new models can be used, similar to feature selection, described in chapter 4.

When different types of estimators are used, the method is called stacking. For the rest, the method is similar to the bagging method.

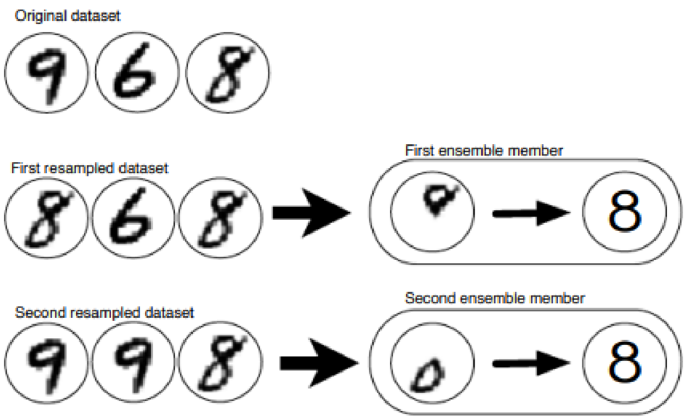


Figure 3.5: Bagging visualized, from [17]

When using bagging or stacking, learners have the same weight when the averaging rule is applied. With boosting however, multiple weak learners only have the same weight at the start of the algorithm, Freud, 1996, [34]. Thereafter, weights of weak learners are updated based on a heuristic. In AdaBoost, datapoints that have a large error, gain an increased importance. Weak learners that estimate the datapoints with high importance relatively well, gain an increased weight. The weight of all other learners is reduced. This way, the datapoints have a lower error in the next iteration. Gradient boosting uses the gradient on the loss function of all weak learners to increase the learners with the highest negative gradient. When gradient boosting is used, often random forests are used as weak learner.
When combining classifiers, the worst outcome is the same error as the best weak learner. Usually however, combining estimators increases the performance of estimators. That is why combining algorithms usually outperform other algorithms, and research suggest to not use combining when introducing a new model, since more computational power often results in a higher performance [17].

Reflection Model ensembles have a great potential to increase the accuracy on trip fuel consumption. However, one should be careful by increasing the computational cost more and more in order to achieve higher accuracy. This way, new changes performed in the research can not properly be compared to other mod-

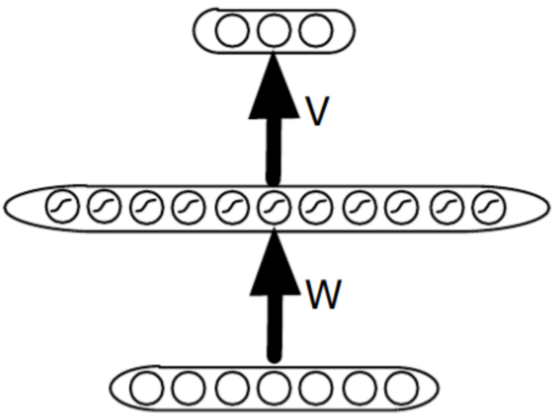


Figure 3.6: Shallow MLP, from [17]

els. Ensemble methods have been used to model fuel consumption [48] or as verification [96]. Compared to other machine learning techniques, ensemble methods improve the accuracy. However, ensembling simple machine learning algorithm have been found to perform worse than an complex neural network. Ensembling neural networks could increase the accuracy even further, but is computationally very expensive.

3.3. Neural Networks

Feedforward supervised neural networks, or MulitLayer Perceptrons (MLP) were among the first non-linear machine learning algorithm introduced by Rumelhart (1986) [71]. A schematic representation of a neural network is shown in Figure 3.6. Input features are connected to nodes by weight and biases. In a hidden layer node, an activation function is used. This activation function introduces non-linear effects. Usually, a sigmoid function is used that maps the resulting feature between 0 and 1.

$$\mathbf{H} = \mathbf{X}\mathbf{W}^{(1)} + \mathbf{b}^{(1)} \tag{3.18}$$

$$\mathbf{O} = \mathbf{H}\mathbf{W}^{(2)} + \mathbf{b}^{(2)} \tag{3.19}$$

$$\mathbf{H} = (\mathbf{X}\mathbf{W}^{(1)} + \mathbf{b}^{(1)})\mathbf{W}^{(2)} + \mathbf{b}^{(2)} \tag{3.20}$$

Key is that the weight and biases of the MLP can be updated efficiently using the Backpropagation algorithm [40]. Using the derivative of the loss function - usually least squares - with respect to the weights and biases, the weights are updated. Common practice is that not all available data is used to determine the average error, but batches. This method is called Stochastic Gradient Descent. One iteration is called an epoch. By training on batches, each epoch in computable faster. The model is trained until new epochs do not reduce the error on the validation set anymore or until the maximum number of epochs has been performed.

BP optimizers Since the derivatives of sigmoid functions can be very low for large positive or negative values, the SGD algorithm converges slow. To overcome this problem, optimizers on SGD are used. Different optimizers are shown below.

- **SGD with momentum**, [69], is the first improved optimzer to can be applied to SGD is. Using this optimizer, the average update of previous time steps is considered to update the weights to the next step. This way, the SGD gets accelerated towards a local optimum. Variations on the method exists, including SDG with Nesterov Momentum. Now, for each time step, the weight get updated as if the averaged momentum of previous time steps is already included
- **Adagrad**, [29], accelerates update steps for parameters that are updated unfrequent and reduces update steps for frequent parameters. This is acheieved by diving the update steps by the squared gradients of

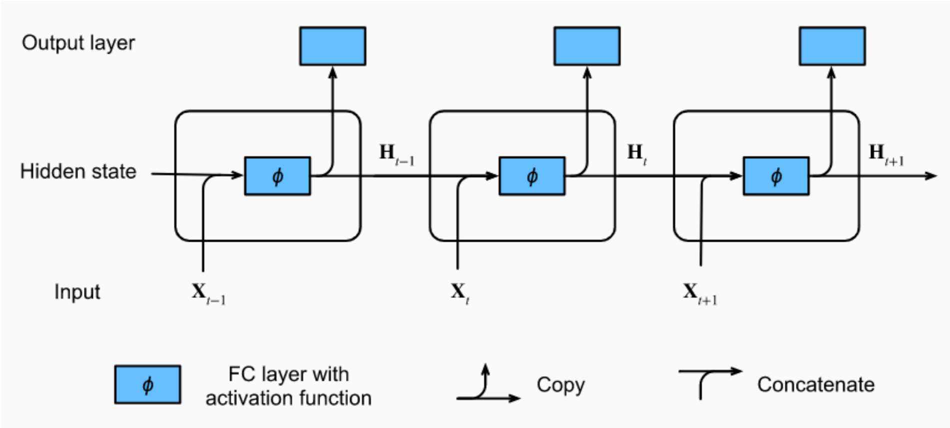


Figure 3.7: Recurrent Neural Network, from [93]

previous update steps. Adadelta, [92] in an extenstion to Adagrad to reduce its decreasing learning rate, by limiting the value of past squared gradients.

- **Root Mean Square prop (RMSprop)** is an extension to the Resilient Backpropagation algorthm (Rprop), [45]. Rprop averages the sign of the gradient of preivous time steps, such that oscillations are reduced. RMSprop improves Rporp by dividing the Rprop part in the update by the moving average of the squared gradients of previous time steps. RMSprop shows similarities to Adadelta, by using the squared gradients of previous time steps [70].
- Combining the RMSProp and SDG with Momentum uses the best of both and is called Adam, introduced by [54]. Adam is a widely used optimizer [70].

Even though it is explainable what the differences are for the optimizers, the advantages and disadvantages are hard to mention [70]. Luckily, all optimizers are easily usable by using the *keras* package in *phyton*. Usually, some optimizers take more time then others, but given enough computational power, they will often converge to the same solution. Therefore, using Adam as optimizer will suffice for the current research.

3.3.1. Recurrent neural networks

For neural networks, it is possible to perform sequential modelling. Similar to time series modelling, the target variable is dependent on the outcome of the previous datapoints. A schematic representation of a recurrent neural network (RNN) is shown in Figure 3.7. The value of the hidden layer and the value of the output layer is calculated using Equation 3.21 and Equation 3.22. Using RNN's, one is able to capture historical information while the number of parameters does not increase by increasing the number of sequential datapoints.

$$\mathbf{H_t} = \phi(\mathbf{X_t W_{xh}}) + \mathbf{H_{t-1} W_{hh}} + \mathbf{b_h} \tag{3.21}$$

$$\mathbf{O_t} = \mathbf{H_t W_{hq}} + \mathbf{b_q} \tag{3.22}$$

Vanishing gradients is an issue with RNN's. Due to averaging of previous time steps, important information of historical data gets lost over time. To overcome this issue, Gated Recurrent Units (GRU) and Long-Short Term Memory RNNs (LSTM) are created. Both methods show similarities. GRU uses less training parameters and so is computationally faster, while LSTM has the potential to get better performance [93]. Since high accuracy is preferred, only the LSTM method will be described. The LSTM is introduced by Hochreiter in 1997 [41]. The LSTM introduces a memory cell as a special hidden state designed to contain information from previous time steps. Using mathematical gates, decisions are made on which information to store in the memorycell. A schematic representation of a LSTM is shown in Figure 3.8. Three different gates are used to decide if the memory cell will be updated, will be used for outputs, or will be cleared. Thereby, a candidate memory cell $\mathbf{\tilde{C}_t}$ is generated which may be added to the memory cell, depending on the output of the input

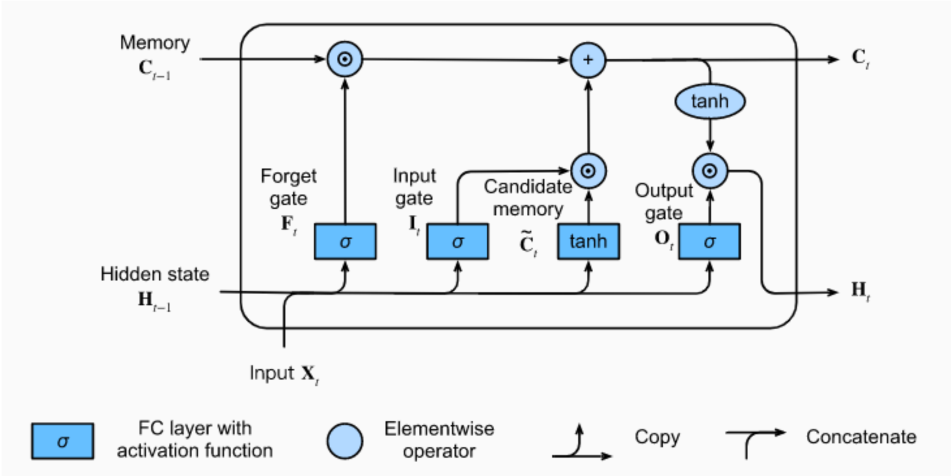


Figure 3.8: Long-Short Term Memory RNN, from [93]

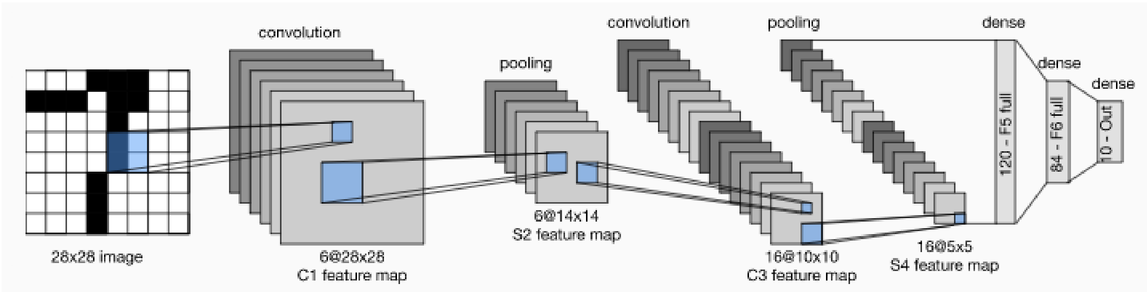


Figure 3.9: Convolutional Neural Network, from [93]

gate. Finally, the result of the output gate gets combined with the result from the memory cell, to get the new output.

3.3.2. Convolutional neural networks

MLP's connect all the input data to all hidden states. When there is prior knowledge in the input data, on can take advantage of that. For example when an image is shown as data sample, one might assume that neighbouring pixels contain more information then sparse data sets. By introducing Convulutional Neural Networks (CNN), one can map a range of input features to the next hidden node. By using a kernel of fixed size, one can determine the amount of input feature that are considered for the next feature. This way, way less parameters have to be optimized, while the accuracy remains high. An example of a CNN network is shown in Figure 3.9. Pooling refers to the action of reducing input feature to a single input feature. From typically four input features, the average or maximum can be taken. This way the input space is reduced, while useful information is stored.

CNN's are useful for input spaces with prior information. This is not directly applicable to trip fuel consumption, but will be useful for congestion and weather modelling described in chapter 4. In this case, data is ordered in a physical space, such that neighbouring elements contain information for features.

3.3.3. Regularization

Neural networks are able to identify complex structures in datasets. The large complexity, or flexibility, makes the model is sensitive to overfitting. Regularization methods are therefore of upmost importance to make a general model that performs well in the test set.

Weight decay is the first regularization method, similar to L_1 or L_2 regularization. In the loss function, the weight are added, such that when training the model, the size of the weights are reduced. Closely related to weight decay are Bayesian Neural Networks (BNN). Using BNN's, one assumes a probability distribution over all weights. The probability distribution is usually high around 0. By multiplying the a priori distribution with the updated weights, a regularized solution is found.

More regularization methods are based on the idea of adding noise to the model, such the model becomes robust to perturbations. Dropout is a commonly used method, where certain nodes are removed from the model. This way, the model does not become reliant on a few nodes, but is trained to capture multiple structural effects. Besides removing nodes, one can also add noise to the weights and biases in the network. Applying noise to input data also reduces the effect of overfitting. Once can select input data to add noise to, for example outliers in the output set.

Using batch normalization, the means and variance of the current batch is used to alter updates of the weights. This way, large updates of weights are limited.

3.3.4. Architecture optimization

The architecture of a neural network can have many possible forms. The amount of hidden layers, the amount of nodes per hidden layer, the type of activation function are just a few of the hyperparameters. Too little networks may not capture al nonlinear effects, while too large networks converge slowly and may overfit to the data. When building a neural network, it is hard to estimate which architecture will result in the most accurate model. To overcome this problem, optimization methods for the architecture of the network are designed. By measuring the importance of certain nodes or connections, the can be removed or new ones will be added.

The cascade correlation learning algorithm added hidden units sequentially to the network. One can also create multiple networks and perform an optimization using the results. Optimization is a widely studied subject [87] and so, many strategies have been implemented. Genetic algorithms are used to create beteer networks from existing networks. Optimization methods have the advantage that a computer does all the work, so the workload it little and no expertise is required, but can computationally be very expensive. Khan [53] used a Constructive neural network for fuel consumption modelling. He highlighted the limitations of the BPNN. The slow learning speed and expertise requirement were argued to be the largest limitations. Thereby, the CNN has potential for better generalization performance on high dimensional data.

3.4. Conclusion supervised methods

Different supervised learning methods have been discussed in the current chapter. A summary of the models is shown in Table 3.1 and a comparison is made in Table 3.2. For the comparison, time series and the RNN method are similar to the regression and NN method, respectively. Using the sequential data can be seen as implementing extra input features. As discussed in subsection 3.1.2, creating sequential data in trip fuel consumption estimations has its challenges, but may increase accuracy substantially. Therefore, the fuel consumption of the last landed flight upon trip fuel estimations is used as precedent data, including a penalty for flights further back. The CNN is not included because of its limited use in the model. Ensembling can be applied to every method to increase the accuracy, so this method is not included, assuming it can be applied on all methods in order to improve performance.

The criteria are ranked from 1 to , indicating 'good' or 'bad' for the application of the model. Per criteria, high or low may be considered good.

- **Model flexibility** The model flexibility determines the models ability to capture complex, non-linear effects. A higher flexibility is advantageous, since there is a lower chance that structures in the data set will be missed. So, an increased flexibility might result in an increased accuracy of the model. The increased flexibility brings an increased chance in overfitting. For now, a well regularized model is assumed.

- **Implementation complexity** The time span of the current project is limited. So, a complex to implement model, containing many hyperparameters is undesired. This way, less time on other research objectives can be spent.

- **Outlier prediction** This criterion corresponds to the inclusion of outliers. For fuel loading decisions, outlier predictions are important regarding the safety of the aircraft. Different models have different potential to estimate the values of outliers .

- **High dimensional data handling** The amount of input features used determines the dimension of the input data. For the current research, high dimensional input spaces are expected, which makes high dimensional data handling of the model important.

Table 3.1: Summary of the supervised learning methods

	Supervised learning category	Used in fuel consumption modelling	Characteristics
-2*Statistical methods	Regression	Yes	Simple Easy to understand
	Time series	Partially	Temporal patterns can be identified
-3* Machine learning methods	RF	Yes	Good at capturing non-linear effects Uses stacking algorithms
	SVM	Yes	Performs well in high-dimensionality Generalizes outliers Good at iterating using new input data
	KNN	Yes	Simple, non-parametric model Sensitive to scales
-3* Neural Networks	NN	Yes	Able to capture non-linear effects well Difficult to interpret Many hyperparameters
	RNN	Yes	Able to capture temporal effects Long term memory possibility
	CNN	No	Takes advantage of prior knowledge on data
Combinations	Ensembling	Yes	Uses the best characteristics of other methods Given enough computational power, will increase the performance of single models

Table 3.2: Comparison of the supervised learning methods

	Regression	**KNN**	**SVM**	**RF**	**NN**
Model flexibility	1	2	2	2	3
Implementation complexity	3	2	2	2	1
Outlier prediction	2	1	2	1	3
Integration with feature extraction	2	1	2	2	3
High dimensional data handling	2	1	3	2	2
Total	**10**	**7**	**11**	**9**	**12**

- **Integration with feature engineering** Feature engineering, described in chapter 4, will be performed to obtain useful features for the high dimensional weather and air traffic data features. Some methods allow for a ranking in input features, which can be used for feature selection. Some methods can be integrated in feature extraction models, which is desirable for high accuracy.

Resulting from the trade-off, the neural networks scores the best. The difference with SVM's and regression, however, is not large. Inspecting the results of the trade-off, it is visible that model flexibility and implementation complexity are inversely related in the current trade-off. So, combining the two criteria, all models score the same. For outlier prediction, KNN and RF score bad since future predictions are always based on historical data points. This way, extreme input feature values will result in a comparable result to histrorical data, which is not desired. Regression and SVM score better in outlier prediction, since an increasing input value will result in an increasing final estimation. For neural networks, abnormal combinations of inputs will be detected the best, as will be described for autoencoders in subsection 4.2.3. Therefore, NN's score the highest. Using regression, SVM and RF, one is able to visualize a ranking in important features. This way, the same model can be used for feature selection as trip fuel consumption estimations. This is not possible for KNN. NN based feature extraction methods can be integrated with the trip fuel consumption estimation NN's. This way, pretrained feature extraction models can be updated, such that better features for the estimation model will be present. If the total amount of flight will be divided per OD pair, day ect., the number of datapoints becomes scarce. SVM's excell in modelling high dimensional dataspaces. KNN performs the worst, since the model is sensitive to all input features. Other models perform well, depending on the quality of regularization.

So, neural networks will be used as resulting model. The model flexibility, outlier prediction ability, and potential in integration with feature extraction make it an well suited method for the current research. Because of the implementation complexity however, intermediate results are harder to obtain. Therefore, the regression method will be used as well as preliminary analysis model and verification method. The simplicity of regressions methods combined with the high accuracy, make it a good verification method. Using stacking on Lasso models result in a high accuracy in literature [49], while the model is regularized well. SVM's scored similar to NN's. Interesting is that SVM's have not been used for data-driven trip fuel consumption estimation yet. If the time-span allows, a simple version of a SVM can be trained, in order to test the difference in accuracy.

3.5. Loss function

The goal of this paper is to increase accuracy in trip fuel consumption estimations, to gain trust of flight dispatchers, such that discretionary fuel loading can be reduced. Keeping this goal in mind, is the standard squared error as loss function not the most applicable method of training a model. Kang (2021) [50] used quantile regression methods to estimate statistical fuel consumption. When performing quantile regression, one is changing to loss function, as described by Koenker et al. (2001) [55]. The loss for underestimating a target value can be penalized harder then overestimating a target value. The resulting optimization problem is shown in Equation 3.23. ρ_τ represents the quantile to estimate and $(y_i - \xi(x_i, \beta))$ represents the error in the estimation of the model. The function of ρ_τ is visualized in Figure 3.10.

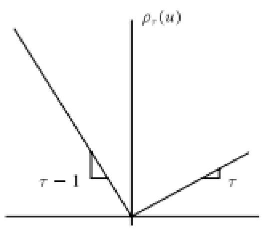


Figure 3.10: Function of ρ_τ, where absolute values are tilted, from [55]

$$\min_{\beta \in \mathbb{R}} \sum_{i-1}^{n} \rho_\tau(y_i - \xi(x_i, \beta)) \tag{3.23}$$

Quantile regression can be desireable in trip fuel consumption estimations. Still, there are limitations. The desired quantile will be around 99.9% for the current model. This way, only 0.1% of the data is above the quantile line, leading to a major decrease in data availability. This might lead to overfitting of the model. Kang (2021) [50] found that extrapolating the 95% quantile model performed better then the 99% model, probably caused by overfitting effects. So, multiple quantile shoudl be used for trained, to identify the differences. Boosting the different models will probably increase the accuracy compared to every single model. For the current research, least squared will be used to train the models. This method can be more easily verified. In the benefit assesment phase, different quantile models can be trained to compare the difference in outcome.

4

Feature engineering

This chapter discusses feature engineering or feature design. Data-driven methods use high dimensional input spaces, which might negatively affect the performance of the model. In general, when using a statistical or machine learning model, one needs to define the input structure more detailed than for a neural network model. Still, one needs to define the model architecture for a neural network model. One can decide to merge features into a smaller feature before passing the information to the rest of the model. In this research, a statistical model and a neural network will be implemented. So, for both models it is important how to handle the input data. En-route weather and traffic data is highly dimensional, while a general state of weather and traffic uncertainty and severity is desired. The goal of this chapter is to investigate methods to engineer features and to apply this methods on the current research. Feature selection and feature extraction methods will be discussed in section 4.1 and section 4.2, respectively. Then, feature engineering methods will be applied on weather and traffic data in section 4.4

4.1. Feature selection

When dealing with a high dimensional input feature space, feature selection is a relatively simple method to reduce the dimensionality [12]. Good features - that contain information on the target variable - will be selected, while bad features will not be selected. Despite the simple idea of feature selection, many approaches on the selection exists. Feature selection methods used for classification problems will not be discussed. When performing feature selection, on can determine a fixed maximum of parameters, or define a stopping criterion for adding new features. Different methods are described below. They are grouped in filter methods - where selection criteria are used to select the best filters -, wrapper methods - which are used to test the performance of feature combinations on a test set to select the best combination - and embedded methods - which select good features and output a result as well -. Testing all possible input combinations for the best result is a NP-hard problem, which is computationally expensive for high dimensional input spaces. Therefore, optimization methods are used in the wrapper method.

- **Filter methods**
    - **Correlation analysis** The pearson's correlation coefficient gives information on the correlation between input variables and the target variable. By computing the correlation between all input features and the target variable. The highest correlated features can be selected. When multiple features are selected, the correlation within the input feature space should be low. By itereratively adding the highest correlated features, one can also calculate the correlation to the existing features before selecting it.
    - **Mutual information** When information of a random variable or input feature reduces uncertainty on the target variable, this is called mutual information. Similar to the correlation analysis, on can select input features with the highest mutual information.
    - **Fishers Score** Using the log-likelihood of input features with respect to the target variable, the fishers score can be determined. Similar to the correlation analysis, on can select input features with the highest Fishers score.

- **Wrapper methods**
  - **Forward Selection** All input features are tested as single input in the model, and the best performing feature is selected. The next iteration, all input feature are combined with the previously selected feature to select the next best feature.
  - **Backward Elimination** Now, all input feature combinations with one feature missing are tested. The combination with the best performance is selected, while the feature missing is removed. This will be iterated. When many selected input features are desired, this algorithm is quicker then forward selection.
  - **Algorithmic search** Many more optimization methods on selecting features exists. The improve the accruacy of the selectin, but increase computatinal costs. Forward-backward selectin, combining the previous two methods, genetic algorithms and particle swarm optimization methods are popular methods.
- **Embedded methods**
  - **Lasso** The Lasso or L_2 regularizer, reduces weights of input features to zero. This way, a set of remaining input features is present which are the selected features. By changing the regularizer parameter, λ, the amount of input features can be changed.
  - **Relative importance** Using other machine learning methods like random forests, one can determine the relative importance of each feature. Similar as using the correlation method, one can select the features.

4.2. Feature extraction

Using feature extraction, one combines input features into a new created feature. This way, information from multiple input features can be used in one single input feature. Different methods for feature extraction have been used.

4.2.1. PCA

The unsupervised method Principal Component Analysis (PCA) or the supervised method Linear Discriminant Analysis are often used linear feature extraction methods [12]. Using this method, the vector along which the variance of the input space is the largest, is selected. This method is sensitive to scaling. Fisher extraction is also a often used feature extraction method. Often, a non-linear feature extraction method is preferred to in order to identify non-linear effects in the feature space. One method for this the kernel PCA, or KPCA. Using this method, a kernel is used to transform the data set. Similar to using kernels in the SVM, important features can be extracted.

4.2.2. Neural Networks

As mentioned in the introduction, a neural network already applies feature extraction. The values of the hidden nodes after activation are representation of the input feature space. By tuning the number of nodes in the hidden layers, the amount of features can be determined. If multiple layers are used, non-linear effect can be captured in the features as well. CNNs using pooling combine input features in a single feature based on the geometrical location of the input feature space.

4.2.3. Autoencoders

A new type of neural networks are autoencoders. These neural networks try to replicate the input features using a neural network with less hidden nodes then input features. This way, important characteristics of the dataset are encoded in the code/bottleneck feature layer, and decoded into the replication of the input feature space. This method is unsupervised since no target variable is needed expect for the input features. A high dimensional dataset can be reduced in size first, before incorporating it in the neural network or other classifier. Autoencoders can be combined with RNN's ans LSTM's, where a sequential dataset is being replicated.

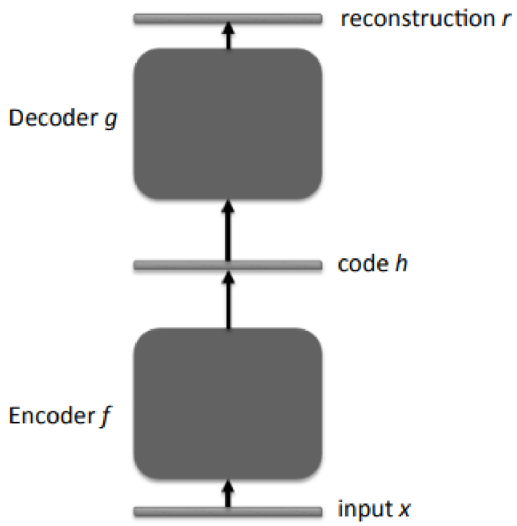


Figure 4.1: Schematic representation of an Autoencoder, from [17]

Autoencoders have been used in studies on traffic flow classification [57]. The autoencoder is able to capture complex effects in the input feature space and compress it to less features. Autoencoders have been used without [95] and in combination with LSTM's for congestion prediction [89]. The combination works well for reducing features and predicting future states. Autoencoders are also used for anomaly detection [68]. Different input vectors can be tested in the autoencoder to compute the resulting error. High errors reflect abnormal input states. This might be useful to investigate difference in normal and abnormal states for air traffic congestion.

4.3. Clustering

Clustering is a form of unsupervised learning and a widely used method with many variations [81]. The goal of clustering is to group similar datapoint together in a group. Input data can be clustered as is done by Kang (2018) [49], in order to indentify similar OD pairs in the dataset. Input feature can be combined this way, reducing the chance of overfitting. In order to identify operational air route and congestion points using air traffic data, the most used regions in the airspace should be grouped. To do this, cluster methods can be used. For air traffic data, the similarity can be geographical location and heading of the aircraft. Not all methods will be discussed extensively in the current chapter, but the major types of clustering algorithms will be briefly discussed. Thereafter, the DBSCAN method will be discussed. This method has been used by Lin (2021) [58] in order to identify congestion points using air traffic data, which is similar to the current research. DBSCAN, introduced by Ester (1996) [30] showed promising results in traffic clustering [25].

Using hierarchical clustering methods, all datapoints are put in cluster iteratively [81]. At the beginning of a hierarchical clustering algorithm, all datapoints are its own cluster. Then, the two most similar objects are clustered in a cluster. This similarity can be for example distance between the datapoints. Then the following two datapoints/cluster will be clustered. To cluster an existing cluster, the mean value of the datapoints can be used, or the lowest/highest value within the cluster. This is repeated until all datapoints are within one single cluster. When the algorithm is completed, a similarity dendrogram can be created, visualising the similarity per created cluster, shown in Figure 4.2. Using such a graph, an appropriate number of clusters can be used. A large jump means the combining of dissimilar clusters/datapoints.

The hierarchical method can be costly for large data sets, since in each iterations, the similarity between all datapoints needs to be computed. K-means clustering is an often used partitional clustering algorithm [12]. In this method, a fixed number of clusters is used as input. Then, on initialization, random means are assigned to each cluster. Each datapoint is clustered to the closest cluster. Then, the means of the clusters are updated using all the datapoints in the cluster. This is iterated until the means do not change anymore. Because of the random initialization, results of the K-means algorithm may vary. Therefore, multiple runs

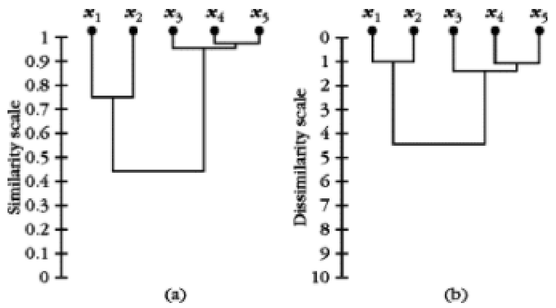


Figure 4.2: A similarity dendrogram, from [81]

should be used in order to select appropriate clusters. The algorithm can be performed for different values of K, such that resulting dissimilarities can be compared. An increasing number of cluster always reduces dissimilarity within the cluster. So, one shoud critically reflect on the reduction in dissimilarity per cluster in order to obtain a good amount of clusters.

DBSCAN is a commonly used clustering method, also in traffic clustering applications. The method clusters datapoints that are densely packed with a lot of closely neighbouring datapoints. Outliers are labeled as noise. The range of close neighbours, indicated with ϵ should be identified along with the number of points to form a dense region. All datapoints are compared with the determined dense packing criterion. These datapoints become part of a cluster, along with all datapoints within its neighbouring range. Then, combine (centre) cluster datapoints that are close to each other into the same cluster. When the algorithm is done, all densely packed regions are identified. For traffic applications, this is useful to identify air route, since air routes contain densely packed aircraft.

4.4. En-route feature modelling

Using the feature engineering methods described in this chapter, approaches per feature item can be described. Flight specific data, airport data and date-time data will also be used, without a personalized feature engineering method. The feature engineering methods for weather forecasts, en-route traffic and route information will be described in the current section.

4.4.1. Weather forecasts

Weather forecasts are different for different geometrical locations and contain many features. Wind speed, wind direction, temperature and humidity are a few examples that very per altitude. But cloud ceiling, chance on rainfall and chance on thunderstorms, are included in weather forecasts for all altitudes above a geometrical location. So, two problems can be identified when incorporating en-route weather forecasts in trip fuel consumption estimations. The first is, which features of a forecast should be considered? The second is, how many weather forecasts along the route should be included and how can all the weather forecasts be combined? For both questions, the previously discussed feature engineering methods are useful. In the current research, historical data from European Centre for Medium-Range Weather Forecasts (ECMWF) will be used. The forecasts include wind speed and direction, temperature and humidity. Terminal Area Forecasts (TAF's) broadcasted by airports contain information on bad weather. Although the TAF is created for low level altitudes, information on for example thunderstorms is still useful for higher altitudes. So, TAF's will be included as well in the weather forecasts. TAF forecast include wind visibility, rain or snow indicators, sky condition, probability on thunderstorm or other special events and forecast change indicators. TAF's are only present at airport locations unlike the better distributed forecasts of ECMWF. The closest TAF should be considered to get the weather prediction at a certain point. The different input features from the weather forecasts could be correlated to trip fuel consumption, or contain no information on trip fuel consumption.

Therefore, feature selection will be performed by ranking all the features based on its correlation coefficient. The correlation ranking works well with statistical methods. By using this, non important features can be removed. Thereafter, manual feature extraction is performed by multiplying forecast change indicators by the other features, to see a possible increase in prediction accuracy. All significant correlated features will be used in the statistical model and the neural network model. The neural network extract useful features itself, while the statistical method does not. If, in the resulting statistical model, the en-route weather forecasts do not contribute to the outcome significantly, features need to be combined using LDA. For the ECMWF features, all features will be used, since all may have an effect on instantaneous fuel consumption.

The second problem is how to use different weather forecasts along the route. A route has many waypoints and so many weather forecast input. Using them all may result in overfitting. Thereby, one can use weather forecasts at a waypoint, but maybe rerouting will take place such that neighbouring weather forecasts should be used as well. Moreover, ECMWF forecasts are used for different heights. For the statistical method, one can determine averages and standard deviations for all features using measurement points along the route. This way, the number of input features is limited. Potentially, the route may be divided into a number of segments - for example 3 - to be able to incorporate varying weather along the route. TAF's are provided at airports only, which does not coincide with the route often. So, the closest TAF at a certain point along the route should be considered, for a fixed number of points. Neural networks have a high potential for this problem. Both CNN's as autoencoders can be trained on such a scenario, where high dimensional data should be compressed into less features. For CNN's, one can use a fixed number of points along the route as center for a kernel, and use neighbouring weather forecasts as inputs for the kernel. However, this method has issues with including TAF forecasts. The fixed sizes of kernels does not work well with the varying locations of TAF's.

Using autoencoders, one can decide on the number of TAF's and ECMWF forecasts to include. Then, using a determined amount of features in the hidden layer, the feature information can be compressed in less features. One can first apply this method for ECMWF forecasts at different heights and locations near the planned route, before including these compressed features in another feature compressor for the entire route. The number of features should be tested for different numbers of weather forecasts to see the changes in accuracy. The pre-trained autoencoder features are useful to put into the statistical method or the neural network. Once implemented however, the weights of the first layers in the autoencoder should be trainable in the entire neural network in order to optimize the final result.

4.4.2. Congestion modelling

Possibly, en-route congestion contains information on trip fuel consumption. The congestion indicates the possibility of holding, the freedom of trajectory alterations and trajectory and route alterations suggested by ATFM. On a network level, congestion has been modelled to identify causes of en-route delay [58]. Here, aircraft data was clustered in order to identify congestion points using a specific congestion metric, such that the congestion points could be modelled as queuing points. Similar to the approach of this study, congestion points can be modelled. From these congestion points, input features are created.

Flight traffic data is collectible from ADS-B data from open sources as OpenSky, FlightRadar24 and FlightAware. For the current research, data from OpenSky will be used. Detailed information on information of flights is given, such as aircraft specific information, location, speed, direction and altitude. This information can be used to locate flights. To group the information, a grid representation will be made from the European airspace, similar to what Lin (2021) [58] did for the Chinese airspace. All flights in the European airspace are hard to gather. Using the busiest airports in Europe however, one can collect departures and arrivals from those airports to reach a proper representation of the airspace.

Using the flight tracking information in a grid representation, the congestion per grid can determined. Different measures are presented in studies for identifying congestion. Lin (2021) [58] determined congestion using Equation 4.1, Equation 4.2 and Equation 4.3, where T is the traffic count, R the number of operational routes and E the entropy, measuring the intended directions of aircraft. Other studies [93], use different measures for congestion. Here, the standard deviations in heading and velocity is used, plus the number of aircraft ascending or descending, used for terminal area's. Both methods showed promising results. Because of the ease of measuring standard deviations in speed and heading plus the traffic count, this method for congestion identification is used.

Another interesting feature extractable from air traffic data, is the comparison of the intended arrival time with the estimated arrival time. This way, more detailed information on flight delays is accesible. Using departure delay only, one is not sure if the delay is compensated by speed increase, or will result in an arrival delay.

$$Score_i = \omega_1 \tilde{T}_i + \omega_2 \tilde{R}_i + \omega_3 \tilde{E}_i \tag{4.1}$$

$$E_i = -\sum_{j}^{N} p_j log p_j \tag{4.2}$$

$$\tilde{T}_i = \frac{T_i - T_{min}}{T_{max} - T_{min}}, \tilde{R}_i = \frac{R_i - R_{min}}{R_{max} - R_{min}}, \tilde{E}_i = \frac{E_i - E_{min}}{E_{max} - E_{min}} \tag{4.3}$$

Once the congestion is determined per grid point, the features need to be processed to get an useful input for the remaining model. This approach is similar to the feature preparation of the weather forecasts. For the statistical method, averages of congestion points and the number of congestion points should be used. For the neural network model, an autoencoder can be trained. This time however, the grid representation can be used beneficially. Using a CNN, features of the grid can be compressed to smaller features at critical points. Using an auto-encoder, useful information on congestion in the entire airspace can be modelled to just a few features. These features will serve as input for the neural network. An LSTM autoencoder is better at predicting future states of congestion. Including this increases accuracy, but implementation complexity as well. For the current research, it is assumed that using no LSTM model in the autoencoder will entail sufficient information, in combination with the time of the day. Depending on interim results, this decision may be altered.

4.4.3. Route information

The last data inputs to use is route information. A route composed by the FPS consists of waypoints along airways. Using aircraft tracking data from OpenSky, the obedience of aircraft to those suggested routes can be measured. From the route inputs, weather and congestion features along the route can be selected. Information from historical flight data can be extracted as well. Using the variation in flight routes, information for future flight is given. In the investigation phase, an indication of reroutings can be given. So, besides the flown routes, the variation in trajectories per OD will be used as input which might entail valuable information.

5

Fuel loading strategy

Using the improved trip fuel consumption estimation, a fuel loading strategy must be implemented to advice on the discretionary fuel to be loaded. This strategy determines the potential in fuel loading. Zhu (2021) [96] uses a fixed time of discretionary fuel on top of the estimated fuel, which is easy to understand and implement. Kang (2018) [49] and Kang (2021) [50] use confidence bounds of the fuel burn estimations and extrapolates them such that the same amount of flights use reserve fuel as in reality, which is the most useful for comparing the influence of the new model with the current used systems. Ayra et al. (2014) [13] uses an optimization model considering the effects of fuel depletion to get to the optimal result. The decision tree used is shown in Figure 5.1. In the study, they estimated the chance of fuel usage in flight and the chance on holding based on historical data. When only the alternate and reserve fuel is left in the simulation, aircraft divert to another airport immediately, unconditioned on the chance that ATC mention that the aircraft is able to land within the near future. This increases the chance of divertion significantly. For the costs, only extra airport fees and extra fuel costs due to diversion is considered. Passenger dissatisfaction and negative pilot experience is not included.

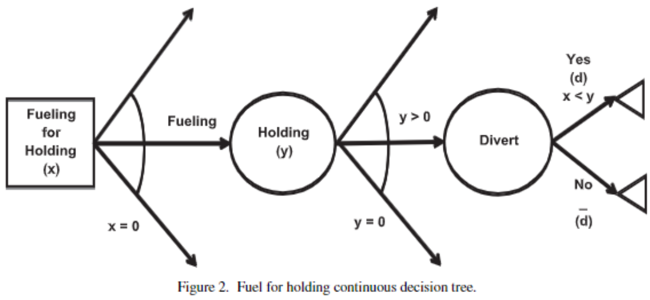


Figure 5.1: Decision tree for a fuel loading strategy, from [13]

So, the existing methods on fuel loading strategies are based on assumptions that might not be realistic. An improved decision tree is shown in Figure 5.2. Chances of going to another state are indicated by $P(X)$. In this decision tree, the chance of landing at the destination while using alternate fuel is included. ATC might decide on this, when the aircraft can land safely using a fraction of the alternate fuel. In case of a missed approach, which does not occur often, the risk of using all reserve fuel is introduced. Then, using a priority emergency landing, there is still a high chance the aircraft will land safely.

It is assumed that usage of the improved decision tree would be needed to identify a good strategy for fuel loading. Probability distributions are needed for several events, without many data points. Also the costs of all the end stages are hard to estimate since an emergency situation is undesired for the passengers, the pilot, the airline and the ATM system, which is hard to express in costs. Thereby, the trip fuel consumption estimation determines the chance of over-burn. Improving the distribution of this probability is key to have an accurate estimation on the following distribution. Since factors are not included in discretionary fuel estimation still, this will be improved in the current research. Moreover, a complex model based on many assumptions might not gain trust of pilots, airlines and ATC. A complex fuel loading optimization will not performed in the current study. First of all since the timespan to perform such an optimization does not fit in the time span of the MSc. thesis and can not be combined well with the other research objectives. Thereby, data of a single airline is used, while data of all flights over Europe complexity including emergency scenario's would be better for determining the probability distributions in emergency events.

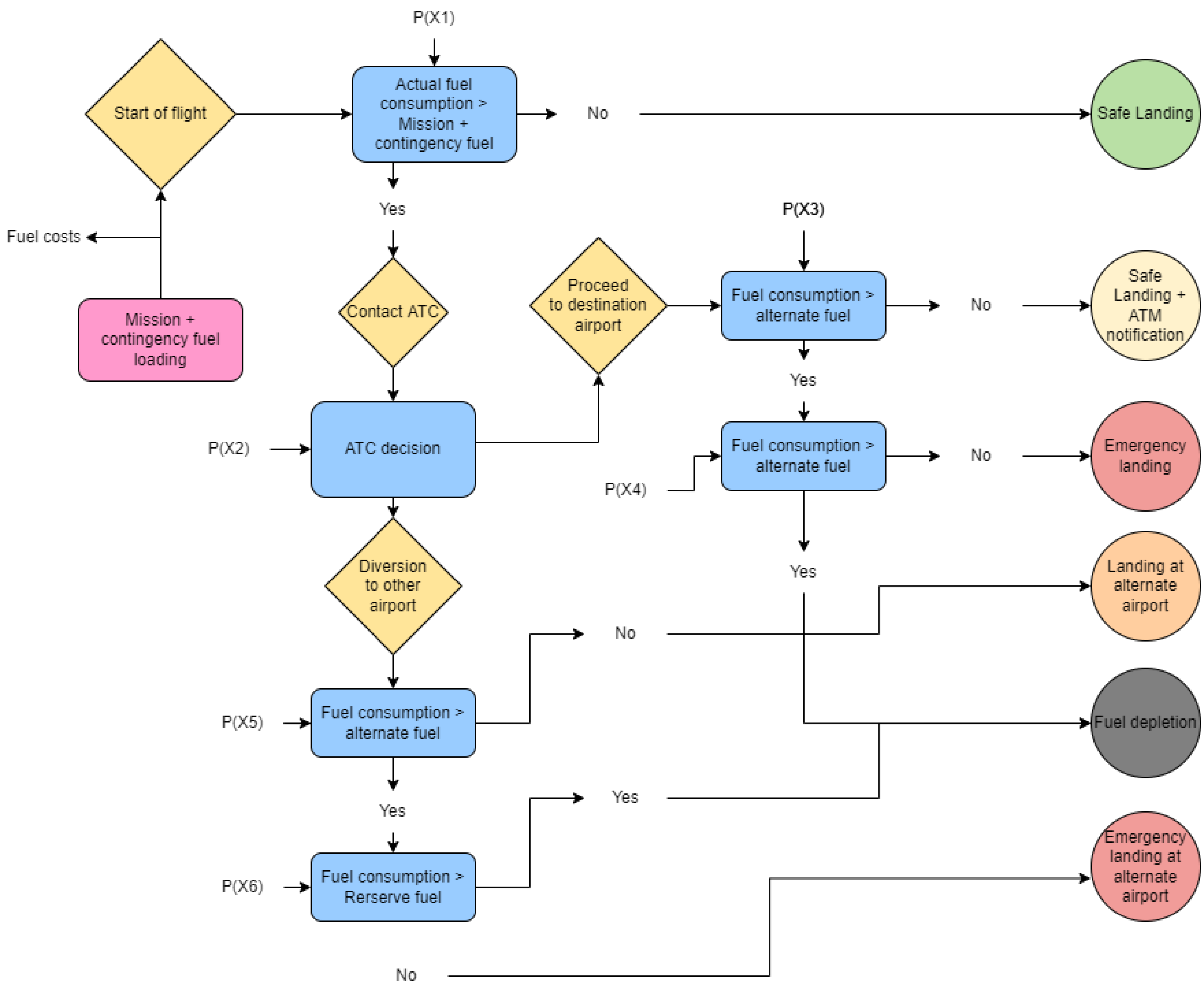


Figure 5.2: New assumed decision tree for a fuel loading strategy

For the current research, a comparison with current practice is made. Based on the literature review, the number of flight using reserve fuel is in the range of 0%-0.1%. From this observation, one may argue that the felt costs by pilots is included in this dataset, and so, costs of reserve fuel usage is included.

Two methods will be used to estimate the amount of fuel to be loaded. First, by applying a base value of discretionary fuel on top of the estimated fuel consumption. The resulting use of discretionary fuel can be compared to practice of airlines. The second is by scaling the trip fuel consumption estimation, such that the same amount of flights land using reserve fuel.

Using the reduced (or increased) weight of the aircraft, fuel savings will be determined using a cost to carry analysis. Ryerson (2015) [73] introduced Equation 5.1, where $b_{i,a}$ represents the block fuel consumed for flight i on aircraft a as a function of weight ($m_{i,a}$) and distance flown d_i. The coefficients can be determined using statistical analysis on the available airline data.

$$b_{i,a} = \beta_{1,a} m_{i,a} + \beta_{2,a} m_{i,a} d_{i,a} + \beta_{1,a} d_{i,a} \tag{5.1}$$

6
Research Proposal

Using the conclusions of previous chapters, a research proposal will be presented in the current chapter.

6.1. Problem statement & research question

In chapter 2, the operations of the industry regarding fuel loading is discussed. Existing research on fuel consumption estimation is presented as well. Concluding from the literature review, trip fuel consumption estimations are increasing in accuracy. Still, factors that influence fuel consumption are not included. Therefore, the problem statement is as follows:

> *Due to uncertainties in flight fuel burn, excess fuel is loaded by airline pilots or dispatchers in order to avoid diversion or fuel depletion, which causes extra fuel burn in return.*

Existing research focused on input data in order to identify correlations between the input data and the trip fuel consumption. Flight specific, time-date and airport traffic and weather data has been included. However, en-route data - including weather, traffic and route information data - has not been included. From the identified research gaps, the research question is created:

> *Can loading discretionary fuel by airlines be reduced by creating a data driven model using en-route information that predicts trip fuel burn?*

6.2. Research Objectives

Using the problem statement and research question, research objective could be stated. The research objectives are found using the conducted literature study on quantitative analysis of fuel burn predictions. The research objective is as follows:

> *To more accurately predict trip fuel burn by creating a data driven model using en-route information such that loading discretionary fuel could be minimized while maintaining the same level of reserve fuel usage.*

The main research objective is divided into sub objectives:

1. Identify the potential in discretionary fuel loading for the case study airline.
   (a) Characterize discretionary fuel loading behaviour for flights of the airline.
   (b) Describe fuel loading differences using a simple model
   (c) Identify extreme trip fuel consumption scenario's, flight times and route deviations.
2. Measure the effect of en-route weather data on trip fuel consumption estimations.
   (a) Collect and pre-process the weather data.
   (b) Perform feature extraction on the weather data.

(c) Identify correlations between weather data and trip fuel consumption.
(d) Include weather data in the trip fuel consumption estimation model.

3. Measure the effect of en-route traffic data on trip fuel consumption estimations.
   (a) Collect and pre-process the air-traffic data.
   (b) Perform feature extraction and identify congestion points using the air traffic data..
   (c) Identify correlations between air traffic data and trip fuel consumption.
   (d) Include air traffic data in the trip fuel consumption estimation model.
4. Include route information in the trip fuel consumption estimation model.
   (a) Collect and cluster historical route data.
   (b) Include route information in the trip fuel consumption model to test changed accuracy in combination with weather and air traffic data.
5. Compute the estimated fuel savings
   (a) Apply a fuel loading strategy using the new trip fuel consumption prediction models.
   (b) Perform a cost-to-carry benefit analysis using the new trip fuel consumption predictions.

6.3. Work Breakdown

The objectives presented in section 6.2 be divided in the into work packages in the following section. In total, six work packages are present. Using the work packages, an estimated planning is provided. Since the project spans over several moths, it is hard to predict the time span of every work package. The work packages are planned sequentially, so from the results of one of the work packages, following work packages may deviate. The work flow is shown in Figure 6.1.

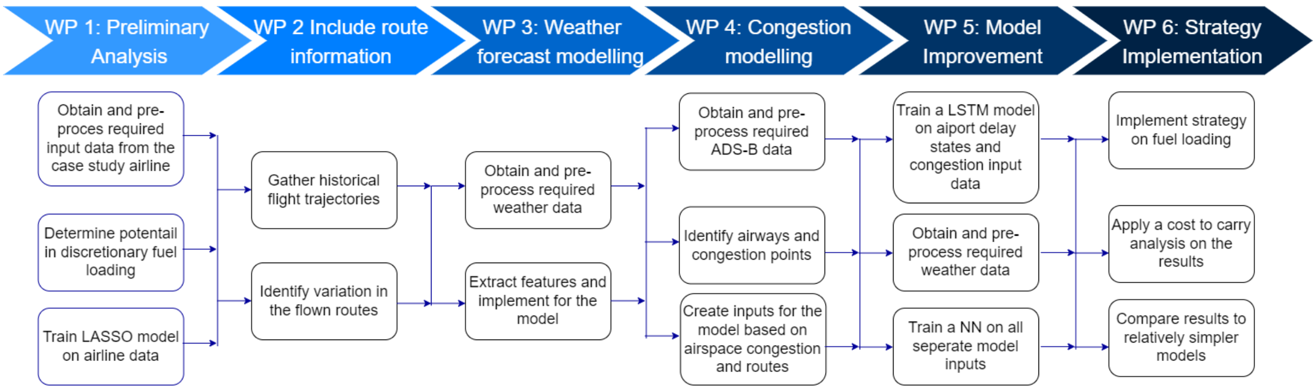


Figure 6.1: Workflow diagram of the research.

6.3.1. WP 1: Preliminary Analysis

Data used to train the model will be gathered from an European airline. The airline fleet consists of roughly 50 aircraft, operating flights within Europe. Previous research has been performed using data from other continents and different airlines. Therefore, the potential in fuel loading may deviate. The quality of ATFM and the constricted airspaces deviates per region. The type of airline and its policies also influences discretionary fuel loading decisions.

So, a preliminary analysis on the fuel consumption data will be performed to identify the potential of discretionary fuel reduction for this dataset. Thereby, the variation in fuel consumption can be investigated. Comparing extreme cases in fuel consumption, the flight time or deviations in the flown trajectories gives a better overview on the trip fuel consumption standards. The higher the flight time is correlated with trip fuel consumption, the stronger the influence of holding is, while correlations with weather severity on a day indicate the impact of weather.

In this work package, a statistical model will be created using flight specific data, date-time data and airport information. The Lasso model will suffice for this. Lasso based stacking will also be performed to measure the increase in accuracy. A neural network that will serve as basis for the inclusion of the other models will be included as well. In this phase, it might be that the accuracy of a neural network is not higher the the statistical model.

6.3.2. WP 2: Include route information

Once a preliminary model has been created, information on the route should be included. Processing of the route information is required to determine accurate locations for weather forecast features and congestion features. For given OD-pairs, historical air traffic data will be used to identify the variation in flown routes. If a high variation is present, one should take this into account when extracting features for weather and traffic, such that the feature actually correspond with the possible flown routes. The variation can be measured by linearly reconstructing the route using the locations per time step. Then, the minimum distance towards a fixed number of waypoints indicates the relative error in flown route. By gathering this information per OD pair, a variation per route can be computed.

The routes will be included by the model by discretizing the possible routes for an OD pair, and include this as input. For this, it is critical that there should be enough flights for all routes, because overfitting or may results otherwise. A feature for the variation in flown trajectories per route will serve as input feature for the model as well. This way, its relevance can be tested.

6.3.3. WP 3: En-route weather forecast modelling

Using the route information, the locations of weather forecasts can be determined. If there is a high variation in flown routes, forecasts in a wider range should be used, to combine the forecasts in one feature. The method for extracting feature for the statistical model and the neural network are described in subsection 4.4.1. For the neural network model, weather features are first generated using an autoencoder structure, such that the feature are pre-trained. Thereafter, the weight can be updated once connected in the entire model.

6.3.4. WP 4: En-route congestion modelling

Similar to the en-route weather forecast inclusion, en-route congestion points will be used as input features for the trip fuel consumption prediction model. The method is described more detailed in subsection 4.4.2. Using congestion identified congestion points, features can be generated for the trip fuel consumption estimation model. Congestion can be computed in various ways. Therefore it is important to test simple feature extraction methods on different congestion measures. High correlated measures might entail more information on trip fuel consumption, useful for the complete model.

6.3.5. WP 5: Model improvement

Using the obtained route information, weather features and congestion features, the complete model can be ran. By comparing a trained model with and without certain features, the difference in accuracy can be tested. The model created by Zhu (2021) [96] improved Lasso based stacking models by 6% in total and 26% for the outlier set for flight time predictions. En-route congestion information and delay states at airports probably are correlated. So, by measuring the inclusion en-route congestion features, the increase in accuracy should be compared to the values obtained by Zhu (2021) [96]. Ideally, RNN layers on traffic delay states are incorporated to obtain a better performing model. At this stage however, time constrictions on the project limit the possibility for this. The resulting model is shown in Figure 6.2. From different data sources, features are generated before combining it in a neural network.

6.3.6. WP 6: Fuel loading strategy and benefit analysis

Finally, a fuel loading strategy must be implemented to advice on the discretionary fuel to be loaded. The strategy implementation is discussed in chapter 5. The amount of fuel resulting from this strategy may be compared to the actual discretionary fuel that is loaded. Then, using a cost to carry analysis, the fuel savings can be computed. Important is to measure the amount of reserve fuel usage using the new strategy. Ideally, a few strategies are compared to each other to determine good aspects of a strategy.

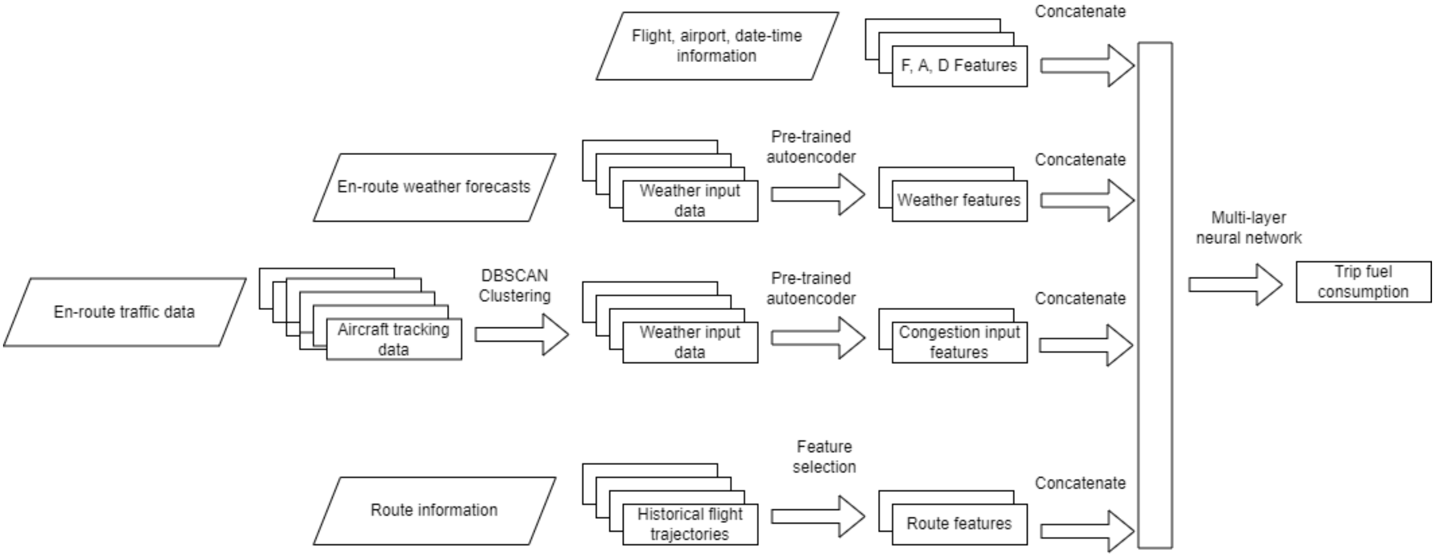


Figure 6.2: Schematic representation of the to be created model.

III

Supporting work

1

Input data: cleaning, computation and verification

The description of obtaining data is eloborated upon in the current chapter. Thereby, the creation of input data and the verification of the input data will be described. First, the data sources used will be described in section 1.1 and their connections in section 1.2. Thereafter, data processing of trip fuel and extra fuel are described in more detail in section 1.3.

1.1. Data sources

The data used in the research originates from different sources. The different sources of data and their connections will be described in the following section.

OFP

The first input source is the Operational Flight Plans (OFP), obtained from the case study airline. Before a flight's departure, an airline must create an OFP and communicate this to Air Traffic Control (ATC). This flight plan consists of all the characteristics of the to-be-performed flight. This data is extracted from the OFP, which may serve as input for a trip fuel prediction model, or may be used for the engineering of other features. The extracted features from the OFP are presented in the paper.

Per flight, multiple OFPs may be present, since the OFP may be revisited when there are significant changes in weather, traffic, or passenger forecasts. For the current research, only the last available OFP is used, which represents the final submitted OFP and the other OFPs from the respective flight are ignored, for simplicity.

FDM

The second input source is the Flight Data Recorder (FDM), measuring important parameters during the actual flight, obtained from the case study airline. The data is available for every second of the flight, starting at the departure from the gate, and ending at the arrival at the gate. Again, the used FDM data is described in the paper. Thereby, in the data source, duplicates of flights are present. Within two duplicates of a flight, minor differences are present, such as at the starting time of a flight. For consistency, only the last created FDM data is used for a specific flight, ignoring the others.

TAF

The last data source is Terminal Aerodrome (TAF) data for the period of interest. The data is gathered from *navlost.eu*, which is a platform providing public access to TAF and METAR data. TAFs are available every 6 hours, sometime with extra updates within those periods. All this information is pre-processed to extract the applicable TAF information per flight. There is no definite method to translate the TAFs to input features, as TAFs are intended for pilots, flight dispatchers and others involved in aviation, and are created to give an image of the weather forecast for humans. Still, the features created in the current research are presented in the paper. A shortened example of a TAF forecast is shown below [1]. TAFs are issued every six hours for most airports. Occasionally, there are updates if the weather forecast is changed which may have an impact on aviation. Only the last available TAF is used during the planning time of a specific flight.

> TAF LEMD 171700Z 1718/1824 21010KT 9999 SCT030 TX27/1815Z TN15/1806Z
> PROB30 TEMPO 1718/1722 3000 SHRA FEW030TCU

With the english translation:

> TAF [TAF] for Madrid/Barajas Airport [LEMD] disseminated at 1700Z of day 17 [171700Z], valid between 1800Z of day 17 to 24Z of day 18 [1718/1824]. The initial forecast is: wind blowing from 210°, with 10 kt; visibility equal or more than 10 km, scattered clouds at 3000 ft AGL [21010KT 9999 SCT030]. The predicted maximum temperature is 27°C, for 1500Z of day 18 [TX27/1815Z]. The predicted minimum temperature is 15°C, for 0600Z of day 18 [TN15/1806Z].
>
> The weather state will temporarily change, with 30% of probability [PROB30], between 1800Z and 2200Z of day 17 [TEMPO 1718/1722] to: visibility equal to 3000 m, moderate shower rain, few clouds at 3000 ft AGL with towering cumulus [3000 SHRA FEW030TCU]. After 2200Z, it'll revert back to initial state.

1.2. Data connections

Once the data is gathered from the different sources, the data sources may be combined, to created the correct features for a specific flight. First the OFP data and FDM data will be combined, and thereafter also the TAF data.

OFP-FDM

It is a seemingly trivial task to connect data from different sources, in order to have data available per flight. There is, however, no identification number present per flight. So, one should connect the data using shared information in the OFP and FDM. Using the shared information, one single combination should be present per flight. Shared information includes the aircraft registration number and route identification. Then, the departure time may be compared. The departure time of the OFP and FDM is often not similar, because of delays of a flight. Since a specific flight has to depart at a certain time, the first available time of the FDM flights can be used to connect the FDM to the OFP a flight. As there could be large departure delays present in the dataset, the FDM time of a flight might be closer the next OFP a flight, with the same aircraft registration number and route identification. To overcome this problem, all flights delayed more than two hours, or depart more than two hours early, are manually checked, if their connection is correct.
The OFP database contains 101,208 flight plans, starting from March 4 2018 until December 31st 2019. In the FDM database, 78.116 flight are recorded on the same time period. From these flights, 1,391 have an unknown origin or destination and route identification. The origin and destination are not reconstructed using the flight longitude and latitude, as it would not be possible to know if the flight is diverted or not. As a result, total of 76,725 flights are used from the FDM datasource. After merging the datasets, some flights did not have a matching OFP, resulting in 75,315 flights that are contained in the dataset used in the research.

OFP-TAF

After merging FDM and OFP, the TAFs are connected to the flights. Each TAF is assigned an ID, which may be connected to a created flight ID. For a TAF, there are periods given for the dissemination and validity of the forecast. Using this information, the most recent TAF for a certain airport during the planning phase is used, ignoring any TAF issued before or thereafter. The TAF may not be more than 18 hours old, to be connected to a flight. Otherwise, a flight has no connected TAF. Then, the connection of TAF information to a flight is as follows, a TAF forecast has a main line for the entire valid period, and updates of that weather over time. Using the main line, each of the following inputs is saved: cloud height, cloud type, horizontal visibility, arrival wind speed, wind direction, wind gust speed, thunder, snow, rain, mist and fog. Thereafter, depending on

[1] https://mediawiki.ivao.aero/index.php?title=TAF_explanation

the estimated arrival time of the flight, values of these parameters get updated. Updates with a temporality (TEMPO) type, are only considered when the arrival time plus or minus one hour is within the time range of the temporary update. A becoming (BECGM) update is always considered if the flight arrives after the first indicated time minus one hour. A from update (FM) is considered if the flight arrives after the first indicated time minus one hour.

Like any method for reading TAFs, there are considerable disadvantages. The severity of weather is not considered as an input, just like to probability of a weather change, which potentially have a large effect on the actual weather. Deterioration or improvement of the weather is treated the same, whereas the case study airline treats it differently. There is also no distinction between a FM update and a BECGM update, where the first is more abrupt than the latter. Still, using the current method, a reasonable picture of the expected weather is present, with information on the expected visibility, precipitation and wind. The features used for the model are described in the paper. In total, 16,533 flights have no TAF forecast. The horizontal and vertical visibility is also presented on the OFP. Therefore, the visibility input features are updated if there is no TAF forecast. This way, each flight can be assigned to a VFR or IFR group.

1.3. Data cleaning (Input feature generation)

From the combined data sources, features may be generated, describing the flight of interest. The computation of most variables are present in the paper, but some critical features are extended upon below. As data often contains errors and missing data, the input data needs to be cleaned before it can be used in a prediction model. A different data-cleaning approach is used for the trip fuel and extra fuel than the other parameters. Since trip fuel is the target variable in the prediction model, it is important that the data that described the target, is representative.

1.3.1. Trip fuel

The trip fuel of a flight is defined as the fuel required for a flight from the start of the runway until the landing. To compute this parameter, two techniques are possible. The first, currently used by the case study airline is based on the weight of the fuel at the two described instances. The second is using the fuel flow of the aircraft, integrated over time. Both methods have their advantages and disadvantages. In the research, the weighing method will be used primarily, and the fuel flow method only, when there is missing data, or an outlier present using the weighing method.

Weighing method

The weighing method is based on measurements of the fuel tank before and after the flight. First, the instance of measurement of the tank is not deterministic. The location of the start of runway varies, and depending on the taxi route, there is uncertainty on the exact location and instance of the start of runway. The landing is more certain since a ground sensor indicates the landing rather accurately. Thereby, the fuel tank sensor does not output the aircraft's weight every second. The recurring interval is one minute, with occasional output data four seconds prior to and after this specific time. A difference of one minute while taxiing may lead lead to a difference of 8-16 kgs, which is relatively small. However, when a measurement point is taken during acceleration, with a larger fuel flow, the difference increases to 100 - 133 kgs, which is already a large difference.

Second, the output data of the sensor may be inaccurate. The largest source of uncertainty arises during acceleration. The difference in pressure in the tank causes deviations in the fuel quantity remaining in the aircraft. The fuel quantity in the tank may rise up to 600 kgs during take-off, which is unrealistic. So, the fuel weight prior to the flight is uncertain, leading to inaccurate results. Also, the first measurement of the weight sensor varies, with occasions were the fuel quantity at the start is 2000 kgs lower than a minute later.

So, the tank weight should be measured a certain time before take-off and after landing, to make sure the amount of sensor errors are little. The fuel used per minute is relatively low for the taxi phase. So, selecting a time instant while taxiing creates more consistent results than a time instant is chosen on the runway. A three minute time interval has been set before being airborne.

Fuel Flow method

In the FDM data, an integrated fuel flow value is present. This fuel is divided in trip fuel and taxi fuel. This value however, does not update roughly 10 seconds before and after take-off, resulting in a low estimated value. Actually integrating the fuel flow also has its downsides. For an average value it is fine, but for integrating it lacks accuracy. The discretization step is 20 seconds, meaning that the fuel flow will be integrated over this time period. As abrupt changes may happen over seconds, the discretization is too large, to determine an accurate total integrated fuel flow. So, when referring to the fuel flow method, the FDM is used, resulting in a trip fuel value that ignores the runway phase of a flight in the computation.

Combination

In conclusion, both methods have their inaccuracies. Ignoring sensor errors, the weighing method approximates real trip fuel the best, as the fuel flow method either ignores the run up phase, or has poor integrated values. For the research, the weighing technique will be used, with an analysis on the outliers. The two methods are compared in Figure 1.2 and Figure 1.1. Visible is that the weighing technique has major outlier, caused by sensor errors. Also, the weighing method results mostly in a higher total fuel used, as the fuel flow method does not take into account the runway phase. Interestingly, the fuel flow method has a higher trip fuel in some cases than the weighing method. This may be caused by a measurement error in either one of the methods. Flights should be included in the dataset as much as possible, also with weighing sensor error. Therefore, outliers in the weighing measurement methods are computed using a linear transformation of the fuel flow method. Univariate outliers may be detected using various metrics [15]. The most common being outliers based on a normal distribution. Using this assumption, outliers are defined as two standards deviations from the mean, shown in Equation 1.1, where TF_{weigh} represent the trip fuel computed using the weighing method and TF_{FF} using the fuel flow method. Visual inspection is possible in the current set-up and will therefore be used as verification. A flight is defined as a regular flight or outlier flight based on Equation 1.1. Thereafter, a linear tranformation is created such that a value is present for each of those flights, using Equation 1.2. As a result, 294 flights are considered outliers, and have other trip fuel methods than the regular flights. 1,091 flights have no tank measurement either before or after take-off. These flights are also computed using the formula show in Equation 1.2, where a_{ff} and b_{ff} are calculated using by creating a linear regression between TF_{weigh} and TF_{FF}.

$$i = \begin{cases} outlierset & ,|TF_{weigh} - TF_{FF}| > 2\sigma^{met} \\ normalset & ,|TF_{weigh} - TF_{FF}| < 2\sigma^{met} \end{cases} \tag{1.1}$$

$$TF_{weigh} = a_{ff} + b_{ff} * TF_{FF} \tag{1.2}$$

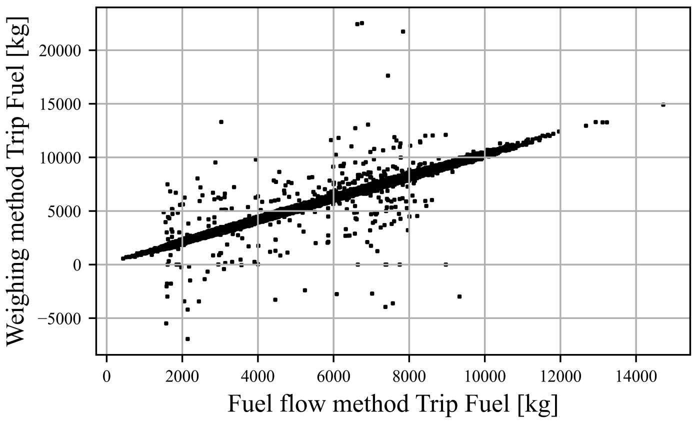


Figure 1.1: Difference in trip fuel from different sources, defined as $TF_{weigh} - TF_{FF}$

1.3.2. Extra fuel

Critical for the benefit assessment, is the extra fuel actually being loaded on aircraft, to compare the newly created strategies. As was discussed in subsection 1.3.1, fuel tank measurement can be inaccurate. So, an

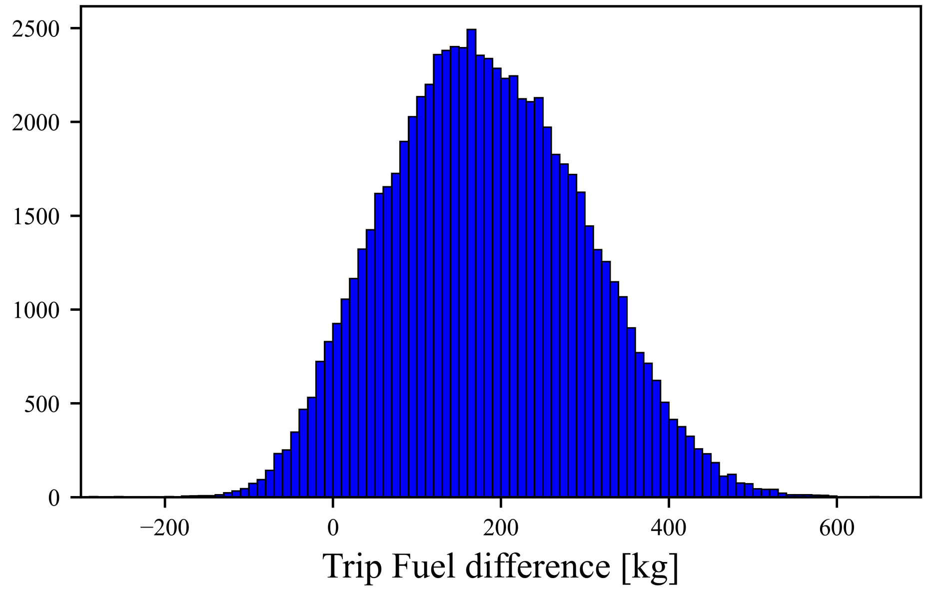


Figure 1.2: Difference in trip fuel from different sources, defined as $TF_{weigh} - TF_{FF}$

analysis on the extra fuel being loaded aircraft is performed as well. The extra fuel is defined as all the extra fuel loaded on top of the planned fuel in the OFP, shown in Equation 1.3. EF represents the extra fuel loaded by pilots, FW the fuel weight upon departure, TF the trip fuel, CF the contingency fuel, $TAXIF$ the taxi fuel, FRF the final reserve fuel, AF the alternate fuel and DDF the discretionary fuel loaded by flight dispatchers.

$$EF = FW - TF - CF - TAXIF - FRF - AF - DDF \tag{1.3}$$

The relationship between the block fuel, as indicated on the OFP and the first measurement of the fuel in the tank, is shown in Figure 1.3. Then, the same relationship is shown in Figure 1.4 for flights without tankering fuel. The flight with tankering fuel clearly have a larger variance in the measure extra fuel loaded. In fact, in the used dataset, 0.6% of the flights without tanker fuel have a negative measured extra fuel, while 12.5% of the flight with tanker fuel have measured extra fuel. So, for many flights, less tanker fuel is loaded than was planned according to the OFP. This makes extra fuel analysis unclear, since it is untraceable which portion for the extra fuel is intended as buffer or as tanker fuel. Therefore, flight with tanker fuel are excluded in the benefit assessment. As fuel usage is still valid for these flights, they are included in trip fuel consumption predictions.
Then, there are extreme data points, probably because of measurement errors. This time, visual inspection in combination with knowledge in airline practice will be used to detect unwanted outliers in the dataset. The distribution does not follow a regular distribution. Still, metrics for skewed datasets exists [42]. The resulting outliers would need to be verified for every metric. As the most additional fuel loaded because of weather uncertainties is 2400 kg, approximately 1 hour of flight time, any value of extra fuel loaded above 3000 kgs is considered to be tanker fuel, and is excluded in the dataset. In total, 131 flights (0.17%) are excluded for benefit assessment using this procedure. Negative values are caused by sensor error, or by dispatcher fuel actually not being loaded on the aircraft. The case study airline uses 600kg as steps for extra fuel to be loaded for discretionary fuel of flight dispatchers. So, extra fuel below 600kg are neglected, because the difference with the flight plan is considered to be too large, or there are suggestions of sensor errors. In total, 34 flights (0.05%) are excluded form the benefit assessment, using this procedure. The other flights, with a negative value for extra fuel are included in the benefit assessment as they are most probably caused by a difference in discretionary fuel loaded by flight dispatchers.

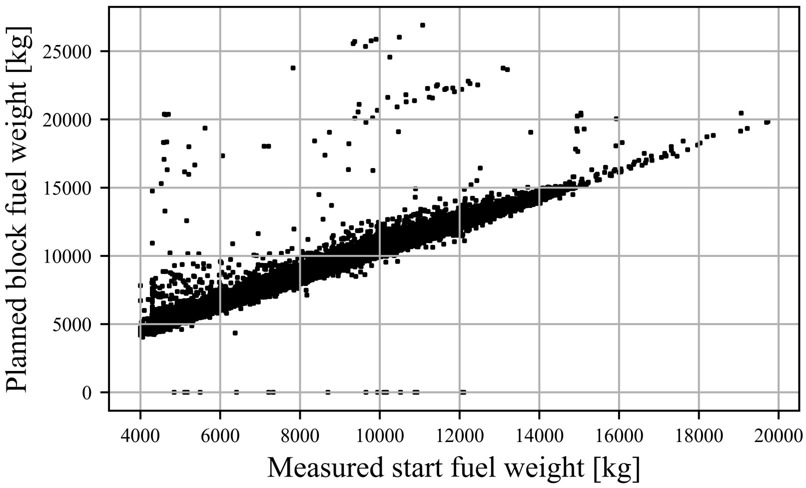


Figure 1.3: Difference in trip fuel from different sources, defined as $TF_{weigh} - TF_{FF}$

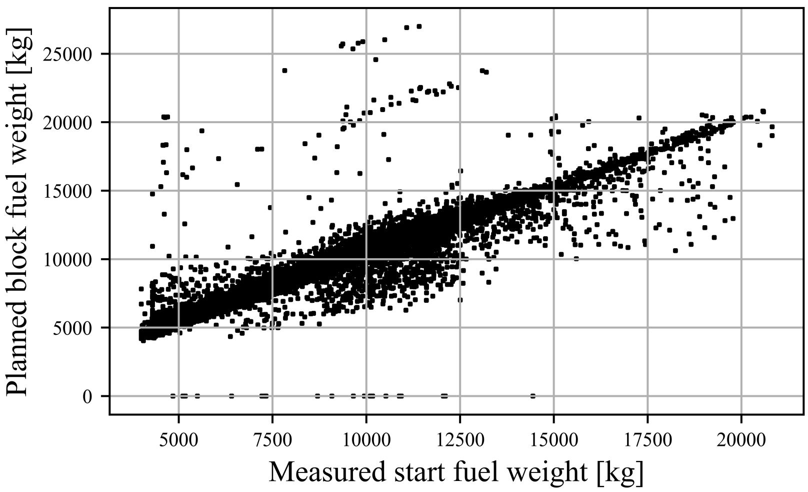


Figure 1.4: Difference in trip fuel from different sources, defined as $TF_{weigh} - TF_{FF}$

2

Results: extra analysis

The current chapter provides extra analysis on the 2TS model proposed in the paper. The advantage of time series modelling will be assessed, using a case study. An analysis on the prediction of outliers will be performed, the effectiveness of quantile regressions will be tested and finally, an ablation study will be presented.

2.1. Case studies

The way the 2TS model reduces the RMSE of the error compared to the GB model can be best explained using case studies of flights. These case studies are sets of flights, grouped on their arrival aiport or OD-pair. The goal is to demonstrate the difference in daily trends in fuel consumption, and the demonstrate the difference in hourly fuel consumption.

Daily deviations in average fuel consumption may be best demonstrated using a set of flights, grouped per OD-pair. This way, critical times trends present in the dataset are not averaged out, such that the GB model performs reasonably good as well. The OD-pair selected for further analysis is an OD-pair, which is one of the higher frequency flight legs. In Figure 2.1, daily average TF deviations are presented per day in the test set for the OD-pair. Also, the GB and 2TS predictions are presented. In the figure, the average TF_{dev} changes over time. Both the GB predictions as the 2TS predictions are able to predict this difference throughout the year. Around day 80, there is a sudden change in average TF_{dev} for a period of a few days. Analyzing the prediction performance of the GB and 2TS model, it is visible that the 2TS model is able to follow this trend more accurately, using the flight data of prior days as input.

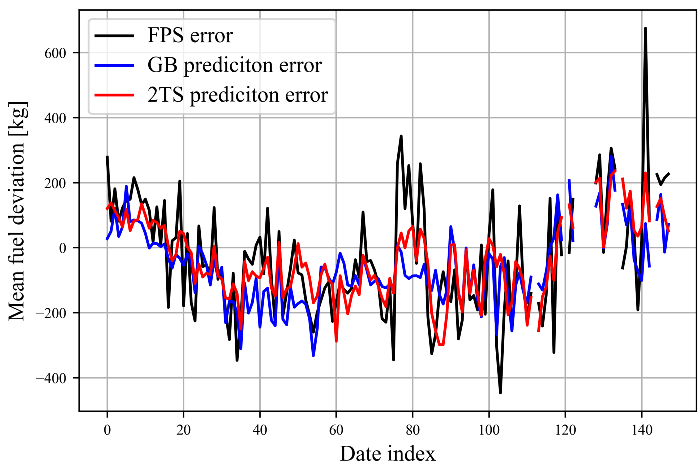


Figure 2.1: Average trip fuel deviations throughout the year

Hourly deviations in average fuel consumption may be best described using a set of flights with a the same arrival airport, such that the resulting number of flights in the analysed dataset is high. Figure 2.2 shows the hourly average TF_{dev} for flights arriving at an arrival airport with a high frequency of arriving aircraft. The values are grouped per blocks of six hours, to enhance interpretability. The figure demonstrates that the 2TS prediction is on average higher than the GB prediction for every hour block in the week. Thereby, a stronger repeating pattern is visible for the GB model, where the prediction of TF_{dev} is low during the night. Apparently, this pattern was present in the training set period, while it isn't in the test set period. So, by using prior days as input data, the 2TS model is able to alter the predictions, based on recent data.

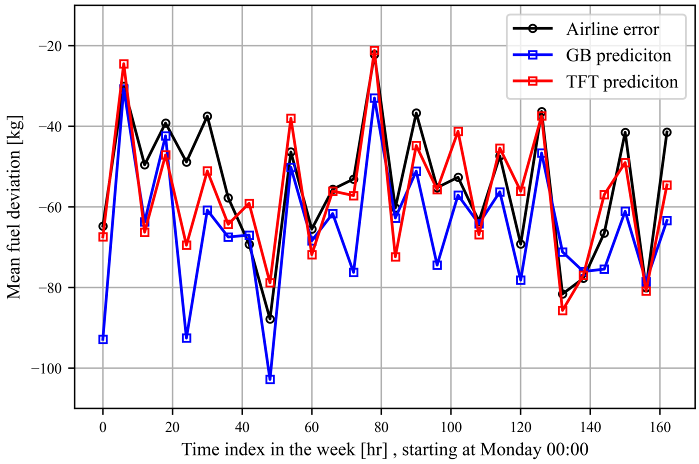


Figure 2.2: Average trip fuel deviations throughout the week

2.2. Outlier analysis

The effectiveness of fuel loading decisions based on the 2TS model depends on its ability to predict outliers. Outliers in TF_{dev} dictate the value of the SPIs using a specific fuel loading strategy. The different sets of outliers have been defined to be at the 5th percentile, the 1st percentile and 0.1st percentile on TF_{dev}. Table 2.1 shows the prediction performance on outliers for the airline practice (FPS) and the GB and 2TS models. The Mean Absolute Error (MAE) is used as the metric, as it may be used to compare errors at different magnitudes. To allow a fair comparison, the mean estimation of a model is added to the MAE. For instance, if the MAE of GB is 302 kgs, but the average estimation of the GB model is -62, the resulting error is 364. This way, the error from the same benchmark between models is compared to each other. In the figure, $\frac{Model}{FPS}$ is also present, to show which factor is remaining of the FPS error for the respective outlier set. Two insights are found, the first being that the 2TS model reduces the MAE of all outliers sets by roughly the same amount. As the MAE of both the FPS and the 2TS increases, the factor of the error remaining increases. For fuel loading decisions, it is important to reduce the error of the major outlier set as much as possible. Even though the error is increased, for the major outlier set, the model performs relatively better for the minor and medium outlier set. For certain SPIs, such as contingency fuel usage, this may be beneficial. Still, to continue prediction for fuel loading decisions, a model excelling in predicting major outliers would be preferred over the current model. The second insight is that compared to the GB model, both the absolute error difference of the 2TS model and the $\frac{Model}{FPS}$ reduces for more extreme outliers. So, the largest improvements of the model are present in the minor outlier set. This shows that the current model suits well for contingency fuel determination or operational optimization, using predicted mission parameters.

Model	**Mean value**	**minor outlier**		**medium outlier**		**major outlier**	
Metric		MAE	$\frac{Model}{FPS}$	MAE	$\frac{Model}{FPS}$	MAE	$\frac{Model}{FPS}$
FPS	0	554	-	848	-	1546	-
2TFT	-48	344	0.62	618	0.73	1332	0.86
GB	-63	381	0.69	661	0.78	1357	0.88

Table 2.1: Model performance in extreme flight sets

2.3. Spatial levels

The following section contains an ablation study, in which certain aspects of the 2TS model are removed, to demonstrate the effectiveness of that component. The FTTS and IFTTS models are already reduced versions of the full 2TS model. The results of these models are present in the paper. The remaining parameters of interest for the ablation study is the spatial level introduction in the FTTS model.

Table 2.2 shows the prediction errors for the FTTS model, using different sets of spatial levels. The FTTS model models a time series over each spatial levels, so the effect of these time series are investigated. The spatial levels OD-pair, dep airport and arrival airport remain as inputs. Only the FTTS-predicted values are removed. The results show that the combination of all spatial levels results in the best performing model. This way, patterns of each group of a flight are included in the model to enhance prediction performance. Thereby, the OD-pair and arrival airport are the best predictors, as a FTTS model using only these spatial levels as inputs, results in the lowest RMSE. The prediction performance using only network FTTS inputs is the worst performing of all of the models.

Spatial level inputs	**RMSE**	**MAE**
All	172.1	128.6
OD-pair only	172.9	130.2
Arr airport	172.8	129.3
Dep airport	173.5	130.4
Network	174.5	131.2

Table 2.2: Ablation study for the spatial levels used in the model

Bibliography

[1] European organisation for the safety of air navigation eurocontrol specification for the initial flight plan. URL www.eurocontrol.int.

[2] Estimation of aircraft fuel consumption. *Journal of Aircraft*, 46(4):1089–1093, 2009.

[3] Khaled Abdelghany, Ahmed Abdelghany, and Sidhartha Raina. A model for the airlines fuel management strategies. *Journal of Air Transport Management*, 11(4):199–206, 2005. ISSN 0969-6997. doi: https://doi.org/10.1016/j.jairtraman.2004.10.002. URL https://www.sciencedirect.com/science/article/pii/S0969699704000705.

[4] Anna Achenbach. *Predictive analytics in airline operations : application of machine learning for arrival time and fuel consumption prediction.* WHU / Chair of Logistics Management, 2018. URL https://opus4.kobv.de/opus4-whu/frontdoor/index/index/docId/831.

[5] Anna Achenbach. *Predictive analytics in airline operations: application of machine learning for arrival time and fuel consumption prediction.* PhD thesis, WHU-Otto Beisheim School of Management, 2021.

[6] Ludovica Adacher, Marta Flamini, and Elpidio Romano. Rerouting algorithms solving the air traffic congestion articles you may be interested in advanced pid tuning based on the modulus optimum method for real systems aip conference a survey of sat solver aip conference rerouting algorithms solving the air traffic congestion. 1836:20059, 2017. doi: 10.1063/1.4981993. URL https://doi.org/10.1063/1.4982000.

[7] Alba Agustín, Antonio Alonso-Ayuso, Laureano F Escudero, Celeste Pizarro, et al. Mathematical optimization models for air traffic flow management: A review. 2010.

[8] Shervin AhmadBeygi, Amy Cohn, Yihan Guan, and Peter Belobaba. Analysis of the potential for delay propagation in passenger airline networks. *Journal of Air Transport Management*, 14:221–236, 9 2008. ISSN 0969-6997. doi: 10.1016/J.JAIRTRAMAN.2008.04.010.

[9] Stefan Reitmannand Sameer Alam and Michael Schultz. (advanced quantification of weather impact on air traffic management - intelligent weather categorization with machine learning. *Thirteenth USA/Europe Air Traffic Management Research and Development Seminar*, 2019.

[10] Richard Alligier and David Gianazza. Learning aircraft operational factors to improve aircraft climb prediction: A large scale multi-airport study. *Transportation research part C: emerging technologies*, 96: 72–95, 2018.

[11] Noor F. Almussa and Abdulaziz Almaktoom. The effects of reducing aircraft weight on profitability and fuel consumption. *PalArch's Journal of Archaeology of Egypt / Egyptology*, 18:206–214, 5 2021. ISSN 1567-214X. URL https://archives.palarch.nl/index.php/jae/article/view/8326.

[12] Yuichiro Anzai. *Pattern recognition and machine learning.* Elsevier, 2012.

[13] Eduardo S. Ayra, David Ríos Insua, and Javier Cano. To fuel or not to fuel? is that the question? *https://doi.org/10.1080/01621459.2013.879060*, 109:465–476, 1 2014. ISSN 1537274X. doi: 10.1080/01621459.2013.879060. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.2013.879060.

[14] Tolga Baklacioglu. doi: https://doi.org/10.1016/j.ast.2015.11.031.

[15] Sharmila Banerjee and Boris Iglewicz. A simple univariate outlier identification procedure designed for large samples. *Communications in Statistics - Simulation and Computation*, 36(2):249–263, 2007. doi: 10.1080/03610910601161264. URL https://www.tandfonline.com/doi/abs/10.1080/03610910601161264.

[16] S. Baumann and U. Klingauf. Modeling of aircraft fuel consumption using machine learning algorithms. *CEAS Aeronautical Journal*, 11:277–287, 1 2020. ISSN 18695590. doi: 10.1007/s13272-019-00422-0.

[17] Yoshua Bengio, Ian Goodfellow, and Aaron Courville. *Deep learning*, volume 1. MIT press Massachusetts, USA:, 2017.

[18] Tatjana Boli and Paul Ravenhill. Sesar: The past, present, and future of european air traffic management research. *Engineering*, 7:448–451, 4 2021. ISSN 2095-8099. doi: 10.1016/J.ENG.2020.08.023.

[19] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152, 1992.

[20] L Breiman, JH Friedman, RA Olshen, and CG Stone. Classification and regression trees, wadsworth international group, belmont, california. *Google Scholar*, 1984.

[21] Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

[22] Gerrit Burmester, Hui Ma, Dietrich Steinmetz, and Sven Hartmannn. Big data and data analytics in aviation. In *Advances in Aeronautical Informatics*, pages 55–65. Springer, 2018.

[23] Sai Ho Chung, Hoi Lam Ma, Mark Hansen, and Tsan Ming Choi. Data science and analytics in aviation. *Transportation Research Part E: Logistics and Transportation Review*, 134:101837, 2 2020. ISSN 1366-5545. doi: 10.1016/J.TRE.2020.101837.

[24] Bela P. Collins. Estimation of aircraft fuel consumption. *Journal of Aircraft*, 19(11):969–975, 1982. doi: https://doi.org/10.2514/3.44799.

[25] Mayara Conde Rocha Murca, Richard DeLaura, R John Hansman, Richard Jordan, Tom Reynolds, and Hamsa Balakrishnan. Trajectory clustering and classification for characterization of air traffic flows. In *16th AIAA Aviation Technology, Integration, and Operations Conference*, page 3760, 2016.

[26] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[27] Jan G De Gooijer and Rob J Hyndman. 25 years of time series forecasting. *International journal of forecasting*, 22(3):443–473, 2006.

[28] Ludwig Drees, Manfred Mueller, Carsten Schmidt-Moll, Patrick Gontar, Kilian Zwirglmaier, Chong Wang, Klaus Bengler, Florian Holzapfel, and Daniel Straub. Risk analysis of the easa minimum fuel requirements considering the acare-defined safety target. *Journal of Air Transport Management*, 65:1–10, 10 2017. ISSN 0969-6997. doi: 10.1016/J.JAIRTRAMAN.2017.07.003.

[29] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.

[30] Martin Ester, Hans-Peter Kriegel, Jiirg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. 1996. URL www.aaai.org.

[31] EU. Reducing emissions from aviation, 2021. URL https://ec.europa.eu/clima/eu-action/transport-emissions/reducing-emissions-aviation_nl.

[32] Eurocontrol. European aviation in 2040 challenges of growth annex 4 network congestion. 2018.

[33] Eurocontrol. Predicting flight routes with a deep neural network in the operational air traffic flow and capacity management system. 2018.

[34] Yoav Freund, Robert E Schapire, et al. Experiments with a new boosting algorithm. In *icml*, volume 96, pages 148–156. Citeseer, 1996.

[35] Christian Eduardo Verdonk Gallego, Víctor Fernando Gómez Comendador, Manuel Angel Amaro Carmona, Rosa María Arnaldo Valdés, Francisco Javier Sáez Nieto, and Miguel García Martínez. A machine learning approach to air traffic interdependency modelling and its application to trajectory prediction. *Transportation Research Part C: Emerging Technologies*, 107:356–386, 2019.

[36] D Gomes, F Oliveira, J Centeno, L Figueiredo, and L Halawi. Issue 3 article 9 2020 part of the air and space law commons, legislation commons, and the other law commons scholarly commons citation scholarly commons citation youssef. *International Journal of Aviation, Aeronautics, and Aerospace*, 7, 2020. doi: 10.15394/ijaaa.2020.1500. URL https://doi.org/10.15394/ijaaa.2020.1500.

[37] Zhang Haifeng, Wang Xu-hui, and Chen Xin-feng. Support vector with roc optimization method based fuel consumption modeling for civil aircraft. *Procedia Engineering*, 99:296–303, 2015.

[38] James Douglas Hamilton. *Time series analysis*. Princeton university press, 2020.

[39] Lu Hao, Mark Hansen, and Megan S. Ryerson. Fueling for contingencies: The hidden cost of unpredictability in the air transportation system. *Transportation Research Part D: Transport and Environment*, 44:199–210, 5 2016. ISSN 13619209. doi: 10.1016/j.trd.2016.02.016.

[40] Robert Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural networks for perception*, pages 65–93. Elsevier, 1992.

[41] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[42] Mia Hubert and Ellen Vandervieren. An adjusted boxplot for skewed distributions. *Computational statistics & data analysis*, 52(12):5186–5201, 2008.

[43] ICAO. Aircraft operations, contruction of visual and istrument flight procedures. 2, 2006.

[44] ICAO. Doc 9976, flight planning and fuel management manual. 2012.

[45] Christian Igel and Michael Hüsken. Improving the rprop learning algorithm. In *Proceedings of the second international ICSC symposium on neural computation (NC 2000)*, volume 2000, pages 115–121. Citeseer, 2000.

[46] Independent. Pilots warn of danger as airlines cut back on fuel, 2013. URL https://www.independent.co.uk/travel/news-and-advice/exclusive-safety-warning-as-budget-airlines-such-as-ryanair-cut-fuel-levels-for-flights-8749046.html.

[47] ME Irrgang. A look at the state of airline fuel conservation, 2011.

[48] Lei Kang. *Changing Fuel Loading Behavior to Improve Airline Fuel Efficiency*. University of California, Berkeley, PhD Thesis, 2017.

[49] Lei Kang and Mark Hansen. Assessing the impact of tactical airport surface operations on airline schedule block time setting. *Transportation Research Part C: Emerging Technologies*, 89:133–147, 4 2018. ISSN 0968-090X. doi: 10.1016/J.TRC.2018.01.018.

[50] Lei Kang and Mark Hansen. Quantile regressionbased estimation of dynamic statistical contingency fuel. *Transportation Science*, 55:257–273, 1 2021. ISSN 15265447. doi: 10.1287/TRSC.2020.0997.

[51] Harshad Khadilkar and Hamsa Balakrishnan. Estimation of aircraft taxi fuel burn using flight data recorder archives. *Transportation Research Part D: Transport and Environment*, 17(7):532–537, 2012.

[52] Waqar Ahmed Khan, Sai-Ho Chung, Hoi-Lam Ma, Shi Qiang Liu, and Ching Yuen Chan. A novel self-organizing constructive neural network for estimating aircraft trip fuel consumption. *Transportation Research Part E: Logistics and Transportation Review*, 132:72–96, 2019. ISSN 1366-5545. doi: https://doi.org/10.1016/j.tre.2019.10.005. URL https://www.sciencedirect.com/science/article/pii/S1366554519303138.

[53] Waqar Ahmed Khan, Hoi Lam Ma, Xu Ouyang, and Daniel Y. Mo. Prediction of aircraft trajectory and the associated fuel consumption using covariance bidirectional extreme learning machines. *Transportation Research Part E: Logistics and Transportation Review*, 145:102189, 1 2021. ISSN 1366-5545. doi: 10.1016/J.TRE.2020.102189.

[54] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[55] Roger Koenker and Kevin F Hallock. Quantile regression. *Journal of economic perspectives*, 15(4):143–156, 2001.

[56] J. Ledolter and B. Abraham. *Statistical Methods for Forecasting*. Wiley Series in Probability and Statistics. Wiley, 2009. ISBN 9780470317297. URL https://books.google.nl/books?id=WIPxdb2P8sAC.

[57] Peng Li, Zhikui Chen, Laurence T Yang, Jing Gao, Qingchen Zhang, and M Jamal Deen. An improved stacked auto-encoder for network traffic flow classification. *IEEE Network*, 32(6):22–27, 2018.

[58] Yu Lin, Lishuai Li, Pan Ren, Yanjun Wang, and WY Szeto. From aircraft tracking data to network delay model: A data-driven approach considering en-route congestion. *Transportation Research Part C: Emerging Technologies*, 131:103329, 2021.

[59] JA Lovegren and RJ Hansman. Quantification of fuel burn reduction in cruise via speed and altitude optimization strategies. *Massachusetts Institute of Technology Cambridge, MA*, 2139, 2011.

[60] Joseph Venuti Luke L. Jensen, R. John Hansman and Tom G. Reynolds. Commercial airline speed optimization strategies for reduced cruise fuel consumption. *AIAA AVIATION Forum*, 2023. doi: 10.2514/6.2013-4289.

[61] Andrew Malwitz, Brian Kim, Gregg G Fleming, Joosung Lee, Sathya N Balasubramanian, Ian A Waitz, A Morales, Kelly Klima, Maryalice Locke, Curtis Holsclaw, et al. System for assessing aviation's global emissions (sage). version 1.5: validation assessment, model assumptions and uncertainties. Technical report, United States. Federal Aviation Administration, 2005.

[62] Michael V. McCrea, Hanif D. Sherali, and Antonio A. Trani. A probabilistic framework for weather-based rerouting and delay estimations within an airspace planning model. *Transportation Research Part C: Emerging Technologies*, 16:410–431, 8 2008. ISSN 0968-090X. doi: 10.1016/J.TRC.2007.09.001.

[63] Tasos Nikoleris, Gautam Gupta, and Matthew Kistler. Detailed estimation of fuel consumption and emissions during aircraft taxi operations at dallas/fort worth international airport. *Transportation Research Part D: Transport and Environment*, 16(4):302–308, 2011.

[64] Jenaro Nosedal, Miquel A Piera, Sergio Ruiz, and Alvaro Nosedal. An efficient algorithm for smoothing airspace congestion by fine-tuning take-off times. *Transportation Research Part C: Emerging Technologies*, 44:171–184, 2014.

[65] Angela Nuic. User manual for the base of aircraft data (bada) revision 3.10. *Atmosphere*, 2010:001, 2010.

[66] Ioanna Pagoni and Voula Psaraki-Kalouptsidi. The impact of carbon emission fees on passenger demand and air fares: A game theoretic approach. *Journal of Air Transport Management*, 55:41–51, 8 2016. ISSN 09696997. doi: 10.1016/J.JAIRTRAMAN.2016.04.004. URL http://dx.doi.org/10.1016/j.jairtraman.2016.04.004.

[67] Ioanna Pagoni and Voula Psaraki-Kalouptsidi. Calculation of aircraft fuel consumption and co2 emissions based on path profile estimation by clustering and registration. *Transportation Research Part D: Transport and Environment*, 54:172–190, 7 2017. ISSN 1361-9209. doi: 10.1016/J.TRD.2017.05.006.

[68] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. *ACM Computing Surveys (CSUR)*, 54(2):1–38, 2021.

[69] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1): 145–151, 1999.

[70] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[71] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[72] Megan S. Ryerson, Mark Hansen, and James Bonn. Time to burn: Flight delay, terminal efficiency, and fuel consumption in the national airspace system. *Transportation Research Part A: Policy and Practice*, 69:286–298, 11 2014. ISSN 09658564. doi: 10.1016/j.tra.2014.08.024.

[73] Megan S. Ryerson, Mark Hansen, Lu Hao, and Michael Seelhorst. Landing on empty: Estimating the benefits from reducing fuel uplift in us civil aviation. *Environmental Research Letters*, 10, 8 2015. ISSN 17489326. doi: 10.1088/1748-9326/10/9/094002.

[74] Megan Smirti Ryerson and Mark Hansen. Capturing the impact of fuel price on jet aircraft operating costs with leontief technology and econometric models. *Transportation Research Part C: Emerging Technologies*, 33:282–296, 8 2013. ISSN 0968-090X. doi: 10.1016/J.TRC.2011.05.015.

[75] Diego Alejandro Salazar, Jorge Iván Vélez, and Juan Carlos Salazar. Comparison between svm and logistic regression: Which one is better to discriminate? *Revista Colombiana de Estadística*, 35(SPE2):223–237, 2012.

[76] Manuel Soler, Alberto Olivares, Ernesto Staffetti, and Daniel Zapata. Framework for aircraft trajectory planning toward an efficient air traffic management. *Journal of Aircraft*, 49(1):341–348, 2012.

[77] Statista. Fuel costs of airlines worldwide from 2011 to 2022, as percentage of expenditure, 2021. URL https://www.statista.com/statistics/591285/aviation-industry-fuel-cost/.

[78] Statsmodels. Introduction statsmodels, 2021. URL https://www.statsmodels.org/stable/index.html.

[79] Lufthansa Systems. Lido flight planning solutions, 2021. URL https://www.lhsystems.com/solutions/flight-operations-solutions/lido-flight-planning-solutions.

[80] Nga Yung Agnes Tang, Cheng Lung Wu, and David Tan. Evaluating the implementation of performance-based fuel uplift regulation for airline operation. *Transportation Research Part A: Policy and Practice*, 133:47–61, 3 2020. ISSN 0965-8564. doi: 10.1016/J.TRA.2019.12.028.

[81] Sergios Theodoridis and Konstantinos Koutroumbas. Pattern recognition. *Elsevier Inc.*, 2009. doi: 10.1016/B978-1-59749-272-0.X0001-2.

[82] E M Tramper. Improving fuel decisions by airline pilots in flight preparation and execution. *TU Delft, MSc. Thesis*, 2014.

[83] A A Trani, F C Wing-Ho, G Schilling, H Baik, and A Seshadri. A neural network model to estimate aircraft fuel consumption. ISSN 2004-6401. doi: 10.2514/6.2004-6401. URL http://arc.aiaa.org.

[84] Anna C Trujiuo. Uncertainties that flight crews and dispatchers must consider when calculating the fuel needed for a flight. 1996.

[85] Mevlut Uzun, Mustafa Umut Demirezen, and Gokhan Inalhan. Physics guided deep learning for data-driven aircraft fuel consumption modeling. *Aerospace 2021, Vol. 8, Page 44*, 8:44, 2 2021. ISSN 22264310. doi: 10.3390/AEROSPACE8020044. URL https://www.mdpi.com/2226-4310/8/2/44/htmhttps://www.mdpi.com/2226-4310/8/2/44.

[86] Rafael Vazquez, Damián Rivas, and Antonio Franco. Stochastic analysis of fuel consumption in aircraft cruise subject to along-track wind uncertainty. *Aerospace Science and Technology*, 66:304–314, 7 2017. ISSN 1270-9638. doi: 10.1016/J.AST.2017.03.027.

[87] Gerhard Venter. Review of optimization techniques. 2010.

[88] Xuhui Wang, Xinfeng Chen, and Zhuming Bi. Support vector machine and roc curves for modeling of aircraft fuel consumption. *Journal of Management Analytics*, 2(1):22–34, 2015.

[89] Wangyang Wei, Honghai Wu, and Huadong Ma. An autoencoder and lstm-based traffic flow prediction method. *Sensors*, 19(13):2946, 2019.

[90] Zhiwei Xing and Hongyue Lu. Departure taxiing route modeling and optimization. In *2015 11th International Conference on Natural Computation (ICNC)*, pages 342–348, 2015. doi: 10.1109/ICNC.2015.7378014.

[91] Jefry Yanto and Rhea P. Liem. Aircraft fuel burn performance study: A data-enhanced modeling approach. *Transportation Research Part D: Transport and Environment*, 65:574–595, 12 2018. ISSN 1361-9209. doi: 10.1016/J.TRD.2018.09.014.

[92] Matthew D Zeiler. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

[93] Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola. *Dive into Deep Learning*. 2020. `https://d2l.ai`.

[94] Hong-hai Zhang, Cheng-peng Jiang, and Lei Yang. Forecasting traffic congestion status in terminal areas based on support vector machine. *Advances in Mechanical Engineering*, 8(9):1687814016667384, 2016.

[95] Sen Zhang, Yong Yao, Jie Hu, Yong Zhao, Shaobo Li, and Jianjun Hu. Deep autoencoder neural networks for short-term traffic congestion prediction of transportation networks. *Sensors*, 19(10):2229, 2019.

[96] Xinting Zhu and Lishuai Li. Flight time prediction for fuel loading decisions with a deep learning approach. *Transportation Research Part C: Emerging Technologies*, 128:103179, 2021.