Root cause analysis of ATC delays

A case study on KLM flights at Schiphol Airport

Lisa Kestens

TU Delft Delft University of Technology

Challenge the future

ROOT CAUSE ANALYSIS OF ATC DELAYS

A CASE STUDY ON KLM FLIGHTS AT SCHIPHOL AIRPORT

by

Lisa Kestens

in partial fulfillment of the requirements for the degree of

Master of Science
in Aerospace Engineering

at the Delft University of Technology,
to be defended publicly on Friday October 15, 2021.

Student number:	4548248	
Project duration:	December 1, 2020 - October 15, 2021	
Supervisor:	Prof. Dr. Ir. J.M. Hoekstra	
Company Supervisors:	F. Huisman	KLM Royal Dutch Airlines
	W. van Miltenburg	KLM Royal Dutch Airlines
	C. Evertse	KLM Royal Dutch Airlines
Thesis committee:	Prof. Dr. Ir. J.M. Hoekstra	TU Delft
	Dr. Ir. J. Ellerbroek	TU Delft
	Ir. P. C. Roling	TU Delft

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.
Cover picture: [1]

TU Delft Delft University of Technology

ACKNOWLEDGEMENTS

Dear reader,

This document marks the official end of my 5 year long studies at the Faculty of Aerospace Engineering at the Delft University of Technology. I would like to thank everyone who helped, taught and guided me during these past 5 incredible years, this would not have been possible without all of you.

Especially, I would like to thank Frans Huisman, Wouter van Miltenburg and Christiaan Evertse, who have granted me the unique opportunity to perform this interesting research as my Master Thesis, and for all the hours they spent on listing, advising and sharing their experience with me. Additionally, I would like to express my gratitude to my supervisor, Professor Hoekstra, for his valuable guidance, insights and feedback during these past 10 months.

Finally, I would like to acknowledge my parents, Ann & Johan, who have made it possible for me to pursue my studies in Delft in becoming an Aerospace Engineer, and who have granted me every opportunity along the way. I would also like to express my gratitude to my boyfriend, Jeff, for all his support and encouragement, and for always listing and believing in me.

Lisa Kestens
Delft, October 2021

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

List of Abbreviations

A-CDM	Airport Collaborative Decision Making
AAD	Absolute Average Deviation
ACC	Area Control Centre
AIBT	Actual In Block Time
AIP	Aeronautical Information Publications
ALDT	Actual Landing Time
AMS	Amsterdam Airport Schiphol
ANSP	Air Navigational Service Provider
AOBT	Actual Off Block Time
APP	Approach Control
ASRT	Actual Start-up Request Time
ATA	Actual Time of Arrival
ATC	Air Traffic Control
ATFM	Air Traffic Flow Management
ATM	Air Traffic Management
ATOT	Actual Take-Off Time
BIC	Bayesian Information Criterion
BN	Bayesian Network
CASA	Computer Assisted Slot Allocation
CDM	Collaborative Decision Making
CFMU	Central Flow Management Unit
CGNN	Causal Generative Neural Network
CPT	Conditional Probability Table
CSA	Cluster-based Sensitivity Analysis
CTA	Control Area
CTOT	Calculated Take-off Time
CTR	Control Zone
DAG	Directed Acyclic Graph
DBSCAN	Density Based Spatial Clustering of Applications in Noise

DSA	Data-based Sensitivity Analysis
EHAM	Amsterdam Airport Schiphol
EIBT	Estimated In Block Time
ELDT	Estimated Landing Time
ETA	Estimated Time of Arrival
ETOT	Estimated Take-Off Time
EUR	European
EXOT	Estimated Taxi Out Time
FIR	Flight Information Region
FP	Frequent Pattern
GNN	Generative Neural Network
GSA	Global Sensitivity Analysis
IATA	International Air Transport Association
ICA	Intercontinental
ICAO	International Civil Aviation Organisation
LL	Log-likelihood
MDL	Minimum Description Length
MGHA	Main Ground Handling Agent
ML	Machine Learning
MMD	Maximum Mean Discrepancy
MSA	Monte-Carlo Sensitivity Analysis
MTT	Minimum Turnaround Time
MUAC	Maastricht Upper Area Control
NN	Neural Network
PCA	Principle Component Analysis
RF	Random Forest
SA	Sensitivity Analysis
SIBT	Scheduled In Block Time
SID	Standard Instrument Departure
SMOTE	Synthetic Minority Oversampling Technique
SOBT	Scheduled Off Block Time
STAR	Standard Arrival Route
SVM	Support Vector Machine
TAF	Terminal Aerodrome Forecast

TMA	Terminal Manoeuvring Area
TOBT	Target Off Block Time
TSAT	Target Start-up Approval Time
TTOT	Target Take-Off Time
TWR	Tower Control
UAC	Upper Area Control
UTA	Upper Control Area
UTC	Coordinated Universal Time
VEC	Variable Effect Characteristic
xAI	Explainable Artificial Intelligence

List of Symbols

$\bar{x}$	Mean value of variable x
$\emptyset$	Empty set
ϵ	Error term
η	Equivalent sample size
η	Learning rate in Neural Network
Γ	Gamma function
$\hat{y}$	Estimation of the target variable
∇	Gradient
ϕ	Information theory scoring function
σ^2	Variance
τ	Learning iteration
E	Error function
h	Activation function
k	Maximum length of an association rule
l	Length of reduced feature set
m	Number of attributes
n	Number of data samples
O	Order
O_i	Parent set of variable X_i
P	Probability distribution
p_{ij}	Lag time
R	Relative importance
r	Correlation coefficient

S	Measure of sensitivity
w	Weight in Neural Network
X_i	Variable or attribute i in data set
D	Data set
G	Directed Acyclic Graph
g	Scoring function

I

SCIENTIFIC ARTICLE

Root cause analysis of ATC delays: A case study on KLM flights at Schiphol Airport

L. Kestens
Supervised by Dr. Ir. J.M. Hoekstra

Control & Operations, Faculty of Aerospace Engineering
Delft University of Technology, Delft, The Netherlands

***Abstract*—Due to the continuous growth of air traffic up to the year 2020, the air transportation network has become more complex, and the airports and airspace busier. However, capacities have not grown at the same rate as air traffic, making Air Traffic Control one of the most encountered primary delays. A data driven approach is taken in order to expose the drivers of the ATC delays for KLM Royal Dutch Airlines flights at Schiphol Airport. The used data consists of public and proprietary data, and contains information related to the weather, KLM flight operations, operational data, and Airport Collaborative Decision Making. To perform the analysis, two causal methods were used, association rule mining as a baseline method and a Bayesian network as the state-of-the-art model. Both methods were able to identify various conditions that trigger and/or prevent ATC delay occurrence, and agreed on the majority of the identified influential factors of the ATC delay. It was found that the main influences of ATC delay are the average startup delay of flights in the 20 minute time interval of the flight's departure, as well as the received pure ATFM delay and the assigned regulation delay key. Additionally, other influential parameters on the ATC delay related both to the amount of traffic volume and congestion at the airport, as well as individual variables of the flight, such as the propagation of arrival delay, the number of updates in the CDM process and the delay in the closure of the doors. The main discrepancies in the results could be attributed to the limitations of both methods. In general, it was found that both methods are suitable to diagnose direct causes or influencing factors on a target variable. The Bayesian network method was found to be more suitable to better understand a system and the dynamics between a large number of variables, as the conditional dependencies can be observed from the learned structure, and are not hidden in a large number of frequent patterns. However, first diagnoses of influential variables can also be done using association rule mining, which could find more indirect effects on the target variable compared to the Bayesian network, in which indirect relations might be lost in the structure learning process.**

***Index Terms*—Air Traffic Control, Flight Delays, Air Traffic Control Delays, Bayesian Network, Association Rule Mining**

I. Introduction

CIVIL air traffic has experienced exponential growth over the last decades. Due to its growth, the aviation network has become increasingly complex, and the capacity of airports and airspace has not grown at the same rate as air travel. This has caused more flight delays, which impact multiple stakeholders, such as the airline itself, its passengers and the airports in their flight network [1]. These delays furthermore translate into large financial and economic consequences for the stakeholders, especially the airline [2]. These consist of direct costs, but also of indirect costs such as the long term effects on passenger loyalty, market share and airline revenue [3]. For all these reasons, it is important to better understand delays, and possibly use this knowledge to further reduce them.

One of the delay sources in air transportation is Air Traffic Control (ATC). ATC is part of the system of Air Traffic Management (ATM), of which the main purpose is to ensure safety as well as efficiency, by keeping aircraft separated both vertically as longitudinally [4]. This type of delay is typically received before departure, while the aircraft is still at the gate. In general, it is issued due to an imbalance between demand and offered capacity of the runway(s), airspace, and gates [5]. However, ATC delay can also be experienced while the aircraft is already en-route. The causes of these en-route delays can mainly be attributed to the ATC capacity according to the Air Navigation Service Providers (ANSP), as well as staffing, weather and disruptions or actions in the ATC system [6].

In essence, ATC delay can be distinguished in two types of delay, startup delay, represented as $d_{startup}$, and ATFM delay d_{ATFM}. When a flight is regulated, its ATC delay is defined by the total received ATFM delay, otherwise the startup delay represents the flight's ATC delay. Where ATFM delay is caused by regulation in the airspace, is startup delay caused by local delay at the departure airport. Therefore, the ATFM delay is defined as the difference between the Calculated Take-off Time (CTOT) and the scheduled take-off time without the regulation, which is the sum of the Scheduled Off-Block Time (SOBT) and the Standard Taxi-Out time (STXO), shown in Equation 1. A regulated flight receives a CTOT, such that the number of aircraft using the regulated airspace can be controlled [7]. Both the CTOT and SOBT have the format of a timestamp, whereas the STXO represents a duration and is thus expressed in minutes.

The Target Start-up Approval Time (TSAT) is a timestamp issued by ATC, which is a function of the issued Target Off Block Time (TOBT), which is translated into a Target Take-Off Time, TTOT, as seen in Equation 3 [8]. The TTOT is established by assigning the earliest possible take-off time, taking into account the Estimated Taxi-Out Time (EXOT) of the aircraft to the runway. The TSAT is then found by reversing the calculation as seen in Equation 4 [8]. The startup delay is defined as the difference between the TSAT and the TOBT,

which is shown in Equation 2 [8].

$$d_{ATFM} = \text{CTOT} - \text{SOBT} - \text{STXO} \tag{1}$$

$$d_{startup} = \text{TSAT} - \text{TOBT} \tag{2}$$

$$\text{TTOT} >= \text{TOBT+EXOT} \tag{3}$$

$$\text{TSAT} = \text{TTOT-EXOT} \tag{4}$$

ATC delays are classified as initial delays, but can have a tremendous impact throughout the flight operation network of an airline due to aircraft, passenger and crew connectivity, which can result in reactionary delays for other flights [9]. However, reactionary delays are not only impacting the operations of the airline, but can also propagate to other airlines and airports [10, 11, 12]. This connectivity in the flight network is said to be the most challenging aspect in any transport system [1]. Therefore, this research aims to find the root causes of the ATC delays with a case study on KLM flights at Amsterdam Schiphol Airport. The ATC delay that is used as a target variable in this research is the specific ATC delay received upon departure from the airport. However, in order to perform a full analysis, one observation represents a turnaround procedure at the airport, such that the columns contain information on both the arrival and departure process at Amsterdam Schiphol Airport. Additionally has the scope of this research project been limited in both time and space. This research focuses on the ATC delays in the Dutch airspace, and the used data relates to the day of operation. The research objective of this study has been formulated as follows.

> *The main research objective is to expose the drivers of the ATC delays encountered by KLM flights at Schiphol Airport, by performing a root cause analysis of these ATC delays.*

A lot of research on flight delay has been done in earlier studies, in which different specific research fields can be distinguished. One of these areas is flight delay propagation, which has been studied thoroughly [2, 11, 13, 14, 15]. Additionally, the use of machine learning algorithms to predict delays has gained research interest in recent years [16, 17, 18, 19]. Another research field is found in the analysis and development of airline/airport disruption models to manage flight delays, for different levels in the operation [20, 21, 22]. The research area of this study is situated in the causal analysis of delays, which has received less attention than the aforementioned research fields, and it is seen as an on-going research area [1].

However, in recent years the causal influences of delays in flight networks have been researched by several studies. Fernandes et al. [23] used machine learning models to predict departure delay, but additionally used data-based sensitivity analysis to further analyse what the most influencing factors are on the made predictions. Compared to most other studies in this research area, which often focused on delays in a network of airports or in a geographical region, Fernandes et al. [23] analysed delays of a single airline in the EU, which is similar to the case study of this research project. In a study by Diana [24], latent constructs of flight delays were researched using Confirmatory Factor Analysis. This method can be used to find causal relationships among variables, but also requires to make prior assumptions about the variables and their underlying relationships, which does not allow to find complex and unexpected relationships in a big data set. Additionally, arrival delays and their patterns were researched by Abdel-Aty et al. [25] using frequency and regression analysis. The results obtained from the regression analysis were found to be poor, and therefore more advanced statistical techniques had to be used, namely logistic regression and the analysis of variance (ANOVA) method, which led to better results.

Sternberg et al. [1] has used the data mining method association rule mining using the Apriori algorithm for finding the main reasons of departure delay at the largest Brazilian airports. Similarly, Proença et al. [26] have also studied flight delay patterns by using diverse subgroup set discovery algorithms, which is another data mining method. This study focused on flight information that is available six months prior to operation, with the aim to be able to avoid delays by making changes in block times and crew schedules. These data mining methods allowed to expose underlying and more complex patterns, compared to a conservative statistical correlation analysis. Additionally, recent research by Rodriguez-Sanz et al. [27] and Truong [28] has focused on the discovery and analysis of the main drivers of delays using a Bayesian network analysis. This is a probabilistic graphical model, which allows to find and quantify complex and hidden dependencies among the variables, by using inference methods and sensitivity analyses.

The novelty of this research work lies in the combination of focusing on ATC delays specifically, whereas most of earlier studies only analysed general arrival and/or departure delays, which is combined with a focus on a single airline, namely KLM Royal Dutch Airlines. This permits to take into account the specific processes, practises and data of the airline. Additionally, two methods are used for causal analysis, namely association rule mining and a Bayesian network, which were found to be most suitable for this application from the performed literature review. By comparing their findings and results, the performance and value of these methods for causal analysis can be investigated.

The content of the remainder of this research paper is as follows. In section II the methodology is presented, which consists out of the description of the data acquisition and processing, and the selection and implementation of the causal methods. This is followed by the presentation of the results in section III, which are discussed in detail under section IV. Finally, the conclusions on this research work are drawn in section V, which is followed by recommendations for future work.

II. Method

This section presents the developed methodology of this research. This could be split into 2 different phases. First, the required data and its sources had to be identified, acquired and preprocessed. Additionally, the methods used for causal analysis had to be selected and developed. As a baseline method,

Association Rule Mining (ARM) was selected, whereas a Bayesian Network (BN) was used as the state-of-the-art model.

A. Data Acquisition & Preprocessing

In order to be able to discover complex patterns in the data, information from several sources, both public and proprietary, was carefully selected and acquired. The available data ranges from the 15th of November 2018, until the 31st of December 2019, which spans a total of 412 days. These dates were selected both due to availability constraints, and as the airspace was at its most congested point in this period, before the relapse of aviation in the year 2020 due to the enormous impact of the Covid-19 pandemic. The following data sources were made available for this research and used:

- KLM flight data
- KLM route data
- CDM data
- Operational data from the Dutch ANSP
- Weather data in the Dutch airspace
- Dutch FIR data

Collaborative Decision Making (CDM) is the process of sharing information on the operational processes between multiple stakeholders and operators at an airport, with the goal of enhancing informed decision making for all parties [8]. This source contained information on changes in flight state, and milestones in the CDM process such as the TOBT and TSAT, which was only available for flights handled by KLM ground services. From the KLM data base, individual flight information was retrieved, ranging from identification data to performance related information. Additionally, this data could be used to link the flight data to the other data sources. The KLM route data was retrieved from a separate data base, which contained all waypoints included in the flight's flight plan. From the Dutch ANSP, the LVNL, information on runway usage, capacity declarations, regulations and demand on arrival, departure and the Initial Approach Fixes (IAF) was retrieved. In these data, all flights at the airport were included, where non-KLM flight information was made anonymous. Finally, hourly weather information was gathered on all available weather stations in the Netherlands from the KNMI, and information on the layout of the Dutch Flight Information Region (FIR) such as waypoint locations and sector definitions were found from the public Aeronautical Information Publication.

Most of the data in these sources required individual preprocessing before the data could be integrated on the KLM flight data. This was mostly the case for the CDM, operational and weather data. The operational data was mostly transformed into data per 20 minute interval, such that this could be merged with the flight data. The CDM data set had a vertical data format, and thus the transformation mostly concerned extracting features that allowed the data set to be represented as a horizontal data set, meaning that each flight represented one row or observation. Additionally, the weather information from all stations were aggregated using the weighted average method to hold weather information on each sector in the Dutch FIR, with separate weather information for Amsterdam Schiphol Airport. After this, all resulting data sets were integrated with each other. Finally, the inbound and outbound flight data sets were linked to each other, such that one data set was obtained, where one observation represented one turnaround operation of a KLM flight at Schiphol Airport. The total number of observations in the data set was 139,177 at this point. This obtained data set then needed to be preprocessed before it could be used as input into the causal methods. This included outlier removal, handling missing values, uni- and multivariate discretization, and finally feature selection.

The first step that was taken was to remove the observations which missed values for a substantial part of the columns. This step removed observations in particular that were missing data from a specific source, which resulted in a large number of missing values for the same columns. This was done for each of the sources, such that the observations that missed operational, CDM and/or weather data could be directly removed. This resulted in a data set with length 131,045 which means that 5.8% of the observations were removed.

Secondly, outliers in the data set needed to be identified and removed, in order to prevent the bias introduced by these values to be retrieved in the found results [29]. A basic and very common method in the detection of outliers is by considering all values that deviate more than 3 times the standard deviation, σ, from the mean of the variable, μ, shown in Equation 5 and 6 [29]. If the underlying data is normally distributed, this will result in keeping 99.7% of the observations. This type of outlier detection is thus parametric, as it makes assumptions on the distribution of the data.

$$\text{outlier} > \mu + 3 \cdot \sigma \tag{5}$$

$$\text{outlier} < \mu - 3 \cdot \sigma \tag{6}$$

However, not all variables that had to be checked for outliers could be assumed to have a normal distribution. Therefore, outlier detection was also done by performing clustering, where the data records that fall outside of the identified clusters are considered as outliers [29]. Multiple types of clustering algorithms exist but for outlier removal in particular, it is beneficial to use a density based cluster, which works by searching for regions with a high density of observations. By exploring the density of the data, these methods do not require the user to set the number of clusters, and observations which are located in low density regions are automatically classified as outliers. Therefore, the DBSCAN clustering algorithm, a commonly used density based method, has been used in order to identify outliers. The disadvantage of density based clustering algorithms is that in contrast to the well-known K-Means clustering algorithm for example, it has a higher computational complexity as it compares all points in the data set to every other point. This disadvantage however did not outweigh the added benefits of using the DBSCAN algorithm for outlier identification [29].

Finally, after all outlier detection and removal steps, which are summarised in Table I, the data set comprised of 122,436 rows, which each represented 1 turnaround procedure of a

TABLE I
SUMMARY OF OUTLIER REMOVAL ACTIONS.

Explanation	Outlier removal method
Air Traffic Control delay	Standard deviation
Difference departure and ATC delay	Standard deviation
Arrival delay	Standard deviation
Slot delay evolution + Slot delay dynamics	DBSCAN algorithm
Pure ATFM delay	Standard deviation
Waiting for departure	Standard deviation
Waiting for ground handling + actual taxi in time	DBSCAN algorithm
Delay induced by regulation	Standard deviation + DBSCAN algorithm

KLM flight at Schiphol Airport. This indicates that 6.6% of the input data set was classified as an outlier.

After outlier removal, 6.9% of the 122,436 flight observations had one or more missing values in the attributes. Missing values in data can be handled in several manners, such as removal or imputation, and depends heavily on the size and type of data set [29]. One of the possibilities is to simply remove the observations with the missing values. This approach can be used if there are sufficient observations available, and when the observations with the missing values are drawn from the same distribution as the complete data set. Therefore, the distribution of the data points with missing values was analysed and it was decided to remove these observations, as they represented the same distribution of values as the full data set. Finally, all data cleaning steps and their impact on the data set size are summarised in Table II.

TABLE II
SUMMARY OF DATA PROCESSING STEPS AND NUMBER OF REMOVED OBSERVATIONS.

Processing step	% removed	Resulting number of observations
Initial cleaning	5.8	131,045
Outlier detection	6.6	122,436
Handling missing values	6.9	113,939

As a next step in the data processing, a redundancy analysis was done among the features, meaning that the pairwise correlation between all features has been computed and analysed. This was done to minimise the redundancy in the features included in the final data set, which should be combined with selecting the most relevant features for the model [30]. This redundancy analysis was performed before the variables were discretized, as the variables hold the most information in their original format. This data processing step allowed to identify variables which held very similar information and thus only one of the similar variables should be further considered in the data set. For the continuous variables, the Spearman correlation coefficient, r_s, using the ranked difference between two variables D_i (Equation 7 [31]), was used, whereas the correlation among the discrete attributes was analysed using the entropy based measure Mutual Information (MI) (Equation 8 [30]). The variables which were found to be redundant as they had a high correlation with another attribute and similar correlation values with respect to other variables are shown in Table III.

$$r_s = 1 - \frac{6 \sum_{i=1}^{N} D_i^2}{N^3 - N} \tag{7}$$

$$MI(X;Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \tag{8}$$

TABLE III
REMOVED REDUNDANT ATTRIBUTES.

Deleted variable(s)	Redundancy with
Difference dew point & temperature	Horizontal view
Number of runway operations+ demand values	The runway usage rate
Delay duration of specific delay codes	Arrival/Departure delay
Doors closed delay inbound flight	Arrival delay
Regulation indicators	Reason for regulation
CFMU-reason for regulation	IATA delay code for regulation
Indicator of arrival, departure or en-route regulation	IATA delay code for regulation
Number of TTOT updates	Number of TSAT updates
Difference TTOT and Target TTOT	ATC delay
First/last waypoint used	Sector used
Number of CTOT updates inbound	Pure ATFM delay inbound
Waiting for departure	ATC delay
Time between ready and start taxi	ATC delay
Difference turnaround time planned and actual	Arrival delay, ATC delay
Hour of operation	Time of day operation
Month of operation	Season
Horizontal view & ceiling	Visibility condition
Demand on IAFs	Capacity and demand imbalance on IAFs

The next step in the data processing phase was to discretize the continuous variables, as both the selected methods are limited to be used with solely discrete data sets. When discretizing continuous variables, it is evident that information will be lost. Therefore, it is important to minimise the information loss by using the most suitable method and number of categories. In this research, the data was discretized using the K-Means clustering algorithm. In K-Means clustering, the centroids of the clusters are initiated randomly, after which the observations are assigned to the cluster of the nearest centroid, measured in euclidean distance. After this, the centroids are updated to be the mean value of all observations belonging to the cluster, which is repeated until the clusters and their centroids are no longer alternated [29]. However, this clustering method requires the number of clusters to be specified by the user. The optimal number of clusters was determined by selecting the number of clusters which returned the maximum point of curvature on the graph in function of the cluster's inertia. The inertia of a clustering measures the total squared distance between the data points and the centre of their assigned cluster, and can thus only decrease for an increasing number

of clusters. By using the point of maximum curvature, the number of clusters is optimised as adding extra clusters does not lead to major improvement in the inertia of the obtained clusters.

K-means clustering was used for univariate and multivariate clustering. Univariate clustering indicates that only 1 variable is used as input into the algorithm, and that thus the output consists of that variable being discretized into bins. For multivariate clustering, several variables are used as an input for the clustering algorithm, meaning that the result is a single clustering label, which now holds information on the values of all input variables. This is also known as variable aggregation, and helps to reduce the amount of features present in the data set. The following aggregated features were created:

- *Wind speed*: aggregation of the average wind speed and the speed of the maximum wind gust.
- *Wake turbulence category*: aggregation of the percentage of light, medium and heavy category flights in 20 minute time interval of the flight's departure.
- *Demand, capacity and difference capacity and demand on the IAFs*: demand, capacity and difference capacity and demand of the three individual IAFs in the Dutch FIR.
- *CDM stability*: aggregation of the percentage of TSAT and TOBT adherence and TOBT stability of the flights in 20 minute time interval of the flight's departure.
- *Average CDM updates*: aggregation of the average number of TSAT and TOBT updates of the flights in 20 minute time interval of the flight's departure.
- *Slot delay information*: aggregation of the difference between the first and last slot delay of the flight, as well as the difference between the maximum and minimum assigned slot delay of a flight.

Additionally to the K-Means clustering method, binning by equal frequency was also used, for specific variables. This method returns bins with a varying width, but each bin contains approximately the same number of observations. This method was applied in particular on the target variable, ATC delay, in order to ensure that each bin would occur frequently enough to be analysed by the causality methods. In this method, the number of bins was determined manually, and in case of the ATC delay variable set to be 4, such that each of the bins of ATC delay would occur approximately 25% of the time. Additionally, the pure ATFM delay and company induced delay were forced into a discretization of 0 and >0. For the variable relating to the regulation rate, the found bins were slightly adjusted to have 0 as a separate category as this indicated that no regulation was active.

The result of the discretization process is either a single label, for multivariate clustering, or in the case of univariate discretization an interval, where this interval denotes the range of values included in the bin.

Finally, the features which held the most relevant information in relation to the ATC delay had to be selected to be used as input for the causal analysis methods, as their computational complexity grows exponentially with the number of variables. The final step in the feature selection process was to select the features with the highest dependence or correlation with ATC delay, in order to discover complex and statistically significant patterns in the data. This was again done using the Mutual Information criterion, only this time the metric was only computed between each feature and the target variable, ATC delay. From these results, the features were ranked on score and the top 30 features were selected, which was a result of a trade-off between including as much information as possible, model complexity, and computational load. However, these attributes were not the final selection. Additional attributes of interest from an airline perspective, and which did not belong to the best 30, were manually selected to be kept in the data set, and are listed below.

- *Weather information at departure (visibility, wind speed & direction) and visibility at arrival*
- *TSAT adherence of individual flight, and of flights 20 minute time interval of the flight's departure (CDM stability)*
- *Difference capacity and demand at arrival/departure*
- *Aircraft type*
- *Weekday of departure*
- *Short turnaround time indicator*
- *Peak indicators at arrival/departure*
- *Waiting for ground services & boarding duration*

The final data set consisted out of 46 variables in total. The majority of the used variables are straightforward and could be directly found in the data sources. However, part of the used variables are constructs of the original variables and require some additional explanation.

- *TOBT adherence:* True when no TOBT update larger than 5 minutes is made in the last 20 minutes before TOBT [32].
- *TSAT adherence:* True when the aircraft has called ready within the TSAT window (+- 5 minutes of TSAT) [32].
- *Actual runway usage rate:* Computed by taking the moving average departure or arrival rate by taking the mean time between the flights in a time window of 20 minutes, and the runway usage rate is then 60 minutes divided by this average timing between flights.
- *Difference capacity and actual runway usage rate:* The difference between the declared capacity in aircraft movements per hour and the actual usage rate of the active runways in aircraft movements per hour.
- *The number of updates of CDM milestones:* This variable exists for multiple milestones, such as the TOBT, TSAT, and CTOT. This variable is simply the count of the number of updates a flight has received for the respective CDM milestone in its turnaround/departure process.
- *Peak runway indicators:* These variables indicate whether the inbound/outbound flight arrived/departed during an outbound, inbound, combination of inbound and outbound or off peak period.
- *Regulation delay key:* For both inbound and outbound flights, a regulation can be assigned. The reason for this regulation is translated in an IATA delay key, which are standardised delay codes, and also exist for specific ATFM delays. The explanation of these codes can be

found in Table IV [33].

- *Average startup delay:* This variable indicates the average startup delay of the departed flights in the 20 minute time window, 10 minutes before and after the observed flight departed.
- *Ratio of regulated flights:* This variable indicates the ratio of regulated flights to the total number of departed flights in the 20 minute time window, 10 minutes before and after the observed flight departed.
- *Pure ATFM delay:* The delay caused by the regulation of the flight, caused by the initial issued CTOT. Additional delay caused by the airline not being able to comply with this CTOT is not counted as pure ATFM delay.
- *Company induced delay:* This is the delay that is caused by the airline, because the flight cannot adhere to its issued CTOT in case of a regulation. The additional delay caused by changing the CTOT is then assigned as company induced delay.
- *Regulation induced delay:* The difference between the departure delay before regulation, and the departure delay after the flight was regulated.
- *All doors closed delay:* The difference between the actual all doors closed moment and scheduled all doors closed moment.
- *Difference TTOT and Optimal TTOT:* The difference between the actual Target Take-off Time and the optimal TTOT of the flight. When positive, this means that the TTOT was later than the optimal TTOT and vice versa. The optimal TTOT in the CDM process is the time that is optimal for that flight, and is issued in order to be able to get an earlier slot [34].
- *Boarding duration:* The duration between the flight state changed to boarding, and the change to flight state gate is closing.
- *Visibility:* This variable is an aggregation of the horizontal view and the ceiling. The states are good, marginal and Low Visibility Procedures (LVP), and are based on the conditions defined by the Dutch ANSP.

TABLE IV
IATA DELAY CODES & REASONS [33].

Delay code	Description
81	ATFM due to ATC en-route demand/capacity
82	ATFM due to ATC staff / equipment en-route
83	ATFM due to restriction at destination airport
84	ATFM due to weather at the destination
98	ATFM due to industrial action outside own airline
99	ATFM due to unknown reason

B. Association Rule Mining

As a baseline method, Association Rule Mining (ARM) also known as Frequent Pattern Identification (FPI) was used. This is a knowledge discovery in databases method, which is done by identifying recurring or frequent patterns in the input data set [29, 35]. ARM has been used in a previous study by Sternberg et al. [1], to determine and analyse frequent patterns in departure delays at Brazilian airports. This data mining method allowed to identify underlying associations and correlations in the data set, which were missed by the statistical correlation analysis in that same research paper.

This data mining method allows the algorithm to discover *association rules*, which can be presented as A $\rightarrow$ B, where A is also known as the antecedent and B the consequent [1]. This example is an association rule of length 2, but also multiple conditions can be present in both the antecedent and consequent, for example A, B$\rightarrow$C, which is a rule with length 3. The standard input data format is a transactional database, which implies that every column is binary, indicating whether the condition is observed for the data record. Not every data set comes in this standard format, and therefore the discretized data set used in this research needed to be one-hot-encoded, before it could be used in the ARM algorithm. When one-hot-encoding the 46 variables included in the final data set, this led to 178 columns. Furthermore can one distinguish two steps in ARM. The first step in this data mining method is to find the frequent item sets, which are sets of variables that frequently occur together in the data set. Once these frequent item sets have been found, the association rules are made by permuting the found conditions in formats of antecedents and consequents, which is the rule generation step. In order to perform ARM, several algorithms have been developed [29]. These algorithms differentiate themselves in the manner they obtain or mine the frequent item sets from the input data. The most known include Apriori, FP-growth and Eclat [36, 37, 38]. Both Apriori and FP-growth are suitable for horizontal data sets, and thus for this application. In FP-growth, the frequent item sets are found by using the observational data as a starting point, and building a *frequent pattern tree* or *FP-tree*. This approach eliminates the need to generate and calculate the parameters for each possible association rule, compared to the Apriori method. Therefore, the FP-growth algorithm is used in this research, as it is more scalable to use on large data sets and computationally faster [39].

This data mining method has been implemented into a Python environment using the open source package *mlxtend* [40]. The final algorithm used in order to mine frequent patterns from the data set is shown in algorithm 1, and the second step in this algorithm, FP-growth, is explained in more detail in algorithm 2. It can be seen that the results of algorithm 1 contain the total set of association rules, R, and the filtered rules, $R_{filtered}$. The filtering is done based on the target variable t, which is the ATC delay, as defined by the scope of this research. The filtering function returns the set of rules in which the consequent consists of a condition of the ATC delay, such that $R_{filtered}$ only contains patterns leading to this variable.

However, one of main challenges in this method remains to manage the search space, as the number of possible rules grows exponentially with the number of variables and their cardinality [35]. Another challenge is the identification of the rules which hold the most information, as the found number of rules is often very large. In order to manage these aforementioned challenges, the method uses different hyperparameters and measures in order to limit the search space and to identify the most relevant association rules.

Algorithm 1: Association rule mining algorithm.

Input: Data set D, minimum support $MinSup$, minimum confidence $MinConf$, maximum length rules K, target variable t

1) $D_{encoded}$ = One-Hot-Encode(D)
2) FI=FP-growth($D_{encoded}$,$MinSup$,K)
3) R,$R_{filtered}$=association rules (FI,$MinConf$,t)

Output:Association rules R, filtered association rules $\mathbf{R_{filtered}}$

Algorithm 2: FP-growth algorithm details.

Input: Encoded Data set $D_{encoded}$, minimum support $MinSup$, maximum length rules K

1) $Support_{variables}$=Calculate support of each variable in data set ($D_{encoded}$)
2) D_{ranked}=rank data observations from high support to low ($D_{encoded}$,$Support_{variables}$,$MinSup$)
3) FP-tree,FI= Build tree (D_{ranked},K)

Output:Frequent item sets FI

Rule length The rule length is the maximum length of a frequent item set, and consequently an association rule. For a low value of this parameter, the rules are more simple and general. For increasing rule length, the discovered rules become more complex, as more conditions can be included, and more rules are discovered. Therefore, this parameter should be optimised carefully as setting it too low may result in only finding simple patterns, whereas a too high value will return too many rules, which makes it hard to filter to the most relevant ones.

Support The support of a condition is simply the frequency of the condition in the data set. Thus, the Support(X=x,Y=y) is the frequency of variable X equals x and variable Y equals y. The mathematical formulation of this measure is shown in Equation 9 [29].

Confidence The confidence of an association rule is the conditional probability of the consequent, given the antecedent. For example, Confidence (X=x$\rightarrow$ Y=y), is the conditional probability of Y equals y, when variable X equals x. The confidence value of a rule can be expressed as the ratio of support, as seen in Equation 10 [29].

$$\begin{aligned} \text{Support}(X = x) = P(X = x) &= \frac{\text{Frequency(X=x)}}{\text{Total records}} \\ \text{Support}(X = x, Y = y) &= P(X = x \cup Y = y) \end{aligned} \tag{9}$$

$$\begin{aligned} \text{Confidence}(X = x \rightarrow Y = y) &= P(Y = y | X = x) \\ &= \frac{\text{Support}(X = x, Y = y)}{\text{Support}(X = x)} \end{aligned} \tag{10}$$

The aforementioned hyperparameters need to be set by the user. As a trade-off between the number of rules generated and the ability to mine specific rules, the maximum length of a rule was set to be 4. The minimum support value determines how often a condition or pattern needs to occur before it can be considered as frequent, and thus be included in an association rule. Higher support values will return patterns with a high frequency in general, and will reduce the number of found rules. In order to determine this hyperparameter, the method applied by Sternberg et al. [1] was used, where the support is a function of how often a pattern should occur on average per day, expressed in Equation 11. The result of applying Equation 11, considering that the data set is made up of 113,939 observations (N) spread over 412 days, is shown in Table V. It was decided to use an average daily occurrence, denoted as $\mu_{occurrence}$, of 25 as the support limit, leading to a value of 0.09. Finally, the minimum confidence threshold is used in the process of making association rules from the found frequent item sets, as its computation requires an antecedent and consequent. Its minimum value was determined based on the minimum probability of any of the conditions of the ATC delay, such that the rules found have a higher conditional probability than purely the occurrence of delay [1]. The used expression is presented in Equation 12.

$$\frac{\mu_{occurrence} \cdot N_{days}}{N} = \text{MinSup} \tag{11}$$

$$\begin{aligned} P(d_{ATC}|conditions) >= P(d_{ATC}), \\ P(d_{ATC}) = \text{MinConf} \end{aligned} \tag{12}$$

TABLE V
MINIMUM AVERAGE OCCURRENCE PER DAY OF RULE AND CORRESPONDING MINIMUM SUPPORT VALUES.

Occurrence per day	Support
10	0.036
20	0.072
30	0.108
27.65	0.1
24.9	0.09

Next to the measures of support and confidence to represent the strength of an association rule, an additional correlation measure can be used, namely lift, presented in Equation 13 [29]. The lift measure can be used to identify rules which do not actual consist of a causal relationship, but are detected as the consequent has a high support [1]. Therefore, the lift of a rule can also be seen as the correlation, as a lift value of 1 indicates that the conditional probability of $Y = y$ given $X = x$ is equal to the probability of $Y = y$, and thus $X = x$ and $Y = y$ can be said to be independent. Lift values larger than 1 thus indicate a positive correlation, whereas values less than 1 show that the antecedent and consequent are actually negatively correlated [29].

$$\text{Lift}(X = x \rightarrow Y = y) = \frac{\text{Confidence } (X = x \rightarrow Y = y)}{\text{Support } (Y = y)} \tag{13}$$

A common method to interpret and analyse the found association rules is by selecting the most significant rules and analyse them. In this study, the metric used for significance is lift. However, it is also common that this selection of

association rules returns patterns with a lot of overlap between them, meaning that they are not diverse enough to discover all relevant patterns [26, 41]. Therefore, additional measures can be used to ensure analysing the most significant patterns, but also the ones having the least amount of redundancy among the rules. Therefore, a redundancy measure is implemented next to the significance measure, in order to ensure diversity in the top discovered patterns [26, 41]. This redundancy measure should represent the degree of overlap with other frequent patterns, which have a higher significance value. Therefore, it counts the number of rules with a higher lift value that have the same condition in the antecedent, for each condition in the rule's antecedent. Finally, this is normalised for the number of antecedents of the rule. This implies that the pattern with the highest lift will have a redundancy measure of 0 by default, as no rule has a higher significance than this one.

C. Bayesian Network

As a state-of-the-art method, the Bayesian Network (BN) model was selected. This method has been used in earlier studies researching causes of flight delays by Rodriguez-Sanz et al. [27] and Truong [28]. A Bayesian network was selected as it has been found that it has excellent properties for causal analysis. This is because it can be constructed based on an input data set and identify the conditional dependencies between the variables. Additionally, by using analysis techniques as Bayesian inference, the main causes or drivers of a variable and their influence on that variable can be analysed and quantified [27, 28]. According to Rodriguez-Sanz et al. [27], Bayesian network analysis is a very suitable method to analyse airport saturation or flight delays for multiple reasons. First of all can the found structure be interpreted in a straightforward manner, and can it be used for causal analysis. Secondly is a BN a probabilistic model, which is beneficial as flight delays are a highly stochastic process. Lastly, BN allows to analyse the influence and relationships between multiple variables in the network, which enables the detection of complex relationships and interactions among the variables [27].

A Bayesian network is a graphical presentation of a joint probability distribution [42]. It comprises of two main elements, a graphical structure, which is also known as a Directed Acyclic Graph (DAG), and Conditional Probability Tables (CPT) for every node/variable in the network [9, 27]. The DAG represents which nodes or variables are conditionally dependent on each other by means of links, whereas the CPTs represent the quantitative information, and contain actual probabilities [27]. When constructing a BN, the term structure learning refers to the process in which the DAG is built, which is followed by parameter learning to obtain the CPTs [43].

$$P\left(X_1, \ldots, X_m\right) = \prod_{i=1}^{m} P\left(X_i \mid X_1, \ldots, X_m\right) = \prod_{i=1}^{m} P\left(X_i \mid O_i\right) \tag{14}$$

In Equation 14, the mathematical expression is shown for the joint probability of a network structure with m variables, $P\left(X_1, \ldots, X_m\right)$, and that this joint distribution is equal to the product of the conditional probabilities of the variables, X_i, on the variables in their parent set, denoted as O_i [28, 42]. X_i can also been seen as the child node of the variables in its parent set O_i.

1) Structure Learning: There are multiple ways on how a Bayesian network graph can be constructed. If there is enough knowledge on the problem, or the causal relationships are already known, a Bayesian network can be constructed manually. However, constructing a large network from reference literature and/or expert knowledge is not sufficient to capture the complex and large amount of relationships between the included variables. In this case, the DAG can be learned from pure observational data, for which multiple methods exist [44]. In score based functions, a scoring function is used to measure the degree that the found structure represents the underlying data [45]. However, different structures need to be found in order to evaluate their fit to the data using the scoring function, which is done by the search method. The search method is often a heuristic method, as the number of possible DAGs is very large, and grows exponentially with the number of variables or nodes. The second type of structure learning method is dependency based or constraint based learning, which works with conditional independence tests on the variables in the input data set, and the most commonly used one is the PC-algorithm [44, 46]. Hybrid methods are then a combination of the score based and constraint based algorithms [47, 48]. In hybrid learning, the main idea is to find the undirected links between the nodes, using conditional test of independence as in constrained based methods. The found undirected structure is then the input for the score based function, which optimises the direction of the found links [47, 48]. The hybrid learning algorithm used in this research is based on the algorithm developed by Tsamardinos et al. [47], and is a combination of the constraint based PC-algorithm and hill climb search score based method.

The number of possible DAG structures grows exponentially with the number of nodes or variables, as shown in Equation 15. Therefore, using exhaustive search methods which consider all possible structures becomes computationally impossible when having more than a few variables. As a heuristic search method, hill climb search was used in combination with tabu search. The hill climb method looks for the best improvement in the neighbouring solutions, and is therefore very sensitive to finding local maxima as the best found solution [49]. The addition of the tabu search method allows to save recently visited solutions and exclude them from the possible next states, which helps the hill climb search algorithm to escape from local maximum solutions [49].

$$N_{DAG} = 2^{m \cdot (m-1)} \tag{15}$$

In the research field of BN, there is no consensus on which structure learning algorithm leads to the best results, as this is often highly dependent on the input data set [43, 46, 50]. Overall, for large networks, it is beneficial to use score based learning methods as they make use of heuristics to learn the DAG structure. Considering the size of the network and

the expected amount of edges, only score based and hybrid learning methods have been considered.

In both hybrid and score based learning methods a scoring function needs to be used to optimise the structure of the BN, of which the most commonly used ones are K2, BDeu and Bayesian Information Criterion (BIC). The former two are Bayesian scoring functions, whereas the BIC method is classified as an Information Theory method, which aims to maximise the log-likelihood function of the model on the input data.

In order to select the best performing structure learning algorithm, their performance had to be checked on an independent and separate data set, which has also been used for verification purposes. The used data set was sampled from an open source available Bayesian network with 37 nodes and 46 links, named ALARM, which has been developed by researchers and has been commonly used to test the performance of different structure learning algorithms [51]. As the actual Bayesian network structure is known, the learned structures by the different algorithms could be assessed on their performance. The data set used as an input for the structure learning algorithms was sampled for 100,000 observations, resulting in a data set size of 100,000 x 37, which is in the same order of magnitude of the size of the data set that will be the input to analyse the ATC delays in this research. The Structural Hamming Distance (SHD) was used to asses the performance of the learning method. This metric counts the number of additional, missing, or wrong oriented edges in the learned structure compared to the actual model, meaning that its value should be minimised [47]. The performance of the different structure learning algorithms on this ALARM data set is shown in Table VI. It can be seen that the hybrid algorithm outperforms the score based method on both run time and finding the correct structure (SHD). However, when the hybrid method was applied to the ATC delay related data set, it was found that its computational complexity is heavily dependent on the data set used as input. This could be explained by the fact that there are a lot more dependencies between the variables in the ATC delay related data set compared to the used data set sampled from the reference ALARM network. This led to an increase in the number of conditional independence test that had to be executed, making the hybrid learning algorithm computationally very heavy. Therefore, the hill climb algorithm was selected. From Table VI, it can be seen that optimising the K2 score had the best performance in terms of SHD, and this scoring function was found to perform best for large data sets in earlier studies [52]. Therefore, the K2 scoring function was selected as an optimiser for the structure learning in this research. The formula of the K2 scoring function is shown in Equation 16 [45]. Here, G denotes the DAG, D the input data set, m the number of variables, r_i the number of states of variable X_i, and q_i the number of configurations of the parents set of variable X_i. N_{ij} and N_{ijk} respectively represent the number of instances that the parent set of variable i takes value j, and the number of instances that this happens and the variable X_i takes value k.

Additional to the structure learning method used, other hyperparameters can be specified in the structure learning process: the maximum amount of parent nodes in the network and the nodes which cannot be included in the network, also known as the black list.

$$g_{K2}(G:D) = \log(P(G)) + \sum_{i=1}^{m}\left[\sum_{j=1}^{q_i}\left[\log\left(\frac{(r_i-1)!}{(N_{ij}+r_i-1)!}\right)+\sum_{k=1}^{r_i}\log\left(N_{ijk}!\right)\right]\right] \quad (16)$$

Maximum number of parents The larger this number, the larger to computational complexity of learning the structure and analysing it. Thus, a trade-off must be made between learning the most correct network and computational complexity. From iterations, it was found that the number of conditional dependence relationships, or links, is high in the data set and associated structure. Therefore, it was decided to limit the number of parents to a maximum of 3, as this balanced the representation of dependencies and complexity in the network. Additionally, this is in line with the maximum length parameter of the baseline method, which was set to be 4, meaning that maximum 3 conditions could lead to another condition.

Black list These include all links that cannot be added to the network, and are set by the user. If the black list is empty, the network can learn links that are in reality not possible. For example, information on the outbound flight cannot cause inbound flight information, or flight information cannot lead to the season or aircraft type used. Additionally, as ATC delay is the target variable due to the defined scope of this research, it was chosen to have this node as an ending node, in order to be able to analyse the structure bottom up. Therefore, the following links are included in the black list of the structure learning algorithm, which consisted in total out of 676 links:

- All variables $\rightarrow$ season
- All variables $\rightarrow$ aircraft type
- All variables $\rightarrow$ weekday
- All variables $\rightarrow$ time of the day, peak indicators
- ATC delay $\rightarrow$ all variables
- Outbound related variables $\rightarrow$ inbound related variables

2) Parameter Learning: There are two methods available to learn the parameters of the BN, also known as the conditional probability tables associated with each node [53]. One way to estimate these CPTs is to use Maximum Likelihood Estimation (MLE), which essentially maximises the probability that the data was created by the network structure, $P(data|model)$ [53]. This method is relatively sensitive to overfitting, as the resulting CPTs are fully based on the learned model and data.

An alternative to MLE is to use Bayesian estimation for the conditional probability tables [53]. In this method, initial conditional probability tables are created based on a Dirichlet distribution, which is very often used as a prior in Bayesian probability theory and applications. These CPTs are then updated using the model and the input data, which is more conservative compared to the MLE method, and thus reduces the risk of overfitting the parameters. A drawback of this method is that the equivalent sample size of the Dirichlet distribution needs to be set, which has a large influence on

TABLE VI
RESULTS OF DIFFERENT STRUCTURE LEARNING ALGORITHMS ON THE ALARM DATA SET, N=100,000, P VALUE=0.05.

Algorithm	Run time (s)	SHD (-)	Scoring function (-)		
			K2	BDeu	BIC
Hill climb K2	934.0	26	-941.798E3	-941.588E3	-944.809E3
Hill climb BDeu	765.1	35	-941.636E3	-941.224E3	-944.282E3
Hill climb BIC	731.8	34	-942.651E3	-942.252E3	-944.317E3
Hybrid K2	576.7	18	-959.115E3	-958.928E3	-959.622E3
Hybrid BDeu	656.5	17	-959.124E3	-958.928E3	-959.622E3
Hybrid BIC	627.7	17	-970.762E3	-970.488E3	-971.404E3

the obtained results [54]. Therefore, it was assumed that in this research there is sufficient data available to prevent overfitting of the conditional probability tables to the data, and the MLE method was adopted.

3) Inference: Once the Bayesian network structure and parameters have been learned from the data set, the network can also be used to better understand and quantify the causal relationships and interactions between the different variables by applying inference [28]. Inference can be used to make predictions on missing variables and to perform a sensitivity analysis by quantifying the effect of fixing certain variables in the network on the probability distribution of other variables in the network [27]. The following definitions are used in inference theory [55]:

Query A set over which the posterior probability must be computed in the inference method, given the evidence. Example : $P(U|X = x)$

Evidence The set of variables that is fixed in the query, and thus serves as input to the inference algorithm. Example: $X = x$ in $P(U|X = x)$

Inference can be done in two ways, namely exact and approximate inference [53]. A commonly used method to perform exact inference is Variable Elimination. In this method, the posterior probability is found by working through the network starting from the evidence. Here, each time a summation is done over all states of a variable, it is eliminated from the distribution [53]. By doing so, the inference algorithm is working itself through the network structure, starting from the nodes present in the evidence. This method can become very computationally expensive for complex networks as an increasing number of joint and conditional probability distributions need to be computed, over a large number of nodes.

When the BN grows in number of nodes and links, exact inference becomes too computationally complex, and approximate inference methods can be used in its analysis. Approximate methods use sampling to perform inference, and one of the most commonly used approximate inference methods is likelihood weighting [55]. In this method, the evidence variables are fixed to a certain state as specified in the input evidence. Additionally, the other nodes are sampled based on the learned conditional probability tables. However, as the evidence variables are fixed to a certain state, they cannot be sampled according to their CPTs based on the values of their (possible) parent nodes. Therefore, the likelihood weight, L, is introduced for each sample, which reflects the probability of the evidence variable taking this state, based on its parent configuration [55]. The posterior probability distribution of a variable given the evidence can then be obtained by dividing the sum of the likelihood per state of the variable by the total sum of the likelihood in the sampled data set, as illustrated in Equation 17 [55]. However, in order to obtain a good approximation of the real probability distribution, the required sample size had to be determined, such that the law of large numbers would apply. The required sample size was determined by producing probability distributions of the target variable for different sample sizes under the same set of evidence. The evidence was varied along all possible states of the variables arrival delay, TOBT updates, the difference between the capacity and actual used runway rate during departure, and the wind speed category during departure. These variables were chosen due to their diverse positions in the learned Bayesian network, as will be seen in the result section. The result for the variable arrival delay at the state [1;39] is plotted in Figure 1. Considering all included variables and states, it was found that for each of these specified evidence values the probability distribution stabilised at 100,000 samples. Therefore this is the required sampling size to obtain valuable results.

$$P'\left(X = x_i \mid E = e\right) = \frac{\sum \prod_j L\left(E_j = e_j\right) \text{ if } X = x_i}{\sum \prod_j L\left(E_j = e_j\right)} \quad (17)$$

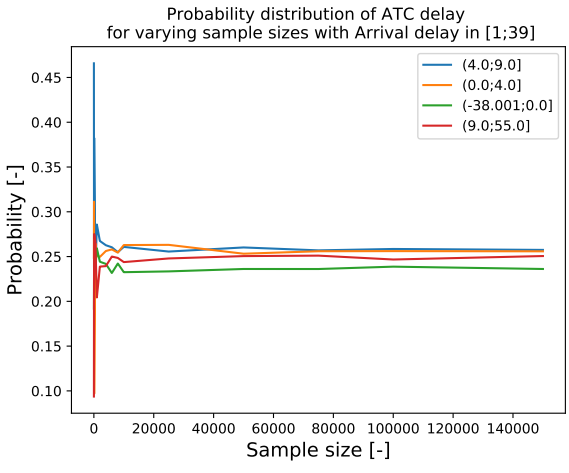


Fig. 1. Determination of the required sample size in inference using arrival delay [1;39] as evidence.

Additional to inference, it is interesting to look at the Markov Blanket (MB) of the variables of interest. The MB is the subset of variables that is needed to infer the value of the variable of interest, and thus the subset of variables

that makes that particular variable independent of all other attributes in the BN [42]. In Bayesian networks, this subset of variables consists out of the parent nodes, the children nodes and their respective parent variables.

Finally, all the aforementioned processes and steps in BN modelling were implemented into the software environment used, Python. This was done making use of the open source module *pgmpy*, which consists of dedicated functions to construct and analyse Bayesian networks [56].

D. Verification

After implementation of both the FP-growth ARM and Bayesian network methods, verification was necessary to ensure correctness.

The verification process of the FP-growth methodology was done by using a very small example data set on the method, which allowed to calculate all support and confidence values and construct the FP-tree manually. The used data set consisted of 3 observations and 5 variables. The manually calculated results could then be compared to the results found by the implemented Python algorithm. It was found that the manually obtained results matched the results of the computer model, meaning that the FP-growth method works correctly and is thus successfully verified.

For the Bayesian network methodology, verification was done using standard BNs available from Bayesian model software modules in R and Python [57]. It was chosen to select two verification models, one named 'Cancer', categorised as a small network with 5 variables and 4 edges, and 'ALARM', categorised as size medium with 37 variables or nodes and 46 edges [51, 55]. By using these standard models, the true Bayesian networks are known and available. By sampling data according to the DAG and the corresponding CPTs, data sets which represent the BNs could be obtained. Subsequently, these sampled data sets were used as input for the BN model learning algorithms, which allowed to compare the learned structures to the true models. Several metrics can be used to assess the quality of the learned structure. Tsamardinos et al. [47] developed and used a measure for assessing the quality of fitted and learned structures from data, namely the Structural Hamming Distance (SHD). This metric counts the number of additional, missing, and wrong oriented edges in the learned structure compared to the actual model [47]. A learned structure representing the actual model well will have a low value for this metric, whereas structures with less good solutions will have higher values for the SHD.

All learning methods were applied to the data sets sampled from the small and medium reference networks. It was found that the algorithms worked as expected. The exhaustive search method on the small network for example had a much higher run time, but also outperformed the hill climb method. Next to this could it be observed that the hybrid and score based functions returned different solutions, as expected. This implies that the algorithms have been correctly implemented and thus that the method is verified.

E. Validation

After verification, validation was required to ensure that the results obtained by the methods are credible. The approach taken in this research is a common method in data mining methods, namely the use of a separate training and test data set, also known as hold-out [29]. These data sets need to be the same for both methods, in order to be able to compare their performance. It was determined to use a training and test data set ratio of 90% to 10%, which resulted in a data set size of 102,545 and 11,394 observations respectively.

Earlier studies using ARM for data mining purposes, albeit with different applications, have used this validation method and found it most suitable for this method [58, 59]. In order to validate the results found by ARM, the test set is also used as an input data set and the same algorithm is run on it. The found association rules from both the training and test data set are compared and if the statistical measures such as support and confidence deviate for more than a set threshold, the rule is seen as unstable and thus not validated. As a threshold, 5% was used, which has been commonly used in the validation process of association rules [60].

The results of the validation process are shown in Table VII. It was found that more than 97% of the rules in both the training and test set were also found in the association rules of the other data set. From these corresponding rules, none of them had a deviation in the support and confidence values of more than 5% between the test and training set, meaning all of the common rules between the test and training set were validated. These results imply that the found association rules are present along the entire data set and thus are actual frequent patterns. From this point, only the validated set of association rules will be used in the result and discussion section.

TABLE VII
RESULTS OF THE ARM VALIDATION.

% rules validated training set	**% rules validated test set**
98.0	97.7

For BN, prediction can be used for validation, which is most commonly used for this method [55]. Here, the test data set is used as input to the prediction process, which is a more classical validation approach for machine learning models [28]. In order to make predictions on a single target variable, the BN uses the Markov blanket of that variable. This is possible as all the variables in the Markov blanket are known from the test data set, which makes the target variable independent of all variables in the structure which are not included in the Markov blanket [42]. As a Bayesian network is inherently a probabilistic method, it cannot return a single value, but always returns a probability distribution. Therefore, the predictions can be made according to the Maximum A Posterior (MAP) method, or stochastic. The MAP method implies that the state of the variable that is being predicted is determined by the state with the highest probability in the posterior distribution resulting from inference. The stochastic prediction method computes the probabilities of the states in the same manner,

however, the predicted state is determined stochastic using the computed probabilities. In this validation process, the predictions are done according to the MAP method, as this generally results in the highest obtained predictive accuracy.

The accuracy of the predictions is expressed as a percentage, as this is an application of classification due to the discretization. For the target variable ATC delay, an overall predictive accuracy of 47.2% was obtained. Additionally, the confusion matrix on the made predictions could be used to analyse the prediction accuracy in more detail, which is presented in Table VIII. The values in the matrix are again expressed as a percentage, such that each row adds up to 100. It can be seen that in the case the BN does not predict the correct class of ATC delay, the highest percentage of falsely made predictions for each of the bins is found in the nearest neighbouring bin.

TABLE VIII
CONFUSION MATRIX FOR THE PREDICTION ON THE ATC DELAY VARIABLE IN PERCENTAGE.

		Predicted values			
		<0	(0,4]	(4,9]	>=9
Actual values	<0	25.5	56.2	15.8	2.5
	(0,4]	0.3	66.7	25.5	7.5
	(4,9]	0.2	41.7	42.0	16.1
	>=9	1.1	7.6	36.0	55.3

However, when all the data is present except for the target variable, which is being predicted, only the variables in the Markov blanket of ATC delay will determine the prediction. Therefore, only a small part of the BN is validated using this method. For this reason, other variables in the DAG have also been predicted using the hold out test set, and the resulting predictive accuracy is shown in Table IX. These variables were selected to be the number of TOBT updates, arrival delay, difference between capacity and actual usage rate during departure, and wind speed during departure due to their different positions in the structure, such that other parts of the structure could be validated. It can be seen that for all these variables, the predictive accuracy exceeds that of the ATC delay. This might be explained by the position of these variables in the network compared to the ATC delay variable. When making predictions using a BN, all variables are known, except the variable that is being predicted. The value of the predicted variable is thus completely defined by the variables in the Markov blanket of the predicted variable, as the definition states. As the ATC delay variable only has three parent nodes and no children nodes, which are two implications of using the black list and limiting the number of parents in the network, these three variables determine the value of the ATC delay completely. The other predicted variables in the network also have only three parent nodes, but they also have children nodes, which are also included in the Markov blanket, as well as their parent nodes. Therefore, many more variables are included in the determination of these variables. It can thus be concluded that the limitations used in the structure learning of the BN have limited the predictive accuracy of the target variable.

TABLE IX
PREDICTIVE ACCURACY OF OTHER KEY VARIABLES IN THE BN.

Variable	**Predictive accuracy %**
TOBT updates	80.2
Arrival delay	79.5
Difference capacity and actual runway usage during departure	71.9
Wind speed during departure	54.0

III. RESULTS

A. Association Rule Mining

After performing validation, the number of validated frequent patterns was found to be 3,341,323. When filtered to only contain patterns leading to the ATC delay variable, 13,002 association rules remained. Figure 2 displays the average metrics of the filtered association rules, which only contain consequents consisting of the ATC delay variable. It can be seen that the average support values are almost equal for all four classes, but in the lift and confidence metrics a trend can be observed. The lowest class of ATC delay both has the highest confidence as well as lift. One of the middle classes, with values between 4 and 9 minutes, has the lowest values in confidence as well as lift, meaning that these patterns are less strong on average than the other classes of ATC delay.

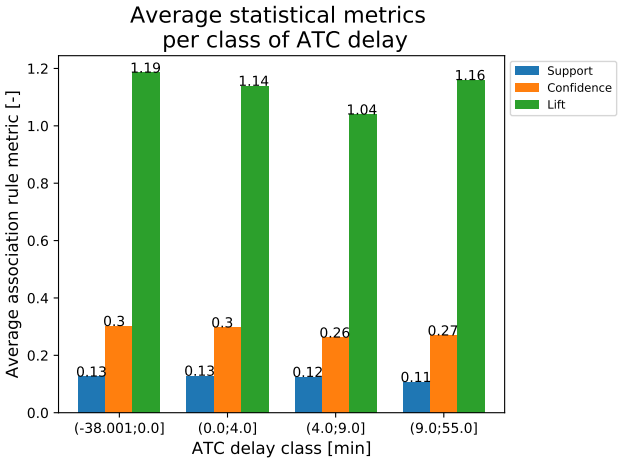


Fig. 2. The average metrics of the filtered association rules per class of ATC delay.

Table X shows the top 10 discovered frequent patterns based on the lift value. It can be observed that these rules indeed have a lot of commonalities in terms of conditions present in the antecedents and consequent. This is a known phenomenon, as it is very common in data mining methods that the most significant association rules or patterns are found to be redundant and overlapping [26, 41]. All these 10 most significant rules contain the conditions of TOBT adherence and no IATA delay key assigned, which in combination with various other conditions lead to an increased probability of having an ATC delay below 0 between 130 and 144%, compared to the probability of an average flight having below 0 minutes ATC delay. The antecedent conditions additional to the TOBT adherence and the delay key mostly relate to

TABLE X
THE 10 ASSOCIATION RULES WITH HIGHEST LIFT VALUES.

Antecedent	Consequent	Support	Confidence	Lift
Delay key category dep = None ^TOBT adherence = True ^TSAT updates = [0.0;14.0]	ATC delay = (-38.001;0.0]	0.10	0.62	2.43
Delay key category dep = None ^TOBT adherence = True ^Average startup delay = [-281.5;5.5]	ATC delay = (-38.001;0.0]	0.09	0.62	2.43
Delay key category dep = None ^TOBT adherence = True ^TOBT updates = [0.0;1.0]	ATC delay = (-38.001;0.0]	0.11	0.60	2.36
Delay key category dep = None ^TOBT adherence = True ^Average CDM updates = 0	ATC delay = (-38.001;0.0]	0.09	0.60	2.34
Delay key category dep = None ^TOBT adherence = True ^Pure ATFM delay dep = 0	ATC delay = (-38.001;0.0]	0.12	0.59	2.33
Delay key category dep = None ^TOBT adherence = True ^Regulation induced delay = [-6.0;2.0]	ATC delay = (-38.001;0.0]	0.12	0.59	2.31
Delay key category dep = None ^TOBT adherence = True ^Company induced delay = 0	ATC delay = (-38.001;0.0]	0.12	0.59	2.30
Delay key category dep = None ^TOBT adherence = True ^All doors closed delay = [-35.0;3.0]	ATC delay = (-38.001;0.0]	0.12	0.59	2.30
Delay key category dep = None ^TOBT adherence = True ^TSAT adherence = True	ATC delay = (-38.001;0.0]	0.12	0.59	2.30
Delay key category dep = None ^TOBT adherence = True	ATC delay = (-38.001;0.0]	0.12	0.59	2.30

TABLE XI
THE 10 ASSOCIATION RULES WITH LOWEST REDUNDANCY VALUES.

Antecedent	Consequents	Support	Confidence	Lift	Redundancy
Delay key category dep = None ^TOBT adherence = True ^TSAT updates = [0.0;14.0]	ATC delay = (-38.001;0.0]	0.10	0.62	2.44	0
Season = S	ATC delay = (9.0;55.0]	0.11	0.28	1.22	0
Arrival delay = [1.0;39.0]	ATC delay = (9.0;55.0]	0.10	0.28	1.20	0
Short turnaround time = 0	ATC delay = (9.0;55.0]	0.09	0.27	1.17	0
Delay key category dep = None ^TOBT adherence = True ^ Average startup delay = [-281.5;5.5]	ATC delay = (-38.001;0.0]	0.09	0.62	2.43	0.67
Average usage rate of runways dep = [60.7;90.1]	ATC delay = (9.0;55.0]	0.10	0.32	1.37	1
TSAT updates = [15.0;30.0]	ATC delay = (9.0;55.0]	0.10	0.31	1.35	1
Time of day = midday	ATC delay = (0.0;4.0]	0.09	0.27	1.04	1
TOBT updates = [0.0;1.0] ^Delay key category dep = None ^TOBT adherence = True	ATC delay = (-38.001;0.0]	0.11	0.60	2.36	1.33
Pure ATFM delay dep = 0 ^Difference TTOT and optimal TTOT = [-5400.0;-239.0] ^ TSAT updates = [0.0;14.0]	ATC delay = (-38.001;0.0]	0.09	0.56	2.20	1.33

the CDM information of that flight, namely TSAT adherence, low number of TSAT and TOBT updates, low average startup delay in 20 minute time interval, and low number of average updates of the TSAT and TOBT in that same 20 minute interval, denoted as the average CDM updates. Additionally, also regulation related variables such as the pure ATFM delay, regulation and company induced delay at departure are present. It can also be seen that the most significant association rules all have the lowest category of ATC delay as a consequent.

Table XI shows the top 10 association rules when filtered on minimal redundancy and secondary maximum lift. It can be observed that the rules displayed differ from the most significant rules in Table X. As expected is the most significant rule also the rule with the lowest redundancy, due to the used definition. Compared to the most significant ones, also patterns leading to the highest category of ATC delay are now included. As expected, other conditions are also present in the antecedents of the patterns. Arrival delay for example leads to an increased probability of 20% of receiving an ATC delay higher than 9 minutes, as well as the number of TSAT updates between 15 and 30 results in approximately 30% additional probability. Also the short turnaround indicator, the season and departure rate are now observed in the antecedent column.

Additionally, the same analysis of the most significant and least redundant patterns was repeated for each set of association rules leading to each category of the ATC delay. The most important antecedent conditions found from this analysis are summarised per ATC delay category in Table XII. For the ATC delay class <0, the found patterns were the same as the ones presented in Table X, meaning that all patterns consisted of the antecedent conditions TOBT adherence and no delay key present, and other conditions related to CDM updates en regulation information. These patterns were found to have the highest significance, with 130-144% in increase in the probability of receiving ATC delay below 0 minutes.

For one higher class of ATC delay, between 0 and 4 minutes, the average startup delay being low, low number of TSAT updates and no regulation delay key were found to be the antecedents of the most significant frequent patterns found. These conditions were combined with other variables such as a low all doors closed delay, low ratio of regulated flights in the 20 minute time window of departure, and a small difference between the actual TTOT and the optimal TTOT of that flight. Additionally, the probability of this delay category is increased during the winter and for Embraer aircraft types.

For the class (4,9], the most influential factors are again average startup delay and the absence of a regulation. However, the average startup delay state has increased compared to the lower classes of ATC delay, to values between 5.5 and 12.5 minutes. Additionally, also the number of TSAT updates was found to be influential, without the addition of other conditions, with values between 15 and 30 updates resulting in an increased probability of 24% for ATC delay in class (4,9]. Also the departure rate or runway usage rate at the airport and

arrival delay above 1 minute were found in the association rules, but both lead to a small increase in the probability of just 3%.

It was found that for the highest class of ATC delay, the most influential factor is a high average startup delay ($\geq$5.5), which leads to an increased probability of he highest class of ATC delay between 47 and 50%. This condition is found in combination with the presence of an outbound peak during departure, denoted by value 2, the outbound flight taking place on a weekday, good visibility conditions on either arrival or departure, or when the flight did adhere to its assigned TSAT. When sorting the rules on redundancy instead of significance, other antecedent conditions could also be identified. As discussed before does an arrival delay between 1 and 39 minutes make it 20% more likely for a flight to receive high ATC delay. Additionally, when the TSAT is updated between 15 and 30 times, the probability rises with even 33%, and a flight in summer during the off-peaks months leads to an increased probability of 22%. Finally, the influence of a short turnaround time being not true, thus the flight having a long turnaround time, was found to increase the probability of the highest ATC delay class with 17%.

B. Bayesian Network

The optimised DAG found by the structure learning algorithm can be found in Figure 3, which consists out of a total of 46 nodes and 115 links, with a total K2 score of $3,248,136.4$. The nodes have been coloured according to the category of the variable: Fixed variables(1-11), flight (delay) data (12-23), CDM data (24-32), and operational data (33-45).

1) Structure Analysis: When looking at the structure, it can be noted that the nodes of the same colours are often grouped together and thus have a lot of dependencies between each other. The CDM (green) and flight delay data (orange) also have a lot of conditional dependencies between several nodes from the groups, showing a strong relationship between the CDM information and different flight delay variables. Additionally, the fixed conditions in pink often influence the operational variables (blue), and can be found at the top of the network, indicating that they are the drivers of several other variables in the network, as expected, and due to the inclusion of part of these variables in the black list.

First of all, the Markov blanket of the ATC delay is analysed. It can be seen that the average startup delay, the regulation delay key and the assigned pure ATFM delay of the flight directly influence the ATC delay. The maximum number of parents in the network was set to 3 in order to limit the network complexity, and together with the use of the black list, this has shown to have implications on the accuracy of the found structure, as discussed in the validation method of section II. Therefore, the ATC delay variable has been split into four binary variables, one for each category, and the structure learning algorithm was run again. The maximum number of parents was still set to be 3, meaning that in total 12 parents nodes could possibly be found if the parent nodes of each ATC delay class would be unique. In the obtained structure, the parent nodes of the 4 ATC delay variables now included 5 distinct variables. For the lowest ATC delay class, the average startup delay was replaced by the departure delay key category. In the two middle classes, both the regulation delay code and average startup delay were still parent nodes, but the third parent node was the number of TSAT updates and the departure delay key category for the ATC delay classes (0;4] and (4;9] respectively. Lastly, in the parent configuration of the highest class, the regulation delay key was replaced by the departure delay key category.

Looking at the structure, some initial observations can be made. It can be directly observed that the weekday variable is not connected to any of the other nodes, meaning that this variable is conditionally independent of all variables, as well as the other way around. Additionally, the nodes number 18

TABLE XII
SUMMARY OF THE MOST INFLUENTIAL CONDITIONS ON ATC DELAY PER CATEGORY.

ATC Delay	Antecedent	Lift
<0	TOBT Adherence = True ^ Delay key category = None ^ TOBT updates = [0;1] / TSAT updates = [0;14] / Average startup delay = [-281.5;5.5] / Average CDM updates = '0'/ Pure ATFM delay = 0 / TSAT adherence = True / All doors closed delay = [-35.0;3.0]	2.30-2.44
(0,4]	Average startup delay = [-281.5;5.5] ^ Regulation delay key departure = 0 ^ Difference TTOT and optimal TTOT = [-181;481] \ Ratio regulated flights = [0.0;0.2] \ Season = W \ Aircraft type = E75/90 \ All doors closed delay = [-35.0;3.0]	1.53-1.57
	Regulation delay key departure = 0 ^ TSAT updates = [0;14] ^ Difference TTOT and optimal TTOT = [-181;481] \ Ratio regulated flights = [0.0;0.2] \ All doors closed delay = [-35.0;3.0]	1.51-1.54
	Difference capacity and usage arr = [10,59]	1.10
	Time of day dep = Midday	1.03
(4,9]	Average startup delay = [5.5,12.6] ^ Regulation delay key departure = 0 ^ Visibility departure = Good \ Company induced delay = 0 \ Visibility arrival = Good \ TSAT adherence = True	1.39-1.43
	TSAT updates = [15,30]	1.24
	Actual runway usage departure = [43.5,60.7]	1.03
	Arrival delay = [1.0;39.0]	1.03
$\geq$9	Average startup delay = [5.5,12.6] ^ weekday = 1/ Peak indicator dep = 2 / TSAT adherence = True / Visibility arr = Good / Visibility dep = Good / Company induced delay = 0	1.45-1.47
	Actual runway usage departure = [60.7,90.1]	1.37
	TSAT updates = [15,30]	1.35
	Season = Summer (off peak)	1.22
	Arrival delay = [1,39]	1.20
	Short turnaround time = 0	1.17

and 43, the delay key category and the regulation delay key of the outbound flight are both central nodes in the network, as they each have the maximum number of parent nodes, 3, but have 8 and 9 child nodes respectively. The delay key of the flight is dependent on the arrival delay of the inbound flight, as well as the received pure ATFM delay of the outbound flight, and lastly whether or not there is a short turnaround period. The delay key category then influences many variables related to the CDM process, namely the TSAT/TOBT updates and adherence, but also the variables of the all door closed delay and the induced regulation delay. The delay key of regulation is conditionally dependent on very different variables, such as the sector used in the Dutch airspace, the season and aircraft type. The nodes influenced by this central variable are mostly related to specific regulation information, as expected, such as the number of CTOT updates, the ratio of regulated flight in the 20 minute time frame, the company induced delay, and the multiple regulation indicator.

Finally, it is interesting to analyse the parent and child nodes of variables that are directly influenced by the performance of the processes of the airline itself. Therefore, nodes 12, 15, 16, 17, 23, 24, 27 and 28 are discussed in more detail. From these eight variables, 3 of them, waiting for ground services, company induced delay and boarding duration are end points in the network, meaning that their value does not influence the probability of any other variable directly when analysing the structure top down. The arrival delay is driven by the assigned ATFM delay of the inbound flight, as well as the difference between the capacity and the actual usage rate of the runways on arrival and the rate of regulation. A child node of the arrival delay is the received ATFM delay of the departure flight, which is directly influencing the ATC delay. Additionally, the waiting for ground services time is dependent on the arrival delay, however this variable has no influence on another variable. Both the TSAT and TOBT adherence are dependent on the assigned departure delay key category. Additionally, the TOBT adherence is influenced by the TSAT adherence, which is dependent on the turnaround time of the flight. The number of TOBT updates a flight receives is also dependent on the turnaround time, as well as the adherence of the TOBT, but influences the all doors closed delay of the flight. Finally, the following relation could be identified from the network, starting from the above discussed variables to the ATC delay:
TSAT adherence $\rightarrow$ TOBT adherence $\rightarrow$ TOBT updates $\rightarrow$ all doors closed delay $\rightarrow$ TSAT updates $\rightarrow$ Average startup delay $\rightarrow$ ATC delay.

2) Inference: The method of inference was used to analyse the obtained BN. As explained in section II, this was done using approximate inference, specifically weighted likelihood, with a sample size of 100,000. The results are shown in Figure 4-12. In these figures, the influence of different variables on the probability distribution of the ATC delay, or other variables of interest, is illustrated by means of a bar plot. Each category of the evidence variable is represented with a different colour, such that bars of the same colour add up to 1, representing the complete probability distribution across the categories of the target variable.

Figure 4 presents the influence of the two direct parent nodes of the ATC delay variable, namely the pure ATFM delay and the regulation delay key on departure. It can be clearly observed that both of these variables have a strong influence on the probability distribution of the ATC delay, as expected due to their position in the network. Thus, if a regulation is present, the probabilities of larger values of ATC delay increase. The result of inference using the third parent node in the obtained network, the average startup delay, is shown in Figure 5. Again, it can be seen that this variable has a profound influence on the ATC delay, where the highest class of the average startup delay even leads to more than 60% probability of encountering more than 9 minutes of ATC delay.

One of the interesting variables that influences the average startup delay is the ratio of regulated flights in the same 20 minute time interval, and its influence on the average startup delay is also shown in Figure 5. Additionally, the influence of the ratio of regulated flights on the ATC delay was analysed and displayed in Figure 6, as well as the influence of the average number of CDM updates received by the flights in the same 20 minute time interval. It can be observed that both variables influence the probabilities of receiving the different classes of ATC delay. When the percentage of regulated flights increases, the probability for a high ATC delay ($\geq$9) also increases, whereas the opposite can be said for the class between 0 and 4 minutes. In terms of the number of CDM updates, class 2, which is the class with a high average number of TSAT updates, the same relation can be seen, again decreasing the probability of low ATC delay values and increasing the high ATC delay probability.

Figure 7 shows the results of performing inference with the inbound and all doors closed delay variables fixed as evidence. It can be observed that the impact on the probability distribution of the ATC delay is less pronounced compared to the variables that are the direct parent nodes of the ATC delay, however an effect can still be observed. Figure 8 represents the probability distribution of the ATC delay for different assigned IATA delay key categories. It can be seen that the ATFM delay reason increases the ATC delay probability enormously, and flights with reactionary or weather delay key categories also have an increased probability to obtain the highest class of ATC delay, which is in line with what was obtained for the arrival delay influence. Additionally, the probability of having less than 0 minutes ATC delay increases when there is no initial delay key assigned ('None').

Figure 9 presents the influence of the variables relating to the number of CDM updates a flight received during its departure process, namely of the TOBT and the TSAT, two key milestones in the determination of the startup delay. It can be observed that both variables have an influence on the probability distribution of the ATC delay, however the number of TSAT updates seems to have a larger impact compared to the TOBT. When a flight has more than 30 TSAT updates, the probability of having the largest category of ATC delay rises to approximately 40%, whereas this is less than 20% for the lowest category of TSAT updates. Looking at the impact of the TOBT updates, the probability of having the

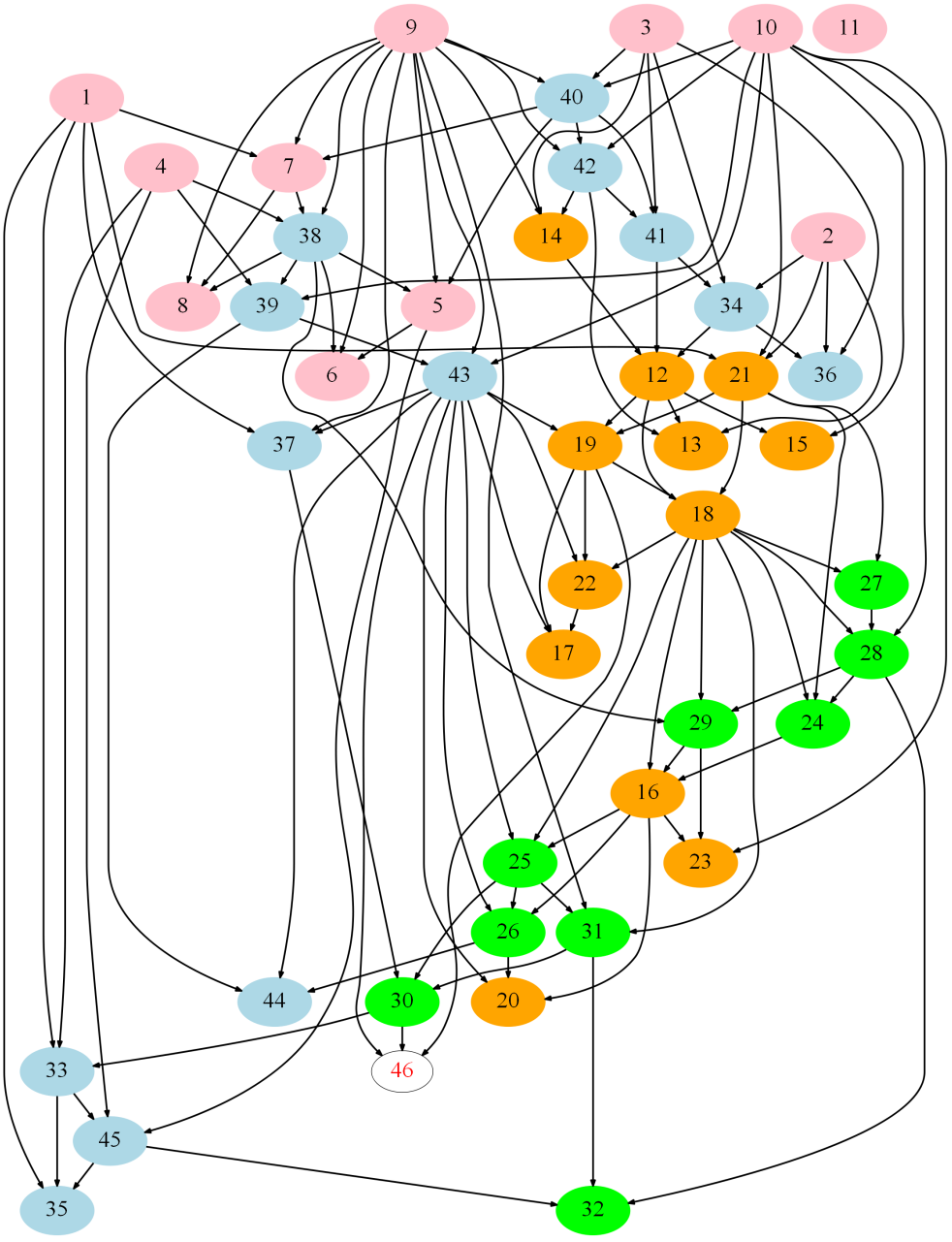
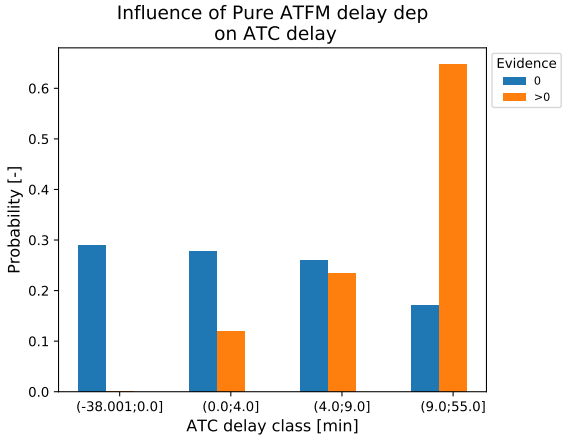


Fig. 3. The resulting learned Bayesian network structure.

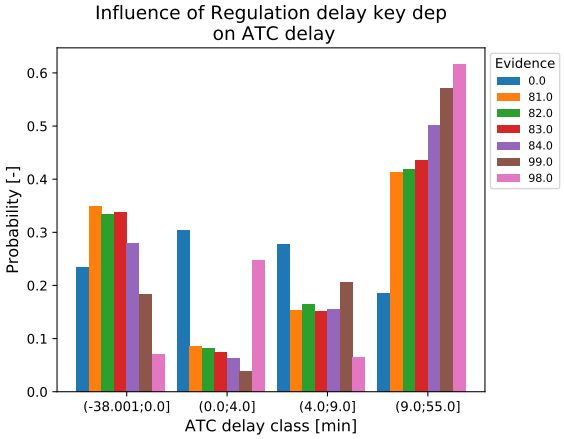
highest ATC delay category is 28% for TOBT updates higher than 5, whereas this probability is 22% for TOBT updates between 0 and 1. The influence of the TOBT and TSAT adherence are also visualised, and can be found in Figure 10. The influence of the adherence to these CDM milestones is less pronounced than the number of updates. Especially for the TOBT adherence, the influence is marginal to none, whereas the impact of the TSAT adherence on the probability distribution remains limited around 5%.

Figure 11 shows the results of the actual runway usage rate or departure rate and the difference between the capacity and the actual departure rate upon departure, both expressed in number of aircraft movements per hour. The resulting probability distributions show that the probability for higher classes of ATC delay increases with increasing departure rate. Additionally, the variable of the difference between the capacity and actual departure rate shows that when the declared capacity is exceeded, leading to negative values, the probability for high values of ATC delay increases.

Finally, the results of inference with the season and short turnaround time indicator fixed as evidence are visualised in Figure 12. For the season, a small change in the probability distribution of the ATC delay category (0.0;4.0] and $\geq$9 can be observed, namely that the former is more probable during

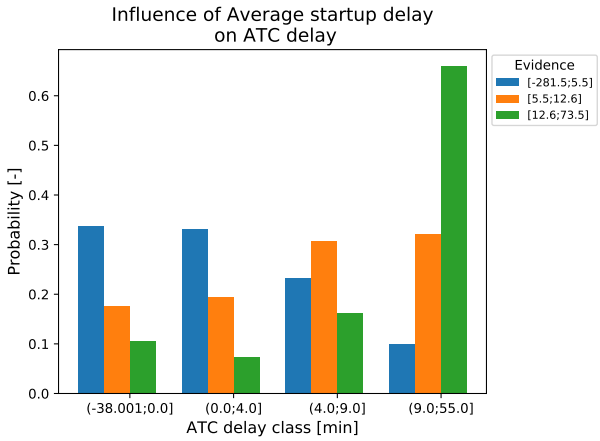


(a) Pure outbound ATFM delay.

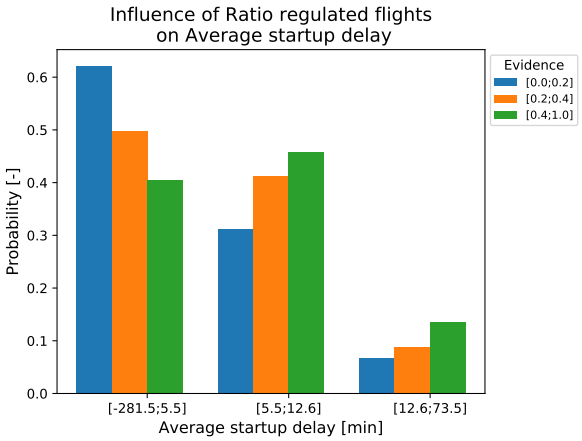


(b) Regulation delay key for departure.

Fig. 4. Influence departure regulation variables on ATC delay.

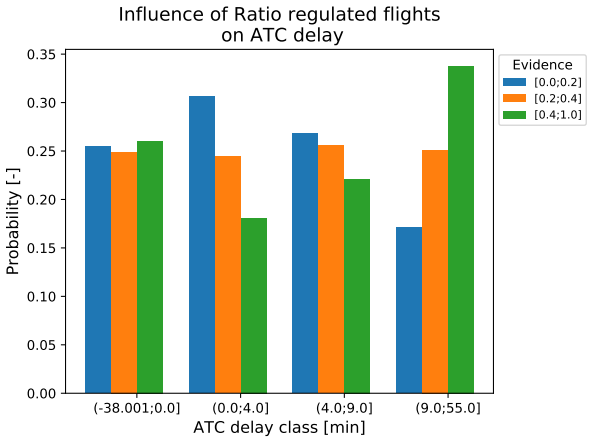


(a) Influence of average startup delay on ATC delay.

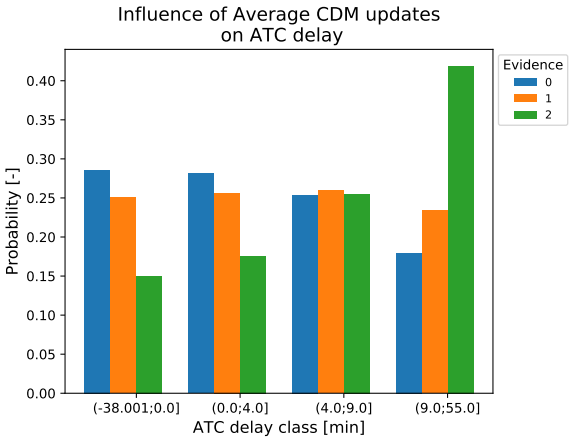


(b) Influence ratio of regulated flights on average startup delay.

Fig. 5. Influence average startup delay on ATC delay and ratio of regulated flights on average startup delay.

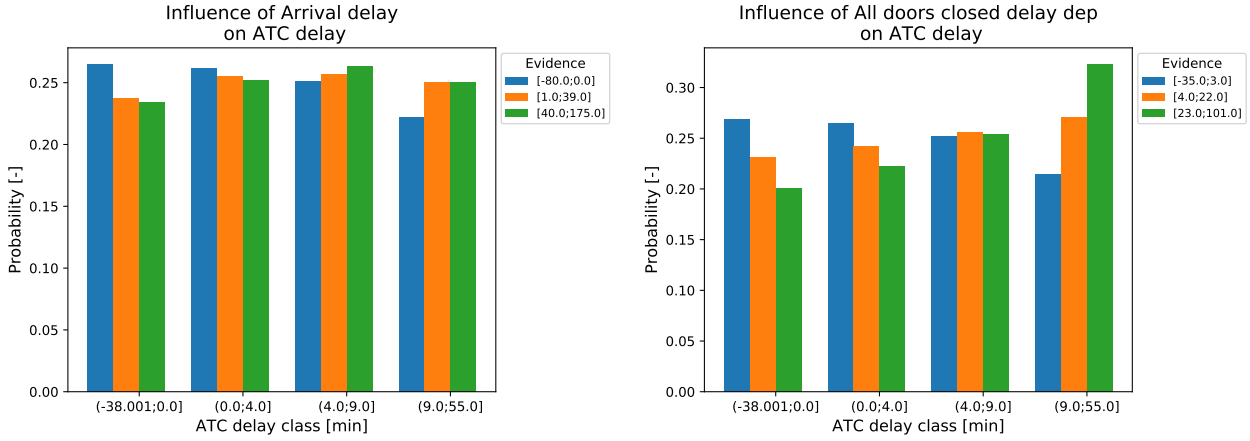


(a) Ratio of regulated flights in 20 minute time frame.



(b) Average number of CDM updates of flights in 20 minute time frame.

Fig. 6. Influence of variables of flights in same 20 minute time window.



(a) Arrival delay.

(b) All doors closed delay.

Fig. 7. Influence of earlier delay information on ATC delay.

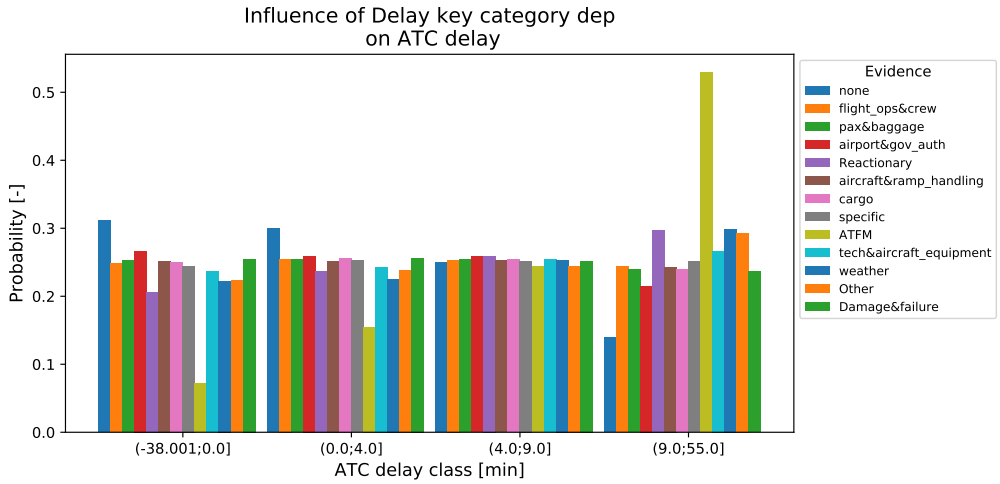
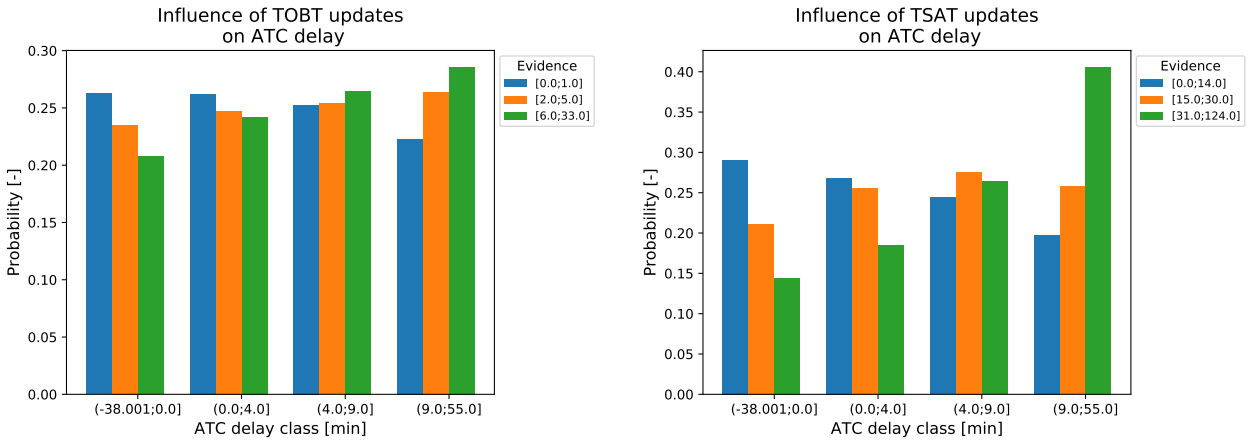


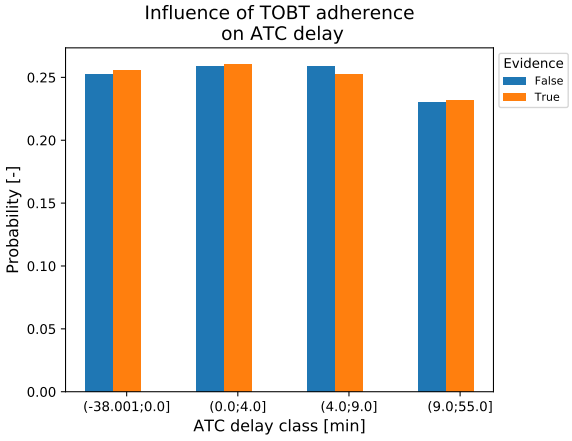
Fig. 8. Influence of the departure delay key category on ATC delay.



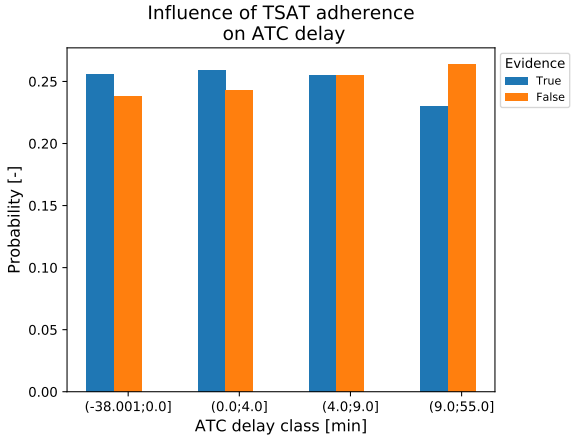
(a) Number of TOBT updates.

(b) Number of TSAT updates.

Fig. 9. Influence number of CDM milestone updates on ATC delay.

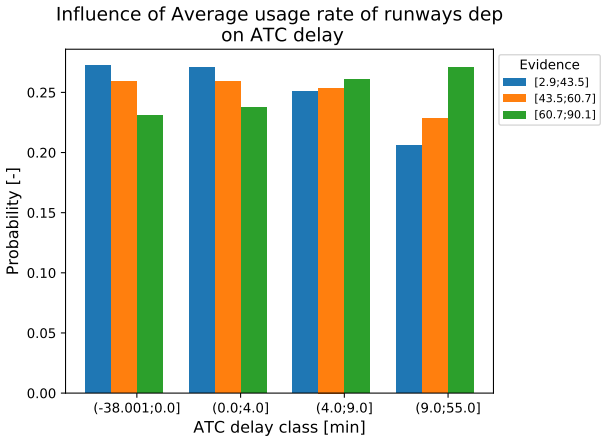


(a) TOBT adherence.

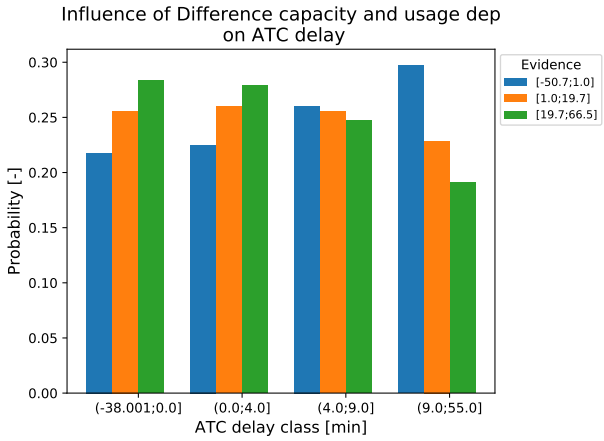


(b) TSAT adherence.

Fig. 10. Influence CDM adherence variables on ATC delay.

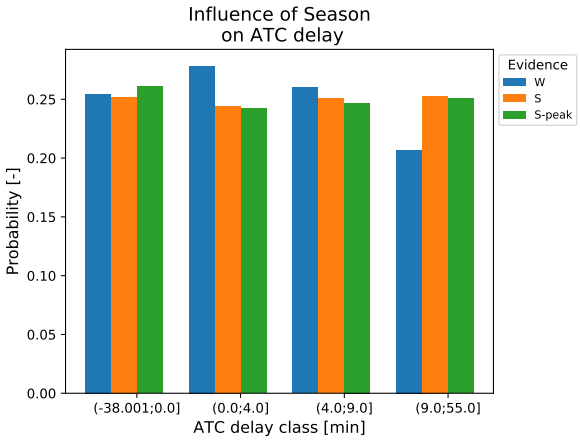


(a) Average departure runway usage rate.

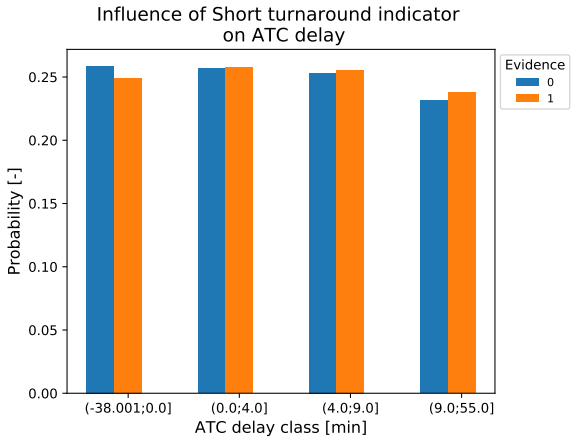


(b) Difference capacity and actual runway usage rate departure.

Fig. 11. Influence of operational variables during departure on ATC delay.



(a) Season.



(b) Short turnaround indicator.

Fig. 12. Influence of fixed variables on ATC delay.

the winter, and the latter during the summer months. For the short turnaround time indicator, 0 meaning a long turnaround and 1 a short turnaround, no influence could be found, as seen in the visualisation.

IV. Discussion

In this section, the presented results from section III are discussed. Additionally, the findings of the two methods are compared, which is followed by a discussion of the limitations of the used methodology and made assumptions.

A. Association Rule Mining

By performing the analysis of the most significant and diverse patterns per ATC delay category, the most influential conditions on the target variable could be found. It can be concluded that mostly the absence of an assigned delay key and TOBT adherence are key in the increase of the probability of receiving no ATC delay. Additionally, the average startup delay in the 20 minutes time interval of departure could be observed in the strongest rules, for almost all categories. The same holds for the TSAT updates of the flight, but in general did these patterns have lower lift values, and thus leading to a less strong pattern. In the patterns of the ATC delay between 0 and 4 or 4 and 9 minutes the condition of having no outbound regulation also led to the most strong patterns, which indicates that flights that are not regulated actually have a higher probability of having this delay category, due to their received startup delay.

Other conditions could also be noticed, such as the all doors closed delay lower than 3 minutes increases the probability of having lower than 4 minutes ATC delay. Also the season being winter was found to increase the probability of ATC delay in category (0,4], as well as having a ratio of regulated flights between 0 and 0.2. This means that a low number of regulated flights actually leads to more chance of having between 0 and 4 minutes ATC delay.

A number of conditions were found in combination with these most influential conditions. For example in the association rules of the ATC delay classes (4,9] and $\geq$9, the variables relating to good visibility on arrival and departure were found in combination with the average startup delay. This implies that even when there is good visibility, which is not expected to influence the ATC delay and congestion at the airport in a negative manner, the probability of having a higher ATC delay increases, despite these good visibility conditions being present. The same could be observed for the TSAT adherence being True. It can be that the variables of visibility and TSAT adherence do influence the ATC delay probability negatively when taking on other states, however ATC delay can also be high when these conditions are not in adverse conditions. This effect also shows that when multiple conditions are present in the antecedent of an association rule, not necessarily all conditions have an influential relationship on the consequent. This is due to the fact that the ARM method adds these additional conditions as they are very frequent, but do not directly cause the consequent.

The data mining method also found that a long turnaround process increases the probability of having an ATC delay of more than 9 minutes. This is an unexpected effect, as a short turnaround process could possibly increase the delay encountered as the scheduled time for all the processes might be shorter. This could be the result of flights with a long turnaround time being not prioritised by the ground handling agent, the operations control centre and air traffic control. In the definition of the short turnaround time indicator, the actual turnaround time was used in the formulation of this variable. Therefore it could also be that the received delay would increase the turnaround time to be over the limit of what is considered a short turnaround time, which was 1.5 and 3 hours for European and Intercontinental flights respectively.

When observing the association rules when sorted on redundancy, patterns are found that have either the smallest or largest ATC delay category as the antecedent, as displayed in Table XI. This implies that the most significant patterns can be found for these two classes, and that the patterns of the other 2 bins of ATC delay have overlap with them and/or have a lower rule strength. It was also observed that part of the conditions that lead to an increased probability of having an ATC delay of (4,9] and $\geq$ 9 are similar. For example, in the frequent patterns of both classes, the most influential conditions are an average startup delay of larger than 5.5 minutes, TSAT updates larger than 14, and arrival delay between 1 and 39 minutes. However, the increased probability is higher for the class $\geq$ 9, compared to the class between (4,9] minutes, as can be seen by the corresponding lift values. This was already found by looking at the average values of the rule strength, shown in Figure 2, where it was found that the class (4,9] of ATC delay leads to less strong patterns on average. This could be explained by the fact that multiple conditions influence the ATC delay of a flight. The class between 4 and 9 minutes is situated between the extreme values of ATC delay. This results in the fact that the most important influences of high ATC delay can lead to this class, but also influences of the lower ATC delay can lead to this value, as flight delay remains a stochastic process. Therefore, the significance of the found patterns is less strong.

It should be noted that not all conditions of a variable can be assessed for their influence on the target variable. An example of this is the following; an arrival delay between 1 and 39 minutes is found to have influence on the probability of occurrence of the highest category of ATC delay. However, the class of arrival delay large than 40 minutes is not found in the identified frequent patterns, as this large delay does not occur frequent enough for the ARM method to include it in the pattern analysis. Therefore, this method is suitable to get an initial understanding of the patterns in the data, but it lacks the ability to perform deep (sensitivity) analyses. This can potentially be solved by using discretization methods such as binning by equal frequency, where each resulting bin has an equal frequency occurrence in the data set. Therefore, the full range of states of each condition can possibly be included in the mining methodology.

Lastly, adverse weather conditions were not seen as an influence in any of the results. This can be an implication of

the minimum support value, which was set to 9%, which might have been too high to include the bad weather conditions in the mining process as well. Considering operational variables that are related to the declared capacity and the difference with the demand and the actual runway usage rates, none of them could be retrieved in the most significant patterns, only the actual departure rate. This association rule showed that higher departure rates (>60 aircraft/hour) relate to a larger probability for the highest values of ATC delay, which shows the impact of traffic volumes on the ATC delay.

B. Bayesian Network

The three nodes that directly influence the ATC delay in the found BN structure are the average startup delay, the pure ATFM delay and the regulation delay key of the outbound flight. The latter two can be explained as, by the used definition, ATC delay consists of the slot or ATFM delay if a flight is regulated, and therefore these two variables directly relate to this delay. If a flight was not regulated, by definition, the ATC delay consists of the startup delay. It is interesting to note that the average startup delay in the 20 minute time interval of the flight is directly influential on the ATC delay of an individual flight, which shows that the ATC delay of a flight is heavily influenced by the situation of congestion at the airport itself, which is more influential than individual performance of the flight. When the input data set included a binary variable of each of the ATC delay categories, allowing the model to identify up to 12 parent variables, the departure delay category and number of TSAT updates could also be found in the parent configurations of these variables. This implies that the three parent nodes in the original network have a very strong influence on the ATC delay for all categories, as the separation of ATC delay into four variables did only add 2 additional parent nodes. However, for each ATC delay class, a unique combination of parent nodes was obtained, meaning that the drivers of each of the classes of ATC delay slightly differ, which was also observed in the results of the baseline method.

An interesting parent node of the average startup delay is the ratio of regulated flights. For increasing ratio of regulated flights in the 20 minute time frame, the probability of high values of average startup delay bins increase compared to the probabilities associated with the lower ratios of regulated flights. This can be explained by the use of the departure manager, which will prioritise flights with a regulation and thus Calculated Take-off Time. This results in other flights in that same time interval to be not prioritised, leading to an increase in their startup delay. Furthermore does this variable also influence the ATC delay. Especially the probability of the highest class of ATC delay is influenced a lot by the number of regulated flights. Again the same explanation can be followed. However, lower ratios of regulated flights increase the probability of receiving between 0 and 4 minutes of ATC delay. Here, less regulations lead to less strict take-off times, due to the absence of an issued CTOT, leading to no or less heavy prioritisation of other flights and thus less extreme values of ATC delay in that time window. Additionally, it was observed that the number of CDM updates of flights in the same 20 minute time interval influences the ATC delay. It was found that state 2 of this variable, meaning a high number of average TSAT updates, returns the highest probability for an ATC delay $\geq$ 9. State 1, which represents a higher number of average TOBT updates, has less influence on the distribution. The number of TSAT updates is high when there are a lot of changes in the departure manager and availability of the runways. Therefore, this parameter shows that a high congestion at the airport indeed leads to increased chances of high ATC delay, as expected, and that the ground handling process of other flights can heavily impact the ATC delay of a flight as well.

From the inference results, no influence of the TOBT adherence could be found on the probability of the target node, and also the adherence to the TSAT was found to have a limited influence. Both these variables are included however included in the chain of variables influencing the ATC delay. When looking at the number of updates of these milestones, larger influences on the ATC delay probability distribution could be found. Furthermore did the results show that the number of TSAT updates have a lot more influence than the number of TOBT updates, but for both variables influence on the ATC delay probability distribution could be found. The difference in influence importance can be explained by the fact that a TSAT update can be triggered when the TOBT is updated, however this can also be due to other factors. When the TSAT is altered without a TOBT update, this can imply that the difference between them is growing as the TSAT delayed, therefore increasing the ATC delay. When a TOBT is updated, the difference with the newly issued TSAT does not change by definition, which results in less influence on the ATC delay.

The arrival delay factor influences the ATC delay as well, where inbound delay of more than 1 minute increases the probability for higher values of ATC delay slightly. Flights with arrival delay are likely to have delay outbound due to the propagation of delay. Another driver of the ATC delay in the network was the all doors closed delay. This variable was found to be influenced by the number of TOBT updates and is thus linked to the ground process of the flight. Additionally, it also influences the TSAT updates, as the longer the flight is not ready, the TSAT will need to be updated. However, the all doors closed delay is not only dependent on the ground and handling processes, but can also be linked to other delay factors, such as (missing) passengers. This can be seen as the delay key category of the flight also directly influences the duration of this delay type. The departure delay key category was found to also influence the ATC delay. A profound increase in the highest class of ATC delay probability was seen in case of an ATFM delay key, again, this is as expected due the definition of ATC delay. Also the reactionary category was found to increase the probability of high ATC delay, which is in agreement with the found influence of the arrival delay on ATC delay, as discussed before. If a flight has no registered delay key, the probability of having low ATC delay, (<4 minutes) increases. This could be explained by the fact that flights with no initial delay have finished all there

turnaround and departing processes in time, such that they can depart during their originally scheduled slot, leading to less overall congestion and delays.

The difference between the offered capacity and the actual used runway rate did not have a direct link with the ATC delay in the BN structure, however, its value was found to influence the probability of the ATC delay categories. It was found that more actual runway movements than the offered capacity lead to increased probability for high ATC delay, and the same relationship could be found between the arrival capacity usage and arrival delay of flights. Thus, larger delays occur when the capacity is being overused. This might indicate that the capacity is actually underdeclared, which leads to more delays.

Inference results with the season fixed as evidence did also alternate the probability distribution of the ATC delay. Here, it could be seen that winter season led to reduced probability in the highest class of ATC delay, $\geq$9, and increased the probability between 0 and 4 minutes. Therefore, it can be said that large delays are more likely to occur during the summer months, whereas small delays are more likely during the winter, which might be a result of higher traffic volumes during the summer months. It is also interesting to notice that the BN inference did find a difference between the winter and summer, but there is no difference between the summer months off peak and on peak, as was expected due to increased volumes of flights during the on peak months of the holiday season (July-August). Additionally, the influence of the short turnaround indicator was analysed using inference, and it was found that this actually did not influence the probability distribution. It was expected that a short turnaround would increase the probability for ATC delay, as the various turnaround procedures have a smaller scheduled time, and there is less time to absorb the inbound delay to prevent reactionary delay for the departing flight. While it was found that the arrival delay and outbound reactionary delay in fact increase the probability for ATC delay, other factors with a larger impact where found which are not directly linked to the turnaround time of a flight. Alternatively, it could also be that the BN did not find a relationship due to the large number of nodes and links that are present between the short turnaround indicator and the target variable in the structure, which might have caused the effect to be lost.

C. Comparison of Methods

First of all, it can be said that both methods recognise the direct influence of the average startup delay in the 20 minute time interval on the ATC delay, as well as the influence of the presence of a regulation. The secondary influences, such as the number of TSAT updates, arrival delay, ratio of regulated flights, all doors closed delay and the actual departure rate could also be seen in the obtained results of both methods. In these observed drivers of ATC delay, both methods showed that the conditions leading to an increased probability for the two highest classes of ATC delay are very similar. However, the strength of the patterns was found to be higher for the highest class, $\geq$9, compared to the ATC delay category of (4;9]. Thus, more extreme values of ATC delay show more distinct patterns, which was expected. As flight delays are a stochastic process, the middle classes of delay can also be a result of conditions that generally lead to high or low values of delay, which reduces the strength of the found patterns.

Furthermore did the two methods agree on the fact that the weather conditions during arrival and departure did not have a direct influence on the ATC delay. Especially the visibility variable was expected to influence the target variable, due to the reduced capacity at marginal or low visibility procedure conditions. It could be verified that marginal and low visibility procedure (LVP) conditions did increase the probability of low realised departure rates, as expected. However, the BN was built such that a relationship was found between the departure rate and the ATC delay, where increasing departure rates led to increasing probability of high classes of ATC delay, indicating the importance of traffic volumes. Due to the restriction in the number of parent nodes and thus links in the structure, the indirect influence of the bad visibility conditions was not included in the learned network. As mentioned before, it could be that the ARM method did not capture the effect of these conditions on the ATC delay due to the used support value, which was likely too high to include the small frequency of bad weather days. It can thus be hypothesised that both methods did not recognise the influence of the weather on the ATC delays, due to the limitations of the made assumptions.

However, also differences in the obtained results between the methods could be found. The ARM method found that there is an influence of the weekday on the ATC delay, whereas the learned Bayesian network does not link this variable to any other variable in the data set. This could also be an implication of the used black list in the structure learning process, which enforced the weekday indicator to only have children nodes. Additionally, the methods found opposite results with respect to the the influence of a short turnaround time. The ARM method identified that a long turnaround would increase the probability of having a high ATC delay, however using inference on the BN, no large influence of this variable could be seen on the ATC delay. This might be caused by the fact that the BN results are dependent on the learned structure, which was limited by the black list and the number of nodes. As the short turnaround time variable is located far from the ATC delay, the relation observed by the ARM might be lost by the learned structure. From these discrepancies, it can thus be said that the ARM method is able to detect more indirect relationships, such as the short turnaround time and the weekday indicator, compared to the Bayesian network. Here, indirect relationships can be lost due to the number of variables that are located between them, and the limitation in the number of links in the structure, due to the use of the black list and the maximum number of parents.

Another discrepancy in the results could be found in the influence of the TOBT adherence. This variable was found to be very influencing for the occurrence of low ATC delay by the baseline method when combined with no delay key assigned upon departure. This effect was less profound in the BN model, where the TOBT adherence was found to have very limited effects on the probability distribution of the ATC delay. It can be suspected that the ARM method identified this

pattern due to the high frequency of this pattern in the data set, but not necessarily implying a causal relationship. This is an effect of having multiple conditions in the antecedent, which can lead to the fact that only part of the antecedent actually has a causal influence on the ATC delay, and the other conditions in the antecedent have a high frequency, which leads to them being included in the frequent pattern. However, the discrepancy found in the influence of the TOBT adherence could also be a result of the structure learning process, which might have lost the relation between the two variables due to the high dependence and correlation of the TOBT adherence with the other CDM related variables.

As mentioned before, a disadvantage of the ARM method is that only the effect of the state of a variable that is present in the frequent patterns can be identified. The influence of other states of that same variable cannot be assessed as they are not found to be frequent enough. An example is the influence of the average startup delay variable. It was found by the ARM method that the middle class increases the probability of receiving a higher ATC delay. When using the BN, the same relationship was found, however it could be observed that the highest class of average startup delay increased that probability even more. This was not possible to assess with the results of the ARM method, as this state was not found as an antecedent condition. The exact same phenomenon could be found in number of TSAT updates, all doors closed delay and arrival delay. By using the BN, it was possible to analyse the influence of having reactionary, ATFM and weather delay key categories assigned, which were not identified by the baseline method, for that same reason.

As both methods agreed on many of the determining factors of ATC delay, both of them are suitable to investigate or diagnose direct causes and influencing factors on certain conditions. When it is the goal to better understand a system and the dynamics between a large number of variables, the Bayesian network method is more suitable, as the conditional dependencies can be observed from the learned structure, and are not hidden in a large number of frequent patterns. However, first diagnoses of influential variables can also be done using the ARM method, and this method can be said to find more indirect effects on the target variable compared to BN, in which indirect relations might be lost. However, it is important to optimise the length of the rules well, as an increasing number of conditions in the antecedent can also add parameters which are not actually influential but simply very frequent. Furthermore has the BN more possibilities to use the model for analysing the impact of certain conditions on the other variables in the network, which cannot be achieved using ARM.

D. Limitations of the Methodology

In the developed methodology, discussed under section II, several assumptions had to be made and parameters had to be determined, which have impacted the results and led to limitations.

First of all, the selected methods, ARM and BN, required discretization of the variables. It is inevitable that discretization leads to loss of information. Next to the loss of information do the used discretized bins of the variables also heavily influence the relationships that can be found by the methodology. This implies that other discretization methods could have led to different patterns, their strength and conditional dependencies between the variables.

Secondly does the computational complexity of both methods grow exponentially with the number of variables included. This constraint has limited the number of features that could be used in the analysis, and therefore a strict feature selection method had to be adopted. By selecting the features which held most dependence with the target variable, as discussed in section II, the strongest patterns could be found. However, it is likely that other patterns could have been obtained using more of the available data, which are left unexplored.

Pure observational data has been used for the purpose of this research. However, pure observational data always contains deviations and imperfections, which cannot always be identified and/or corrected. The faults in the data have been removed in the data processing part of the methodology, by applying outlier detection and handling missing values accordingly. In addition to this was there only CDM data available for the flights at Amsterdam Schiphol Airport which have been handled by KLM ground services. This does not give complete information of the situation of the airport, due to proprietary data constraints.

Finally, the obtained results from both methods were affected by the made assumptions and selected hyperparameters. In order to limit the complexity of the BN, the number of parents was limited to 3 in the structure learning process. This limits the complexity of the network, which was necessary in order to perform analysis on the learned structure. However, if the number of parents was not limited, the number of actual links in the network would increase, representing the actual number of conditional dependencies present between the variables, which might lead to a better prediction accuracy. Moreover are the performed analysis using the BN and the obtained results dependent on the performance of the structure and parameter learning algorithms used. In the baseline method ARM the minimum frequency or support, used to make a trade-off between limiting the number of identified rules and finding the most significant ones, resulted in the method to discard specific states of variables. These conditions did not occur frequently enough, hence it was not possible to analyse their influence on the target variable using only this method.

V. Conclusion

This paper has presented the methodology and findings of the performed research with the aim to analyse the root causes of air traffic control delays, with a case study on KLM flights at Amsterdam Schiphol Airport. Data was gathered and integrated from multiple sources, spanning a period from 15 November 2018 to 31 December 2019, including flight (delay) data, operational data on capacity, demand and actual traffic volumes, CDM data and weather information. Two methods were used for causal analysis, namely association rule mining

using the FP-growth algorithm and a Bayesian network, using heuristic hill climbing with tabu search as a structure learning method.

The used methods agreed on the majority of the identified influential factors of the ATC delay. It was found that the main influences of ATC delay are the average startup delay of flights in the 20 minute time interval of the flight's departure, as well as the received pure ATFM delay and the regulation delay key. The two regulation regulated parameters were expected to be found, as ATFM delay is included in the used definition of ATC delay. The other variable, the average startup delay in the 20 minute time interval, shows that the ATC delay is profoundly impacted by the congestion at the airport of departure.

In general, it was found that ATC delay is partially caused by the situation at the airport. The traffic volume does influence the ATC delay, as a higher departure rate increased the probability of high ATC delay categories. Also the congestion at the airport, namely the average startup delay and the average number of CDM updates, as well as the number of regulations that have been issued in the 20 minute time interval of departure impact the probability on receiving ATC delay. The influence of the ratio of regulated flights might be due to the prioritisation of regulated flights due to their fixed CTOT, resulting in high startup delays for the remainder of the flights. Furthermore, it could be found that delays are more likely to occur when the actual departure rate exceeds the declared capacity, both for general arrival delay as for ATC delay upon departure. However, also individual performance of a flight could be found as causes of increased probability in ATC delay, such as reactionary delay, which is caused by propagated delay from the arrival flight. Additionally, the number of TSAT & TOBT updates a flight receives was found among one of the causal influences, as well as the flight's delay in closing its doors, which was found to lead to an increasing number of TSAT updates and therefore to a higher ATC delay.

In these identified influences, both methods found that in general, the strongest patterns could be found for the lowest and highest category of ATC delay, which was as expected due to the stochastic nature of flight delays. Furthermore did the two methods not find an influence of weather conditions on the probability distribution of the ATC delay. This was most likely a result of the limitations of both methods, namely the required frequency or minimum support in the baseline method, and the limitation of the number of parent nodes in the Bayesian network structure learning.

The methods did not agree on certain patterns found. The ARM method found that the TOBT adherence is an important driver for having low ATC delay. However, this pattern could not be validated when looking at the results found from inference on the BN, although the TOBT adherence is included in the chain in the network leading to the target. When multiple conditions are present in patterns found by the baseline method, it can be that only part of the conditions are actually influencing the consequent, and that the additional conditions in the antecedent only have a minor influence and are mostly added in the pattern due to their high frequency. Both methods also found opposite results with respect to the short turnaround indicator. In the association rules, it was found that a long turnaround actually increases the probability of having a high ATC delay, whereas the BN could hardly find a relationship between the two variables. These differences could be explained by comparing the two methods. First of all, the results of the Bayesian network are dependent on the performance of the structure and parameter learning algorithms. Additionally, the structure learning was limited as certain links between nodes were forbidden and the maximum number of parents was set to be 3 in order to reduce the complexity of the BN. These limitations could explain the different results as influential relationships could be lost due to these parameters. This also influenced the prediction accuracy of the BN, which was used as a validation method. The prediction accuracy was found to only be 47.2% on the ATC delay, whereas up to 80% accuracy was achieved for other variables in the network. Again, this is an implication of making use of the black list and the maximum number of parent nodes, which constrained the ATC delay to only have three parent nodes, and no child nodes, and thus a limited amount of variables were used in the prediction process of the target variable.

When using ARM to perform causal analysis, the disadvantage is that this method did not allow to perform a full analysis of each of the variables in the data set. If a condition is not seen as frequent enough, it is excluded from the analysis and thus cannot be analysed. This phenomenon was seen for several variables in the data, as some states have a high influence on the target variable according to the BN, but did not occur sufficiently frequent to be found in the association rules.

It can be concluded that both methods are fit to perform a diagnostic analysis of a system to determine causal or influential relationships between variables. The BN is more suitable to get a better understanding of an entire system, which also allows to perform analysis on scenarios that did not occur in the used data set, and assess the impact on the other variables.

For future work, several recommendations can be made. First of all, it is recommended to further investigate the influence of setting less restrictive number of parents in the structure learning of the BN on the prediction accuracy and the extra found relationships. Additionally, as the discretization method determines for a large part which relationships can be identified from the data set, it is interesting to investigate the influence of other possible discretization methods, such as manual, equal frequency and equal width binning methods, on the found patterns and relationships. Also different aggregation of the available features, such as crosswind, the combination of peaks and time of the day could possibly reveal more info on the influence on ATC delay. The level of detail in the analysis could potentially be further increased by creating separate models and thus analyses for various groups of data with different underlying dynamics, such as European/Intercontinental flights, summer and winter, flights with a short and long turnaround time, and possibly regulated and non-regulated flights. By doing so, it can be possible to discover different and more detailed underlying drivers of ATC delay per data group. However, building separate models can

only be achieved when enough data is available per data group. Furthermore is it recommended to generalise and validate the applicability of the methodology by applying it to different case studies, which can consist out of different airports and/or airlines, or different aggregation levels, such as the ATC delay per outbound peak or day. Finally, the obtained knowledge and insights on ATC delay from this research can be used to create or integrate ATC delay in a decision support model to optimise KLM's operations at their hub. In this context, the BN can also be used to design, simulate and asses the effectiveness and impact of mitigation strategies on the other variables and the ATC delay.

REFERENCES

[1] Alice Sternberg, Diego Carvalho, Leonardo Murta, Jorge Soares, and Eduardo Ogasawara. An analysis of Brazilian flight delays based on frequent patterns. *Transportation Research Part E: Logistics and Transportation Review*, 95:282–298, November 2016. ISSN 1366-5545. doi: 10.1016/j.tre.2016.09.013. URL http://www.sciencedirect.com/science/article/pii/S1366554516301740.

[2] Yanjun Wang, Yakun Cao, Chenping Zhu, Fan Wu, Minghua Hu, Vu Van Duong, Michael Watkins, Baruch Barzel, and H. Eugene Stanley. Universal patterns in passenger flight departure delays. *Scientific Reports*, 10 (1), 2020. doi: 10.1038/s41598-020-62871-6.

[3] Ilias Vlachos and Zhibin Lin. Drivers of airline loyalty: Evidence from the business travelers in China. *Transportation Research Part E: Logistics and Transportation Review*, 71:1–17, November 2014. ISSN 1366-5545. doi: 10.1016/j.tre.2014.07.011. URL http://www.sciencedirect.com/science/article/pii/S136655451400132X.

[4] Ayesha Sadiq, Farooq Ahmad, Sher afzal Khan, Jose Valverde, Tabbasum Naz, and Muhammad Anwar. Modeling and analysis of departure routine in air traffic control based on Petri nets. *Neural Computing and Applications*, 25:1099–1109, October 2014. doi: 10.1007/s00521-014-1590-4.

[5] Daniel Alberto Pamplona and Claudio Jorge Pinto Alves. An overview of air delay: A case study of the Brazilian scenario. *Transportation Research Interdisciplinary Perspectives*, 7:100189, September 2020. ISSN 2590-1982. doi: 10.1016/j.trip.2020.100189. URL http://www.sciencedirect.com/science/article/pii/S2590198220301007.

[6] Performance Review Report :An Assessment of Air Traffic Management in Europe during the Calendar Year 2019. Technical report, Eurocontrol, 2019. URL https://www.eurocontrol.int/publication/performance-review-report-prr-2019.

[7] Stephanie Stolz and Patrick Ky. Reducing Traffic bunching through a more Flexible Air Traffic Flow Management. In *Proceedings of the USA/FAA Air Traffic Management R&D Seminar 2001*, Sante Fe, New mexico USA, 2001.

[8] Royal Schiphol Group. Schiphol Airport CDM Operations Manual, 2019. URL https://www.schiphol.nl/nl/download/b2b/1569488978/7ERl8iHeLELDtgFsnK0mGi.pdf.

[9] Cheng-Lung Wu and Kristie Law. Modelling the delay propagation effects of multiple resource connections in an airline network using a Bayesian network model. *Transportation Research Part E: Logistics and Transportation Review*, 122:62–77, February 2019. ISSN 1366-5545. doi: 10.1016/j.tre.2018.11.004. URL http://www.sciencedirect.com/science/article/pii/S1366554517309857.

[10] Ning Xu, George Donohue, Kathryn Blackmond Laskey, and Chun-Hung Chen. Estimation of delay propagation in the national aviation system using bayesian networks. In *Proceedings of the 6th USA/ Europe Air Traffic Management Research and Development Seminar*, page 11, Baltimore, MD, USA, 2005.

[11] Nikolas Pyrgiotis, Kerry M. Malone, and Amedeo Odoni. Modelling delay propagation within an airport network. *Transportation Research Part C: Emerging Technologies*, 27:60–75, February 2013. ISSN 0968-090X. doi: 10.1016/j.trc.2011.05.017. URL http://www.sciencedirect.com/science/article/pii/S0968090X11000878.

[12] Bruno Campanelli, Jose Javier Ramasco, Pablo Fleurquin, Victor M Eguíluz, Izaro Etxebarria, and Andres Arranz. Data-driven modelling of the tree of reactionary delays. In *International Conference on Research in Air Transportation*, page 8, Istanbul Technical University, Turkey, May 2014.

[13] Bruno Campanelli, Pablo Fleurquin, Víctor Eguíluz, Jose Javier Ramasco, Andres Arranz, Izaro Extebarria, and Carlo Ciruelos. Modeling reactionary delays in the European air transport network. In *SIDs 2014 - Proceedings of the SESAR Innovation Days*, November 2014. Journal Abbreviation: SIDs 2014 - Proceedings of the SESAR Innovation Days Publication Title: SIDs 2014 - Proceedings of the SESAR Innovation Days.

[14] 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(5):221–236, September 2008. ISSN 0969-6997. doi: 10.1016/j.jairtraman.2008.04.010. URL http://www.sciencedirect.com/science/article/pii/S0969699708000550.

[15] Karthik Gopalakrishnan, Hamsa Balakrishnan, and Richard Jordan. Deconstructing Delay Dynamics. In *Proceedings of the International Conference on Research in Air Transportation*, page 8, Drexel University,Philadelphia, United States of America, June 2016.

[16] Banavar Sridhar, Yao Wang, and Alexander Klein. Modeling Flight Delays and Cancellations at the National, Regional and Airport Levels in the United States. In *Eighth USA/Europe Air Traffic Management Research and Development Seminar*, Napa, California, USA, 2009.

[17] Juan Jose Rebollo and Hamsa Balakrishnan. Characterization and prediction of air traffic delays. *Transportation Research Part C: Emerging Technologies*, 44:231–241, July 2014. ISSN 0968-090X. doi: 10.1016/j.trc.2014.04.007. URL http://www.sciencedirect.com/science/article/

pii/S0968090X14001041.

[18] Philippe Monmousseau, Daniel Delahaye, Aude Marzuoli, and Eric Feron. Predicting and Analyzing US Air Traffic Delays using Passenger-centric Data-sources. In *Thirteenth USA/Europe Air Traffic Management Research and Development SeminarManagement Research & Development Seminar*, page 10, 2019.

[19] Bin Yu, Zhen Guo, Sobhan Asian, Huaizhu Wang, and Gang Chen. Flight delay prediction for commercial air transport: A deep learning approach. *Transportation Research Part E: Logistics and Transportation Review*, 125:203–221, May 2019. ISSN 1366-5545. doi: 10.1016/j.tre.2019.03.013. URL http://www.sciencedirect.com/science/article/pii/S1366554518311979.

[20] Guglielmo Lulli and Amedeo Odoni. The European Air Traffic Flow Management Problem. *Transportation Science*, 41(4):431–443, November 2007. ISSN 0041-1655. doi: 10.1287/trsc.1070.0214. URL https://pubsonline.informs.org/doi/abs/10.1287/trsc.1070.0214. Publisher: INFORMS.

[21] Tatjana Bolić, Lorenzo Castelli, Luca Corolli, and Desirée Rigonat. Reducing ATFM delays through strategic flight planning. *Transportation Research Part E: Logistics and Transportation Review*, 98:42–59, February 2017. ISSN 1366-5545. doi: 10.1016/j.tre.2016.12.001. URL http://www.sciencedirect.com/science/article/pii/S1366554516305427.

[22] Bruno F. Santos, Maarten M. E. C. Wormer, Thomas A. O. Achola, and Richard Curran. Airline delay management problem with airport capacity constraints and priority decisions. *Journal of Air Transport Management*, 63:34–44, August 2017. ISSN 0969-6997. doi: 10.1016/j.jairtraman.2017.05.003. URL http://www.sciencedirect.com/science/article/pii/S0969699716302514.

[23] Nuno Fernandes, Sérgio Moro, Carlos J. Costa, and Manuela Aparício. Factors influencing charter flight departure delay. *Research in Transportation Business & Management*, 34:100413, March 2020. ISSN 2210-5395. doi: 10.1016/j.rtbm.2019.100413. URL http://www.sciencedirect.com/science/article/pii/S2210539519300495.

[24] Tony Diana. Validating delay constructs: An application of confirmatory factor analysis. *Journal of Air Transport Management*, 35:87–91, March 2014. ISSN 0969-6997. doi: 10.1016/j.jairtraman.2013.11.014. URL http://www.sciencedirect.com/science/article/pii/S0969699713001440.

[25] Mohamed Abdel-Aty, Chris Lee, Yuqiong Bai, Xin Li, and Martin Michalak. Detecting periodic patterns of arrival delay. *Journal of Air Transport Management*, 13(6):355–361, November 2007. ISSN 0969-6997. doi: 10.1016/j.jairtraman.2007.06.002. URL http://www.sciencedirect.com/science/article/pii/S0969699707000646.

[26] Hugo M. Proença, Ruben Klijn, Thomas Bäck, and Matthijs van Leeuwen. Identifying flight delay patterns using diverse subgroup discovery. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 60–67, November 2018. doi: 10.1109/SSCI.2018.8628933.

[27] Alvaro Rodriguez-Sanz, Fernando Gomez Comendador, Rosa Arnaldo Valdes, and Javier A. Perez-Castan. Characterization and prediction of the airport operational saturation. *Journal of Air Transport Management*, 69:147–172, June 2018. ISSN 0969-6997. doi: 10.1016/j.jairtraman.2018.03.002. URL http://www.sciencedirect.com/science/article/pii/S0969699717303691.

[28] Dothang Truong. Using causal machine learning for predicting the risk of flight delays in air transportation. *Journal of Air Transport Management*, 91:101993, March 2021. ISSN 0969-6997. doi: 10.1016/j.jairtraman.2020.101993. URL https://www.sciencedirect.com/science/article/pii/S0969699720305755.

[29] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 3 edition, 2012.

[30] Hanchuan Peng, Fuhui Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, August 2005. ISSN 1939-3539. doi: 10.1109/TPAMI.2005.159. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[31] Andrew P. King and Robert J. Eckersley. Chapter 2 - Descriptive Statistics II: Bivariate and Multivariate Statistics. In Andrew P. King and Robert J. Eckersley, editors, *Statistics for Biomedical Engineers and Scientists*, pages 23–56. Academic Press, January 2019. ISBN 978-0-08-102939-8. doi: 10.1016/B978-0-08-102939-8.00011-6. URL https://www.sciencedirect.com/science/article/pii/B9780081029398000116.

[32] Schiphol Group. Practical tools to improve Predictability Performance, March 2021. KLM Internal document.

[33] Eurocontrol. Standard IATA Delay Codes-Airport Handling Manual. URL https://ansperformance.eu/library/iata-delay-codes.pdf.

[34] Eurocontrol. A-CDM Impact Assessment - Final Report, March 2016. URL https://www.eurocontrol.int/sites/default/files/2019-04/a-cdm-impact-assessment-2016.pdf.

[35] Jochen Hipp, Ulrich Güntzer, and Gholamreza Nakhaeizadeh. Algorithms for association rule mining - a general survey and comparison. *Association for Computing Machinery SIGKDD Explorations Newsletter*, 2(1):58–64, June 2000. ISSN 1931-0145. doi: 10.1145/360402.360421. URL https://doi.org/10.1145/360402.360421.

[36] Rakesh Agrawal and Ramakrishnan Srikant. Fast Algorithms for Mining Association Rules in Large Databases. In *Proceedings of the 20th International Conference on Very Large Data Bases*, VLDB '94, pages 487–499, San Francisco, CA, USA, September 1994. Morgan Kaufmann Publishers Inc. ISBN 978-1-55860-153-6.

[37] Jiawei Han, Jian Pei, and Yiwen Yin. Mining frequent patterns without candidate generation. *ACM SIGMOD*

Record, 29(2):1–12, May 2000. ISSN 0163-5808. doi: 10.1145/335191.335372. URL https://doi.org/10.1145/335191.335372.

[38] Mohammed Zaki, Srinivasan Parthasarathy, Mitsunori Ogihara, and Wei li. New Algorithms for Fast Discovery of Association Rules. In *Conference on Knowlegde Discovery and Data Mining*, pages 283–286, January 1997.

[39] Jiawei Han, Hong Cheng, Dong Xin, and Xifeng Yan. Frequent pattern mining: Current status and future directions. *Data Mining and Knowledge Discovery*, 15:55–86, July 2007. doi: 10.1007/s10618-006-0059-1.

[40] Sebastian Raschka. MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open Source Software*, 3(24), April 2018. ISSN 2475-9066. doi: 10.21105/joss.00638. URL https://joss.theoj.org/papers/10.21105/joss.00638.

[41] Dong Xin, Hong Cheng, Xifeng Yan, and Jiawei Han. Extracting redundancy-aware top-k patterns. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '06, pages 444–453, New York, NY, USA, August 2006. Association for Computing Machinery. ISBN 978-1-59593-339-3. doi: 10.1145/1150402.1150452. URL https://doi.org/10.1145/1150402.1150452.

[42] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Elsevier Science & Technology, 1988. ISBN 978-0-08-051489-5.

[43] Marco Scutari, Catharina Graafland, and J. Gutiérrez. Who Learns Better Bayesian Network Structures: Constraint-Based, Score-based or Hybrid Algorithms? In *Proceedings of Machine Learning Research*, volume 72, pages 416–427, 2018.

[44] Jie Cheng, Russell Greiner, Jonathan Kelly, David Bell, and Weiru Liu. Learning Bayesian networks from data: An information-theory based approach. *Artificial Intelligence*, 137(1):43–90, May 2002. ISSN 0004-3702. doi: 10.1016/S0004-3702(02)00191-1. URL http://www.sciencedirect.com/science/article/pii/S0004370202001911.

[45] Luis de Campos. A Scoring Function for Learning Bayesian Networks based on Mutual Information and Conditional Independence Tests. *Journal of Machine Learning Research*, 7:2149–2187, October 2006.

[46] Kazuki Natori, Masaki Uto, Yu Nishiyama, Shuichi Kawano, and Maomi Ueno. Constraint-Based Learning Bayesian Networks Using Bayes Factor. In Joe Suzuki and Maomi Ueno, editors, *Advanced Methodologies for Bayesian Networks*, Lecture Notes in Computer Science, pages 15–31, Cham, 2015. Springer International Publishing. ISBN 978-3-319-28379-1. doi: 10.1007/978-3-319-28379-1_2.

[47] Ioannis Tsamardinos, Laura E. Brown, and Constantin F. Aliferis. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, 65(1):31–78, October 2006. ISSN 1573-0565. doi: 10.1007/s10994-006-6889-7. URL https://doi.org/10.1007/s10994-006-6889-7.

[48] Hongru Li and Huiping Guo. A Hybrid Structure Learning Algorithm for Bayesian Network Using Experts' Knowledge. *Entropy*, 20:620, August 2018. doi: 10.3390/e20080620.

[49] Frederick S. Hillier and Gerald J. Lieberman. *Introduction to Operations Research*. Tenth edition,. McGraw-Hill Education, 2015.

[50] Silvia Acid, Luis M. de Campos, Juan M. Fernandez-Luna, Susana Rodriguez, Jose Maria Rodriguez, and Jose Luis Salcedo. A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service. *Artificial Intelligence in Medicine*, 30(3):215–232, March 2004. ISSN 0933-3657. doi: 10.1016/j.artmed.2003.11.002. URL https://www.sciencedirect.com/science/article/pii/S0933365703001325.

[51] Ingo A. Beinlich, H. J. Suermondt, R. Martin Chavez, and Gregory F. Cooper. The ALARM Monitoring System: A Case Study with two Probabilistic Inference Techniques for Belief Networks. In Jim Hunter, John Cookson, and Jeremy Wyatt, editors, *AIME 89*, Lecture Notes in Medical Informatics, pages 247–256, Berlin, Heidelberg, 1989. Springer. ISBN 978-3-642-93437-7. doi: 10.1007/978-3-642-93437-7_28.

[52] Alexandra M Carvalho. Scoring functions for learning bayesian networks. *Inesc-id Tec. Rep*, 12, 2009.

[53] David Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, UK, 2012.

[54] Harald Steck. Learning the Bayesian network structure: dirichlet prior versus data. In *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 511–518, Arlington, Virginia, USA, July 2008. AUAI Press. ISBN 978-0-9749039-4-1.

[55] Korb B. Kevin and Ann E. Nicholson. *Bayesian Artificial Intelligence*. CRC Press, 2 edition, 2010.

[56] Ankur Ankan and Abinash Panda. pgmpy: Probabilistic graphical models using python. In *Proceedings of the 14th Python in Science Conference (SCIPY 2015)*. Citeseer, 2015.

[57] Erdogan Taskesen. bnlearn, 2019. URL https://github.com/erdogant/bnlearn. original-date: 2020-01-01T21:00:30Z.

[58] Dion H. Goh and Rebecca P. Ang. An introduction to association rule mining: An application in counseling and help-seeking behavior of adolescents. *Behavior Research Methods*, 39(2):259–266, May 2007. ISSN 1554-3528. doi: 10.3758/BF03193156. URL https://doi.org/10.3758/BF03193156.

[59] Carlos Ordonez. Association Rule Discovery With the Train and Test Approach for Heart Disease Prediction. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, 10(2):334–343, May 2006. doi: 10.1109/TITB.2006.864475.

[60] Mingkai Peng, Sangmin Lee, Adam G. D'Souza, Chelsea T. A. Doktorchik, and Hude Quan. Development and validation of data quality rules in administrative health

data using association rule mining. *BMC Medical Informatics and Decision Making*, 20(1):75, April 2020. ISSN 1472-6947. doi: 10.1186/s12911-020-1089-0. URL https://doi.org/10.1186/s12911-020-1089-0.

II

SCIENTIFIC ARTICLE APPENDICES

A

Data Integration

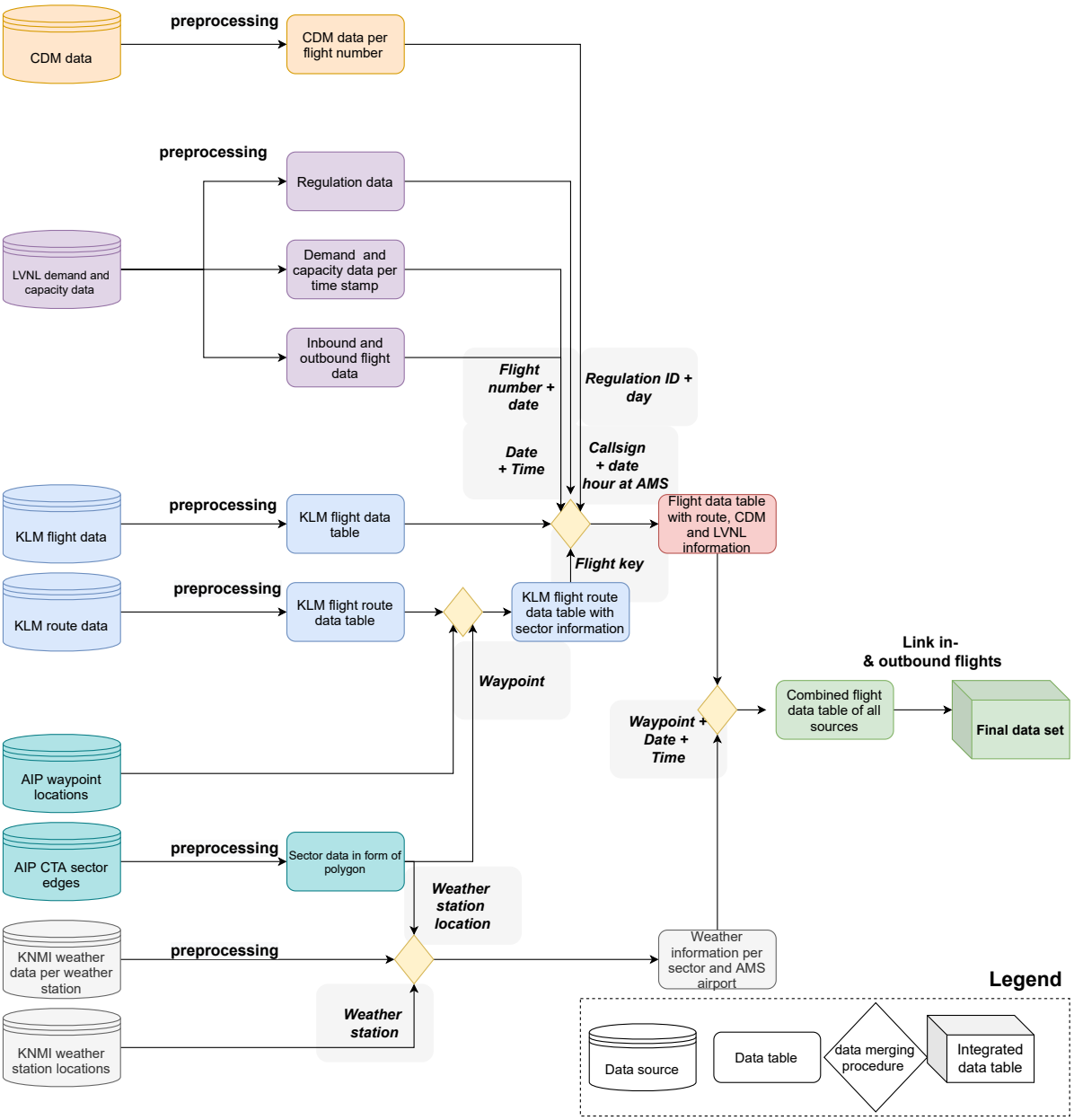


Figure A.1 – Data pipeline and integration diagram.

B

CORRELATION MATRICES

B.1. SPEARMAN CORRELATION MATRIX

	atxi_inbound	wags_inbound	adcdla_inbound	arrdla_inbound	cindla_inbound	dly1_dur_inbound	dly2_dur_inbound	nrctot_inbound	purdla_inbound	reg_induced_dla_inbound	RATE_inbound	DD_inbound	VV_inbound	P_sector_inbound	EHAMARR_inbound
atxi_inbound	1	0.008	-0.064	0.152	-0.002	-0.068	-0.041	-0.078	-0.081	0.039	-0.039	0.018	-0.14	-0.153	-0.094
wags_inbound	0.008	1	-0.085	-0.12	-0.01	-0.101	-0.075	-0.21	-0.171	0.052	-0.165	-0.019	-0.05	-0.016	-0.13
adcdla_inbound	-0.064	-0.085	1	0.579	0.086	0.72	0.531	0.263	0.344	0.177	0.069	0.057	0.123	-0.019	-0.057
arrdla_inbound	0.152	-0.12	0.579	1	0.085	0.715	0.403	0.387	0.544	-0.036	0.204	0.067	0.073	-0.102	-0.1
cindla_inbound	-0.002	-0.01	0.086	0.085	1	0.085	0.068	0.069	0.015	-0.028	-0.001	0.008	0.002	0.005	-0.035
dly1_dur_inbound	-0.068	-0.101	0.72	0.715	0.085	1	0.363	0.417	0.612	0.048	0.223	0.052	0.104	-0.044	-0.062
dly2_dur_inbound	-0.041	-0.075	0.531	0.403	0.068	0.363	1	0.242	0.321	0.098	0.065	0.044	0.092	-0.028	-0.069
nrctot_inbound	-0.078	-0.21	0.263	0.387	0.069	0.417	0.242	1	0.729	-0.161	0.631	0.055	0.113	-0.066	0.127
purdla_inbound	-0.081	-0.171	0.344	0.544	0.015	0.612	0.321	0.729	1	-0.29	0.497	0.061	0.112	-0.065	0.016
reg_induced_dla_inbound	0.039	0.052	0.177	-0.036	-0.028	0.048	0.098	-0.161	-0.29	1	-0.169	-0.001	0.002	0.014	-0.042
RATE_inbound	-0.039	-0.165	0.069	0.204	-0.001	0.223	0.065	0.631	0.497	-0.169	1	0.025	0.052	-0.051	0.237
DD_inbound	0.018	-0.019	0.057	0.067	0.008	0.052	0.044	0.055	0.061	-0.001	0.025	1	0.056	-0.026	-0.008
VV_inbound	-0.14	-0.05	0.123	0.073	0.002	0.104	0.092	0.113	0.112	0.002	0.052	0.056	1	0.193	0.07
P_sector_inbound	-0.153	-0.016	-0.019	-0.102	0.005	-0.044	-0.028	-0.066	-0.065	0.014	-0.051	-0.026	0.193	1	0.026
EHAMARR_inbound	-0.094	-0.13	-0.057	-0.1	-0.035	-0.062	-0.069	0.127	0.016	-0.042	0.237	-0.008	0.07	0.026	1
c_EHARR_inbound	-0.102	-0.122	-0.026	-0.148	-0.018	-0.082	-0.039	-0.045	-0.097	0.009	0.078	-0.015	0.082	0.096	0.514
nr_fl_inbound_inbound	-0.109	-0.171	0.019	0.029	-0.03	0.076	0.011	0.248	0.198	-0.094	0.367	0	0.108	0.037	0.687
Average Rate Arr_inbound	-0.11	-0.175	0.074	0.088	-0.014	0.142	0.061	0.291	0.254	-0.095	0.375	0.008	0.122	0.036	0.537
mdlg_arr_inbound	-0.062	0	0.015	-0.016	0	0.016	0.014	0.037	0.032	-0.007	0.037	-0.081	0.493	0.151	0.046
wfdep_outbound	0.004	-0.028	0.024	0.024	0.004	0.032	0.032	0.071	0.057	-0.013	0.06	0.001	0.006	-0.037	0.005
adcdla_outbound	0.079	0.028	0.216	0.341	0.024	0.279	0.162	0.167	0.246	-0.016	0.09	0.045	0.051	-0.076	-0.05
depdla_outbound	0.073	0.014	0.22	0.336	0.024	0.284	0.172	0.2	0.267	-0.024	0.122	0.042	0.055	-0.091	-0.044
cindla_outbound	0.006	0.004	0.011	0.019	0.002	0.014	0.009	0.012	0.018	0	0.009	0.005	0.007	-0.003	0.01
dly1_dur_outbound	0.065	0.01	0.205	0.31	0.024	0.265	0.16	0.187	0.251	-0.022	0.114	0.04	0.05	-0.082	-0.042
dly2_dur_outbound	0.056	0.017	0.133	0.22	0.01	0.173	0.105	0.1	0.151	-0.01	0.057	0.022	0.031	-0.055	-0.039
lstdla_outbound	0.009	-0.022	0.078	0.107	0.008	0.1	0.062	0.094	0.109	-0.007	0.054	0.022	0.068	-0.028	0.012
nrctot_outbound	-0.006	-0.027	0.051	0.059	0.005	0.066	0.042	0.085	0.083	-0.007	0.05	0.007	0.074	-0.011	0.03
purdla_outbound	0.008	-0.023	0.077	0.105	0.009	0.099	0.061	0.094	0.108	-0.006	0.053	0.023	0.068	-0.027	0.011
locatc_outbound	0.013	0.01	0.004	0.002	-0.001	0.009	0.003	0.031	0.022	-0.009	0.049	-0.001	-0.008	-0.025	0.015
reg_induced_dla_outbound	0.005	-0.013	0.02	0.027	-0.004	0.027	0.019	0.025	0.031	-0.001	0.018	-0.019	0.003	-0.009	0.007
#TOBT_outbound	0.069	-0.021	0.171	0.328	0.026	0.245	0.138	0.111	0.205	-0.016	0.044	0.027	0.027	-0.028	-0.094
#TSAT_outbound	0.025	0.018	0.113	0.166	0.01	0.149	0.089	0.145	0.169	-0.034	0.1	0.048	0.043	-0.038	-0.05
#TTOT_outbound	0.025	0.025	0.108	0.16	0.009	0.144	0.085	0.137	0.161	-0.032	0.094	0.046	0.038	-0.037	-0.054
#CTOT_outbound	-0.004	-0.023	0.058	0.071	0.006	0.074	0.047	0.081	0.088	-0.007	0.044	0.015	0.068	-0.015	0.016
#EXOT_outbound	-0.029	-0.037	0.032	0.021	0.007	0.03	0.028	0.017	0.022	-0.006	-0.019	0.013	-0.002	0.014	-0.003
#ASRT_outbound	-0.001	0.012	0.004	0.004	0.002	0.003	0.005	-0.005	0.001	0.004	-0.008	0	0.001	0.004	-0.005
STARTUP DLA_outbound	0.008	0.023	0.075	0.087	0.004	0.088	0.062	0.095	0.101	-0.007	0.078	0.038	0.042	-0.034	-0.011
D TTOT TARGET_outbound	0.028	0.038	0.074	0.091	0.003	0.088	0.06	0.088	0.097	-0.005	0.072	0.037	0.034	-0.045	-0.02
D TTOT OPTIMAL_outbound	0.035	0.069	0.06	0.087	0.003	0.066	0.041	-0.013	0.02	0.021	-0.023	0.016	0.001	-0.032	-0.047
% REG FLIGHTS AROUND_outbound	-0.013	-0.009	0.098	0.118	0.007	0.122	0.073	0.152	0.146	-0.023	0.074	0.02	0.199	-0.002	0.011
AVG STARTUP DLA AROUND_outbound	0.035	0.054	0.116	0.148	0.01	0.149	0.092	0.154	0.173	-0.023	0.113	0.065	0.081	-0.067	-0.063
DD_outbound	0.042	-0.001	0.041	0.06	0.005	0.039	0.031	0.039	0.046	-0.001	0.014	0.786	0.022	-0.052	-0.019
VV_outbound	-0.111	0.008	0.077	0.046	-0.003	0.074	0.052	0.062	0.077	-0.004	0.021	0.046	0.773	0.194	0.045
P_sector_outbound	-0.155	-0.013	-0.02	-0.101	0.006	-0.042	-0.028	-0.064	-0.062	0.013	-0.046	0.019	0.183	0.97	0.028
EHAMDEP_outbound	-0.035	-0.052	0.012	-0.014	-0.009	0.049	0.013	0.139	0.116	-0.044	0.198	-0.008	0.038	0.022	0.223
c_EHDEP_outbound	-0.068	-0.034	0.01	-0.08	0	-0.003	0.008	0.042	0.016	-0.001	0.079	0.032	0.15	0.089	0.179
nr_fl_outbound_outbound	-0.026	-0.023	0.04	0.006	0.002	0.063	0.035	0.134	0.116	-0.03	0.182	0.01	0.06	0.031	0.169
Average Rate Dep_outbound	-0.004	0.017	0.059	0.026	0.007	0.076	0.048	0.11	0.108	-0.017	0.141	0.015	0.07	0.031	0.099
mdlg_dep_outbound	-0.065	0.002	0.028	0.003	0	0.032	0.019	0.049	0.05	-0.011	0.049	-0.066	0.454	0.137	0.039
c_d_EHAMDEP_outbound	-0.019	0.014	0.01	-0.031	0.005	-0.029	0.006	-0.074	-0.071	0.036	-0.114	0.034	0.073	0.011	-0.063
c_d_EHAMARR_inbound	0.007	0.039	0.017	0.009	0.015	0.03	0.022	-0.098	-0.022	0.004	-0.084	-0.011	-0.034	0.023	-0.62
atcdla_outbound	0.015	0.015	0.102	0.134	0.011	0.126	0.081	0.111	0.133	-0.011	0.081	0.043	0.044	-0.043	-0.02
d_depdla_atcdla	0.08	0.004	0.162	0.274	0.017	0.216	0.129	0.132	0.192	-0.017	0.07	0.007	0.017	-0.075	-0.038
pln_tat	0.101	0.265	0.029	0.008	0.018	0.009	-0.002	-0.16	-0.108	0.07	-0.171	-0.005	-0.091	-0.011	-0.165
act_tat	0.054	0.291	-0.118	-0.245	-0.016	-0.177	-0.112	-0.252	-0.252	0.079	-0.211	-0.019	-0.092	0.008	-0.116
d_tat	-0.097	0.125	-0.428	-0.746	-0.073	-0.517	-0.287	-0.252	-0.363	0.022	-0.12	-0.037	-0.032	0.043	0.074
tat_delay	0.008	0.068	-0.207	-0.305	-0.058	-0.243	-0.154	-0.082	-0.15	0.004	0.018	-0.008	0.017	-0.024	0.081
boarding_duration	0.067	0.215	-0.051	-0.073	-0.011	-0.068	-0.06	-0.18	-0.139	0.049	-0.148	-0.007	-0.034	0.008	-0.096
waiting_ready	0.039	0.062	0.002	0.021	-0.004	0.012	0.009	-0.004	0.004	0.004	0.011	-0.014	-0.018	-0.025	-0.02
time_in_fir	0.01	0.012	-0.029	0.121	-0.011	-0.009	-0.008	0.014	0.022	-0.023	0.088	-0.042	-0.116	-0.213	0.026
time_in_tma	0.012	-0.052	0.028	0.144	0.003	0.053	0.016	0.051	0.066	-0.026	0.008	0.032	0.093	0.19	-0.06
D_T_TD_inbound	-0.16	-0.062	0.136	0.069	0.003	0.107	0.095	0.114	0.109	-0.001	0.021	0.021	0.835	0.213	0.071
D_T_TD_outbound	-0.118	0.012	0.059	0.023	-0.006	0.058	0.033	0.05	0.062	-0.012	-0.008	-0.025	0.66	0.224	0.031
diff_capacity_actual_outbound	-0.061	-0.061	-0.027	-0.072	-0.012	-0.036	-0.021	-0.005	-0.027	-0.003	-0.006	0.011	0.057	0.012	0.122
diff_capacity_actual_inbound	0.018	0.082	-0.134	-0.229	-0.015	-0.219	-0.128	-0.286	-0.309	0.073	-0.214	-0.032	-0.074	0.027	-0.019

	c_EHARR_inbound	nr_fl_inbound_inbound	Average Rate Arr_inbound	mclg_arr_inbound	wfdep_outbound	adcdla_outbound	depdla_outbound	cindla_outbound	dly1_dur_outbound	dly2_dur_outbound	lstdla_outbound	nrctot_outbound	purdla_outbound	locatc_outbound
atxi_inbound	-0.102	-0.109	-0.11	-0.062	0.004	0.079	0.073	0.006	0.065	0.056	0.009	-0.006	0.008	0.013
wags_inbound	-0.122	-0.171	-0.175	0	-0.028	0.028	0.014	0.004	0.01	0.017	-0.022	-0.027	-0.023	0.01
adcdla_inbound	-0.026	0.019	0.074	0.015	0.024	0.216	0.22	0.011	0.205	0.133	0.078	0.051	0.077	0.004
arrdla_inbound	-0.148	0.029	0.088	-0.016	0.024	0.341	0.336	0.019	0.31	0.22	0.107	0.059	0.105	0.002
cindla_inbound	-0.018	-0.03	-0.014	0	0.004	0.024	0.024	0.002	0.024	0.01	0.008	0.005	0.009	-0.001
dly1_dur_inbound	-0.082	0.076	0.142	0.016	0.032	0.279	0.284	0.014	0.265	0.173	0.1	0.066	0.099	0.009
dly2_dur_inbound	-0.039	0.011	0.061	0.014	0.032	0.162	0.172	0.009	0.16	0.105	0.062	0.042	0.061	0.003
nrctot_inbound	-0.045	0.248	0.291	0.037	0.071	0.167	0.2	0.012	0.187	0.1	0.094	0.085	0.094	0.031
purdla_inbound	-0.097	0.198	0.254	0.032	0.057	0.246	0.267	0.018	0.251	0.151	0.109	0.083	0.108	0.022
reg_induced_dla_inbound	0.009	-0.094	-0.095	-0.007	-0.013	-0.016	-0.024	0	-0.022	-0.01	-0.007	-0.007	-0.006	-0.009
RATE_inbound	0.078	0.367	0.375	0.037	0.06	0.09	0.122	0.009	0.114	0.057	0.054	0.05	0.053	0.049
DD_inbound	-0.015	0	0.008	-0.081	0.001	0.045	0.042	0.005	0.04	0.022	0.022	0.007	0.023	-0.001
VV_inbound	0.082	0.108	0.122	0.493	0.006	0.051	0.055	0.007	0.05	0.031	0.068	0.074	0.068	-0.008
P_sector_inbound	0.096	0.037	0.036	0.151	-0.037	-0.076	-0.091	-0.003	-0.082	-0.055	-0.028	-0.011	-0.027	-0.025
EHAMARR_inbound	0.514	0.687	0.537	0.046	0.005	-0.05	-0.044	0.01	-0.042	-0.039	0.012	0.03	0.011	0.015
c_EHARR_inbound	1	0.593	0.56	0.059	0.016	-0.131	-0.113	0	-0.105	-0.077	-0.025	-0.014	-0.025	0.022
nr_fl_inbound_inbound	0.593	1	0.895	0.068	0.045	-0.01	0.02	0.009	0.014	0.007	0.028	0.035	0.028	0.05
Average Rate Arr_inbound	0.56	0.895	1	0.071	0.068	0.015	0.057	0.009	0.048	0.032	0.036	0.038	0.036	0.067
mclg_arr_inbound	0.059	0.068	0.071	1	-0.001	-0.004	-0.003	0.003	-0.004	-0.003	0.037	0.048	0.037	-0.003
wfdep_outbound	0.016	0.045	0.068	-0.001	1	-0.218	0.307	0.031	0.256	0.166	0.182	0.129	0.181	0.457
adcdla_outbound	-0.131	-0.01	0.015	-0.004	-0.218	1	0.8	0.052	0.727	0.534	0.202	0.095	0.191	0.001
depdla_outbound	-0.113	0.02	0.057	-0.003	0.307	0.8	1	0.064	0.932	0.546	0.316	0.167	0.306	0.26
cindla_outbound	0	0.009	0.009	0.003	0.031	0.052	0.064	1	0.055	0.045	0.21	0.145	0.063	-0.009
dly1_dur_outbound	-0.105	0.014	0.048	-0.004	0.256	0.727	0.932	0.055	1	0.301	0.288	0.148	0.281	0.224
dly2_dur_outbound	-0.077	0.007	0.032	-0.003	0.166	0.534	0.546	0.045	0.301	1	0.174	0.085	0.167	0.292
lstdla_outbound	-0.025	0.028	0.036	0.037	0.182	0.202	0.316	0.21	0.288	0.174	1	0.728	0.977	-0.121
nrctot_outbound	-0.014	0.035	0.038	0.048	0.129	0.095	0.167	0.145	0.148	0.085	0.728	1	0.732	-0.126
purdla_outbound	-0.025	0.028	0.036	0.037	0.181	0.191	0.306	0.063	0.281	0.167	0.977	0.732	1	-0.125
locatc_outbound	0.022	0.05	0.067	-0.003	0.457	0.001	0.26	-0.009	0.224	0.292	-0.121	-0.126	-0.125	1
reg_induced_dla_outbound	0.005	0.017	0.022	-0.001	-0.032	0.087	0.069	-0.094	0.075	0.034	-0.038	0.111	-0.016	0.041
#TOBT_outbound	-0.12	-0.057	-0.033	-0.017	0	0.399	0.404	0.027	0.386	0.281	0.102	0.001	0.102	-0.115
#TSAT_outbound	-0.08	0.014	0.058	-0.014	0.287	0.247	0.418	0.004	0.384	0.235	-0.096	-0.144	-0.106	0.303
#TTOT_outbound	-0.082	0.008	0.051	-0.016	0.279	0.241	0.408	0.002	0.374	0.23	-0.124	-0.18	-0.133	0.305
#CTOT_outbound	-0.021	0.023	0.028	0.041	0.112	0.121	0.183	0.109	0.164	0.102	0.696	0.873	0.709	-0.119
#EXOT_outbound	-0.028	-0.019	-0.023	-0.037	-0.01	0.078	0.062	0.011	0.055	0.047	0.014	0.004	0.011	-0.025
#ASRT_outbound	-0.012	-0.009	-0.012	0.001	0.05	0.034	0.063	0.023	0.054	0.039	0.015	0.008	0.011	0.029
STARTUP DLA_outbound	-0.013	0.05	0.081	0.016	0.501	0.217	0.529	0.04	0.478	0.319	0.216	0.108	0.213	0.365
D TTOT TARGET_outbound	-0.023	0.04	0.073	0.011	0.505	0.22	0.532	0.043	0.48	0.321	0.219	0.116	0.216	0.37
D TTOT OPTIMAL_outbound	-0.074	-0.051	-0.048	0	0.095	0.451	0.513	0.047	0.452	0.342	0.135	0.014	0.127	0.1
% REG FLIGHTS AROUND_outbound	-0.064	0.027	0.04	0.12	0.068	0.108	0.142	0.054	0.128	0.08	0.288	0.393	0.29	-0.019
AVG STARTUP DLA AROUND_outbound	-0.074	0.032	0.089	0.035	0.357	0.198	0.413	0.032	0.373	0.234	0.152	0.104	0.149	0.271
DD_outbound	-0.054	-0.024	-0.022	-0.094	-0.001	0.059	0.053	0.008	0.051	0.031	0.028	0.011	0.028	0.001
VV_outbound	0.006	0.059	0.069	0.449	-0.013	0.085	0.071	0.011	0.063	0.051	0.068	0.072	0.067	-0.009
P_sector_outbound	0.094	0.038	0.037	0.148	-0.035	-0.073	-0.087	-0.002	-0.078	-0.053	-0.023	-0.006	-0.022	-0.024
EHAMDEP_outbound	0.254	0.358	0.381	0.027	0.042	0.02	0.04	0	0.035	0.019	0.017	0.029	0.018	0.085
c_EHDEP_outbound	0.291	0.282	0.296	0.139	-0.004	-0.043	-0.042	0	-0.041	-0.025	-0.001	0.016	-0.001	0.029
nr_fl_outbound_outbound	0.237	0.34	0.387	0.038	0.141	0.041	0.127	0.001	0.113	0.066	0.016	0.004	0.015	0.16
Average Rate Dep_outbound	0.186	0.283	0.354	0.041	0.156	0.065	0.168	0.001	0.155	0.084	0.023	-0.003	0.023	0.166
mclg_dep_outbound	0.039	0.063	0.066	0.757	0.002	0.009	0.011	0.007	0.009	0.004	0.043	0.049	0.042	-0.006
c_d_EHAMDEP_outbound	-0.02	-0.104	-0.113	0.053	-0.049	-0.053	-0.071	-0.005	-0.066	-0.039	-0.023	-0.023	-0.023	-0.061
c_d_EHAMARR_inbound	0.244	-0.204	-0.088	-0.015	-0.006	-0.016	-0.019	-0.007	-0.019	-0.002	-0.024	-0.037	-0.023	0.001
atcdla_outbound	-0.028	0.042	0.073	0.012	0.432	0.302	0.571	0.089	0.521	0.341	0.433	0.102	0.422	0.297
d_depdla_atcdla	-0.109	-0.017	-0.001	-0.019	-0.049	0.674	0.621	0.001	0.578	0.362	-0.041	0.059	-0.045	0.023
pln_tat	-0.145	-0.204	-0.21	-0.037	0.05	-0.139	-0.088	-0.006	-0.073	-0.083	-0.026	0.002	-0.024	0.049
act_tat	-0.095	-0.185	-0.206	-0.027	0.123	-0.113	-0.017	0.009	-0.016	-0.037	0.008	0.024	0.008	0.09
d_tat	0.073	-0.013	-0.047	0.014	0.18	0.156	0.275	0.034	0.242	0.151	0.096	0.051	0.089	0.127
tat_delay	0.043	0.059	0.05	0.021	0.266	0.453	0.627	0.051	0.569	0.358	0.205	0.112	0.196	0.235
boarding_duration	-0.089	-0.124	-0.13	-0.01	-0.028	0.082	0.065	0.007	0.054	0.059	0.004	-0.005	0.002	0.02
waiting_ready	-0.016	-0.003	0.011	-0.01	0.339	0.039	0.22	0.03	0.19	0.137	0.127	0.111	0.125	0.207
time_in_fir	-0.011	0.093	0.079	-0.035	0.015	0.044	0.048	-0.005	0.044	0.031	-0.013	-0.036	-0.014	0.018
time_in_tma	-0.094	-0.033	-0.015	-0.029	-0.009	0.067	0.06	0.012	0.054	0.047	0.043	0.049	0.042	-0.013
D_T_TD_inbound	0.076	0.102	0.117	0.527	-0.003	0.055	0.052	0.005	0.048	0.033	0.07	0.083	0.07	-0.016
D_T_TD_outbound	-0.03	0.035	0.044	0.495	-0.021	0.089	0.07	0.008	0.06	0.054	0.069	0.082	0.068	-0.011
diff_capacity_actual_outbound	0.119	0.071	0.023	0.049	-0.157	-0.084	-0.183	-0.004	-0.17	-0.095	-0.023	0.02	-0.022	-0.125
diff_capacity_actual_inbound	0.307	-0.352	-0.525	-0.032	-0.076	-0.105	-0.145	-0.007	-0.131	-0.089	-0.054	-0.049	-0.054	-0.054

	reg_induced_dla_outbound	#TOBT_outbound	#TSAT_outbound	#TTOT_outbound	#CTOT_outbound	#EXOT_outbound	#ASRT_outbound	STARTUP DLA_outbound	D TTOT TARGET_outbound	D TTOT OPTIMAL_outbound	% REG FLIGHTS AROUND_outbound
atxi_inbound	0.005	0.069	0.025	0.025	-0.004	-0.029	-0.001	0.008	0.028	0.035	-0.013
wags_inbound	-0.013	-0.021	0.018	0.025	-0.023	-0.037	0.012	0.023	0.038	0.069	-0.009
adcdla_inbound	0.02	0.171	0.113	0.108	0.058	0.032	0.004	0.074	0.074	0.06	0.098
arrdla_inbound	0.027	0.328	0.166	0.16	0.071	0.021	0.004	0.087	0.091	0.087	0.118
cindla_inbound	-0.004	0.026	0.01	0.009	0.006	0.007	0.002	0.004	0.003	0.003	0.007
dly1_dur_inbound	0.027	0.245	0.149	0.144	0.074	0.03	0.003	0.088	0.088	0.066	0.122
dly2_dur_inbound	0.019	0.138	0.089	0.085	0.047	0.028	0.005	0.062	0.06	0.041	0.073
nrctot_inbound	0.025	0.111	0.145	0.137	0.081	0.017	-0.005	0.095	0.088	-0.013	0.152
purdla_inbound	0.031	0.205	0.169	0.161	0.088	0.022	0.001	0.101	0.097	0.02	0.146
reg_induced_dla_inbound	-0.001	-0.016	-0.034	-0.032	-0.007	-0.006	0.004	-0.007	-0.005	0.021	-0.023
RATE_inbound	0.018	0.044	0.1	0.094	0.044	-0.019	-0.008	0.078	0.072	-0.023	0.074
DD_inbound	-0.019	0.027	0.048	0.046	0.015	0.013	0	0.038	0.037	0.016	0.02
VV_inbound	0.003	0.027	0.043	0.038	0.068	-0.002	0.001	0.042	0.034	0.001	0.199
P_sector_inbound	-0.009	-0.028	-0.038	-0.037	-0.015	0.014	0.004	-0.034	-0.045	-0.032	-0.002
EHAMARR_inbound	0.007	-0.094	-0.05	-0.054	0.016	-0.003	-0.005	-0.011	-0.02	-0.047	0.011
c_EHARR_inbound	0.005	-0.12	-0.08	-0.082	-0.021	-0.028	-0.012	-0.013	-0.023	-0.074	-0.064
nr_fl_inbound_inbound	0.017	-0.057	0.014	0.008	0.023	-0.019	-0.009	0.05	0.04	-0.051	0.027
Average Rate Arr_inbound	0.022	-0.033	0.058	0.051	0.028	-0.023	-0.012	0.081	0.073	-0.048	0.04
mdg_arr_inbound	-0.001	-0.017	-0.014	-0.016	0.041	-0.037	0.001	0.016	0.011	0	0.12
wfdep_outbound	-0.032	0	0.287	0.279	0.112	-0.01	0.05	0.501	0.505	0.095	0.068
adcdla_outbound	0.087	0.399	0.247	0.241	0.121	0.078	0.034	0.217	0.22	0.451	0.108
depdla_outbound	0.069	0.404	0.418	0.408	0.183	0.062	0.063	0.529	0.532	0.513	0.142
cindla_outbound	-0.094	0.027	0.004	0.002	0.109	0.011	0.023	0.04	0.043	0.047	0.054
dly1_dur_outbound	0.075	0.386	0.384	0.374	0.164	0.055	0.054	0.478	0.48	0.452	0.128
dly2_dur_outbound	0.034	0.281	0.235	0.23	0.102	0.047	0.039	0.319	0.321	0.342	0.08
lstdla_outbound	-0.038	0.102	-0.096	-0.124	0.696	0.014	0.015	0.216	0.219	0.135	0.288
nrctot_outbound	0.111	0.001	-0.144	-0.18	0.873	0.004	0.008	0.108	0.116	0.014	0.393
purdla_outbound	-0.016	0.102	-0.106	-0.133	0.709	0.011	0.011	0.213	0.216	0.127	0.29
locatc_outbound	0.041	-0.115	0.303	0.305	-0.119	-0.025	0.029	0.365	0.37	0.1	-0.019
reg_induced_dla_outbound	1	0	-0.007	-0.013	0.084	0.005	0.003	-0.007	0	-0.021	0.041
#TOBT_outbound	0	1	0.19	0.189	0.047	0.065	0.028	0.093	0.095	0.149	0.016
#TSAT_outbound	-0.007	0.19	1	0.989	-0.085	0.127	0.055	0.413	0.421	0.038	0.049
#TTOT_outbound	-0.013	0.189	0.989	1	-0.119	0.122	0.055	0.404	0.42	0.046	0.032
#CTOT_outbound	0.084	0.047	-0.085	-0.119	1	0.033	0.015	0.106	0.113	0.031	0.333
#EXOT_outbound	0.005	0.065	0.127	0.122	0.033	1	0.044	-0.015	0.047	0.074	-0.004
#ASRT_outbound	0.003	0.028	0.055	0.055	0.015	0.044	1	0.047	0.048	0.055	0.009
STARTUP DLA_outbound	-0.007	0.093	0.413	0.404	0.106	-0.015	0.047	1	0.93	0.358	0.092
D TTOT TARGET_outbound	0	0.095	0.421	0.42	0.113	-0.012	0.048	0.93	1	0.405	0.095
D TTOT OPTIMAL_outbound	-0.021	0.149	0.038	0.046	0.031	0.074	0.055	0.358	0.405	1	0.028
% REG FLIGHTS AROUND_outbound	0.041	0.016	0.049	0.032	0.333	-0.004	0.009	0.092	0.095	0.028	1
AVG STARTUP DLA AROUND_outbound	0.017	0.086	0.48	0.473	0.11	-0.044	0.028	0.56	0.556	0.11	0.255
DD_outbound	-0.018	0.032	0.06	0.057	0.02	0.026	0.004	0.039	0.04	0.029	0.031
VV_outbound	0.002	0.036	0.064	0.06	0.072	0.034	0.005	0.036	0.033	0.035	0.19
P_sector_outbound	-0.008	-0.029	-0.033	-0.032	-0.011	0.014	0.005	-0.03	-0.041	-0.032	-0.006
EHAMDEP_outbound	0.042	-0.069	0.046	0.038	0.014	-0.072	-0.023	0.071	0.067	-0.077	0.078
c_EHDEP_outbound	0.027	-0.094	-0.025	-0.03	0.003	-0.024	-0.009	0.014	0.006	-0.044	0.055
nr_fl_outbound_outbound	0.033	-0.048	0.208	0.202	-0.002	-0.09	-0.016	0.198	0.208	-0.05	0.041
Average Rate Dep_outbound	0.025	-0.013	0.277	0.27	-0.003	-0.099	-0.018	0.229	0.238	-0.028	0.023
mdg_dep_outbound	-0.002	-0.005	-0.004	-0.007	0.045	-0.024	0	0.018	0.013	0.002	0.117
c_d_EHAMDEP_outbound	-0.015	-0.01	-0.07	-0.066	-0.019	0.044	0.014	-0.07	-0.073	0.037	-0.045
c_d_EHAMARR_inbound	-0.001	0.019	-0.001	0.001	-0.028	-0.006	-0.002	-0.002	0	0.008	-0.054
atcdla_outbound	-0.097	0.176	0.376	0.365	0.168	-0.004	0.045	0.841	0.787	0.356	0.083
d_depdla_atcdla	0.309	0.311	0.142	0.14	0.03	0.077	0.031	-0.1	-0.055	0.291	0.077
pln_tat	-0.015	-0.179	-0.001	0.004	-0.014	-0.131	0.012	0.036	0.059	-0.022	0.046
act_tat	-0.016	-0.222	0.017	0.022	0.005	-0.118	0.027	0.128	0.15	0.068	0.039
d_tat	0.006	-0.052	0.101	0.101	0.05	0.028	0.04	0.245	0.243	0.243	-0.023
tat_delay	0.042	0.114	0.248	0.242	0.115	0.04	0.056	0.429	0.428	0.421	0.054
boarding_duration	0.007	0.006	0.035	0.039	-0.001	-0.026	0.125	0.041	0.056	0.095	0.015
waiting_ready	0.033	0.003	0.134	0.13	0.096	-0.028	0.013	0.274	0.31	0.119	0.039
time_in_fir	-0.005	0.028	0.028	0.03	-0.026	0.02	0.001	0.019	0.023	0.022	-0.108
time_in_tma	0.021	0.075	0.053	0.049	0.046	0.008	0.001	0.011	0.007	0.004	0.122
D_T_TD_inbound	0.01	0.036	0.039	0.034	0.076	0.025	0	0.031	0.022	0.002	0.222
D_T_TD_outbound	0.011	0.037	0.059	0.055	0.08	0.057	0.008	0.027	0.023	0.043	0.215
diff_capacity_actual_outbound	0.016	-0.08	-0.295	-0.295	0.003	0.057	0.004	-0.212	-0.227	-0.026	0.043
diff_capacity_actual_inbound	-0.016	-0.063	-0.133	-0.127	-0.045	0.003	0.002	-0.107	-0.106	-0.001	-0.097

	AVG STARTUP DLA AROUND_outbound	DD_outbound	VV_outbound	P_sector_outbound	EHAMDEP_outbound	c_EHDEP_outbound	nr_fl_outbound_outbound	Average Rate Dep_outbound	mclg_dep_outbound	c_d_EHAMDEP_outbound	c_d_EHAMARR_inbound
atxi_inbound	0.035	0.042	-0.111	-0.155	-0.035	-0.068	-0.026	-0.004	-0.065	-0.019	0.007
wags_inbound	0.054	-0.001	0.008	-0.013	-0.052	-0.034	-0.023	0.017	0.002	0.014	0.039
adcdla_inbound	0.116	0.041	0.077	-0.02	0.012	0.01	0.04	0.059	0.028	0.01	0.017
arrdla_inbound	0.148	0.06	0.046	-0.101	-0.014	-0.08	0.006	0.026	0.003	-0.031	0.009
cindla_inbound	0.01	0.005	-0.003	0.006	-0.009	0	0.002	0.007	0	0.005	0.015
dly1_dur_inbound	0.149	0.039	0.074	-0.042	0.049	-0.003	0.063	0.076	0.032	-0.029	0.03
dly2_dur_inbound	0.092	0.031	0.052	-0.028	0.013	0.008	0.035	0.048	0.019	0.006	0.022
nrctot_inbound	0.154	0.039	0.062	-0.064	0.139	0.042	0.134	0.11	0.049	-0.074	-0.098
purdla_inbound	0.173	0.046	0.077	-0.062	0.116	0.016	0.116	0.108	0.05	-0.071	-0.022
reg_induced_dla_inbound	-0.023	-0.001	-0.004	0.013	-0.044	-0.001	-0.03	-0.017	-0.011	0.036	0.004
RATE_inbound	0.113	0.014	0.021	-0.046	0.198	0.079	0.182	0.141	0.049	-0.114	-0.084
DD_inbound	0.065	0.786	0.046	0.019	-0.008	0.032	0.01	0.015	-0.066	0.034	-0.011
VV_inbound	0.081	0.022	0.773	0.183	0.038	0.15	0.06	0.07	0.454	0.073	-0.034
P_sector_inbound	-0.067	-0.052	0.194	0.97	0.022	0.089	0.031	0.031	0.137	0.011	0.023
EHAMARR_inbound	-0.063	-0.019	0.045	0.028	0.223	0.179	0.169	0.099	0.039	-0.063	-0.62
c_EHARR_inbound	-0.074	-0.054	0.006	0.094	0.254	0.291	0.237	0.186	0.039	-0.02	0.244
nr_fl_inbound_inbound	0.032	-0.024	0.059	0.038	0.358	0.282	0.34	0.283	0.063	-0.104	-0.204
Average Rate Arr_inbound	0.089	-0.022	0.069	0.037	0.381	0.296	0.387	0.354	0.066	-0.113	-0.088
mclg_arr_inbound	0.035	-0.094	0.449	0.148	0.027	0.139	0.038	0.041	0.757	0.053	-0.015
wfdep_outbound	0.357	-0.001	-0.013	-0.035	0.042	-0.004	0.141	0.156	0.002	-0.049	-0.006
adcdla_outbound	0.198	0.059	0.085	-0.073	0.02	-0.043	0.041	0.065	0.009	-0.053	-0.016
depdla_outbound	0.413	0.053	0.071	-0.087	0.04	-0.042	0.127	0.168	0.011	-0.071	-0.019
cindla_outbound	0.032	0.008	0.011	-0.002	0	0	0.001	0.001	0.007	-0.005	-0.007
dly1_dur_outbound	0.373	0.051	0.063	-0.078	0.035	-0.041	0.113	0.155	0.009	-0.066	-0.019
dly2_dur_outbound	0.234	0.031	0.051	-0.053	0.019	-0.025	0.066	0.084	0.004	-0.039	-0.002
lstdla_outbound	0.152	0.028	0.068	-0.023	0.017	-0.001	0.016	0.023	0.043	-0.023	-0.024
nrctot_outbound	0.104	0.011	0.072	-0.006	0.029	0.016	0.004	-0.003	0.049	-0.023	-0.037
purdla_outbound	0.149	0.028	0.067	-0.022	0.018	-0.001	0.015	0.023	0.042	-0.023	-0.023
locatc_outbound	0.271	0.001	-0.009	-0.024	0.085	0.029	0.16	0.166	-0.006	-0.061	0.001
reg_induced_dla_outbound	0.017	-0.018	0.002	-0.008	0.042	0.027	0.033	0.025	-0.002	-0.015	-0.001
#TOBT_outbound	0.086	0.032	0.036	-0.029	-0.069	-0.094	-0.048	-0.013	-0.005	-0.01	0.019
#TSAT_outbound	0.48	0.06	0.064	-0.033	0.046	-0.025	0.208	0.277	-0.004	-0.07	-0.001
#TTOT_outbound	0.473	0.057	0.06	-0.032	0.038	-0.03	0.202	0.27	-0.007	-0.066	0.001
#CTOT_outbound	0.11	0.02	0.072	-0.011	0.014	0.003	-0.002	-0.003	0.045	-0.019	-0.028
#EXOT_outbound	-0.044	0.026	0.034	0.014	-0.072	-0.024	-0.09	-0.099	-0.024	0.044	-0.006
#ASRT_outbound	0.028	0.004	0.005	0.005	-0.023	-0.009	-0.016	-0.018	0	0.014	-0.002
STARTUP DLA_outbound	0.56	0.039	0.036	-0.03	0.071	0.014	0.198	0.229	0.018	-0.07	-0.002
D TTOT TARGET_outbound	0.556	0.04	0.033	-0.041	0.067	0.006	0.208	0.238	0.013	-0.073	0
D TTOT OPTIMAL_outbound	0.11	0.029	0.035	-0.032	-0.077	-0.044	-0.05	-0.028	0.002	0.037	0.008
% REG FLIGHTS AROUND_outbound	0.255	0.031	0.19	-0.006	0.078	0.055	0.041	0.023	0.117	-0.045	-0.054
AVG STARTUP DLA AROUND_outbound	1	0.074	0.076	-0.063	0.095	0.018	0.303	0.372	0.039	-0.091	-0.001
DD_outbound	0.074	1	0.048	-0.008	-0.02	0.025	-0.005	0.002	-0.08	0.038	-0.017
VV_outbound	0.076	0.048	1	0.195	0.014	0.142	0.049	0.069	0.493	0.089	-0.04
P_sector_outbound	-0.063	-0.008	0.195	1	0.027	0.097	0.038	0.037	0.147	0.009	0.02
EHAMDEP_outbound	0.095	-0.02	0.014	0.027	1	0.531	0.727	0.565	0.028	-0.557	-0.014
c_EHDEP_outbound	0.018	0.025	0.142	0.097	0.531	1	0.536	0.465	0.14	0.304	0.045
nr_fl_outbound_outbound	0.303	-0.005	0.049	0.038	0.727	0.536	1	0.86	0.042	-0.221	0.019
Average Rate Dep_outbound	0.372	0.002	0.069	0.037	0.565	0.465	0.86	1	0.047	-0.113	0.049
mclg_dep_outbound	0.039	-0.08	0.493	0.147	0.028	0.14	0.042	0.047	1	0.05	-0.016
c_d_EHAMDEP_outbound	-0.091	0.038	0.089	0.009	-0.557	0.304	-0.221	-0.113	0.05	1	0.05
c_d_EHAMARR_inbound	-0.001	-0.017	-0.04	0.02	-0.014	0.045	0.019	0.049	-0.016	0.05	1
atcdla_outbound	0.503	0.045	0.041	-0.039	0.063	0.001	0.172	0.206	0.017	-0.072	0.001
d_depdla_atcdla	0.031	0.018	0.041	-0.073	0.004	-0.046	0	0.011	-0.009	-0.03	-0.02
pln_tat	0.115	-0.006	-0.099	-0.005	0.015	-0.002	0.037	0.077	-0.039	-0.017	0.032
act_tat	0.139	-0.019	-0.094	0.014	0.02	0.017	0.05	0.085	-0.033	-0.014	0.016
d_tat	0.116	-0.023	0.003	0.044	0.024	0.043	0.062	0.069	0.005	-0.007	-0.024
tat_delay	0.261	0.009	0.051	-0.023	0.05	0.02	0.11	0.127	0.019	-0.044	-0.029
boarding_duration	0.061	0.012	0.02	0.01	-0.01	0	0.013	0.043	-0.004	0.006	0.032
waiting_ready	0.212	-0.014	-0.012	-0.025	0.025	-0.004	0.097	0.116	-0.009	-0.032	0.004
time_in_fir	0.011	-0.037	-0.118	-0.197	0.012	-0.024	0.017	0.011	-0.03	-0.023	-0.034
time_in_tma	0.032	0.032	0.111	0.167	-0.007	-0.019	-0.003	0.004	-0.019	0	-0.002
D_T_TD_inbound	0.067	-0.009	0.692	0.198	0.013	0.149	0.037	0.05	0.493	0.104	-0.04
D_T_TD_outbound	0.061	0.003	0.822	0.216	-0.006	0.142	0.02	0.044	0.545	0.113	-0.045
diff_capacity_actual_outbound	-0.34	0.012	0.04	0.012	0.145	0.553	-0.137	-0.359	0.042	0.392	-0.014
diff_capacity_actual_inbound	-0.179	-0.024	-0.06	0.025	-0.15	-0.049	-0.18	-0.198	-0.042	0.101	0.334

	atcdla_outbound	d_depdla_atcdla	pln_tat	act_tat	d_tat	tat_delay	boarding_duration	waiting_ready	time_in_fir	time_in_tma	D_T_TD_inbound	D_T_TD_outbound	diff_capacity_actual_outbound	diff_capacity_actual_inbound
atxi_inbound	0.015	0.08	0.101	0.054	-0.097	0.008	0.067	0.039	0.01	0.012	-0.16	-0.118	-0.061	0.018
wags_inbound	0.015	0.004	0.265	0.291	0.125	0.068	0.215	0.062	0.012	-0.052	-0.062	0.012	-0.061	0.082
adcdla_inbound	0.102	0.162	0.029	-0.118	-0.428	-0.207	-0.051	0.002	-0.029	0.028	0.136	0.059	-0.027	-0.134
arrdla_inbound	0.134	0.274	0.008	-0.245	-0.746	-0.305	-0.073	0.021	0.121	0.144	0.069	0.023	-0.072	-0.229
cindla_inbound	0.011	0.017	0.018	-0.016	-0.073	-0.058	-0.011	-0.004	-0.011	0.003	0.003	-0.006	-0.012	-0.015
dly1_dur_inbound	0.126	0.216	0.009	-0.177	-0.517	-0.243	-0.068	0.012	-0.009	0.053	0.107	0.058	-0.036	-0.219
dly2_dur_inbound	0.081	0.129	-0.002	-0.112	-0.287	-0.154	-0.06	0.009	-0.008	0.016	0.095	0.033	-0.021	-0.128
nrctot_inbound	0.111	0.132	-0.16	-0.252	-0.252	-0.082	-0.18	-0.004	0.014	0.051	0.114	0.05	-0.005	-0.286
purdla_inbound	0.133	0.192	-0.108	-0.252	-0.363	-0.15	-0.139	0.004	0.022	0.066	0.109	0.062	-0.027	-0.309
reg_induced_dla_inbound	-0.011	-0.017	0.07	0.079	0.022	0.004	0.049	0.004	-0.023	-0.026	-0.001	-0.012	-0.003	0.073
RATE_inbound	0.081	0.07	-0.171	-0.211	-0.12	0.018	-0.148	0.011	0.088	0.008	0.021	-0.008	-0.006	-0.214
DD_inbound	0.043	0.007	-0.005	-0.019	-0.037	-0.008	-0.007	-0.014	-0.042	0.032	0.021	-0.025	0.011	-0.032
VV_inbound	0.044	0.017	-0.091	-0.092	-0.032	0.017	-0.034	-0.018	-0.116	0.093	0.835	0.66	0.057	-0.074
P_sector_inbound	-0.043	-0.075	-0.011	0.008	0.043	-0.024	0.008	-0.025	-0.213	0.19	0.213	0.224	0.012	0.027
EHAMARR_inbound	-0.02	-0.038	-0.165	-0.116	0.074	0.081	-0.096	-0.02	0.026	-0.06	0.071	0.031	0.122	-0.019
c_EHARR_inbound	-0.028	-0.109	-0.145	-0.095	0.073	0.043	-0.089	-0.016	-0.011	-0.094	0.076	-0.03	0.119	0.307
nr_fl_inbound_inbound	0.042	-0.017	-0.204	-0.185	-0.013	0.059	-0.124	-0.003	0.093	-0.033	0.102	0.035	0.071	-0.352
Average Rate Arr_inbound	0.073	-0.001	-0.21	-0.206	-0.047	0.05	-0.13	0.011	0.079	-0.015	0.117	0.044	0.023	-0.525
mdlg_arr_inbound	0.012	-0.019	-0.037	-0.027	0.014	0.021	-0.01	-0.01	-0.035	-0.029	0.527	0.495	0.049	-0.032
wfdep_outbound	0.432	-0.049	0.05	0.123	0.18	0.266	-0.028	0.339	0.015	-0.009	-0.003	-0.021	-0.157	-0.076
adcdla_outbound	0.302	0.674	-0.139	-0.113	0.156	0.453	0.082	0.039	0.044	0.067	0.055	0.089	-0.084	-0.105
depdla_outbound	0.571	0.621	-0.088	-0.017	0.275	0.627	0.065	0.22	0.048	0.06	0.052	0.07	-0.183	-0.145
cindla_outbound	0.089	0.001	-0.006	0.009	0.034	0.051	0.007	0.03	-0.005	0.012	0.005	0.008	-0.004	-0.007
dly1_dur_outbound	0.521	0.578	-0.073	-0.016	0.242	0.569	0.054	0.19	0.044	0.054	0.048	0.06	-0.17	-0.131
dly2_dur_outbound	0.341	0.362	-0.083	-0.037	0.151	0.358	0.059	0.137	0.031	0.047	0.033	0.054	-0.095	-0.089
lstdla_outbound	0.433	-0.041	-0.026	0.008	0.096	0.205	0.004	0.127	-0.013	0.043	0.07	0.069	-0.023	-0.054
nrctot_outbound	0.102	0.059	0.002	0.024	0.051	0.112	-0.005	0.111	-0.036	0.049	0.083	0.082	0.02	-0.049
purdla_outbound	0.422	-0.045	-0.024	0.008	0.089	0.196	0.002	0.125	-0.014	0.042	0.07	0.068	-0.022	-0.054
locatc_outbound	0.297	0.023	0.049	0.09	0.127	0.235	0.02	0.207	0.018	-0.013	-0.016	-0.011	-0.125	-0.054
reg_induced_dla_outbound	-0.097	0.309	-0.015	-0.016	0.006	0.042	0.007	0.033	-0.005	0.021	0.01	0.011	0.016	-0.016
#TOBT_outbound	0.176	0.311	-0.179	-0.222	-0.052	0.114	0.006	0.003	0.028	0.075	0.036	0.037	-0.08	-0.063
#TSAT_outbound	0.376	0.142	-0.001	0.017	0.101	0.248	0.035	0.134	0.028	0.053	0.039	0.059	-0.295	-0.133
#TTOT_outbound	0.365	0.14	0.004	0.022	0.101	0.242	0.039	0.13	0.03	0.049	0.034	0.055	-0.295	-0.127
#CTOT_outbound	0.168	0.03	-0.014	0.005	0.05	0.115	-0.001	0.096	-0.026	0.046	0.076	0.08	0.003	-0.045
#EXOT_outbound	-0.004	0.077	-0.131	-0.118	0.028	0.04	-0.026	-0.028	0.02	0.008	0.025	0.057	0.057	0.003
#ASRT_outbound	0.045	0.031	0.012	0.027	0.04	0.056	0.125	0.013	0.001	0.001	0	0.008	0.004	0.002
STARTUP DLA_outbound	0.841	-0.1	0.036	0.128	0.245	0.429	0.041	0.274	0.019	0.011	0.031	0.027	-0.212	-0.107
D TTOT TARGET_outbound	0.787	-0.055	0.059	0.15	0.243	0.428	0.056	0.31	0.023	0.007	0.022	0.023	-0.227	-0.106
D TTOT OPTIMAL_outbound	0.356	0.291	-0.022	0.068	0.243	0.421	0.095	0.119	0.022	0.004	0.002	0.043	-0.026	-0.001
% REG FLIGHTS AROUND_outbound	0.083	0.077	0.046	0.039	-0.023	0.054	0.015	0.039	-0.108	0.122	0.222	0.215	0.043	-0.097
AVG STARTUP DLA AROUND_outbound	0.503	0.031	0.115	0.139	0.116	0.261	0.061	0.212	0.011	0.032	0.067	0.061	-0.34	-0.179
DD_outbound	0.045	0.018	-0.006	-0.019	-0.023	0.009	0.012	-0.014	-0.037	0.032	-0.009	0.003	0.012	-0.024
VV_outbound	0.041	0.041	-0.099	-0.094	0.003	0.051	0.02	-0.012	-0.118	0.111	0.692	0.822	0.04	-0.06
P_sector_outbound	-0.039	-0.073	-0.005	0.014	0.044	-0.023	0.01	-0.025	-0.197	0.167	0.198	0.216	0.012	0.025
EHAMDEP_outbound	0.063	0.004	0.015	0.02	0.024	0.05	-0.01	0.025	0.012	-0.007	0.013	-0.006	0.145	-0.15
c_EHDEP_outbound	0.001	-0.046	-0.002	0.017	0.043	0.02	0	-0.004	-0.024	-0.019	0.149	0.142	0.553	-0.049
nr_fl_outbound_outbound	0.172	0	0.037	0.05	0.062	0.11	0.013	0.097	0.017	-0.003	0.037	0.02	-0.137	-0.18
Average Rate Dep_outbound	0.206	0.011	0.077	0.085	0.069	0.127	0.043	0.116	0.011	0.004	0.05	0.044	-0.359	-0.198
mdlg_dep_outbound	0.017	-0.009	-0.039	-0.033	0.005	0.019	-0.004	-0.009	-0.03	-0.019	0.493	0.545	0.042	-0.042
c_d_EHAMDEP_outbound	-0.072	-0.03	-0.017	-0.014	-0.007	-0.044	0.006	-0.032	-0.023	0	0.104	0.113	0.392	0.101
c_d_EHAMARR_inbound	0.001	-0.02	0.032	0.016	-0.024	-0.029	0.032	0.004	-0.034	-0.002	-0.04	-0.045	-0.014	0.334
atcdla_outbound	1	-0.18	0.003	0.081	0.223	0.42	0.036	0.241	0.027	0.016	0.033	0.03	-0.202	-0.105
d_depdla_atcdla	-0.18	1	-0.114	-0.1	0.111	0.341	0.047	0.059	0.033	0.058	0.024	0.05	-0.026	-0.075
pln_tat	0.003	-0.114	1	0.893	-0.084	-0.174	0.236	0.048	-0.024	-0.004	-0.125	-0.115	-0.066	0.069
act_tat	0.081	-0.1	0.893	1	0.259	0.109	0.262	0.088	-0.059	-0.041	-0.121	-0.102	-0.066	0.112
d_tat	0.223	0.111	-0.084	0.259	1	0.784	0.114	0.113	-0.09	-0.1	-0.029	0.023	-0.044	0.129
tat_delay	0.42	0.341	-0.174	0.109	0.784	1	0.096	0.184	-0.022	-0.029	0.014	0.059	-0.105	0.018
boarding_duration	0.036	0.047	0.236	0.262	0.114	0.096	1	0.033	-0.03	0.011	-0.033	0.036	-0.044	0.066
waiting_ready	0.241	0.059	0.048	0.088	0.113	0.184	0.033	1	-0.003	0.017	-0.021	-0.01	-0.117	-0.032
time_in_fir	0.027	0.033	-0.024	-0.059	-0.09	-0.022	-0.03	-0.003	1	-0.443	-0.173	-0.182	-0.021	-0.111
time_in_tma	0.016	0.058	-0.004	-0.041	-0.1	-0.029	0.011	0.017	-0.443	1	0.158	0.183	-0.013	-0.066
D_T_TD_inbound	0.033	0.024	-0.125	-0.121	-0.029	0.014	-0.033	-0.021	-0.173	0.158	1	0.809	0.078	-0.086
D_T_TD_outbound	0.03	0.05	-0.115	-0.102	0.023	0.059	0.036	-0.01	-0.182	0.183	0.809	1	0.067	-0.065
diff_capacity_actual_outbound	-0.202	-0.026	-0.066	-0.066	-0.044	-0.105	-0.044	-0.117	-0.021	-0.013	0.078	0.067	1	0.105
diff_capacity_actual_inbound	-0.105	-0.075	0.069	0.112	0.129	0.018	0.066	-0.032	-0.111	-0.066	-0.086	-0.065	0.105	1

B.2. MUTUAL INFORMATION MATRIX

Variable	ac_type	arr_rwy	dep_rwy	fenpnt_inbound	fexpnt_outbound	dominant sector_inbound	dominant sector_outbound	aircraft_swap_inbound	aircraft_swap_outbound	regcse1_inbound	regcse2_inbound	regcse1_outbound	regcse2_outbound	GOA_inbound	STACK_inbound	TV_inbound
ac_type	1.004	0.013	0.008	0.068	0.038	0.022	0.024	0.017	0.018	0.121	0.121	0.01	0.01	0	0.022	0.08
arr_rwy	0.013	1.735	0.386	0.159	0.012	0.157	0.003	0.001	0.001	0.052	0.024	0.003	0.001	0	0.158	0.028
dep_rwy	0.008	0.386	1.629	0.003	0.249	0.002	0.252	0	0	0.032	0.008	0.033	0.03	0	0.002	0.007
fenpnt_inbound	0.068	0.159	0.003	1.951	0.009	1.068	0.006	0.001	0.001	0.068	0.059	0.002	0.001	0	1.056	0.02
fexpnt_outbound	0.038	0.012	0.249	0.009	2.068	0.005	1.485	0.001	0.001	0.018	0.01	0.096	0.087	0	0.005	0.006
dominant sector_inbound	0.022	0.157	0.002	1.068	0.005	1.068	0.004	0	0	0.041	0.035	0.001	0	0	1.055	0.017
dominant sector_outbound	0.024	0.003	0.252	0.006	1.485	0.004	1.569	0	0	0.007	0.007	0.092	0.088	0	0.004	0.004
aircraft_swap_inbound	0.017	0.001	0	0.001	0.001	0	0	0.626	0.096	0.005	0.004	0	0	0	0	0.003
aircraft_swap_outbound	0.018	0.001	0	0.001	0.001	0	0	0.096	0.62	0.005	0.004	0	0	0	0	0.002
regcse1_inbound	0.121	0.052	0.032	0.068	0.018	0.041	0.007	0.005	0.005	1.457	0.916	0.007	0.004	0	0.041	0.575
regcse2_inbound	0.121	0.024	0.008	0.059	0.01	0.035	0.007	0.004	0.004	0.916	0.994	0.004	0.003	0	0.035	0.63
regcse1_outbound	0.01	0.003	0.033	0.002	0.096	0.001	0.092	0	0	0.007	0.004	0.807	0.54	0	0.001	0.002
regcse2_outbound	0.01	0.001	0.03	0.001	0.087	0	0.088	0	0	0.004	0.003	0.54	0.594	0	0	0.001
GOA_inbound	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0
STACK_inbound	0.022	0.158	0.002	1.056	0.005	1.055	0.004	0	0	0.041	0.035	0.001	0	0	1.068	0.018
TV_inbound	0.08	0.028	0.007	0.02	0.006	0.017	0.004	0.003	0.002	0.575	0.63	0.002	0.001	0	0.018	0.761
LOC_ID_inbound	0.08	0.021	0.007	0.006	0.006	0.003	0.004	0.003	0.002	0.573	0.629	0.002	0.001	0	0.003	0.745
REG_REASON_x_inbound	0.08	0.05	0.033	0.006	0.012	0.003	0.004	0.004	0.003	0.899	0.605	0.003	0.002	0	0.003	0.722
REG_REASON_y_inbound	0.004	0.067	0.048	0.004	0.014	0.001	0.001	0.002	0.002	0.518	0.244	0.008	0.003	0.001	0.001	0.255
REG_REASON_DETAILS_inbound	0.08	0.064	0.045	0.004	0.015	0.001	0.004	0.004	0.004	0.856	0.569	0.005	0.002	0	0.001	0.684
TOBT STABLE_outbound	0.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOBT ADH_outbound	0.006	0	0	0	0	0	0	0	0	0.002	0.002	0	0	0	0	0.002
TSAT ADH_outbound	0.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
regulated_outbound	0.005	0	0.034	0.001	0.083	0	0.083	0	0	0.003	0.003	0.396	0.396	0	0	0.001
season	0	0.016	0.046	0.001	0.064	0	0.001	0	0	0.087	0.044	0.013	0.009	0	0	0.032
time_of_day_inbound	0.048	0.004	0.003	0.012	0.012	0.007	0.008	0.005	0.009	0.029	0.027	0.005	0.001	0	0.007	0.017
time_of_day_outbound	0.009	0.004	0.005	0.003	0.009	0	0.005	0.004	0.008	0.024	0.021	0.004	0	0	0	0.02
Visibility_outbound	0	0.008	0.016	0	0.002	0	0	0	0	0.011	0.001	0.001	0.001	0	0	0.001
Visibility_inbound	0.001	0.012	0.013	0	0.002	0	0	0	0	0.017	0.002	0.001	0.001	0	0	0.001
regyn_inbound	0.119	0.021	0.007	0.02	0.009	0.012	0.006	0.004	0.004	0.693	0.693	0.003	0.003	0	0.012	0.363
mregyn_inbound	0.024	0.007	0.001	0.066	0.004	0.038	0.002	0	0.001	0.121	0.128	0.002	0.002	0	0.038	0.006
mregyn_outbound	0.002	0	0.01	0	0.031	0	0.028	0	0	0.001	0.001	0.092	0.092	0	0	0
regcse3_outbound	0.01	0.001	0.031	0.001	0.091	0	0.09	0	0	0.004	0.003	0.665	0.593	0	0	0.002
regcse3_inbound	0.12	0.041	0.022	0.069	0.014	0.043	0.007	0.004	0.005	1.204	0.984	0.005	0.003	0	0.043	0.632
dly1_key_inbound	0.028	0.014	0.008	0.022	0.006	0.009	0.003	0.003	0.008	0.148	0.142	0.003	0.002	0	0.009	0.086
dly2_key_inbound	0.018	0.005	0.002	0.011	0.003	0.003	0.001	0.002	0.006	0.051	0.048	0.002	0.001	0	0.003	0.018
dly1_key_outbound	0.015	0.008	0.01	0.003	0.017	0.001	0.014	0.001	0.005	0.03	0.025	0.061	0.059	0	0.001	0.019
dly2_key_outbound	0.007	0.003	0.005	0.002	0.01	0.001	0.009	0	0.001	0.006	0.005	0.045	0.044	0	0.001	0.004
month_inbound	0	0.124	0.158	0.002	0.137	0.001	0.003	0.002	0.001	0.169	0.055	0.024	0.013	0	0.001	0.043
month_outbound	0	0.125	0.16	0.002	0.137	0.001	0.003	0.002	0.001	0.169	0.055	0.024	0.013	0	0.001	0.043
hour_inbound	0.136	0.066	0.019	0.086	0.055	0.039	0.044	0.013	0.016	0.113	0.098	0.014	0.007	0	0.039	0.078
hour_outbound	0.08	0.024	0.037	0.054	0.089	0.03	0.068	0.011	0.012	0.053	0.045	0.02	0.012	0	0.03	0.042
weekday_inbound	0	0.001	0.002	0.001	0.012	0	0.002	0	0	0.013	0.006	0.005	0	0	0	0.007
weekday_outbound	0	0.001	0.002	0.001	0.013	0	0.002	0	0	0.013	0.005	0.005	0.001	0	0	0.006
peaks_rwy_inbound	0.047	0.047	0.008	0.012	0.009	0.001	0.007	0.001	0.001	0.05	0.039	0.001	0	0	0.001	0.033
peaks_rwy_outbound	0.003	0.007	0.034	0.006	0.014	0.002	0.01	0	0.001	0.015	0.009	0.003	0.003	0	0.002	0.008
eur	0.413	0.012	0.007	0.03	0.022	0.016	0.019	0.011	0.012	0.118	0.118	0.004	0.007	0	0.016	0.08
short_tat	0.049	0.006	0	0.005	0.002	0.003	0.001	0	0	0.02	0.018	0	0	0	0.003	0.015
O_sector_inbound	0	0.002	0.001	0.001	0.001	0.001	0	0	0	0.005	0.003	0.002	0.001	0	0.001	0.002
O_sector_outbound	0	0.002	0.001	0	0.002	0	0.001	0	0	0.002	0.001	0.003	0.001	0	0	0
O_outbound	0	0.002	0.002	0	0	0	0	0	0	0.002	0.001	0.001	0	0	0	0
O_inbound	0	0.002	0.001	0	0	0	0	0	0	0.003	0.001	0.001	0	0	0	0.001
M_inbound	0	0.001	0.004	0	0.001	0	0	0	0	0.008	0.001	0	0	0	0	0.001
M_outbound	0	0.001	0.004	0	0	0	0	0	0	0.003	0	0	0	0	0	0
R_inbound	0	0.025	0.01	0	0.002	0	0	0	0	0.008	0.001	0.001	0	0	0	0.001
R_outbound	0	0.023	0.011	0	0.002	0	0	0	0	0.006	0.001	0.001	0	0	0	0.001
Y_inbound	0	0.001	0.001	0	0.001	0	0	0	0	0.002	0	0	0	0	0	0
Y_outbound	0	0.001	0.001	0	0	0	0	0	0	0.001	0	0	0	0	0	0
S_inbound	0	0	0.001	0	0	0	0	0	0	0.002	0	0	0	0	0	0
S_outbound	0	0	0.001	0	0	0	0	0	0	0.001	0	0	0	0	0	0
wtc_inbound	0.114	0.017	0.004	0.008	0.016	0.002	0.016	0.003	0.008	0.041	0.041	0.002	0.001	0	0.002	0.023
wtc_outbound	0.06	0.005	0.001	0.009	0.003	0.005	0.002	0.005	0.001	0.014	0.014	0.001	0.001	0	0.005	0.009
c_d_IAFS	0.001	0.005	0	0.004	0	0.001	0	0	0.001	0.003	0.002	0	0	0	0.001	0.003
c_IAFS	0	0.004	0.002	0.001	0.001	0	0	0	0	0.006	0.003	0.002	0.001	0	0	0.006
d_IAFS	0.001	0.005	0	0.005	0	0.001	0	0	0.001	0.003	0.001	0	0	0	0.001	0.002
wind_speed_inbound	0.001	0.034	0.022	0	0.002	0	0	0	0.001	0.015	0.001	0.001	0	0	0	0.002
wind_speed_sector_inbound	0.002	0.032	0.02	0.08	0.004	0.08	0.001	0	0	0.016	0.004	0.001	0	0	0.08	0.003
wind_speed_outbound	0.001	0.032	0.029	0	0.002	0	0	0	0	0.011	0	0.001	0	0	0	0.001
wind_speed_sector_outbound	0.004	0.021	0.039	0.001	0.071	0.001	0.075	0	0	0.011	0.001	0.004	0.004	0	0.001	0.001
CDM_stability	0.005	0.001	0.001	0.001	0.001	0	0.001	0	0	0.004	0.002	0	0	0	0	0.002
avg_cdm_updates	0.004	0.013	0.009	0.002	0.006	0.001	0.003	0	0.002	0.024	0.018	0.001	0.001	0	0.001	0.017
slotdla_info	0.001	0	0.006	0	0.018	0	0.018	0	0	0.002	0.002	0.057	0.056	0	0	0.001

Variable	LOC_ID_inbound	REG_REASON_x_inbound	REG_REASON_y_inbound	REG_REASON_DETAILS_inbound	TOBT STABLE_outbound	TOBT ADH_outbound	TSAT ADH_outbound	regulated_outbound	season	time_of_day_inbound	time_of_day_outbound	Visibility_outbound	Visibility_inbound
ac_type	0.08	0.08	0.004	0.08	0.001	0.006	0.001	0.005	0	0.048	0.009	0	0.001
arr_rwy	0.021	0.05	0.067	0.064	0	0	0	0	0.016	0.004	0.004	0.008	0.012
dep_rwy	0.007	0.033	0.048	0.045	0	0	0	0.034	0.046	0.003	0.005	0.016	0.013
fenpnt_inbound	0.006	0.006	0.004	0.004	0	0	0	0.001	0.001	0.012	0.003	0	0
fexpnt_outbound	0.006	0.012	0.014	0.015	0	0	0	0.083	0.064	0.012	0.009	0.002	0.002
dominant sector_inbound	0.003	0.003	0.001	0.001	0	0	0	0	0	0.007	0	0	0
dominant sector_outbound	0.004	0.004	0.001	0.004	0	0	0	0.083	0.001	0.008	0.005	0	0
aircraft_swap_inbound	0.003	0.004	0.002	0.004	0	0	0	0	0	0.005	0.004	0	0
aircraft_swap_outbound	0.002	0.003	0.002	0.004	0	0	0	0	0	0.009	0.008	0	0
regcse1_inbound	0.573	0.899	0.518	0.856	0	0.002	0	0.003	0.087	0.029	0.024	0.011	0.017
regcse2_inbound	0.629	0.605	0.244	0.569	0	0.002	0	0.003	0.044	0.027	0.021	0.001	0.002
regcse1_outbound	0.002	0.003	0.008	0.005	0	0	0	0.396	0.013	0.005	0.004	0.001	0.001
regcse2_outbound	0.001	0.002	0.003	0.002	0	0	0	0.396	0.009	0.001	0	0.001	0.001
GOA_inbound	0	0	0.001	0	0	0	0	0	0	0	0	0	0
STACK_inbound	0.003	0.003	0.001	0.001	0	0	0	0	0	0.007	0	0	0
TV_inbound	0.745	0.722	0.255	0.684	0	0.002	0	0.001	0.032	0.017	0.02	0.001	0.001
LOC_ID_inbound	0.745	0.721	0.252	0.684	0	0.001	0	0.001	0.031	0.009	0.014	0.001	0.001
REG_REASON_x_inbound	0.721	1.104	0.598	1.038	0	0.002	0	0.001	0.081	0.012	0.017	0.014	0.02
REG_REASON_y_inbound	0.252	0.598	1.478	0.562	0.001	0.001	0	0.003	0.131	0.029	0.029	0.02	0.028
REG_REASON_DETAILS_inbound	0.684	1.038	0.562	1.279	0.001	0.002	0	0.002	0.116	0.028	0.035	0.028	0.037
TOBT STABLE_outbound	0	0	0.001	0.001	0.068	0.004	0.001	0	0	0	0	0	0
TOBT ADH_outbound	0.001	0.002	0.001	0.002	0.004	0.486	0.045	0	0	0	0	0	0
TSAT ADH_outbound	0	0	0	0	0.001	0.045	0.15	0	0	0	0	0	0
regulated_outbound	0.001	0.001	0.003	0.002	0	0	0	0.539	0.009	0	0	0	0.001
season	0.031	0.081	0.131	0.116	0	0	0	0.009	1.029	0.001	0	0.021	0.025
time_of_day_inbound	0.009	0.012	0.029	0.028	0	0	0	0	0.001	1.057	0.491	0.001	0.004
time_of_day_outbound	0.014	0.017	0.029	0.035	0	0	0	0	0	0.491	1.063	0.002	0.003
Visibility_outbound	0.001	0.014	0.02	0.028	0	0	0	0	0.021	0.001	0.002	0.316	0.109
Visibility_inbound	0.001	0.02	0.028	0.037	0	0	0	0.001	0.025	0.004	0.003	0.109	0.337
regyn_inbound	0.363	0.363	0.168	0.363	0	0.002	0	0.003	0.036	0.01	0.012	0.001	0.001
mregyn_inbound	0.005	0.005	0.023	0.006	0	0	0	0.002	0.016	0.008	0.005	0	0.001
mregyn_outbound	0	0	0.001	0.001	0	0	0	0.08	0.006	0	0	0	0
regcse3_outbound	0.001	0.002	0.004	0.003	0	0	0	0.396	0.011	0.001	0.001	0.001	0.001
regcse3_inbound	0.631	0.775	0.394	0.735	0	0.002	0	0.003	0.071	0.028	0.022	0.011	0.018
dly1_key_inbound	0.084	0.091	0.067	0.092	0.001	0	0	0.002	0.028	0.043	0.032	0.001	0.001
dly2_key_inbound	0.017	0.02	0.019	0.021	0	0	0	0.001	0.01	0.025	0.016	0.001	0.001
dly1_key_outbound	0.018	0.025	0.03	0.029	0.007	0.011	0.015	0.054	0.024	0.009	0.012	0.001	0.001
dly2_key_outbound	0.004	0.005	0.009	0.007	0.002	0.01	0.01	0.039	0.008	0.003	0.004	0.001	0.001
month_inbound	0.041	0.16	0.259	0.243	0	0	0	0.011	0.985	0.001	0	0.028	0.031
month_outbound	0.041	0.16	0.26	0.243	0	0	0	0.011	0.981	0.001	0	0.028	0.031
hour_inbound	0.068	0.081	0.138	0.105	0.001	0.001	0.001	0.004	0.006	1.057	0.692	0.003	0.007
hour_outbound	0.033	0.041	0.073	0.061	0.001	0.006	0.002	0.007	0.006	0.674	1.063	0.005	0.005
weekday_inbound	0.006	0.012	0.022	0.015	0	0	0	0	0	0	0	0	0
weekday_outbound	0.006	0.012	0.02	0.014	0	0	0	0	0	0	0	0	0
peaks_rwy_inbound	0.033	0.044	0.05	0.052	0	0.001	0	0	0.009	0.016	0.04	0.001	0.002
peaks_rwy_outbound	0.008	0.013	0.013	0.016	0	0.001	0	0.002	0.014	0.027	0.044	0	0
eur	0.08	0.08	0.004	0.08	0	0.004	0.001	0.001	0	0.047	0.008	0	0.001
short_tat	0.015	0.016	0.005	0.019	0	0.003	0	0	0.005	0.018	0.01	0.001	0.001
O_sector_inbound	0.001	0.002	0.008	0.007	0	0	0	0.001	0.006	0.002	0.001	0	0
O_sector_outbound	0	0.001	0.005	0.003	0	0	0	0.001	0.005	0	0.001	0	0.001
O_outbound	0	0.001	0.005	0.004	0	0	0	0	0.002	0.001	0.001	0	0
O_inbound	0.001	0.002	0.006	0.005	0	0	0	0	0.002	0.001	0	0	0.001
M_inbound	0.001	0.008	0.012	0.013	0	0	0	0	0.006	0.001	0.001	0.018	0.036
M_outbound	0	0.003	0.005	0.006	0	0	0	0	0.004	0	0.001	0.025	0.012
R_inbound	0.001	0.012	0.018	0.017	0	0	0	0	0.005	0	0	0.008	0.014
R_outbound	0.001	0.008	0.013	0.012	0	0	0	0	0.005	0	0	0.014	0.009
Y_inbound	0	0.002	0.004	0.003	0	0	0	0	0.002	0.001	0.001	0.005	0.01
Y_outbound	0	0.001	0.001	0.001	0	0	0	0	0.002	0	0.001	0.009	0.004
S_inbound	0	0.001	0.002	0.004	0	0	0	0	0.003	0	0	0.004	0.004
S_outbound	0	0.001	0.002	0.003	0	0	0	0	0.003	0	0	0.003	0.001
wtc_inbound	0.021	0.022	0.014	0.022	0	0.001	0	0	0.001	0.164	0.066	0	0.002
wtc_outbound	0.008	0.009	0.006	0.011	0	0.002	0.001	0	0	0.057	0.042	0	0
c_d_IAFS	0.002	0.003	0.035	0.004	0	0	0	0	0.003	0.005	0.003	0	0
c_IAFS	0.005	0.005	0.043	0.013	0	0	0	0	0.007	0.002	0.002	0	0
d_IAFS	0.002	0.003	0.033	0.004	0	0	0	0	0.003	0.003	0.002	0	0
wind_speed_inbound	0.001	0.022	0.039	0.046	0	0	0	0	0.017	0.014	0.014	0.002	0.003
wind_speed_sector_inbound	0.001	0.017	0.027	0.037	0	0	0	0	0.015	0.004	0.007	0.001	0.001
wind_speed_outbound	0.001	0.018	0.031	0.039	0	0	0	0	0.018	0.01	0.015	0.002	0.002
wind_speed_sector_outbound	0.001	0.014	0.023	0.029	0	0	0	0.003	0.013	0.011	0.011	0.001	0.001
CDM_stability	0.002	0.004	0.005	0.004	0.001	0.047	0.009	0	0.001	0.001	0.003	0	0
avg_cdm_updates	0.017	0.025	0.044	0.034	0.003	0.001	0.001	0	0.036	0.006	0.017	0.003	0.003
slotdla_info	0.001	0.001	0.003	0.002	0	0.002	0.003	0.091	0.004	0	0.001	0	0

Variable	regyn_inbound	mregyn_inbound	mregyn_outbound	regcse3_outbound	regcse3_inbound	dly1_key_inbound	dly2_key_inbound	dly1_key_outbound	dly2_key_outbound	month_inbound	month_outbound	hour_inbound	hour_outbound	weekday_inbound	weekday_outbound
ac_type	0.119	0.024	0.002	0.01	0.12	0.028	0.018	0.015	0.007	0	0	0.136	0.08	0	0
arr_rwy	0.021	0.007	0	0.001	0.041	0.014	0.005	0.008	0.003	0.124	0.125	0.066	0.024	0.001	0.001
dep_rwy	0.007	0.001	0.01	0.031	0.022	0.008	0.002	0.01	0.005	0.158	0.16	0.019	0.037	0.002	0.002
fenprt_inbound	0.02	0.066	0	0.001	0.069	0.022	0.011	0.003	0.002	0.002	0.002	0.086	0.054	0.001	0.001
fexprt_outbound	0.009	0.004	0.031	0.091	0.014	0.006	0.003	0.017	0.01	0.137	0.137	0.055	0.055	0.089	0.012	0.013
dominant sector_inbound	0.012	0.038	0	0	0.043	0.009	0.003	0.001	0.001	0.001	0.001	0.039	0.03	0	0
dominant sector_outbound	0.006	0.002	0.028	0.09	0.007	0.003	0.001	0.014	0.009	0.003	0.003	0.044	0.068	0.002	0.002
aircraft_swap_inbound	0.004	0	0	0	0.004	0.003	0.002	0.001	0	0.002	0.002	0.013	0.011	0	0
aircraft_swap_outbound	0.004	0.001	0	0	0.005	0.008	0.006	0.005	0.001	0.001	0.001	0.016	0.012	0	0
regcse1_inbound	0.693	0.121	0.001	0.004	1.204	0.148	0.051	0.03	0.006	0.006	0.169	0.113	0.053	0.013	0.013
regcse2_inbound	0.693	0.128	0.001	0.003	0.984	0.142	0.048	0.025	0.005	0.055	0.055	0.098	0.045	0.006	0.005
regcse1_outbound	0.003	0.002	0.092	0.665	0.005	0.003	0.002	0.061	0.045	0.024	0.024	0.014	0.02	0.005	0.005
regcse2_outbound	0.003	0.002	0.092	0.593	0.002	0.002	0.001	0.059	0.044	0.013	0.013	0.007	0.012	0	0.001
GOA_inbound	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STACK_inbound	0.012	0.038	0	0	0.043	0.009	0.003	0.001	0.001	0.001	0.001	0.039	0.03	0	0
TV_inbound	0.363	0.006	0	0.002	0.632	0.086	0.018	0.019	0.004	0.004	0.043	0.078	0.042	0.007	0.006
LOC_ID_inbound	0.363	0.005	0	0.001	0.631	0.084	0.017	0.018	0.004	0.004	0.041	0.068	0.033	0.006	0.006
REG_REASON_x_inbound	0.363	0.005	0	0.002	0.775	0.091	0.02	0.025	0.005	0.16	0.16	0.081	0.041	0.012	0.012
REG_REASON_y_inbound	0.168	0.023	0.001	0.004	0.394	0.067	0.019	0.03	0.009	0.259	0.26	0.138	0.073	0.022	0.02
REG_REASON_DETAILS_inbound	0.363	0.006	0.001	0.003	0.735	0.092	0.021	0.029	0.007	0.243	0.243	0.105	0.061	0.015	0.014
TOBT STABLE_outbound	0	0	0	0	0	0.001	0	0.007	0.002	0	0	0.001	0.001	0	0
TOBT ADH_outbound	0.002	0	0	0	0.002	0	0	0.011	0.01	0	0	0.001	0.006	0	0
TSAT ADH_outbound	0	0	0	0	0	0	0	0.015	0.01	0	0	0.001	0.002	0	0
regulated_outbound	0.003	0.002	0.08	0.396	0.003	0.002	0.001	0.054	0.039	0.011	0.011	0.004	0.007	0	0
season	0.036	0.016	0.006	0.011	0.071	0.028	0.01	0.024	0.008	0.985	0.981	0.006	0.006	0	0
time_of_day_inbound	0.01	0.008	0	0.001	0.028	0.043	0.025	0.009	0.003	0.001	0.001	1.057	0.674	0	0
time_of_day_outbound	0.012	0.005	0	0.001	0.022	0.032	0.016	0.012	0.004	0	0	0.692	1.063	0	0
Visibility_outbound	0.001	0	0	0.001	0.011	0.001	0.001	0.001	0.001	0.028	0.028	0.003	0.005	0	0
Visibility_inbound	0.001	0.001	0	0.001	0.018	0.001	0.001	0.001	0.001	0.031	0.031	0.007	0.005	0	0
regyn_inbound	0.693	0.097	0.001	0.003	0.682	0.139	0.047	0.023	0.004	0.045	0.045	0.079	0.028	0.003	0.003
mregyn_inbound	0.097	0.391	0.001	0.002	0.126	0.039	0.016	0.007	0.002	0.02	0.02	0.023	0.009	0	0
mregyn_outbound	0.001	0.001	0.202	0.092	0.001	0.001	0	0.015	0.012	0.008	0.008	0.002	0.003	0	0
regcse3_outbound	0.003	0.002	0.092	0.72	0.004	0.002	0.001	0.06	0.044	0.017	0.017	0.008	0.014	0.001	0.002
regcse3_inbound	0.682	0.126	0.001	0.004	1.282	0.144	0.05	0.027	0.006	0.095	0.095	0.104	0.047	0.009	0.009
dly1_key_inbound	0.139	0.039	0.001	0.002	0.144	1.688	0.22	0.048	0.013	0.037	0.037	0.115	0.066	0.002	0.002
dly2_key_inbound	0.047	0.016	0	0.001	0.05	0.22	0.804	0.021	0.006	0.015	0.015	0.054	0.028	0.001	0.001
dly1_key_outbound	0.023	0.007	0.015	0.06	0.027	0.048	0.021	1.96	0.201	0.033	0.033	0.044	0.046	0.002	0.002
dly2_key_outbound	0.004	0.002	0.012	0.044	0.006	0.013	0.006	0.201	0.978	0.012	0.012	0.014	0.015	0	0
month_inbound	0.045	0.02	0.008	0.017	0.095	0.037	0.015	0.033	0.012	2.471	2.435	0.01	0.007	0.002	0.002
month_outbound	0.045	0.02	0.008	0.017	0.095	0.037	0.015	0.033	0.012	2.435	2.471	0.01	0.007	0.002	0.002
hour_inbound	0.079	0.023	0.002	0.008	0.104	0.115	0.054	0.044	0.014	0.01	0.01	2.712	1.311	0.001	0.001
hour_outbound	0.028	0.009	0.003	0.014	0.047	0.066	0.028	0.046	0.015	0.007	0.007	1.311	2.645	0.001	0.001
weekday_inbound	0.003	0	0	0.001	0.009	0.002	0.001	0.002	0	0.002	0.002	0.002	0.001	0.589	0.478
weekday_outbound	0.003	0	0	0.002	0.009	0.002	0.001	0.002	0	0.002	0.002	0.001	0.001	0.478	0.589
peaks_rwy_inbound	0.038	0.006	0	0.001	0.047	0.014	0.002	0.008	0.001	0.012	0.012	0.455	0.15	0	0
peaks_rwy_outbound	0.009	0.003	0.001	0.003	0.014	0.006	0.002	0.006	0.001	0.015	0.015	0.121	0.315	0.001	0.001
eur	0.118	0.022	0	0.007	0.118	0.024	0.015	0.011	0.002	0	0	0.119	0.062	0	0
short_tat	0.017	0.002	0	0	0.02	0.014	0.003	0.029	0.004	0.005	0.005	0.106	0.084	0.002	0.002
O_sector_inbound	0.001	0.002	0	0.001	0.004	0.001	0.001	0.001	0	0.007	0.007	0.002	0.002	0	0
O_sector_outbound	0	0.001	0.001	0.001	0.001	0	0	0.001	0	0.006	0.006	0.001	0.001	0	0
O_outbound	0.001	0.001	0	0.001	0.001	0	0	0.001	0	0.003	0.003	0.002	0.002	0	0
O_inbound	0.001	0	0	0	0.002	0	0	0.001	0	0.004	0.004	0.001	0.001	0	0
M_inbound	0	0	0	0	0.009	0	0	0	0	0.006	0.006	0.003	0.002	0	0
M_outbound	0	0	0	0	0.004	0	0	0	0.001	0.005	0.005	0.001	0.002	0	0
R_inbound	0	0	0	0	0.008	0	0	0.001	0	0.021	0.021	0.001	0	0	0
R_outbound	0	0	0	0	0.006	0	0	0.001	0	0.022	0.022	0	0	0	0
Y_inbound	0	0	0	0	0.002	0	0	0	0	0.004	0.004	0.001	0.001	0	0
Y_outbound	0	0	0	0	0.001	0	0	0	0	0.004	0.004	0.001	0.001	0	0
S_inbound	0	0	0	0	0.002	0	0	0.001	0.001	0.006	0.006	0.001	0	0.001	0.001
S_outbound	0	0	0	0	0.001	0	0	0.001	0.001	0.006	0.006	0	0.001	0.001	0.001
wtc_inbound	0.039	0.012	0	0.002	0.041	0.026	0.016	0.007	0.001	0.001	0.001	0.331	0.119	0.001	0.001
wtc_outbound	0.013	0.002	0.001	0.001	0.014	0.002	0.001	0.004	0.002	0	0	0.112	0.202	0.001	0.002
c_d_IAFS	0.002	0	0	0	0.003	0.006	0.005	0.002	0.001	0.003	0.003	0.133	0.048	0	0
c_IAFS	0.001	0.002	0	0.001	0.005	0.001	0.001	0.001	0	0.014	0.014	0.004	0.003	0.001	0.001
d_IAFS	0.001	0	0	0	0.002	0.007	0.005	0.002	0.001	0.003	0.003	0.138	0.049	0	0
wind_speed_inbound	0.001	0.001	0	0.001	0.016	0.002	0.002	0.004	0.001	0.033	0.033	0.027	0.018	0.004	0.003
wind_speed_sector_inbound	0	0.003	0	0.001	0.018	0.001	0.001	0.002	0.001	0.03	0.03	0.02	0.017	0.001	0.001
wind_speed_outbound	0	0	0	0.001	0.013	0	0	0.003	0.002	0.037	0.037	0.019	0.024	0.003	0.004
wind_speed_sector_outbound	0.001	0	0.002	0.005	0.012	0	0	0.002	0.001	0.028	0.028	0.017	0.022	0.001	0.001
CDM_stability	0.002	0	0	0	0.003	0.001	0	0.002	0.002	0.002	0.002	0.012	0.045	0	0
avg_cdm_updates	0.013	0.003	0	0.021	0.021	0.02	0.009	0.068	0.022	0.05	0.05	0.042	0.114	0.002	0.002
slotdla_info	0.002	0.001	0.021	0.056	0.002	0.002	0.001	0.017	0.012	0.006	0.006	0.002	0.005	0	0

Variable	peaks_rwy_inbound	peaks_rwy_outbound	eur	short_tat	O_sector_inbound	O_sector_outbound	O_outbound	O_inbound	M_inbound	M_outbound	R_inbound	R_outbound	Y_inbound	Y_outbound	S_inbound	S_outbound	wtc_inbound	wtc_outbound	c_d_IAFS	c_IAFS	d_IAFS
ac_type	0.047	0.003	0.413	0.049	0	0	0	0	0	0	0	0	0	0	0	0	0.114	0.06	0.001	0	0.001
arr_rwy	0.047	0.007	0.012	0.006	0.002	0.002	0.002	0.002	0.001	0.001	0.025	0.023	0.001	0.001	0	0	0.017	0.005	0.005	0.004	0.005
dep_rwy	0.008	0.034	0.007	0	0.001	0.001	0.002	0.001	0.004	0.004	0.01	0.011	0.001	0.001	0.001	0.001	0.004	0.001	0	0.002	0
fenpnt_inbound	0.012	0.006	0.03	0.005	0.001	0	0	0	0	0	0	0	0	0	0	0	0.008	0.009	0.004	0.001	0.005
fexpnt_outbound	0.009	0.014	0.022	0.002	0.001	0.002	0	0	0.001	0	0.002	0.002	0.001	0	0	0	0.016	0.003	0	0.001	0
dominant sector_inbound	0.001	0.002	0.016	0.003	0.001	0	0	0	0	0	0	0	0	0	0	0	0.002	0.005	0.001	0	0.001
dominant sector_outbound	0.007	0.01	0.019	0.001	0	0.001	0	0	0	0	0	0	0	0	0	0	0.016	0.002	0	0	0
aircraft_swap_inbound	0.001	0	0.011	0	0	0	0	0	0	0	0	0	0	0	0	0	0.003	0.005	0	0	0
aircraft_swap_outbound	0.001	0.001	0.012	0	0	0	0	0	0	0	0	0	0	0	0	0	0.008	0.001	0.001	0	0.001
regcse1_inbound	0.05	0.015	0.118	0.02	0.005	0.002	0.002	0.003	0.008	0.003	0.008	0.006	0.002	0.001	0.002	0.001	0.041	0.014	0.003	0.006	0.003
regcse2_inbound	0.039	0.009	0.118	0.018	0.003	0.001	0.001	0.001	0.001	0	0.001	0.001	0	0	0	0	0.041	0.014	0.002	0.003	0.001
regcse1_outbound	0.001	0.003	0.004	0	0.002	0.003	0.001	0.001	0	0	0.001	0.001	0	0	0	0	0.002	0.001	0	0.002	0
regcse2_outbound	0	0.003	0.007	0	0.001	0.001	0	0	0	0	0	0	0	0	0	0	0.001	0.001	0	0.001	0
GOA_inbound	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STACK_inbound	0.001	0.002	0.016	0.003	0.001	0	0	0	0	0	0	0	0	0	0	0	0.002	0.005	0.001	0	0.001
TV_inbound	0.033	0.008	0.08	0.015	0.002	0	0	0.001	0.001	0	0.001	0.001	0	0	0	0	0.023	0.009	0.003	0.006	0.002
LOC_ID_inbound	0.033	0.008	0.08	0.015	0.001	0	0	0.001	0.001	0	0.001	0.001	0	0	0	0	0.021	0.008	0.002	0.005	0.002
REG_REASON_x_inbound	0.044	0.013	0.08	0.016	0.002	0.001	0.001	0.002	0.008	0.003	0.012	0.008	0.002	0.001	0.001	0.001	0.022	0.009	0.003	0.005	0.003
REG_REASON_y_inbound	0.05	0.013	0.004	0.005	0.008	0.005	0.005	0.006	0.012	0.005	0.018	0.013	0.004	0.001	0.002	0.002	0.014	0.006	0.035	0.043	0.033
REG_REASON_DETAILS_inbound	0.052	0.016	0.08	0.019	0.007	0.003	0.004	0.005	0.013	0.006	0.017	0.012	0.003	0.001	0.004	0.003	0.022	0.011	0.004	0.013	0.004
TOBT STABLE_outbound	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOBT ADH_outbound	0.001	0.001	0.004	0.003	0	0	0	0	0	0	0	0	0	0	0	0	0.001	0.002	0	0	0
TSAT ADH_outbound	0	0	0.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.001	0	0	0
regulated_outbound	0	0.002	0.001	0	0.001	0.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
season	0.009	0.014	0	0.005	0.006	0.005	0.002	0.002	0.006	0.004	0.005	0.005	0.002	0.002	0.003	0.003	0.001	0	0.003	0.007	0.003
time_of_day_inbound	0.016	0.027	0.047	0.018	0.002	0	0.001	0.001	0.001	0	0	0	0.001	0	0	0	0.164	0.057	0.005	0.002	0.003
time_of_day_outbound	0.04	0.044	0.008	0.01	0.001	0.001	0.001	0	0.001	0.001	0	0	0.001	0.001	0	0	0.066	0.042	0.003	0.002	0.002
Visibility_outbound	0.001	0	0	0.001	0	0	0	0	0.018	0.025	0.008	0.014	0.005	0.009	0.004	0.003	0	0	0	0	0
Visibility_inbound	0.002	0	0.001	0.001	0	0.001	0	0.001	0.036	0.012	0.014	0.009	0.01	0.004	0.004	0.001	0.002	0	0	0	0
regyn_inbound	0.038	0.009	0.118	0.017	0.001	0	0.001	0.001	0	0	0	0	0	0	0	0	0.039	0.013	0.002	0.001	0.001
mregyn_inbound	0.006	0.003	0.022	0.002	0.002	0.001	0.001	0	0	0	0	0	0	0	0	0	0.012	0.002	0	0.002	0
mregyn_outbound	0	0.001	0	0	0	0.001	0	0	0	0	0	0	0	0	0	0	0	0.001	0	0	0
regcse3_outbound	0.001	0.003	0.007	0	0.001	0.001	0.001	0	0	0	0	0	0	0	0	0	0.002	0.001	0	0.001	0
regcse3_inbound	0.047	0.014	0.118	0.02	0.004	0.001	0.001	0.002	0.009	0.004	0.008	0.006	0.002	0.001	0.002	0.001	0.041	0.014	0.003	0.005	0.002
dly1_key_inbound	0.014	0.006	0.024	0.014	0.001	0	0	0	0	0	0	0	0	0	0	0	0.026	0.002	0.006	0.001	0.007
dly2_key_inbound	0.002	0.002	0.015	0.003	0.001	0	0	0	0	0	0	0	0	0	0	0	0.016	0.001	0.005	0.001	0.005
dly1_key_outbound	0.008	0.006	0.011	0.029	0.001	0.001	0.001	0.001	0.001	0	0.001	0.001	0	0	0.001	0.001	0.007	0.004	0.002	0.001	0.002
dly2_key_outbound	0.001	0.001	0.002	0.004	0	0	0	0	0	0.001	0	0	0	0	0.001	0.001	0.001	0.002	0.001	0	0.001
month_inbound	0.012	0.015	0	0.005	0.007	0.006	0.003	0.004	0.006	0.005	0.021	0.022	0.004	0.004	0.006	0.006	0.001	0	0.003	0.014	0.003
month_outbound	0.012	0.015	0	0.005	0.007	0.006	0.003	0.004	0.006	0.005	0.021	0.022	0.004	0.004	0.006	0.006	0.001	0	0.003	0.014	0.003
hour_inbound	0.455	0.121	0.119	0.106	0.002	0.001	0.002	0.001	0.003	0.001	0.001	0	0.001	0.001	0.001	0	0.331	0.112	0.133	0.004	0.138
hour_outbound	0.15	0.315	0.062	0.084	0.002	0.001	0.002	0.001	0.002	0.002	0	0	0.001	0.001	0	0.001	0.119	0.202	0.048	0.003	0.049
weekday_inbound	0	0.001	0	0.002	0	0	0	0	0	0	0	0	0	0	0.001	0.001	0.001	0.001	0	0.001	0
weekday_outbound	0	0.001	0	0.002	0	0	0	0	0	0	0	0	0	0	0.001	0.001	0.001	0.002	0	0.001	0
peaks_rwy_inbound	0.918	0.058	0.047	0.033	0	0	0	0	0.001	0	0.001	0.001	0	0	0.001	0	0.13	0.017	0.03	0.001	0.029
peaks_rwy_outbound	0.058	0.79	0.002	0.003	0	0	0	0	0	0	0.001	0.001	0	0	0	0	0.032	0.012	0.002	0	0.001
eur	0.047	0.002	0.413	0.047	0	0	0	0	0	0	0	0	0	0	0	0	0.114	0.059	0.001	0	0.001
short_tat	0.033	0.003	0.047	0.635	0	0	0	0	0	0	0	0	0	0	0	0	0.026	0.003	0	0	0
O_sector_inbound	0	0	0	0	0.071	0.008	0.006	0.018	0	0	0.005	0.002	0	0	0	0	0.001	0	0	0.005	0
O_sector_outbound	0	0	0	0	0.008	0.064	0.014	0.007	0	0	0.003	0.005	0	0	0	0	0	0	0	0.002	0
O_outbound	0	0	0	0	0.006	0.014	0.054	0.009	0	0	0.004	0.009	0	0	0	0	0	0	0	0.002	0
O_inbound	0	0	0	0	0.018	0.007	0.009	0.052	0	0	0.009	0.002	0	0	0	0	0	0	0	0.004	0
M_inbound	0.001	0	0	0	0	0	0	0	0.05	0.012	0.001	0	0.011	0.004	0	0	0.001	0	0	0	0
M_outbound	0	0	0	0	0	0	0	0	0.012	0.035	0	0.001	0.005	0.01	0	0	0	0	0	0	0
R_inbound	0.001	0.001	0	0	0.005	0.003	0.004	0.009	0.001	0	0.542	0.138	0	0	0	0	0	0	0	0.001	0
R_outbound	0.001	0.001	0	0	0.002	0.005	0.009	0.002	0	0.001	0.138	0.542	0	0	0	0	0	0	0	0	0
Y_inbound	0	0	0	0	0	0	0	0	0.011	0.005	0	0	0.017	0.006	0	0	0.001	0	0	0	0
Y_outbound	0	0	0	0	0	0	0	0	0.004	0.01	0	0	0.006	0.014	0	0	0	0	0	0	0
S_inbound	0.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0.004	0	0	0	0	0
S_outbound	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.004	0.021	0	0	0	0	0
wtc_inbound	0.13	0.032	0.114	0.026	0.001	0	0	0	0.001	0	0	0	0.001	0	0	0	0.753	0.026	0.013	0.001	0.013
wtc_outbound	0.017	0.012	0.059	0.003	0	0	0	0	0	0	0	0	0	0	0	0	0.026	0.615	0.008	0	0.008
c_d_IAFS	0.03	0.002	0.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0.013	0.008	0.649	0.002	0.518
c_IAFS	0.001	0	0	0	0.005	0.002	0.002	0.004	0	0	0.001	0	0	0	0	0	0.001	0	0.002	0.066	0
d_IAFS	0.029	0.001	0.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0.013	0.008	0.518	0	0.64
wind_speed_inbound	0.003	0.003	0.001	0.002	0	0	0	0.001	0.003	0.001	0.036	0.03	0.001	0.001	0	0	0.004	0	0.001	0.001	0.001
wind_speed_sector_inbound	0.002	0.004	0.001	0.001	0	0	0	0	0.001	0.001	0.028	0.022	0	0.001	0	0	0.002	0.002	0.001	0.001	0.002
wind_speed_outbound	0.003	0.003	0.001	0.001	0.001	0.001	0.001	0	0.002	0.002	0.033	0.035	0	0	0	0	0.001	0.004	0.001	0.001	0.001
wind_speed_sector_outbound	0.005	0.005	0.004	0.001	0	0	0	0	0.001	0.001	0.027	0.028	0	0	0	0	0.004	0.005	0	0	0.001
CDM_stability	0.005	0.01	0.005	0.003	0	0	0	0	0	0	0	0	0	0	0	0	0.001	0.011	0.004	0	0.003
avg_cdm_updates	0.008	0.009	0.003	0.01	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.001	0.001	0	0.001	0	0.001	0.01	0.002	0.001	0.002
slotdla_info	0	0	0	0	0.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.001	0

Variable	wind_speed_inbound	wind_speed_sector_inbound	wind_speed_outbound	wind_speed_sector_outbound	CDM_stability	avg_cdm_updates	slotdla_info
ac_type	0.001	0.002	0.001	0.004	0.005	0.004	0.001
arr_rwy	0.034	0.032	0.032	0.021	0.001	0.013	0
dep_rwy	0.022	0.02	0.029	0.039	0.001	0.009	0.006
fenpnt_inbound	0	0.08	0	0.001	0.001	0.002	0
fexpnt_outbound	0.002	0.004	0.002	0.071	0.001	0.006	0.018
dominant sector_inbound	0	0.08	0	0.001	0	0.001	0
dominant sector_outbound	0	0.001	0	0.075	0.001	0.003	0.018
aircraft_swap_inbound	0	0	0	0	0	0	0
aircraft_swap_outbound	0.001	0	0	0	0	0.002	0
regcse1_inbound	0.015	0.016	0.011	0.011	0.004	0.024	0.002
regcse2_inbound	0.001	0.004	0	0.001	0.002	0.018	0.002
regcse1_outbound	0.001	0.001	0.001	0.004	0	0.001	0.057
regcse2_outbound	0	0	0	0.004	0	0.001	0.056
GOA_inbound	0	0	0	0	0	0	0
STACK_inbound	0	0.08	0	0	0	0.001	0
TV_inbound	0.002	0.003	0.001	0.001	0.002	0.017	0.001
LOC_ID_inbound	0.001	0.001	0.001	0.001	0.002	0.017	0.001
REG_REASON_x_inbound	0.022	0.017	0.018	0.014	0.004	0.025	0.001
REG_REASON_y_inbound	0.039	0.027	0.031	0.023	0.005	0.044	0.003
REG_REASON_DETAILS_inbound	0.046	0.037	0.039	0.029	0.004	0.034	0.002
TOBT STABLE_outbound	0	0	0	0	0.001	0.003	0
TOBT ADH_outbound	0	0	0	0	0.047	0.001	0.002
TSAT ADH_outbound	0	0	0	0	0.009	0.001	0.003
regulated_outbound	0	0	0	0.003	0	0	0.091
season	0.017	0.015	0.018	0.013	0.001	0.036	0.004
time_of_day_inbound	0.014	0.004	0.01	0.011	0.001	0.006	0
time_of_day_outbound	0.014	0.007	0.015	0.011	0.003	0.017	0.001
Visibility_outbound	0.002	0.001	0.002	0.001	0	0.003	0
Visibility_inbound	0.003	0.001	0.002	0.001	0	0.003	0
regyn_inbound	0.001	0	0	0.001	0.002	0.013	0.002
mregyn_inbound	0.001	0.003	0	0	0	0.003	0.001
mregyn_outbound	0	0	0	0.002	0	0	0.021
regcse3_outbound	0.001	0.001	0.001	0.005	0	0.001	0.056
regcse3_inbound	0.016	0.018	0.013	0.012	0.003	0.021	0.002
dly1_key_inbound	0.002	0.001	0	0	0.001	0.02	0.002
dly2_key_inbound	0.002	0.001	0	0	0	0.009	0.001
dly1_key_outbound	0.004	0.002	0.003	0.002	0.002	0.068	0.017
dly2_key_outbound	0.001	0.001	0.002	0.001	0.002	0.022	0.012
month_inbound	0.033	0.03	0.037	0.028	0.002	0.05	0.006
month_outbound	0.033	0.03	0.037	0.028	0.002	0.05	0.006
hour_inbound	0.027	0.02	0.019	0.017	0.012	0.042	0.002
hour_outbound	0.018	0.017	0.024	0.022	0.045	0.114	0.005
weekday_inbound	0.004	0.001	0.003	0.001	0	0.002	0
weekday_outbound	0.003	0.001	0.004	0.001	0	0.002	0
peaks_rwy_inbound	0.003	0.002	0.003	0.005	0.005	0.008	0
peaks_rwy_outbound	0.003	0.004	0.003	0.005	0.01	0.009	0
eur	0.001	0.001	0.001	0.004	0.005	0.003	0
short_tat	0.002	0.001	0.001	0.001	0.003	0.01	0
O_sector_inbound	0	0	0.001	0	0	0.001	0.001
O_sector_outbound	0	0	0.001	0	0	0.001	0
O_outbound	0	0	0.001	0	0	0.001	0
O_inbound	0.001	0	0	0	0	0.001	0
M_inbound	0.003	0.001	0.002	0.001	0	0.002	0
M_outbound	0.001	0.001	0.002	0.001	0	0.002	0
R_inbound	0.036	0.028	0.033	0.027	0	0.002	0
R_outbound	0.03	0.022	0.035	0.028	0	0.001	0
Y_inbound	0.001	0	0	0	0	0.001	0
Y_outbound	0.001	0.001	0	0	0	0	0
S_inbound	0	0	0	0	0	0.001	0
S_outbound	0	0	0	0	0	0	0
wtc_inbound	0.004	0.002	0.001	0.004	0.001	0.001	0
wtc_outbound	0	0.002	0.004	0.005	0.011	0.01	0
c_d_IAFS	0.001	0.001	0.001	0	0.004	0.002	0
c_IAFS	0.001	0.001	0.001	0	0	0.001	0.001
d_IAFS	0.001	0.002	0.001	0.001	0.003	0.002	0
wind_speed_inbound	1.016	0.236	0.364	0.179	0	0.003	0
wind_speed_sector_inbound	0.236	0.985	0.192	0.154	0	0.002	0
wind_speed_outbound	0.364	0.192	1.028	0.248	0	0.004	0
wind_speed_sector_outbound	0.179	0.154	0.248	0.996	0	0.004	0
CDM_stability	0	0	0	0	0.978	0.015	0.001
avg_cdm_updates	0.003	0.002	0.004	0.004	0.015	0.947	0.004
slotdla_info	0	0	0	0	0.001	0.004	0.26

C

MUTUAL INFORMATION FEATURE SELECTION & FINAL VARIABLES

C.1. MUTUAL INFORMATION WITH ATC DELAY

Table C.1 – MI scores of each feature with the target variable ATC delay, features in bold were selected into the top 30, underlined features were manually added as they were of interest.

Feature	MI score	MI score normalised
avg_startup_delay_outbound	0.1096	1.0000
dly1_key_outbound	0.1062	0.9688
purdla_outbound	0.0883	0.8055
TSAT_updates_outbound	0.0583	0.5318
diff_TTOT_optimal_outbound	0.0529	0.4824
reg_dly_key_outbound	0.0522	0.4765
avg_cdm_updates	0.0474	0.4327
slotdla_info	0.0381	0.3477
adcdla_outbound	0.0381	0.3474
CTOT_updates_outbound	0.0369	0.3365
diff_capacity_actual_outbound	0.0191	0.1745
reg_induced_dla_outbound	0.0191	0.1744
avg_rate_dep_outbound	0.0163	0.1488
mregyn_outbound	0.0161	0.1468
sector_outbound	0.0143	0.1305
TOBT_updates_outbound	0.0114	0.1036
ratio_reg_flights_outbound	0.0098	0.0891
dep_rwy	0.0092	0.0841
time_of_day_outbound	0.0092	0.0840
arrdla_inbound	0.0079	0.0718
purdla_inbound	0.0074	0.0677
dly1_key_inbound	0.0068	0.0622
season	0.0064	0.0583
reg_dly_key_inbound	0.0055	0.0498
reg_rate_inbound	0.0046	0.0421
reg_reason_inbound	0.0045	0.0408
diff_capacity_actual_inbound	0.0043	0.0396
TOBT_adh_outbound	0.0042	0.0383
cindla_outbound	0.0041	0.0376
time_of_day_inbound	0.0037	0.0342
wtc_outbound	0.0031	0.0282
cdm_stability	0.0030	0.0272

Continued on next page

Table C.1 – continued from previous page

Feature	MI score	MI score normalized
arr_rwy	0.0029	0.0260
short_tat	0.0025	0.0226
c_d_EHAMDEP_outbound	0.0024	0.0215
Average Rate Arr_inbound	0.0023	0.0211
reg_induced_dla_inbound	0.0023	0.0208
boarding_duration	0.0018	0.0168
mregyn_inbound	0.0017	0.0152
ac_type	0.0016	0.0149
TSAT_adh_outbound	0.0013	0.0116
DD_outbound	0.0012	0.0112
P_sector_inbound	0.0011	0.0101
DD_inbound	0.0011	0.0097
#ASRT_outbound	0.0011	0.0096
P_sector_outbound	0.0010	0.0093
wtc_inbound	0.0010	0.0092
c_d_IAFS	0.0010	0.0092
aircraft_swap_outbound	0.0008	0.0069
Visibility_outbound	0.0007	0.0065
time_in_tma	0.0007	0.0063
atxi_inbound	0.0006	0.0058
O_outbound	0.0006	0.0058
wind_speed_sector_outbound	0.0006	0.0054
R_inbound	0.0006	0.0054
peaks_rwy_outbound	0.0006	0.0053
O_sector_inbound	0.0005	0.0049
c_EHDEP_outbound	0.0005	0.0048
wind_speed_sector_inbound	0.0005	0.0045
O_sector_outbound	0.0005	0.0044
c_IAFS	0.0005	0.0044
dominant sector_inbound	0.0004	0.0040
STACK_inbound	0.0004	0.0040
wind_speed_inbound	0.0004	0.0039
peaks_rwy_inbound	0.0004	0.0036
R_outbound	0.0004	0.0035
time_in_fir	0.0004	0.0035
O_inbound	0.0004	0.0035
eur	0.0004	0.0034
c_d_EHAMARR_inbound	0.0004	0.0033
#EXOT_outbound	0.0003	0.0032
M_outbound	0.0003	0.0031
Visibility_inbound	0.0003	0.0029
weekday_outbound	0.0003	0.0028
wind_speed_outbound	0.0003	0.0027
M_inbound	0.0003	0.0025
weekday_inbound	0.0003	0.0025
TOBT STABLE_outbound	0.0003	0.0023
c_EHARR_inbound	0.0002	0.0019
S_inbound	0.0002	0.0017
Y_outbound	0.0001	0.0012
aircraft_swap_inbound	0.0001	0.0010
wags_inbound	0.0001	0.0009
Y_inbound	0.0001	0.0009

Continued on next page

Table C.1 – continued from previous page

Feature	MI score	MI score normalized
S_outbound	0.0001	0.0007
cindla_inbound	0.0000	0.0003
GOA_inbound	0.0000	0.0001

C.2. FINAL LIST OF VARIABLES

Table C.2 – Final list of included variables, their unit and discretized values.

Variable	Unit	Discretization
Time of day departure	-	morning (<13h)/midday(13h-18h)/evening(>18h) in Local time
Time of day arrival	-	morning (<13h)/midday(13h-18h)/evening(>18h) in Local time
Peak indicator arrival	-	0 (no)/1(inbound)/2(outbound)/3(combination)
Peak indicator departure	-	0 (no)/1(inbound)/2(outbound)/3(combination)
Visibility arrival	-	Good/Marginal/LVP
Visibility departure	-	Good/Marginal/LVP
Wind speed departure	-	0/1/2
Wind direction departure	degrees	[0.0;130.0][140.0;240.0]/[250.0;360.0]
Season	-	W(Winter)/S(Summer off-peak)/S-peak(July/August)
Aircraft type	-	widebody/737/E75-90
Weekday departure	-	0(no weekday)/1(weekday)
Arrival delay	min	[-80.0;0.0]/[1.0;39.0]/[40.0;175.0]
Delay key category arr	-	none/aircraft&ramp_handling/Reactionary/specific/ flight_ops&crew/ATFM/pax&baggage/airport&gov_auth/ tech&aircraft_equipment/Damage&failure/cargo/weather/Other
Pure ATFM delay arrival	min	[0.0;13.0]/[14.0;50.0]/[51.0;193.0]
Waiting for ground services	min	[0.0;1.0]/[2.0;3.0]/[4.0;19.0]
All doors closed delay dep	min	[-35.0;3.0]/[4.0;22.0]/[23.0;101.0]
Company induced delay	min	0/>0
Delay key category dep	-	none/aircraft&ramp_handling/Reactionary/specific/ flight_ops&crew/ATFM/pax&baggage/airport&gov_auth/ tech&aircraft_equipment/Damage&failure/cargo/weather/Other
Pure ATFM delay dep	min	0/>0
Info on slot delay	-	0/1/2
Short turnaround indicator	-	0(long turnaround)/1(short turnaround)
Regulation induced delay dep	min	[-49.0;-7.0]/[-6.0;2.0]/[3.0;15.0]/[16.0;58.0]
Boarding duration	min	[0.0;10.9]/[10.9;24.0]/[24.0;121.5]
TOBT updates	-	[0.0;1.0]/[2.0;5.0]/[6.0;33.0]
TSAT updates	-	[0.0;14.0]/[15.0;30.0]/[31.0;124.0]
CTOT updates	-	[0.0;1.0]/[2.0;4.0]/[5.0;30.0]
TSAT adherence	-	True/False
TOBT adherence	-	True/False
Difference TTOT and optimal TTOT	sec	[-5400.0;-239.0]/[-181.0;481.0]/[539.0;5280.0]
Average startup delay	min	[-281.5;5.5]/[5.5;12.6]/[12.6;73.5]
Average CDM updates	-	0/1/2
Average CDM adherence	-	0/1/2
Difference capacity and usage dep	aircraft/hour	[-50.7;1.0]/[1.0;19.7]/[19.7;66.5]
Difference capacity and usage arr	aircraft/hour	[-54.8;-12.5] /[-12.5;10.0]/[10.0;59.5]
Difference capacity and demand dep	aircraft/20 min	[-24.3;0.0]/[0.3;7.0]/[7.3;24.7]
Difference capacity and demand arr	aircraft/20 min	[-31.7;-2.3]/[-2.0;5.3]/[5.7;21.7]
Ratio regulated flights	-	[0.0;0.2]/[0.2;0.4]/[0.4;1.0]
Departure runway	-	18L/24/18C/09/36L/36C/27
Sector used during dep	-	sector1/sector2/sector3/sector4/sector5
Regulation reason arrival	-	0/C-ATCCapacity/G-AerodromeCapacity/W-Weather/ O-Other/V-EnvironmentalIssues/P-SpecialEvent/E-AerodromeServices /N-IndActionnon-ATC/S-ATCStaffing/T-ATCEquipment
Regulation rate arrival	aircraft/hour	0(no regulation)/[10;49]/[50;68]
Regulation delay key arr	-	0.0(no regulation)/81.0/82.0/83.0/84.0/89.0/98.0/99.0
Regulation delay key dep	-	0.0(no regulation)/81.0/82.0/83.0/84.0/98.0/99.0
Multiple regulation indicator dep	-	0(no multiple regulations) /1(multiple regulations)
Average usage rate of runways dep	aircaft/hour	[2.9;43.5]/[43.5;60.7]/[60.7;90.1]
ATC delay	min	(-38.001;0.0]/(0.0;4.0]/(4.0;9.0]/(9.0;55.0]

Table C.3 – The clustering centres of the different variables included into the aggregation variable average CDM updates of flights in 20 minute time frame of departing flight.

Labels Average CDM updates	**Average #TSAT updates**	**Average #TOBT updates**
0	10.97	1.00
1	14.86	2.78
2	28.31	1.87

Table C.4 – The clustering centres of the different variables included into the aggregation variable average CDM adherence of flights in 20 minute time frame of departing flight.

Labels Average CDM adherence	**% TSAT adherence**	**% TOBT adherence**	**% TOBT stability**
0	95	75	99
1	98	92	99
2	88	53	97

Table C.5 – The clustering centres of the different variables included into the aggregation variable slot delay information.

Labels Slot delay information	**Dynamics slot delay (min)**	**Evolution slot delay (min)**
0	1.24	-0.23
1	52.43	20.78
2	36.39	-31.58

Table C.6 – The clustering centres of the different variables included into the aggregation variable wind speed departure.

Labels Wind speed departure	**Average wind speed (0.1 m/s)**	**Maximum wind gust (0.1 m/s)**
0	29.40186	50.97019
1	94.60235	146.3851
2	57.75105	92.22337

Table C.7 – The definition used in the states of visibility.

Visibility	**Horizontal view (m)**	**Ceiling (ft)**
Good	≥5000	≥1000
Marginal	<5000 & ≥1500	<1000 & ≥300
LVP	<1500	<300

Table C.8 – The definition used in short turnaround time indicator

Labels	**Actual turnaround time (h)**	
Short turnaround time	**Intercontinental flights**	**European flights**
0	>3	>1.5
1	≤3	≤1.5

D

VERIFICATION RESULTS

D.1. ASSOCIATION RULE MINING

Table D.1 presents the small sample data set used as input to verify the developed association rule mining method using the FP-growth algorithm. The result of one-hot-encoding this data set can be seen in Table D.3, which was based on the individual support or frequency of the conditions, shown in Table D.2. Finally, the manually constructed FP-tree that was built to find the association rules is visualised in Figure D.1.

Table D.1 – Data set used for verification.

TOBT STABLE_outbound	TOBT ADH_outbound	TSAT ADH_outbound	regulated_outbound	season
TRUE	TRUE	TRUE	FALSE	S
TRUE	TRUE	TRUE	TRUE	W
TRUE	FALSE	FALSE	TRUE	S

Table D.2 – The support values of all the one-hot-encoded variables in the verification data set.

Variable	Support
TOBT STABLE_outbound -True	1
TOBT ADH_outbound - True	0.667
TSAT ADH_outbound - True	0.667
Season -S	0.667
regulated - True	0.667
TOBT ADH_outbound - False	0.333
TSAT ADH_outbound - False	0.333
Season - W	0.33
Regulated - False	0.333

Table D.3 – Ordered on support and one-hot encoded data of the verification data set.

Ordered one-hot-encoded data
TOBT Stable True, TSAT ADH True, TOBT ADH True, Summer, Non regulated
TOBT Stable True, TSAT ADH True, TOBT ADH True, Winter, Regulated
TOBT Stable True, Summer, Regulated, TSAT ADH False, TOBT ADH False

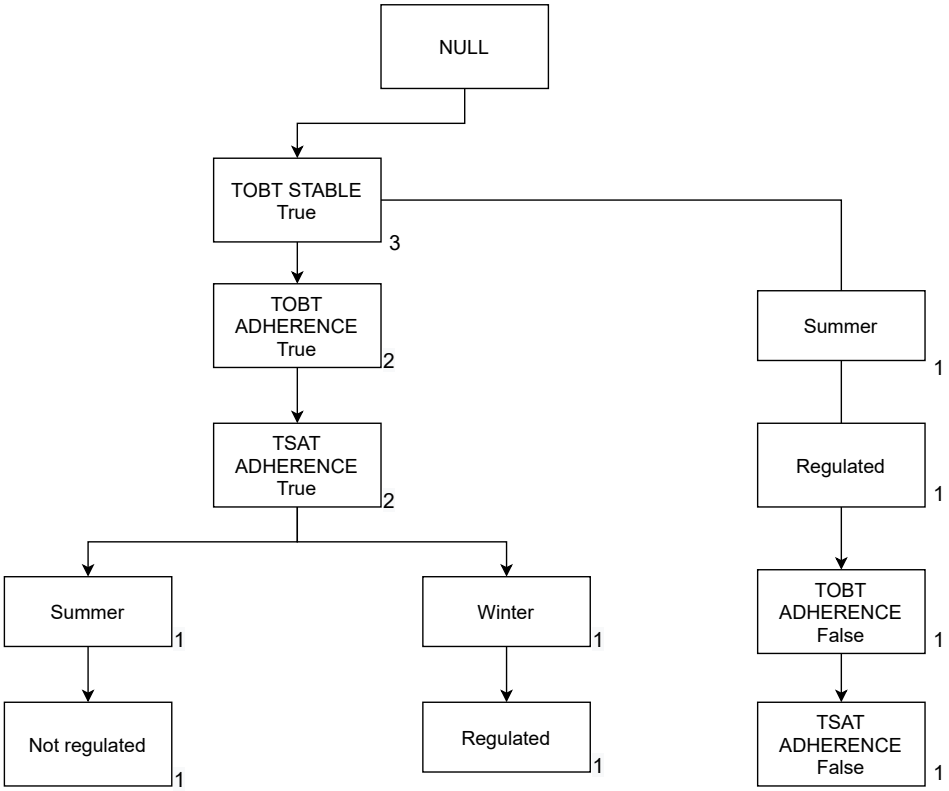


Figure D.1 – FP-tree constructed manually for verification purposes.

D.2. BAYESIAN NETWORK

Table D.4 displays the required run time and structural hamming distance for the different Bayesian network structure learning algorithms on the Cancer Bayesian network. Table D.5 shows the same performance indicators, but a data set sampled from Alarm Bayesian network, with a sample size of 100,000. Additionally, the values of all available scoring functions are added.

Table D.4 – Run time and SHD results for the different structure learning algorithms on the Cancer verification data set.

Algorithm	**Run time (s)**		**SHD (-)**	
	sample size 100,000	sample size 1 M	sample size 100,000	sample size 1M
Exhaustive K2	62.5	95.0	3	3
Exhaustive BIC	54.2	109.9	0	0
Exhaustive BDeu	59.5	96.3	0	0
Hill climb K2	3.1	26.4	3	3
Hill climb BIC	3.0	26.3	3	0
Hill climb BDeu	3.0	26.6	4	4
PC	2.8	24	1	0

Table D.5 – Results of different structure learning algorithms on the sampled ALARM verification data set, n=100,000.

Algorithm	**Run time (s)**	**SHD (-)**	**Scoring function (-)**		
			K2	BDeu	BIC
Exhaustive K2	/	/	/	/	/
Exhaustive BDeu	/	/	/	/	/
Exhaustive BIC	/	/	/	/	/
Hill climb K2	934.0	26	-941.798E3	-941.588E3	-944.809E3
Hill climb BDeu	765.1	35	-941.636E3	-941.224E3	-944.282E3
Hill climb BIC	731.8	34	-942.651E3	-942.252E3	-944.317E3
Hybrid K2	576.7	18	-959.115E3	-958.928E3	-959.622E3
Hybrid BDeu	656.5	17	-959.124E3	-958.928E3	-959.622E3
Hybrid BIC	627.7	17	-970.762E3	-970.488E3	-971.404E3
PC	/	/	/	/	/

E

INFERENCE BAYESIAN NETWORK

E.1. SAMPLING SIZE DETERMINATION

In approximate inference, sampling is used to determine the influence of the fixed evidence variables on other nodes in the network. When sampling, the correct sample size needs to be determined in order to obtain stable results, and that the obtained probability distribution of the nodes would approximate the actual distribution as good as possible. Therefore, four variables with very different positions in the network were used as evidence in order to determine the required sample size. In Figure E.1-E.4 the resulting probability distributions of the ATC delay variable is shown, in function of a varying sample size up to and including 150,000. It can be seen that for all inference cases the probability distribution stabilised when using 100,000 data points in the sampling process.

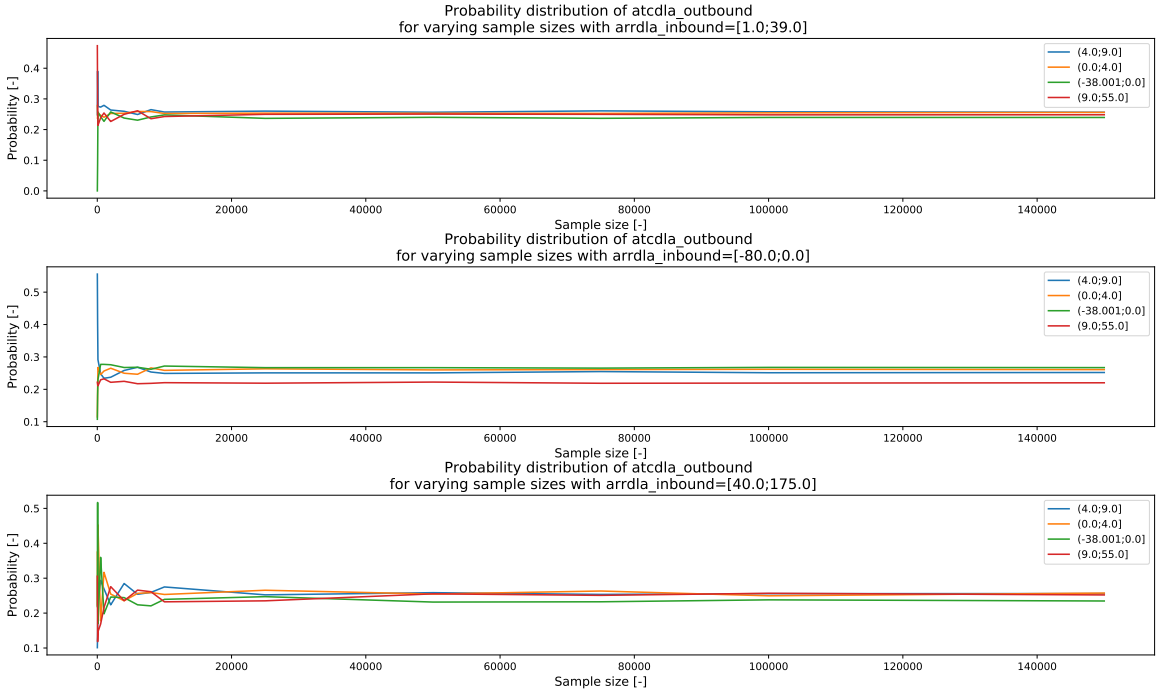


Figure E.1 – The probability distribution of target variable ATC delay in function of the sampling size in approximate inference, with arrival delay as specified evidence.

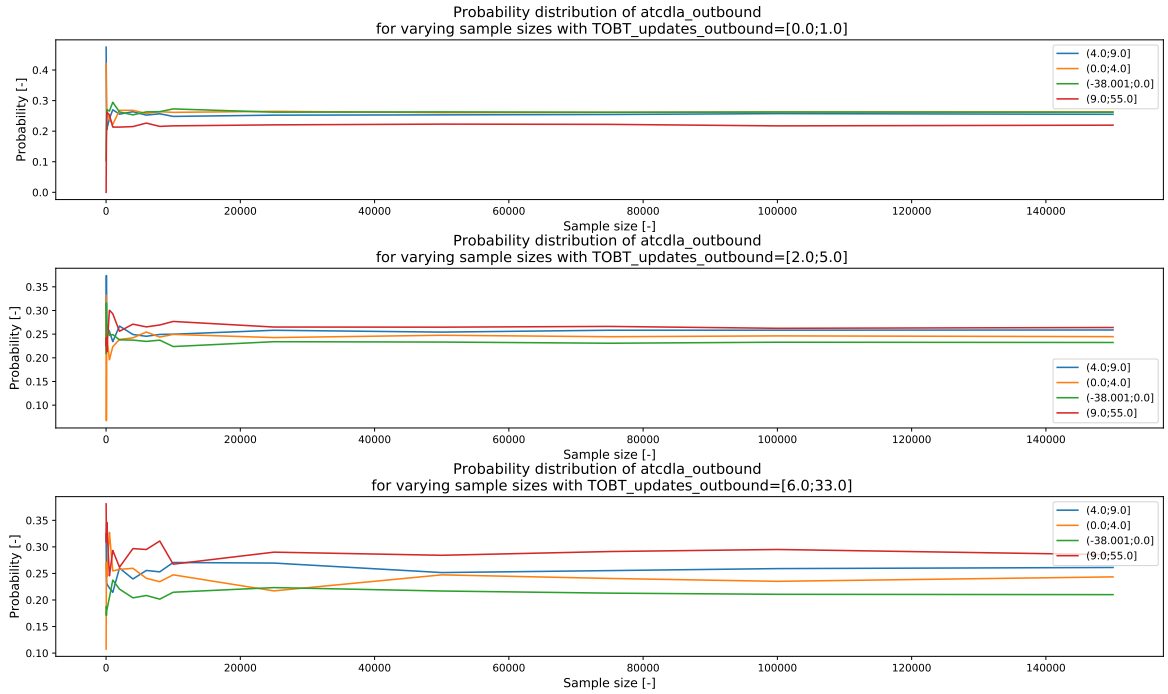


Figure E.2 – The probability distribution of target variable ATC delay in function of the sampling size in approximate inference, with number of TOBT updates as specified evidence.

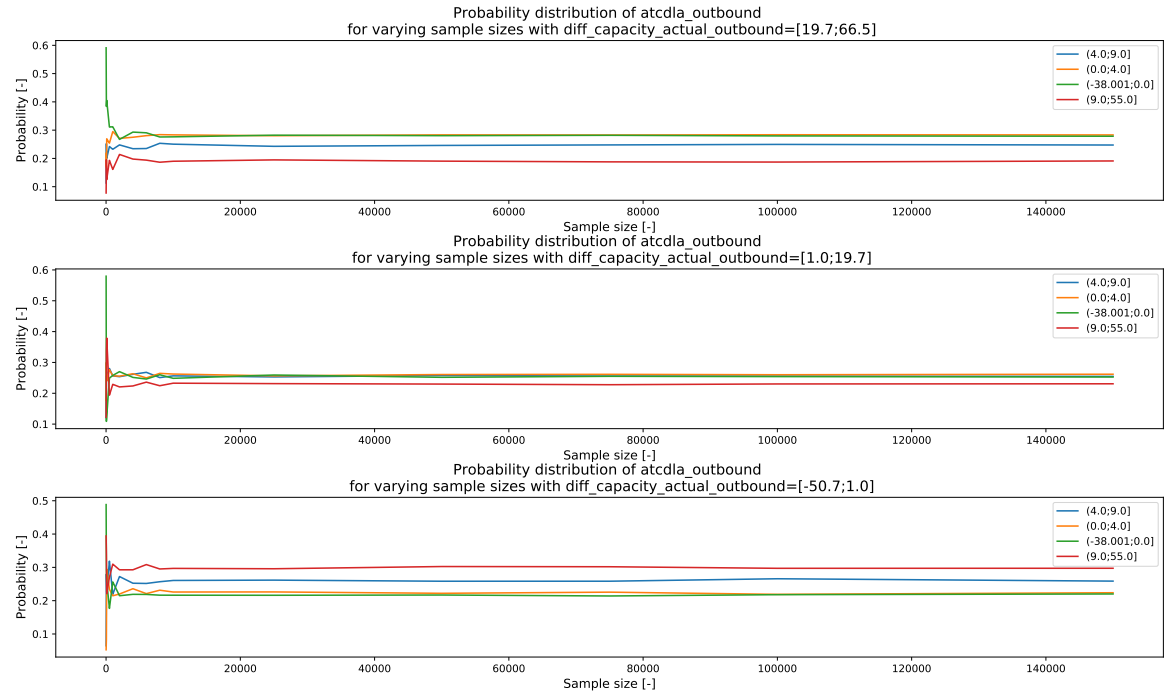


Figure E.3 – The probability distribution of target variable ATC delay in function of the sampling size in approximate inference, with the difference between the capacity and the actual departure rate as specified evidence.

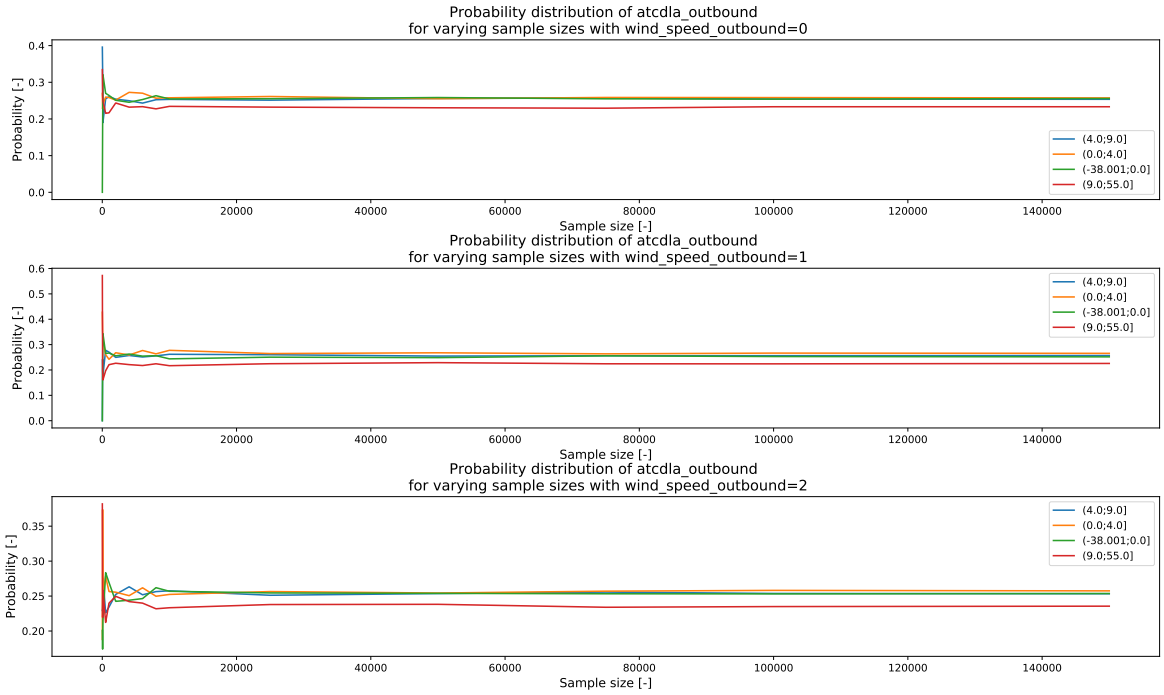


Figure E.4 – The probability distribution of target variable ATC delay in function of the sampling size in approximate inference, with the wind speed during departure as specified evidence.

E.2. INFERENCE RESULTS

This section presents additional results obtained from the inference analysis on the constructed Bayesian network.

E.2.1. AIRLINE INFLUENCE VARIABLES

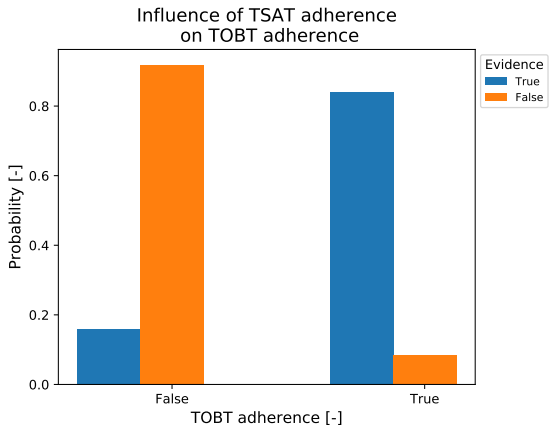


Figure E.5 – Influence of TSAT adherence on TOBT adherence.

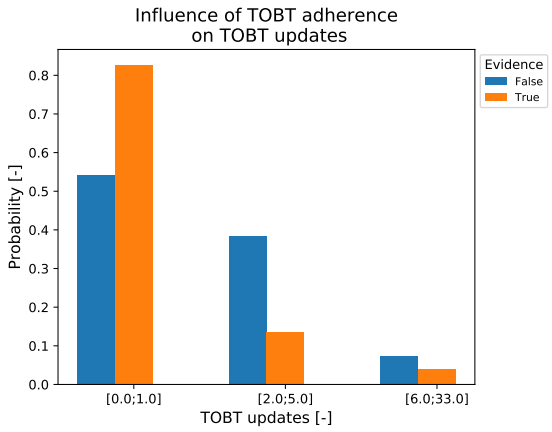


Figure E.6 – Influence of TOBT adherence on TOBT updates.

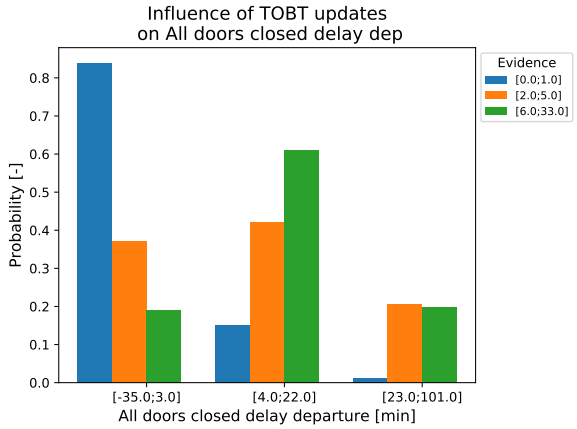


Figure E.7 – Influence of TOBT updates on all doors closed delay.

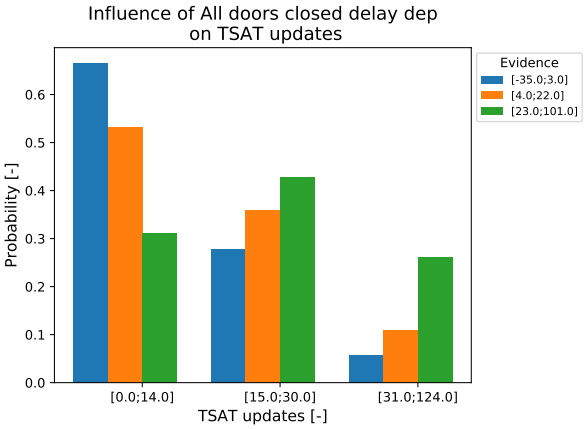


Figure E.8 – Influence of all doors closed delay on TSAT updates.

E.2.2. VERIFICATION OF MODEL

The inference method could also be used to additionally verify the obtained Bayesian network. The results of this are presented in Figure E.9-E.12.

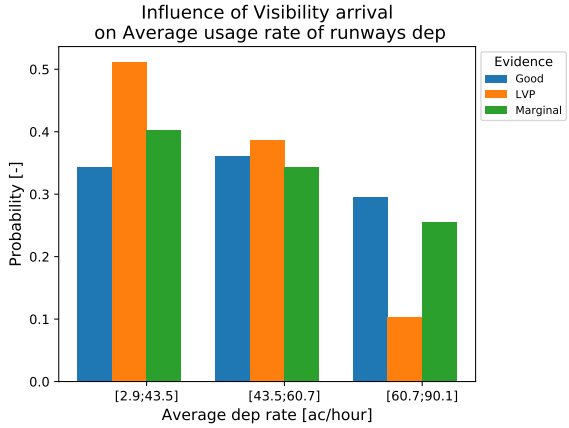


Figure E.9 – Influence of visibility during arrival on the departure rate in aircraft per hour during departure of the flight.

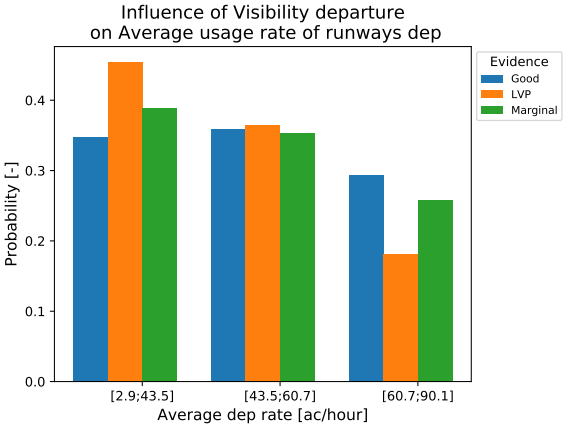


Figure E.10 – Influence of visibility during departure on the departure rate in aircraft per hour during departure of the flight.

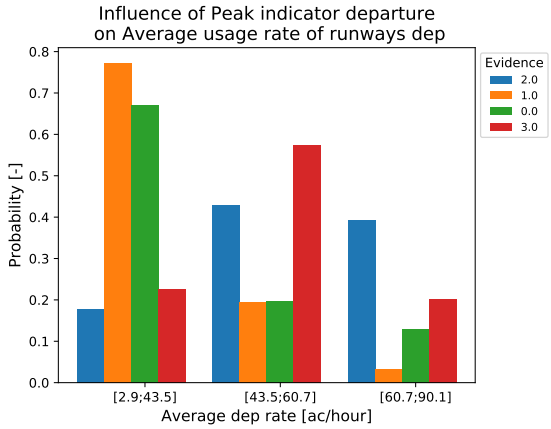


Figure E.11 – Influence of peak indicator during departure on the departure rate in aircraft per hour.

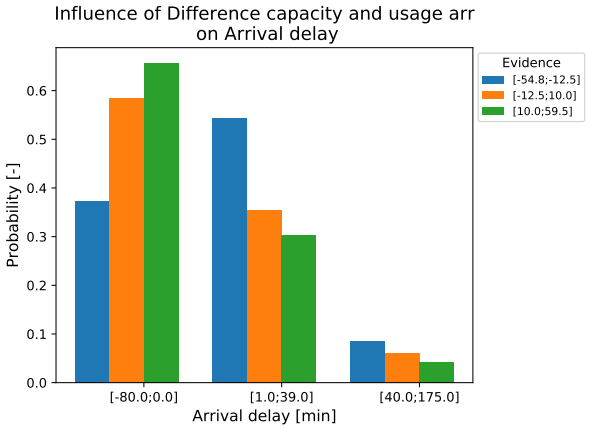


Figure E.12 – Influence of the difference between capacity and actual arrival rate on arrival delay.

E.2.3. INFLUENCE ON ATC DELAY

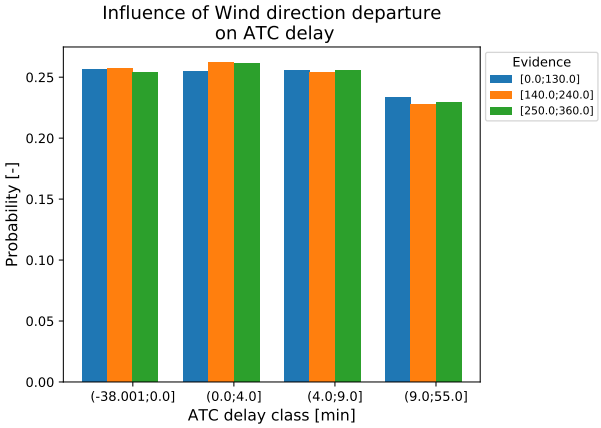


Figure E.13 – Influence of wind direction during departure on ATC delay.

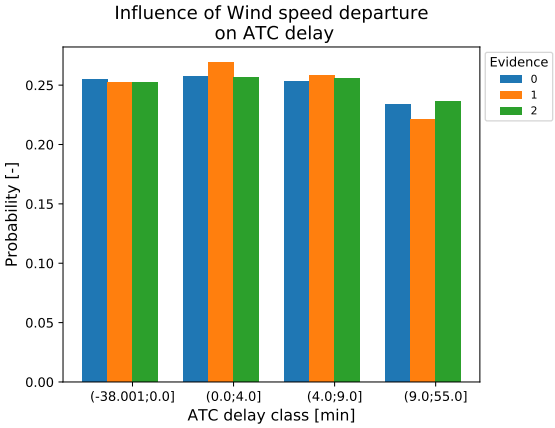


Figure E.14 – Influence of wind speed during departure on ATC delay.

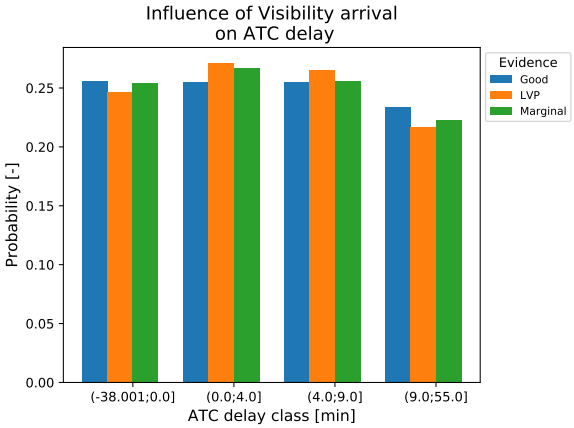


Figure E.15 – Influence of visibility during arrival on the ATC delay.

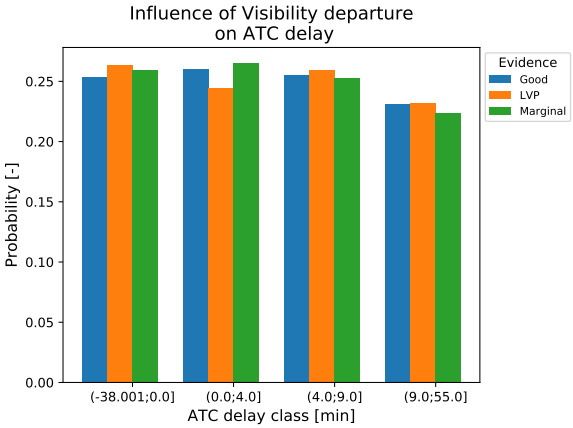


Figure E.16 – Influence of visibility during departure on the ATC delay.

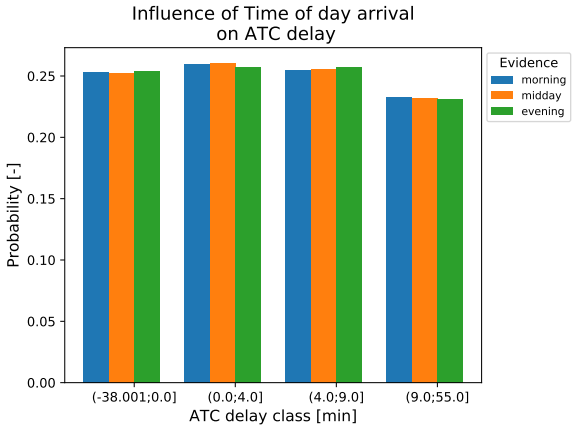


Figure E.17 – Influence of time of day during arrival on ATC delay.

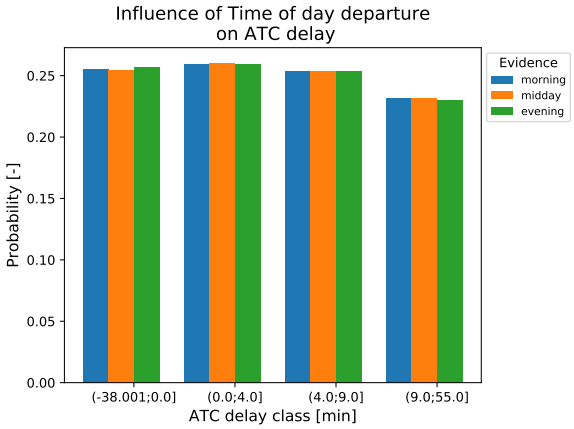


Figure E.18 – Influence of time of day during departure on ATC delay.

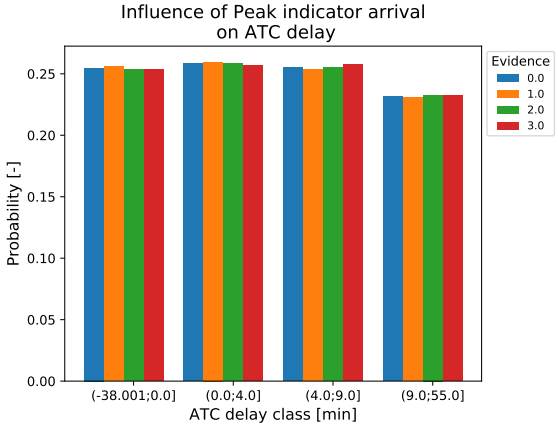


Figure E.19 – Influence of the peak indicator during arrival on ATC delay.

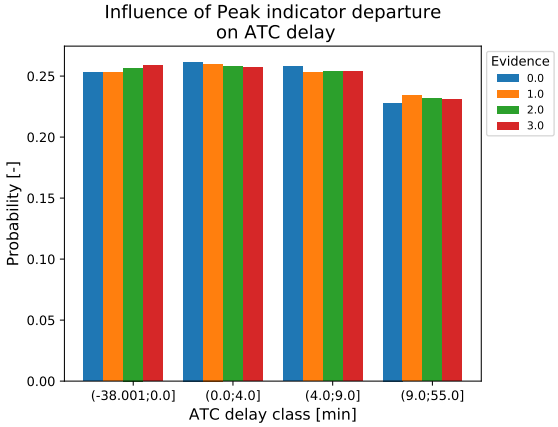


Figure E.20 – Influence of the peak indicator during departure on ATC delay.

III

PRELIMINARY REPORT [ALREADY GRADED]

1

INTRODUCTION

Civil air traffic has experienced exponential growth over the last years. The International Civil Aviation organisation, ICAO [2], reported that the sector has grown at an average rate of 5% per year since 1995, in contrast with an average growth of 2.8% for the world's economy. Due to this growth in air traffic, the system and operations have become more complex, resulting in busier airports and airspace, leading to more flight delays. Delays impact multiple stakeholders, such as the airline, its passengers and the airports in their flight network [3]. Furthermore do these delays have large financial and economic consequences for the involved stakeholders, especially the operating airline [4]. This is not just due to the direct costs of delays, but also because of the long term effects of delays on passenger loyalty, market share and airline revenue [5]. For all of these consequences, it is important to better understand delays, and possibly use this knowledge to further reduce them.

Air Traffic Control (ATC) is a critical element in the system of Air Traffic Management. Its main purpose is to ensure a safe and efficient flight, by keeping aircraft separated both vertically as longitudinally [6]. Air traffic control can be described as a sequential timeline or process that consists out of several control and decision points, which facilitates the necessary communication between the controller and aircraft in an efficient manner [6]. Unfortunately, Air Traffic Control is also a source of delay in civil aviation. ATC delays are initial delays, but they still have an enormous impact as they can trigger reactionary delays throughout the operations of an airline. This is due to aircraft, passenger and crew connectivity [7]. However, these initial ATC delays do not only impact the operations of the involved airline, but can also propagate to other airlines and airports by causing an imbalance between capacity and demand in the airspace [8–10]. ATC delays can thus result in very large impacts throughout the entire network, originating from a single flight delay. Therefore, delays are said to be the most challenging aspect in any transport system [3].

This document contains the literature review for a research project investigating the root causes of the ATC delays, with a case study on KLM Royal Dutch Airline flights at Amsterdam Airport Schiphol. ATC delay is one of the most encountered primary delays due to the increase in air traffic, and Amsterdam Airport Schiphol is even seen as the driver of Air Traffic Flow Management (ATFM) delay in the Eurocontrol Network [11]. Therefore, the main objective of this research has been formulated as follows.

> *The main research objective of this master thesis is to expose the drivers of the ATC delays encountered by KLM flights at Schiphol Airport, by performing a root cause analysis of these ATC delays and their impact on the KLM network.*

Next to the research objective, several research questions have been established:

1. Which data of the different stakeholders is relevant for the ATC process at Schiphol Airport and KLM flights?
2. What root cause analysis / causal analysis techniques are valuable in the context of the research?
3. What are the root causes of ATC delays for KLM flights at Schiphol airport?

The scope of this research project is limited both in time and space. This research will focus on the ATC delays in the Dutch airspace, and data incurred on the day of operation. Additionally, the perspective on delays is

not limited to single arrival or departure ATC delays, but will focus on a turnaround procedure. This means that the flights will be analysed from a cruise-to-cruise perspective, in comparison to the common gate-to-gate representation. This means that the inbound and outbound flight at Amsterdam Airport Schiphol (AMS) are connected in a single data sample.

This research work is novel as the causes of ATC delays have never been researched using a data driven approach. For the majority of the comparable research studies, the work focused solely on the drivers of the general departure and/or arrival flight delays. Additionally, the root cause analysis of pure departure/arrival delay has almost never been combined with a focus on just one airline. This perspective makes it possible to take into account the specific processes and airline data in the analysis. Hence, this research is novel as it aims to combine the root cause analysis of specific Air Traffic Control delays together with a focus on a single airline. To determine these causes, a state-of-the-art method will be used in combination with a baseline model. This approach allows to assess the added value of the state-of-the-art method. Additionally, the results of this research will reveal the processes and situations that lead to ATC delays for KLM flights at their hub airport. By obtaining this information, steps can be taken to improve this, which can potentially reduce the ATC delays and the high costs related to them.

The remainder of this literature review is structured as follows. In chapter 2, the current Air Traffic Management system is reviewed. This is followed by an analysis of the occurrence of delays in air transport networks, presented in chapter 3. In chapter 4, an extensive description of the available causal models is presented, together with a review of their applicability for this research project. Subsequently, the required data, sources and available processing methods are discussed in chapter 5. Finally, a work flow diagram illustrating the research approach and work packages is presented in chapter 6, which is followed by the conclusions in chapter 7.

2

AIR TRAFFIC CONTROL

This chapter introduces the concept of Air Traffic Control (ATC). Air Traffic Control is actually part of the Air Traffic Management system. Additionally, The determinants and influencing factors of the different capacities in the airspace are discussed.

2.1. AIR TRAFFIC MANAGEMENT

ATC is actually a subdivision of Air Traffic Management (ATM), which is introduced in this section. First of all, the division and layout of the airspace is presented, followed by an explanation of the flight plan and planned trajectory of a flight. Additionally, the concepts of regulations and Airport Collaborative Decision Making are discussed.

2.1.1. AIRSPACE DIVISION

Airspace is subdivided into different control zones. The division of airspace in the Amsterdam Flight Information region (FIR) can be seen in Figure 2.1 [12]. A Flight Information Region can be seen as the largest division of airspace, which is again subdivided into different airspace zones. The FIR often contains the entire airspace of a country for smaller countries, such as The Netherlands. For larger countries, their airspace can be divided into several FIRs. In general, a FIR can be subdivided into four different areas or zones of airspace, and each are controlled by other air traffic controllers and centres. These division areas within a FIR are discussed below in more detail and are illustrated in Figure 2.1 [12].

Control Zone (CTR)

This zone in the airspace is the airspace directly surrounding the airport. This area is controlled by Tower Control (TWR), which controls the aircraft on the ground and on the runways as well.

Terminal Manoeuvring Area (TMA)

This part of the airspace is managed by approach control (APP). The TMA connects the Initial Approach Fixes (IAF) to the final approach in the Control Area or CTR, but also connects departing flights from the CTR to the airways in higher parts of the airspace. This area is thus characterised by climbing and descending aircraft.

Control Area (CTA)

The CTA, or Control Area, is managed by the Area Control Centre (ACC). The CTA is bound at an upper flight level of 24,500 feet, as seen in Figure 2.1 [12]. The control area contains the Initial Approach Fixes and thus also the holding areas or stacks.

Stacks

Stacks are areas in the airspace where aircraft are held before they are allowed to make their final approach to the runway via the Initial Approach Fix (IAF). These areas can be seen as the 'waiting areas', before receiving clearance from approach control to enter the TMA for final approach.

Upper Airspace (UTA)

The upper airspace is defined to be above an altitude of 24,500 foot or FL245. This area is controlled by upper area control, which is Maastricht Upper Area Control (MUAC) for the Amsterdam FIR. In this part of the airspace, aircraft are in their en-route phase and follow fixed airways.

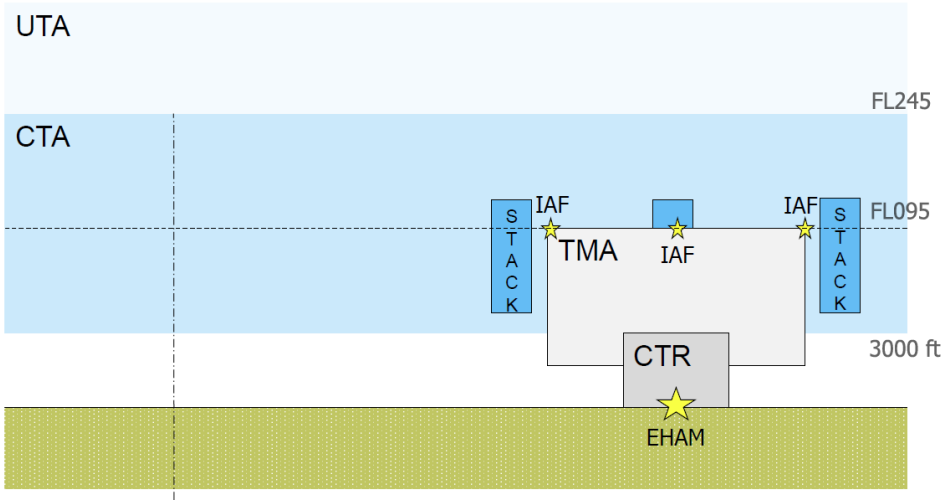


Figure 2.1 – A schematic overview of the airspace in the Amsterdam FIR [12].

In Figure 2.2 the Control Area sectors in the Dutch airspace are plotted. This information has been found from the Aeronautical Information Package (AIP) for the Netherlands, provided by LVNL, the Air Navigational Service Provider (ANSP) in the Dutch Airspace [13]. As can be seen, the areas in the North and South-East of The Netherlands are not part of any control area. This is because this airspace is reserved for military use. In general, the layout of these sectors is dynamic, and dependent on the demand at that moment in time [14]. When demand is high, the CTA is split in the highest possible number of sectors, such that the offered capacity is maximised. However, when demand is below capacity, sectors can be merged such that less controllers are necessary.

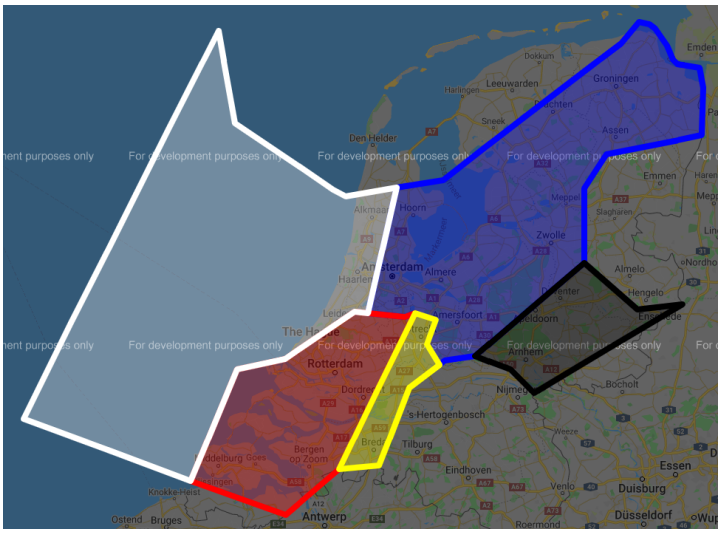


Figure 2.2 – The different sectors in the Amsterdam CTA according to the AIP.

In order to regulate the incoming and outbound traffic in the TMA and CTA, Standard Instrument Departures, SIDs, and Standard Arrival Routes, STARs, are present in the airspace. SIDs and STARs are fixed routes which connect the departure runway to the en-route airway,and the en-route airway to the initial approach fix respectively. As mentioned before, the STARs are designed to connect aircraft from their en-route airway to the initial approach fix and final approach. For Amsterdam Airport Schiphol, three Initial Approach fixes are present [13]:

- ARTIP

- SUGOL
- RIVER

Each of these Initial Approach Fixes have corresponding waiting areas, also better known as stacks, which have been discussed previously.

2.1.2. TRAJECTORY & PLANNING

For each flight, the planned trajectory has to be made and described in a flight plan, which is done 6 to 3 hours before take-off [15]. In ATM, there are different flight trajectories that can be distinguished [16]:

- User preferred trajectory
- Procedure optimal trajectory
- Network optimised trajectory
- Actual flown trajectory

The user preferred trajectory can be seen as the path the aircraft would fly in case of a free route airspace. In practise, this would mean that all flights would be executed at minimal costs by taking the shortest/fastest path to their destination, also known better as free flight. However, the airspace consists out of navigation points, and an aircraft should follow a path connected by these nodes [14]. Therefore, a flight's flight plan consist out of a sequence of navigation points and a specified altitude, which can also be seen as the procedure optimised trajectory [14, 16]. The network optimised trajectory is the flight plan with the incorporated feedback of the network manager, in order to control the capacity in the airspace. Finally, the actual flown trajectory is the actual flight path taken.

Based on the aforementioned definitions of the trajectories, three different stages of flight efficiency that have been defined by Bronsvoort *et al.* [16]:

- Strategic efficiency
- Pre-tactical efficiency
- Tactical efficiency

The tactical efficiency is defined as the difference between the actual flown trajectory and the flight plan route optimised for the airspace situation, also known as the network-optimised trajectory. The pre-tactical and strategic efficiency are then the difference between the network optimised and flight plan trajectory and the user preferred trajectory. The pre-tactical and tactical efficiency can be seen as measures relating to the impact of demand and capacity imbalances before departure and en-route respectively, because of ground holdings and en-route ATC interventions [16].

2.1.3. REGULATIONS

Regulations is a mechanism that is managed by the Central Flow Management Unit (CFMU) and is part of Air Traffic Flow Control Management system [17]. Regulations are installed for flights which are likely to experience en-route congestion end thus airborne delay. In order to avoid this, these flights are regulated and kept on the ground to reduce the airspace congestion and resulting delays [17].

When a flight is regulated, it will receive a Calculated Take-off Time (CTOT), which is issued by the CFMU. The aim of using the calculated take-off time for each of the regulated flights, is to manage the number of aircraft using the regulated airspace at a specific time, and thus ensuring that capacity is not exceeded [15]. This CTOT allocation is based on the filled flight plan, which entails the planned or predicted trajectory of the aircraft and assigns a CTOT based on a 'First Planned First Served' principle [18]. These slots are calculated using CASA, the Computer Assisted Slot Allocation.

In a study by Ruiz *et al.* [18], it has been shown that this principle of allocating Air Traffic Flow Management (ATFM) slots does not take into account the complex network that air transport is, and thus does not provide the overall most efficient solution. By regulating flights, slots of other flights are also impacted, and this can become problematic when multiple regulations are active for a flight. The regulation which has the biggest impact on the flight's trajectory is also known as the most penalising regulation, which can have a positive or negative impact on other flights via interactions due to multiple regulations along the flight's flight path [18].

Regulations are used with an aim to make use of the airspace's capacity as efficient as possible, however, some of the taken actions to manage ATFM is in contradiction with what is seen efficiency by airports and airlines. For example, flights who receive a regulation induced delay might speed up during the flight to absorb this delay, and thus their initial filed flight plan does not accurately represent their trajectory anymore [15]. This unpredictable behaviour can potentially lead to an incident of exceeding the capacity, which can be followed by time period of capacity underusage.

The phenomena of unanticipated peaks in arrival of air traffic in congested or regulated areas of the airspace is also known as 'air traffic bunching' [15]. The bunching peaks experienced are in fact a result of 'operational behaviour', which leads to operational noise, and thus can deviate from what was initially planned by the CFMU. Examples of operational behaviour and noise are for example airline and airport practises to absorb the regulated ground delay [15]. An illustration of the regulation process by the CFMU and how this can lead to unexpected peaks and lows in the actual operations is shown in Figure 2.3 [15]. Stolz and Ky [15] state that in practise, this risk leads to an underdeclaration of capacity by the network controllers and ANSPs, which is again inefficient and leads to more ground holding ATFM delays than what would be necessary in theory.

Stolz and Ky [15] also acknowledges the effect of hub operations on regulations, and therefore also on air traffic bunching. The operational model of a hub is actually in contrast to the idea of Air Traffic Flow Management, which aims to spread out the demand evenly such that the airspace capacity is not exceeded. However, hubs are characterised by their departure/arrival banks, which are time blocks with a high number of inbound/outbound flights, in order to be able to make as many connections between flights as possible.

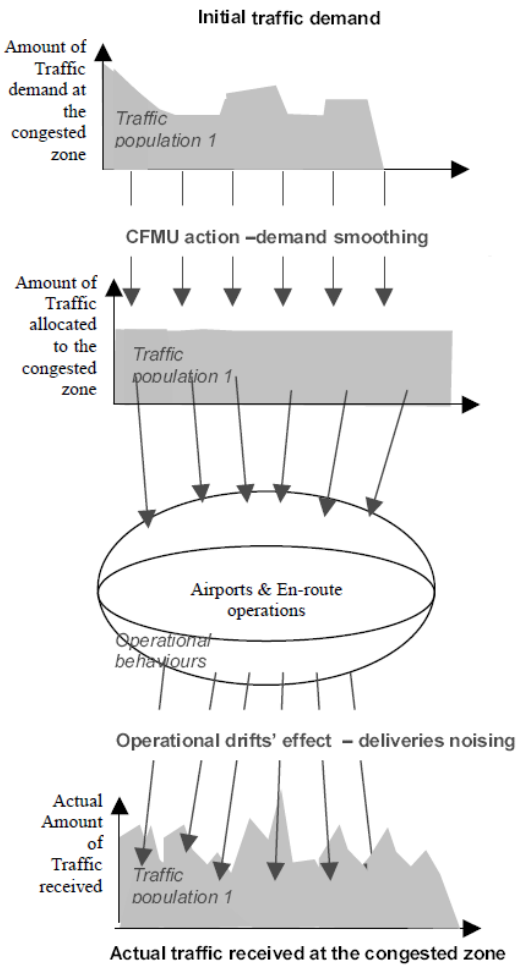


Figure 2.3 – Schematic overview of an air traffic regulations and its possible bunching effects [15].

However, keeping flights at the ground is not guaranteed to be the most optimal solution, as this can also cause congestion at the airport of departure [19]. Additionally, the ground holding programs can not take into account developments on the airspace or arrival airport capacity when assigned the regulation, and can therefore be found to be unnecessary. Therefore, Carlier *et al.* [19] claimed that the ground delay resulting from regulations is not always more efficient than en-route delay.

2.1.4. AIRPORT COLLABORATIVE DECISION MAKING

Flight delay is a manifestation of the interactions between various stakeholders involved in the flight process, such as the airline, airport operator, the slot coordinators and the air navigational service providers [20]. Therefore, Airport Collaborative Decision Making, A-CDM, has been introduced. This is a concept in which multiple stakeholders and operators at an airport share information on the operational processes and thus enhance informed decision making for all parties [21]. The following parties are involved in the information sharing process [22]:

- Airport operations
- Airlines operations

- Ground handling
- Network controller (Eurocontrol in EU)
- Air Traffic Control

By implementing Airport Collaborative Decision Making, there are a number of common goals for these involved stakeholders. The most important ones related to airline punctuality and ATC delays are as follows [22]:

1. Improve the predictability
2. Improve on-time performance
3. Reduce the amount of ATFM slot missed and wasted
4. More flexible planning of the pre-departure process
5. Reduce apron and taxiway congestion
6. Optimise the use of airport infrastructure and resources such as runways, gates, ramps and parking spots.

A-CDM or simply CDM, is a system that works with a milestone approach. These milestones are characterised by different events in the turnaround process. In this approach, the different parties or stakeholders are each responsible for several milestones, of which the updates are shared with all stakeholders, such that the flight process information is coherent and complete for all [22]. A graphical and sequential illustration of the milestones in the turnaround process used in the A-CDM process is displayed in Figure 2.4 [22].

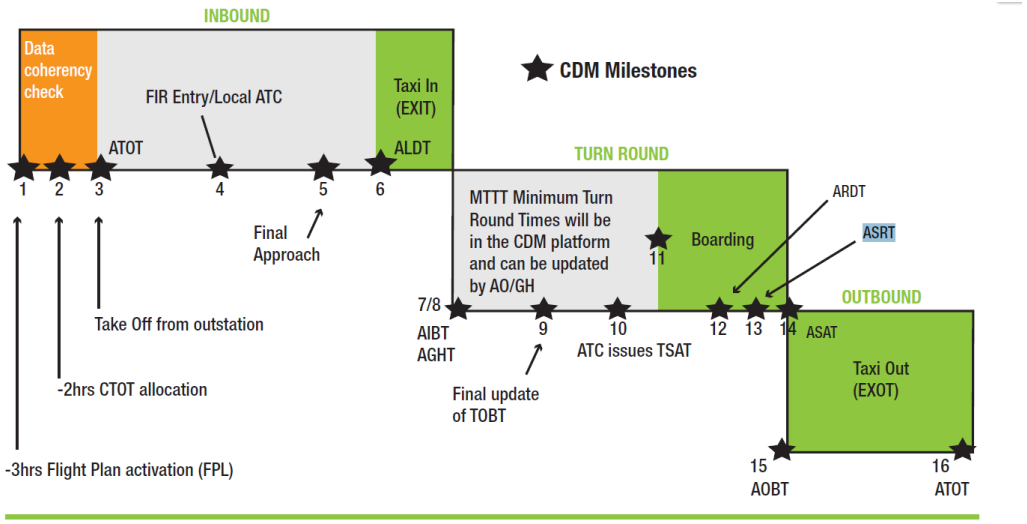


Figure 2.4 – Overview of flight phases & CDM milestones [22].

As one of the goals of CDM is to increase the flexibility and predictability of the turnaround and departure process, estimations have to be made for different milestones. Equation 2.1 shows how the Target Off Block Time (TOBT) is estimated before the aircraft has arrived at the airport [21]. The Minimum Turn Around Time (MTT), is issued by the Main Ground Handling Agent (MGHA) of the respective flight.

$$\text{TOBT} = \text{EIBT} + \text{MTT} \tag{2.1}$$

After the aircraft is in-block, the Main Ground Handling Agent (MGHA) is responsible for updating the TOBT. The used definition at Amsterdam Airport is that at the moment of TOBT, all ground handling activities have been finished, all doors are closed and all boarding equipment have been removed from the aircraft, and thus is the MGHA the stakeholder which can define this milestone.

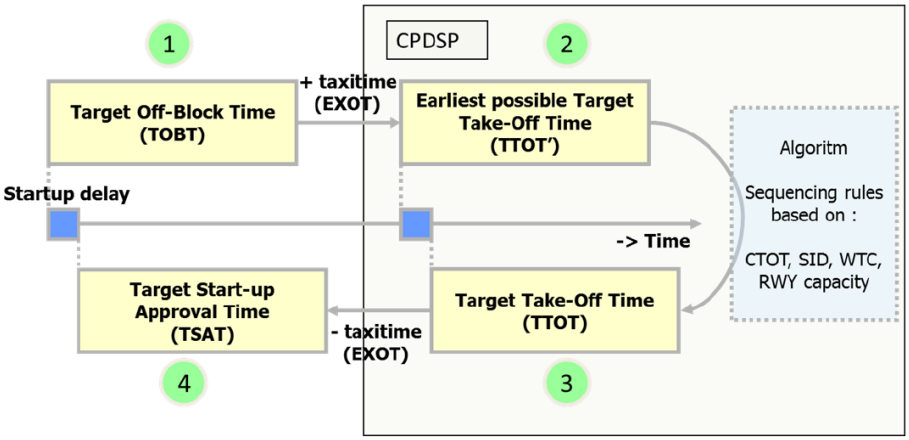


Figure 2.5 – Determination of TSAT in CDM pre-departure planning [21].

During the turnaround process, the Target Start-up Approval Time (TSAT) and the Target Take-off Time (TTOT) are supplied by Air Traffic Control [21]. The TSAT is a function of the issued TOBT, which is translated into a Target Take-Off Time, TTOT, as seen in Equation 2.2. The TTOT is established by assigning the earliest possible take-off time, taking into account the estimated taxi-out time (EXOT). The TSAT is then found by reversing the calculation as seen in Equation 2.3 [21]. The TSAT is thus subject to change, as it depends on TOBT, TTOT, or CTOT for regulated flight, and EXOT changes and runway usage and/or capacity. The interdependencies between these milestones and how they are used to update each other are illustrated in Figure 2.5 [21].

$$\text{TTOT} > \text{TOBT} + \text{EXOT} \tag{2.2}$$

$$\text{TSAT} = \text{TTOT} - \text{EXOT} \tag{2.3}$$

Also the Actual Start-up Request Time (ASRT) is issued by ATC, which is done at the moment the cockpit has declared it is ready, on the condition that this happens in the flight's TSAT [21]. The TSAT window is defined to be -/+ 5 minutes from the declared TSAT. The CDM process comes thus with certain responsibilities for the involved parties. ATC is responsible for delivering a TSAT based on the calculated TOBT or CTOT to the pilot. It is the pilots responsibility to then declare that the aircraft is ready within the received TSAT window [21].

Additional to the milestones, the CDM system also registers changes in the flight states of a flight at the airport. These flight states and their link with the different flight phases and CDM milestones are depicted in Figure 2.6 [21]. It can be seen that for the different flight phases, inbound, turnaround and outbound, several flight states are defined, which are shown and explained in more detail in Table 2.1 [21]. One should note that these flight states have been defined particularly for the implementation of A-CDM at Amsterdam Airport Schiphol, and might not be applicable to other airports.

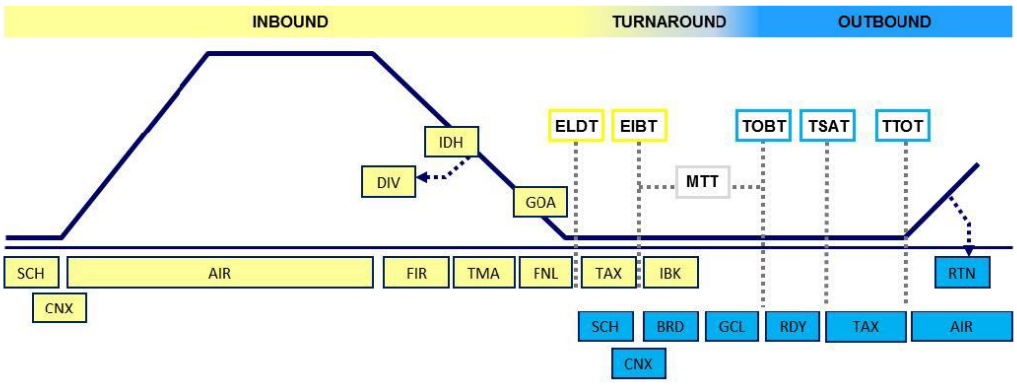


Figure 2.6 – Flight states in CDM system at Amsterdam Airport [21].

Table 2.1 – Flight States per flight phase for CDM@AMS [21].

Inbound	Turnaround	Outbound
SCH = Flight scheduled	SCH = Flight scheduled	RDY = Flight ready
CNX = Flight cancelled	CNX = Flight cancelled	TAX = Flight taxiing
AIR = Flight airborne	BRD = Flight boarding	AIR = Flight airborne
FIR = Flight airborne in Dutch airspace	GCL = Gate closing	RTN = Flight returning
TMA = Flight in approach		
FNL = Flight on final		
TAX = Flight taxiing		
IBK = Flight in-blocks		
IDH = Flight in indefinite holding, unable to continue approach		
DIV = Flight diverting		
GOA = Flight go-around		

2.2. CAPACITY

In the previous section, Air Traffic Management has been introduced, in which capacity is a crucial part. In this section, the different capacities that are present in the airspace are introduced and discussed. In general each part of the airspace is assigned a nominal capacity, however, these can vary dynamically based on the weather and other operational conditions [17].

2.2.1. AIRPORTS

The capacity for airports is often expressed in the number of arrivals and/or departures per hour [23]. The capacity of an airport is constrained by the use of the runways and availability of ramps and gates. Additionally, the airport's capacity is determined under a set of operating conditions [23]:

- Weather conditions: ceiling & visibility
- Aircraft mix
- Air Traffic Control capacity
- Nature of operations

Runway capacity and airport capacity are not equal to each other, as runway capacity is part of the total airport capacity [22]. In general, the runway capacity is the most limiting factor in terms of airport capacity. Again, the runway capacity can be expressed as a function of the following elements [23]:

- Air Traffic Control capacity
- Environment
- Demand
- Runway configuration

Air Traffic Control capacity will be treated in the upcoming subsection. Demand is an influencing factor as the types of aircraft requesting to use the runway have an influence on the capacity, and is actually equal to the airport capacity constraint 'aircraft mix'. As the wing tip vortices strength and speed vary with the size of the aircraft, different separation minima exist for consecutive operations of aircraft of different sizes. This is not only due to the wake turbulence, but also due to different runway occupation times due to varying speeds. For Amsterdam airport, it is reported by Eurocontrol that the mix of aircraft is actually the most constraining factor for the runway throughput [24].

The environment constraint contains factors such as wind conditions, visibility and ceiling, but also noise abatement measures, which can limit certain runway combinations or limit runway usage during specific hours of the day.

Lastly, also the runway configuration at the airport determines the capacity. Here, the number of runway crossings and parallel/intersecting runway operations are taken into account. Parallel operations offer the highest capacity [25]. However, due to the presence of (cross)wind, and to optimise taxi times, intersecting runway operations can be beneficial. However, due to the associated risks of intersecting runway operations, this reduces the offered capacity. The runway configuration at Amsterdam Airport is shown in Figure 2.7 [26].

Amsterdam Airport Schiphol is, like most hub airports in Europe, a slot coordinated airport [27]. The slot coordination is handled by an external party, the Airport Coordination the Netherlands [28]. Amsterdam

Figure 2.7 – Runway layout at Amsterdam Airport [26].

Airport Schiphol has been completely saturated for the last years. The airport's slots have been limited at a yearly amount of 500,000, which was reached in 2017.

As mentioned before, AMS is the major hub airport for KLM Royal Dutch airlines. This means that the airport is characterised by inbound and outbound banks, in order to ensure as many connection possibilities as possible. For illustrative purposes, the network model of Air France-KLM is illustrated in Figure 2.8 [29]. As mentioned before is the hub and spoke model inherently in contrast with the demand capacity balance, as there are large peaks of inbound and outbound traffic during small time periods [15]. Additionally, AMS was one of the driving airports of ATFM delay in Europe during 2019, according to Eurocontrol [11].

Figure 2.8 – Hub and spoke network model of the Air France-KLM group[29].

During the inbound and outbound peaks, the runway configurations change, and thus influence the capacity of the airport. During the inbound peaks, two landing and one take-off runways are active, whereas during the outbound peak 2 runways are used for departure and one for landing [30]. In transition phases between the in- and outbound peaks are sometimes four runways at the same time in usage [30].

2.2.2. AIR TRAFFIC CONTROL

In Air Traffic Control, there are a number of control centres that can be distinguished, such as TWR, APP, ACC and UAC, as discussed in the previous section. Each of these control centres control a different part of the airspace. Therefore, in this section, a distinction is made between the capacity managed by the ANSP, LVNL in the Netherlands, and the Upper Area Controller or Network Controller, Eurocontrol.

AIR NAVIGATIONAL SERVICE PROVIDER

Air Traffic Control capacity determines the capacity of the runway throughput as they ensure that all flights are kept at a safe distance from each other. This is translated into minimum separation minima, which are again dependent on a number of factors (aircraft size, radar availability, operational sequence, and the time spent on the runway) [23]. Next to the separation minima, there are other factors related to ATC determining the capacity, such as the level of technology of the ATC system, the strategy used for sequencing and the length of the common path of the Instrument Landing System [23].

Aircraft type influences the runway occupation time, and thus do slower aircraft reduce the capacity of the runway. Additionally, the wake turbulence induced by aircraft influence the separation minima between different aircraft types, and thus also capacity [23].

The Initial Approach Fixes and the corresponding stacks also have a capacity. The nominal capacities are expressed in aircraft per hour, and are the following: ARTIP:30, RIVER:21 & SUGOL:26. These capacities are dynamically managed by imposing regulations on them. The regulations of the stacks are indirectly also related to the arrival rate at AMS, as these regulations are installed to reduce the number of holdings at the stacks when the demand is higher than the runway capacity.

Additionally, Air Traffic Control actions can not always be predictable or consistent [31]. The consistency of air traffic control decisions to install ground delays has been studied by Kulkarni *et al.* [31]. In this research, it was found that on days where the weather forecast, the Terminal Aerodrome Forecast (TAF), was accurate, the decision consistency for installing ground delay programs had higher consistency than for days where the TAF was inaccurate.

NETWORK CONTROLLER

The Network controller and Upper Area Controller are for the Dutch Airspace the same organisation, Eurocontrol. The upper airspace in the Netherlands and its capacity is managed by Maastricht Upper Area Control (MUAC). As this research project focuses on flights arriving and departing from Schiphol Airport, the upper area airspace will not to subject to analysis.

Eurocontrol is also in charge of the Air Traffic Flow Management, and is thus also consists out of the Central Flow Management Unit. As mentioned before is the CFMU responsible for imposing regulations [17]. The main task of this unit is to manage the imbalance between the demand and capacity, at strategic, pre-tactical and tactical levels in the operation.

3

DELAYS IN AIR TRANSPORT NETWORKS

This chapter discusses the different types of delays that can occur in air transportation. Additionally, the delay dynamics and potential impacts are discussed, as well as the strategies that have been implemented to avoid, manage and mitigate these delays.

3.1. TYPES OF DELAYS

In this section, the different types of delays which can be encountered in air transportation are discussed. The most general definitions of delay are arrival and departure delay. However, there are multiple more specific types of delays, which are part of the general arrival or departure delays, such as reactionary delays, en-route delays and Air Traffic Control (ATC) or Air Traffic Flow Management (ATFM) delays.

3.1.1. TOP LEVEL DELAY TYPES

The most general types of delays are related to when they occur in the flight process, such as departure, arrival and turnaround delay. These top level delay types however do not address the reason of the delay, such as the other delay types discussed further in this section.

Departure delay can be easily defined as the difference between the actual time of departure and the scheduled time of departure. These terms are interchangeable with the actual off block time (AOBT) and the scheduled off block time (SOBT), shown in Equation 3.1.

Arrival delay can then be defined as the difference between the actual time of arrival and the scheduled time of arrival. These are also known as the actual in block time (AIBT) and the scheduled in block time (SIBT), which is formulated in Equation 3.2.

$$\text{Departure delay} = \text{AOBT} - \text{SOBT} \tag{3.1}$$

$$\text{Arrival delay} = \text{AIBT} - \text{SIBT} \tag{3.2}$$

In some studies, turnaround delay has been used as a measure of flight delay [32]. The turnaround delay of a flight can be defined as the additional delay encountered during the turnaround process of a flight, but the definition of what is actually defined as departure delay and turnaround delay remains rather unclear and hard to separate from each other. The definition of turnaround delay will not be further used in this research project and the work will just distinguish between arrival and departure delay.

3.1.2. REACTIONARY DELAY

Reactionary delay was the most common delay reason in both 2018 and 2019, causing 46.4% and 44.4% of the delays respectively [33]. Reactionary delay is a flight delay that is caused by the delay of previous flights. A graphical illustration of this kind of delay is presented in Figure 3.1 [34]. As can be seen from Figure 3.1, delay has a high probability to result in reactionary delays if multiple flight legs follow closely upon each other. Therefore, especially for flights in Europe, the first flight of the day has a high impact on the rest of the day, as there are no overnight flights and the night stop can thus absorb all accumulated delay of the previous day. Therefore, the dynamics of reactionary flight delay is different for Intercontinental (ICA) flights, as there are no night stopovers, longer flight times, and longer turnaround times in general.

3.1.3. AIRSPACE/EN-ROUTE DELAY

Airspace delay is the result of a number of Air Traffic Control (ATC) actions, with the primary aim of keeping all aircraft separated at a safe distance [35]. These actions consist of queuing, rerouting, holding, speed control

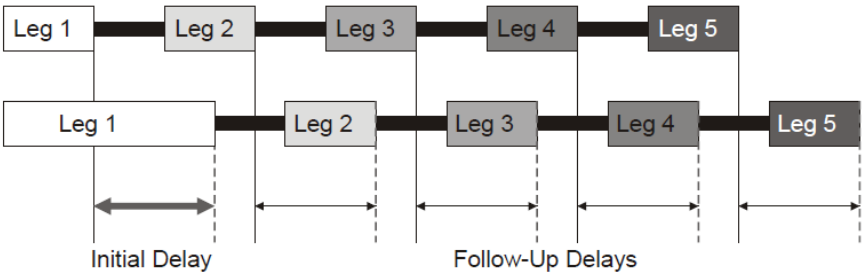


Figure 3.1 – Reactionary delays throughout a sequence of flight legs [34].

and aircraft vectoring [19, 35]. Additional to keeping the aircraft separated, also weather events can result in deviations, and possible resulting congestion, adding to the en-route delay.

Airborne delays are often a result of congested airspace. These mostly occur in the Terminal Manoeuvring Area of large, mostly hub, airports [36]. In these busy TMA's, the air traffic management consists out of two tasks, one being the separation of the aircraft, and the second the management of the air traffic flow through the TMA [36]. In order to efficiently make use of the capacity and therefore minimise airborne delays, Arrival Sequencing and Scheduling is used. If done effectively, it has been shown that this reduces the experienced airborne delays significantly [36].

3.1.4. ATC DELAY

Air Traffic Control (ATC) is a critical element in the system of Air Traffic Management. Its main purpose is to ensure a safe and efficient flight, by keeping aircraft separated both vertically as longitudinally [6]. Air traffic control can be described as a sequential timeline or process that consists out of several control and decision points, which facilitates the necessary communication between the controller and aircraft in an efficient manner [6]. Unfortunately, Air Traffic Control is also a source of delay in civil aviation. ATC delay consists out of two different types of delays, Air Traffic Flow Management delay (ATFM) and start-up delay.

ATC delay of a flight can be received at the gate, before departure, which is better known as start-up delay. In essence, there are four main possibilities for this delay, which are all a result of an imbalance between the demand and the capacity, shown below [20].

- Runway capacity
- Airspace capacity
- Availability of gates
- Arrival acceptance rate

However, ATC delays also include ATFM delays. ATFM delays are the result of regulations in the airspace, as discussed in chapter 2. ATFM delays are mostly allocated on the ground, but can also be experienced en-route, as discussed above. In 2019, 9.9% of the flights in the Eurocontrol area experienced en-route Air Traffic Flow Management (ATFM) delay, with an average of 1.57 minutes per flight [33]. The causes of these delays can mainly be attributed to the Air Traffic Control (ATC) capacity according to the Air Navigation Service Providers (ANSP) [33]. Other causes are said to be staffing, weather and disruptions or actions in the ATC system [33]. Overall, the percentage of delayed ATFM flights reduced slightly between 2018 and 2019, while the number of delays caused by ATC capacity restrictions actually increased by 6.6% [33].

At the base of ATFM delays lies the discrepancy between the offered capacity and the demand. In order to manage the demand at airports, slot coordination has been installed across European Airport [27], such that the demand is known at a strategic level in the operation. For en-route airspace, the demand is only known in the tactical phase of the operations, as the filed flight plans denote the intended route taken. The flight plan is typically filed between 6 and 3 hours before departure [15], which is a very short time frame for the Air Navigational Service Providers and Network Controllers to supply the needed capacity [37]. In order to match the demand and the offered capacity, regulations are imposed, meaning that flights are assigned ground delay and a CTOT slot. In a more mathematical way, this ground ATFM delay can be expressed as the difference in the Calculated Take-Off Time (CTOT), which is issued by the CFMU, and the initially Estimated Take-Off Time (ETOT), as shown in Equation 3.3 [38].

$$\text{ATFM delay} = \text{CTOT} - \text{ETOT} \tag{3.3}$$

These ground delays are installed as airborne congestion is unsafe due to increased controller workload. Additionally, ground delay is also environmentally friendlier than en-route delay, as investigated by Carlier *et al.* [19]. However, the total minutes of ground delay due to ATFM restrictions was estimated to be higher than the delay minutes in case of airborne holdings and rerouting [19]. Although ground delays are more cost efficient in terms of fuel and emissions, the increase in delay minutes due to ground holdings results in the fact that ground delays are actually more expensive than the en-route or airborne manoeuvres, according to Lehouillier *et al.* [17].

3.1.5. IATA DELAY CODES

IATA, the International Air Transport Association, has established delay codes and descriptions, such that the same codes are used among all airports and airlines. Table 3.1 presents the codes related to ATFM restrictions (81-84), airport facilities (87-89), and reactionary delays (91-96), and are the ones deemed to be most relevant in the context of this research. The use of these delay codes are useful as they allow to trace the delay reason more in-depth than using the types of delays discussed earlier in this section.

Table 3.1 – IATA delay codes & reasons [39].

Delay code	Description	Delay code	Description
81	ATFM DUE TO ATC EN-ROUTE DEMAND / CAPACITY	83	ATFM DUE TO RESTRICTION AT DESTINATION AIRPORT
82	ATFM DUE TO ATC STAFF / EQUIPMENT ENROUTE	84	ATFM DUE TO WEATHER AT DESTINATION
87	AIRPORT FACILITIES	88	RESTRICTIONS AT DESTINATION AIRPORT
89	RESTRICTIONS AT AIRPORT OF DEPARTURE		
91	LOAD CONNECTION	92	THROUGH CHECK-IN ERROR
93	AIRCRAFT ROTATION	94	CABIN CREW ROTATION
95	CREW ROTATION	96	OPERATIONS CONTROL

3.2. DELAY PROPAGATION

As discussed in the previous section, a flight delay can cause subsequent flights to also be delayed, which is generally denoted as reactionary delay. This phenomena is better known as delay propagation. Flight delay propagates through the operation of the airline due to aircraft, crew and passenger connectivity [7]. If a flight arrives late, it is evident that the subsequent flight with that same aircraft will have a high probability of also being delayed, and the same principle holds for the flight scheduled with the late arrived cabin/cockpit crew. Lastly, subsequent flights can also be delayed due to the large amount of connecting passengers from the late flight. This is especially the case for hub airports of airlines, where the operation is fit to ensure a large possibility of flight connections. Additionally, due to legal crew and aircraft restrictions, such as minimum rest, duty times and maintenance intervals, initial delays can have large implications on subsequent flights, as valuable resources become unavailable for operation. Additional to the flight delay propagation in the airline's operation, the delay can also propagate to other airports and airlines, as the demand capacity balance in the airspace can be shifted and disturbed [40].

In order to model and better understand the flight delay propagation across different flights and airports, several approaches have been used. One of these methods is the usage of queuing models [9, 41, 42]. Additionally, agent based models have received an increased attention as well [10], as well as dynamic analysis models [43]. Lastly, Bayesian networks have been widely studied for the application of delay propagation, as it can represent network structures and links between the nodes. The propagation of delays across different airports has been modelled by Xu *et al.* [8], whereas the propagation effects within one airline's operation has been studied by Wu and Law [7]. A comparable but simpler approach than Bayesian networks has been taken by AhmadBeygi *et al.* [44], where propagation trees have been used to model the propagation of delay

throughout flight legs due to aircraft and crew connectivity. An illustration of delay propagation within the operations of a single airline is given in Figure 3.2 [44].

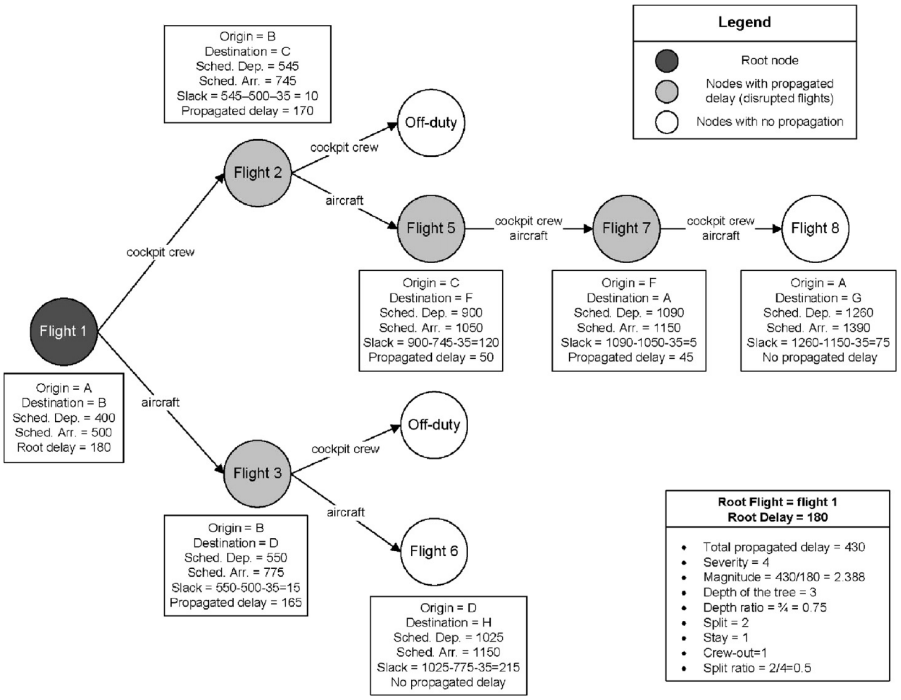


Figure 3.2 – Illustration of flight delay propagation across flights of a single airline [44].

3.3. DELAY MANAGEMENT

As delays often occur in air traffic and can have large implications on the entire operation, airlines have delay management and mitigation strategies. These strategies can be planned, meaning that they are taken into account in the strategical or tactical phase of the operation. Examples of these are as follows [45].

- Reserve crews
- Turnaround time buffers
- Flight time buffers

Additionally, airlines also have strategies that can be executed on the day of operation in case of disruptions. These include the following.

- Aircraft swaps.
- Slot swaps on coordinated airports.

The actions taken to manage the delays and disruptions in the flight network are not taken randomly, but the different components are handled and recovered in a sequential manner, due to the influencing factors between these components: aircraft → crew → ground operations → passengers [46].

In the research area of delay management and mitigation, linear programming models have been mostly applied. Bolić *et al.* [37] have worked on a linear programming model, which can manage and prevent delays on a strategical level, meaning that the planning is done several months before the operation. The delay management is achieved by the redistribution of flights, such that the capacities at the airports and airspace are not exceeded. Another linear programming model was developed by Santos *et al.* [47], which aims at optimising the operational decision making process of delaying flights at a hub or not, while minimising the incurred costs. Lulli and Odoni [48] recognised that in Europe, Air Traffic Flow Management is a complex problem due to the capacity constraints that are both present at the airports as in en-route airspace, and has aimed to reduce the delays by using a linear programming model as well. This model aims to optimise the cost of delay on a strategic level, and achieves this by assigning both ground and airborne delays to flights.

4

CAUSAL ANALYSIS MODELS

Traditionally, causal relations are discovered and identified by performing experiments [49, 50]. However, performing experiments to find causal relations is not possible for every process or application. Experiments are often expensive, time-consuming or not feasible [50]. Therefore, interest has increased in methods that can find causal relations from pure observational data, eliminating the need for practical experiments [49]. Therefore, in this chapter, different causal models are presented and discussed.

4.1. STATISTICAL METHODS

This section discusses different statistical methods which have been used in the past to identify influential factors or causal relationships between the independent and dependent variables.

4.1.1. CORRELATION COEFFICIENTS

Correlation is not equal to causation [51]. However, when there is a sequential trend in the data, correlation between variables can indicate a causal relationship as well. There are several correlation coefficients that can be used, described below [52].

Pearson correlation This is a coefficient which denotes the linear relationship between variables, with a range between +1 and -1. +1 is a perfect linear relationship, 0 is no relationship and -1 denotes a perfect negative linear relationship between the variables [52]. Its formula is shown below in Equation 4.1 [52]. One should be noted that the main assumption for this coefficient is that the variables should be normally distributed, and thus cannot be used on all variables.

Spearman correlation This coefficient denotes the level of association between two variables, eliminating the linearity from the Pearson coefficient. It is a non-parametric statistical test, such that there are no requirements on the data distribution, in contrast with the Pearson coefficient [52]. The formula is shown below in Equation 4.2, where R and S are the ranked values of x and y [52].

$$r_p = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\left(y_i-\bar{y}\right)}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}} \tag{4.1}$$

$$r_s = 1-\frac{6\sum_{i=1}^{n}D_i^2}{n^3-n}, \quad \text{where} \quad D_i = R_i - S_i \tag{4.2}$$

These coefficients have been used in research studies with the aim of finding causal relationships among variables for various applications [3, 53, 54]. In the context of flight delay, Sternberg *et al.* [3] used the Pearson coefficient to initially investigate the correlation of the attributes with flight delay. This however did not lead to any strong correlations, whereas other methods did find strong influential factors. Therefore, these coefficients can be used for initial data exploration and processing, but are less adequate to be used for a full causal analysis in a big data context.

4.1.2. REGRESSION ANALYSIS

Regression analysis is another statistical method that has been widely applied in the research field in order to better understand the influences on flight delays. Regression can be used to predict the target value, but also to find and quantify the influence of the independent variable(s) on the value of the target variable [52]. The formulas of bivariate and multiple regression are shown in Equation 4.3 and 4.4 [52]. In bivariate regression, there is only 1 independent variable, whereas for multiple regression, m attributes are used.

$$y_i = a + bx_i + \epsilon_i \tag{4.3}$$

$$y_i = a + \sum_{i=1}^{m} bx_i + \epsilon_i \tag{4.4}$$

Initiated by Mayer and Sinai [55], regression analysis has been applied to the problem of airline delays, which has been built upon by Santos and Robin [27] and Aydemir *et al.* [56]. In these research papers, the regression analysis focuses on the economic drivers of air traffic delay, and therefore uses independent variables such as market concentration, slot coordination, and the presence of a hub airline/airport.

Additionally, Liu *et al.* [57] used linear regression and the Multinominal Logit model to estimate the causes of en-route flight inefficiency. Additional to linear regression, an attempt has been made by Mohammadian *et al.* [58] to capture non linear relations between flight delay and independent variables, by making use of log linearity. In another study by Abdel-Aty *et al.* [59], a (mathematical) frequency analysis to discover patterns in flight delay was performed. Next to this, statistical regression analysis was used, aiming to find the influencing factors of the detected delay patterns. Initially, simple linear and logarithmic regression were used, but these led to poor results. Therefore, Abdel-Aty *et al.* used logistic regression and the analysis of variance (ANOVA) method, which led to a better model fit.

In essence, the influence of independent parameters or variables can be captured by performing a regression analysis [60]. However, Yu *et al.* [60] acknowledged that for high-dimensional data, regression analysis cannot capture the influence of certain independent variables on the dependent variable, as this correlation or causal relation is often indirect, unclear, or very complex.

4.1.3. GRANGER CAUSALITY

A technique often used in past research is the Granger causality statistical test [61, 62]. This test takes two time series as input, and determines whether there is a causal relationship between the two inputs. This statistical test and its application have two main ideas [63]:

1. The cause of an effect occurs earlier than the effect.

2. The information relating to the cause improves the ability to predict the behaviour of the effect, compared to using all available past information to predict the effect.

Additional to the above mentioned principles, it should be noted that Granger causality is limited to detect linear relationships between the two time series [45, 63]. Additionally, the metric analyses entire time series. This has the implication that if a causal effect is only present in a smaller time window, this effect will be averaged over the entire time window and thus will result in less significant results [63].

Equation 4.5 [63], can be used to asses whether Y_j causes Y_i according to the Granger test. $\sigma^2(Y_i \mid U^-)$ denotes the variance of the residuals when all information U is available. $\sigma^2\left(Y_i \mid U^- \setminus Y_j^-\right)$ then denotes again the variance of the residuals, but now the information of Y_j is excluded from U. If the expression in Equation 4.5 holds, this means that the values of Y_i can thus better be predicted when the information of Y_j is taken into account, and thus it can be said that Y_j Granger causes Y_i [63]. One of the reasons this causality test has been widely applied in past research, is due to its ability to distinguish causal relations from simple correlations between variables [63].

$$\sigma^2(Y_i \mid U^-) < \sigma^2\left(Y_i \mid U^- \setminus Y_j^-\right) \tag{4.5}$$

Du *et al.* [64] have used an alternative mathematical and statistical approach than Belkoura and Zanin. Equation 4.6 and Equation 4.7 defined by Du *et al.* [64], show the regression methods used to determine the residuals, which are used in Equation 4.5. The null hypothesis of the test is that Y_j does not cause Y_i, and therefore, its coefficients in Equation 4.6, $b^1, b^2, ..., b^{p_{ij}}$ should all be zero [64]. It should be noted that variable p_{ij} denotes the lag taken into account, as the cause should always occur earlier in time than the effect.

$$y_i^T = \sum_{m=1}^{p_{ij}} a^m y_i^{T-m} + \sum_{m=1}^{p_{ij}} b^m y_j^{T-m} + \epsilon^T \tag{4.6}$$

$$y_i^T = \sum_{m=1}^{p_{lj}} a^m y_i^{T-m} + \epsilon^T \tag{4.7}$$

As it determines the causal relationships between time series, this method has often been used to determine causal relations between airport delay and delay propagation through multiple airports, such as done by Du *et al.* [64], Belkoura and Zanin [63] and Mazzarisi *et al.* [65]. Due to the restriction to capture only linear causality, many studies have combined Granger causality with non-linear causality metrics and tests [45, 63].

As Granger causality is a temporal causality method, and is a linear model, this method is not applicable in the big data context of this research project, which aims to find complex and non-linear causal relations between not necessarily temporal attributes.

4.2. FREQUENT PATTERN IDENTIFICATION

Frequent pattern identification is a form of Knowledge Discovery in Databases, which is achieved by identifying recurring patterns in the data set [66, 67]. This mining method allows the algorithm to detect and learn association rules in the data set, which can be presented as A -→ B, where A is denoted as the antecedent and B the consequent [3]. Here, the antecedent can be a set of items which occur in the data set, whereas the consequent always is a single item which does not occur in the antecedent. In general, there are two main challenges to this method [66]. The first one is to manage the number of possible rules and thus the search space, as this grows exponentially with the number of variables and the number of values these variables can take. Additionally, identifying the rules which hold the most information remains challenging, as the found number of rules is often very large.

The advantage of using rule generation as data mining method over for example regression analysis, is that it can identify underlying and complex associations and correlations in the data set, as proven in a study by Sternberg *et al.* [3].

Support and confidence are both measures which can be used to asses the validity of the mined rules, and thus act as a filter on the identified rules [68]. Therefore, minimum thresholds need to be defined for both the support as confidence measures. These measures can be defined as follows [3, 69].

Support This measure is simply the percentage of the data records that contain the items. Thus, Support(X=x,Y=y) is the percentage of records in the data set, where variable X equals value x and at the same time, variable Y is equal to y. The mathematical formulation of this measure is shown in Equation 4.8 [67].

Confidence This measure represents the conditional probability of two variables. For example, Confidence (X=x→ Y=y), is the conditional probability that variable Y will equal y when it is given that variable X is equal to x. From the data set, the confidence of a certain rule can be expressed as the ratio of support, as seen in Equation 4.9 [67]. The confidence of a rule is in fact the support of both the antecedent and consequent, divided by the support of the antecedent, in analogy with Bayes' conditional probability theorem.

$$\begin{aligned} \text{Support}(X = x) = P(X = x) &= \frac{\text{Frequency(X=x)}}{\text{Total records}} \\ \text{Support}(X = x, Y = y) &= P(X = x \cup Y = y) \end{aligned} \tag{4.8}$$

$$\text{Confidence}(X = x \rightarrow Y = y) = P(Y = y | X = x) = \frac{\text{Support}(X = x, Y = y)}{\text{Support}(X = x)} \tag{4.9}$$

At first sight, these measures might appear to be similar, however, they represent different characteristics of a rule. The confidence of a rule is a measure of the strength of the rule, whereas the support denotes the statistical significance of a rule [69].

As mentioned before, the identified rules are filtered by setting minimum values for the support and confidence of the association rules. In a study by Sternberg *et al.* [3], Equation 4.10 was used to set the minimum value of the confidence, as it was required to identify rules which had a higher probability for delay due to the antecedents than the general probability for delay from the used data set. In that same study, the minimum support threshold was set by setting thresholds on the amount of times the rule should occur per day on average. The minimum support is then equal this value, divided by the total number of days in the data set, as illustrated in Equation 4.11.

$$P(\text{ delay } \mid \text{ antecedentconditions }) \geqslant P(\text{ delay }) = \text{min confidence} \tag{4.10}$$

$$\text{min support} = \frac{\text{Required frequency in a day}}{\text{Total number of days}} \tag{4.11}$$

However, even with the measures of support and confidence, the number of rules detected by the algorithm can still be very high. In order to further reduce the number of found rules in a data set, several methods

have been developed to only retain the "interesting" rules found from the data set. One of these methods is constraint-based mining, where only the rules are kept which satisfy the user's specifications [68]. Additionally, the found rules' support and confidence can be used to express correlation. Some of the available correlation measures are Lift, Cosine and All-confidence [68]. Sternberg *et al.* [3] has used an additional correlation measure on the extracted rules [67, 68], the Lift measure, presented in Equation 4.12 [3]. By using the lift measure, rules which do not actual consist of a causal relationship, but are detected as the consequent is very frequent in the data set can be filtered out [3]. When a consequent occurs very frequently in the data set, this will lead to a high support value for this consequent, ultimately reducing the lift measure of this rule. The lift measure can also be seen as a correlation measure [67]. In case that a rule has a lift value of 1, this indicates that the conditional probability of $Y = y$ when $X = x$ is equal of the probability of $Y = y$, and thus $X = x$ and $Y = y$ can be said to be independent. A value of less than 1 then indicates that the antecedent and consequent are actually negatively correlated, while a value > 1 indicates a positive correlation [67].

As an example of the general method, Table 4.1 contains five data records from a fictitious data set. Here, the association rule between *Weather=fog* → *Type of Delay = ATC delay* is investigated. The support of *Weather=fog* is in this example 60%, as it occurs for three records out of five. The same applies for the support of *Type of Delay = ATC delay*, which is also 60%. The support of *Weather=fog* → *Type of Delay = ATC delay* together is in this case 40%. Thus, using Equation 4.9, the confidence of this rule, *Weather=fog* → *Type of Delay = ATC delay*, is equal to 66.67%, which leads to a lift value of 1.11. The latter value can be interpreted as the probability of having an ATC delay is 11% higher in the presence of fog [3].

Table 4.1 – Example data set for frequent pattern mining.

Weather	**Type of Delay**
Fog	ATC delay
Rain	ATC delay
Fog	Maintenance
Fog	ATC delay
Cloudy	Crew rotation

$$\text{Lift}(X = x \rightarrow Y = y) = \frac{\text{Confidence } (X = x \rightarrow Y = y)}{\text{Support } (Y = y)} \tag{4.12}$$

An implication of using this method, is that the input data needs to be discretized in order to obtain rules which satisfy the support and confidence minima [3]. This means that binning and categorisation are crucial steps in the process, as working with continuous data leads to very low support values in the data, making it harder to detect the frequent patterns. The available data processing and transformation methods are discussed in more detail in chapter 5.

Association rule mining has originally been developed for retail applications, where dependencies between products could be detected using this method. However, the use of this method has been expanded successfully to other industries and applications, such as flight delay analysis [3]. For the process of association rule mining, multiple methods exist [68]:

- Apriori algorithm
- FP-growth algorithm
- Eclat algorithm

The first two algorithms are used for horizontal data sets, whereas the Eclat algorithm, developed by Zaki *et al.* [70], has been developed for vertical data sets, which is not applicable in this research project.

The Apriori algorithm was developed by Agrawal and Srikant [71] and is commonly used. It is based on the assumption that all subsets of a frequent item set are also frequent, meaning that an item set containing a sub item set that is not frequent can also be seen as not frequent and therefore discarded. However, the Apriori algorithm still performs candidate generation, as it lists all possible rules first, after which is will prune the rules in order to reduce them. This algorithm will thus start with the analysis of all 'rules' with length

one. This will be pruned according to the thresholds set for support. After this, the rules of length 2 will be formed and their support will be calculated. Again, the rules which do not fulfil the support minimum will be filtered out. This process is continued up to a user defined variable , k, denoting the maximum length of the rules, which makes it a *bottom up* approach, as it generates all possible rules starting from length 1 [72]. The Frequent Pattern growth, or FP-growth, algorithm is an algorithm that can perform association rule mining without the generation of each 'rule' candidate, and has been developed by Han *et al.* [72]. The method works following a *'divide-and-conquer'* method, and constructs a so-called Frequent Pattern tree, or FP-tree, from the compressed data [67]. This does not require to generate every potential rule, as done in the Apriori method. Han *et al.* [68] claims that this algorithm can reduce the search time significantly, as the algorithm searches in a more targeted manner, which makes it more scalable and computationally faster [67]. The FP-growth algorithm works in the following steps:

1. Start with 1-length rules and determine the support, rank the occurrence according to the support and filter the variables for minimum support.

2. Rank the attributes of all the data records from highest support to lowest support above the minimum threshold.

3. Build the FP-tree based on all data records in the data base: start from the root node and work down by adding the frequent patterns from the data set. This is done by checking whether the pattern is already in the tree. If it is, the counter is updated, and if it is not, an extra node is added and the count is set to 1.

This process is illustrated using the same example used by the developers of the algorithm Han *et al.* [72], and is shown in Table 4.2 and Figure 4.1. By using this approach, the FP-growth algorithm captures all frequent patterns which are present in the data set, and the height or size of the tree is limited by the maximum length of frequent pattern in the data set. Especially for longer patterns does this method outperform the Apriori algorithm in terms of computational time.

Table 4.2 – Example data set for the FP-growth algorithm [72].

Original Data Record	Ordered Da
f,a,c,d,g,i,m,p	f,c,a,m,p
a,b,c,f,l,m,o	f,c,a,b,m
b,f,h,j,o	f,b,
b,c,k,s,p	c,b,p,
a,f,c,e,l,p,m,n	f,c,a,m,p

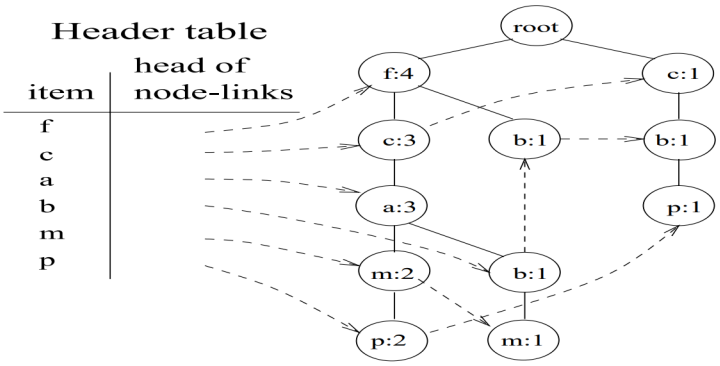


Figure 4.1 – Example of a frequent pattern tree [72].

4.3. BAYESIAN NETWORKS

This section presents the method of Bayesian Networks (BN). This method has been widely applied in the research area of flight delay, especially for delay propagation throughout a network of airports or flights [7, 8]. Rodriguez-Sanz *et al.* [32] have used a BN to perform an analysis of the factors influencing flight delays. This application of the method allows to investigate causal relationships between the variables, which is in line with the objective of this research.

4.3.1. THEORETICAL BACKGROUND

In essence, a Bayesian Network is a graphical presentation of a joint probability distribution [51]. Bayesian networks consist out of two elements, a graphical tree structure, also known as a Directed Acyclic Graph (DAG), and a Conditional Probability Table (CPT) for each node in the network [7, 32]. The graphical tree contains all information regarding the qualitative information on the relationships between the nodes, whereas each conditional probability table contains quantitative information. A node in the structure represents a

variable, whereas the links between the nodes represent direct dependency between the variables connected by the link [32].

Equation 4.13 is the formula representing the joint probability of a network structure , $P(X_1,\ldots,X_n)$, and how this is equal to the product of the conditional probabilities of the variables, X_i, on the variables in their parent set, O_i [51, 73].

Equation 4.14 shows a practical example of Equation 4.13, used by Wu and Law [7] in the context of flight delay. Here, t is the departure delay for flight j, and k,q,c,p,g are all independent variables possibly influencing the departure delay probability of flight j. It should be noted that $P_j(t \mid k,q,c,p,g)$ is also known as the posterior probability of delay for flight j [7].

$$P(X_1,\ldots,X_m) = \prod_{i=1}^{m} P(X_i \mid O_i) = \prod_{i=1}^{m} \theta_{X_i|YO_i} \tag{4.13}$$

$$P_j(t \mid k,q,c,p,g) = \frac{P(t,k,q,c,p,g)}{P(k,q,c,p,g)} = \frac{P(k,q,c,p,g \mid t)P_j(t)}{P(k,q,c,p,g)} \quad \forall j \tag{4.14}$$

By applying Bayes' theorem, all variables on the right side in Equation 4.14 can be computed using historical flight data, and therefore it makes it possible to get all probability distributions, as done and proven in Wu and Law [7].

An example Bayesian Network is shown in Figure 4.2 [32], where variables x_2 and x_3 are the parent variables of the variable x_4, and x_4 is again a parent variable of x_5. Alternatively, x_5 can also been seen as the child node of x_4.

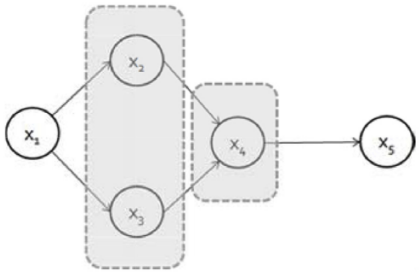


Figure 4.2 – Illustrative example of causal inference in a Bayesian Network [32].

4.3.2. CONSTRUCTING BAYESIAN NETWORKS

There are multiple ways on how a Bayesian Network can be constructed. If there is enough knowledge on the problem, or the causal relationships are already known, a Bayesian Network can be constructed manually, such as done in the work by Wu and Law [7] and Xu *et al.* [8]. However, constructing a large network from reference literature and/or expert knowledge, is not sufficient to capture the complex and large amount of relationships between the attributes. Therefore, when the necessary knowledge is not present or sufficient, a BN can also be constructed from historical data, as done by Rodriguez-Sanz *et al.* [32] and Truong [73]. There are multiple steps when generating a BN from data, which are the following [32, 74]:

1. Create the variables
2. Learn the structure
3. Learn the parameters

Structure learning concerns finding the most suitable DAG for the underlying data, whereas parameter learning relates to finding the parameters associated with the found DAG, also known as the Conditional Probability Tables [74].

For learning BN structures from data, multiple algorithms have been developed [75]. In general, the algorithms can be divided into two groups, one where the variables or nodes should be ordered already, and one where the node ordering is left unknown [75]. The latter requires no prior knowledge about the relationship between the variables, but also makes the learning algorithm more complex and computationally heavier, as relationships between all the variables need to be investigated [75].

Additional to the division of algorithms based on the presence of node ordering, the algorithms can be further split into search & scoring based algorithms, and dependency or constraint based ones [75]. A schematic overview of the different types of Bayesian Network Learning methods is presented in Figure 4.3.

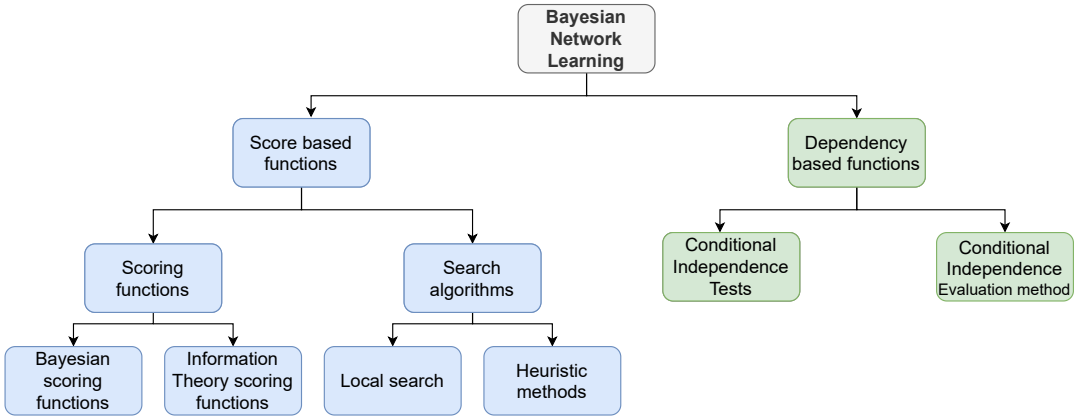


Figure 4.3 – Types of Bayesian Network learning algorithms.

SCORE BASED

For scoring based algorithms, the algorithm uses a scoring function in order to represent the fit of the Bayesian Network to the data [76]. Therefore, these algorithms use a scoring function, which is optimised to maximise the fit of the BN to the data. In order to evaluate the found structures using the scoring method, these algorithms always use a search algorithm to find the possible Bayesian Network structures. The scoring functions can again be grouped into two major types, the Bayesian scoring function and Information Theory based functions [76]. Bayesian Scoring functions are scoring functions which represent and maximise the *posterior* probability, $p(G|D)$, and the most commonly used ones are K2, BD and BDeu [77]. Information theory functions are likelihood based, where the scoring function represents the log-likelihood, $LL(G|D)$ [77].

As for the scoring functions, there are many different search algorithms available and researched for application in Bayesian Network learning [76]. Local search algorithms have been popular, but due to the exponential growth of the search space of possible DAGs, heuristic search methods have been used as well. Heuristic methods are methods which are likely to find a good solution to the problem, but they do not guarantee to find the most optimal solution [78]. In essence, heuristic methods are iterative, where each iteration, the algorithm searches for a better feasible solution than currently found in previous iterations. Well-known examples of heuristic methods are tabu search, branch and bound, simulated annealing and genetic or evolutionary algorithms. Tabu search was used in a study on airport delay by Truong [73], as it was found by earlier studies to discover the global optimum solution, as it has the ability to flee local optimums. Rodriguez-Sanz *et al.* [32] used a hill climbing with random restarts search method, which is an extended local search algorithm.

Equation 4.15 shows the objective function, where $G*$ denotes best fitted structure [76]. Equation 4.17-4.20 show commonly used Bayesian scoring functions, K2, BD and BDeu,respectively [76]. The symbols used in these functions are explained in Table 4.3. The prior probability distribution, $p(G)$, is often assumed to be uniform, which makes it the same value for every possible DAG [76, 77]. Therefore, this part of the scoring function becomes a constant factor, and can be removed in the optimisation procedure [76].

It should be noted that scoring functions are often developed for the usage on solely discrete data, as r_i denotes the number of states of variable. Therefore, these scoring function require the discretization of continuous data. However, in a study by Dojer [79], scoring functions have been transformed to handle a mix of discrete and continuous data.

$$G^* = \arg\max_{G \in G_n} g(G:D) \tag{4.15}$$

$$N_{ij} = \sum_{k=1}^{r_i} N_{ijk} \tag{4.16}$$

$$g_{K2}(G:D) = \log(p(G)) + \sum_{i=1}^{n} \left[\sum_{j=1}^{q_i} \left[\log\left(\frac{(r_i-1)!}{(N_{ij}+r_i-1)!} \right) + \sum_{k=1}^{r_i} \log\left(N_{ijk}!\right) \right] \right] \tag{4.17}$$

For the BD and BDeu functions, the prior distributions of the variables are assumed to be Dirichlet distributed, which is a generalisation of the K2 function [76]. Γ represents the Gamma function, and η_{ijk} are the

Table 4.3 – Explanation of the variables used in Bayesian Scoring Functions, retrieved from the expressions used in work by de Campos [76].

Variable	Meaning
g	Scoring function
D	Data set
G	The Directed Acyclic Graph
$p(G)$	Prior probability of DAG G
n	Number of variables in data set D
r_i	Number of states of variable X_i
q_i	Number of configurations of parent set $Pa_g(X_i)$
$w_{ij}, j = 1, ..., q_i$	Configuration of parent set of X_i
N_{ijk}	Number of instances in data set where variable X_i takes on value x_{ik} and parent set is takes configuration w_{ij}
N_{ij}	Number if instances in data set where parent set of variable X_i takes configuration w_{ij}
N_{ik}	Number if instances in data set where variable X_i takes value x_{ik}
η	Equivalent sample size
$p(.\|G_0)$	probability distribution prior Bayesian Network G_0

hyperparameters of the Dirichlet prior distributions. However, the determination of these hyperparameters is not straightforward [76], and therefore, the BDeu scoring function has been developed. In the BDeu scoring function, the hyperparameters can simply be estimated using the expression in Equation 4.18, where η denotes the equivalent sample size [76]. Therefore, the BDeu scoring function is the most commonly used in learning Bayesian networks [80].

$$\eta_{ijk} = \eta \times p\left(x_{ik}, w_{ij} \mid G_0\right) \tag{4.18}$$

$$g_{BD}(G:D) = \log(p(G)) + \sum_{i=1}^{n}\left[\sum_{j=1}^{q_i}\left[\log\left(\frac{\Gamma\left(\eta_{ij}\right)}{\Gamma\left(N_{ij}+\eta_{ij}\right)}\right) + \sum_{k=1}^{r_i}\log\left(\frac{\Gamma\left(N_{ijk}+\eta_{ijk}\right)}{\Gamma\left(\eta_{ijk}\right)}\right)\right]\right] \tag{4.19}$$

$$g_{BDeu}(G:D) = \log(p(G)) + \sum_{i=1}^{n}\left[\sum_{j=1}^{q_i}\left[\log\left(\frac{\Gamma\left(\frac{\eta}{q_i}\right)}{\Gamma\left(N_{ij}+\frac{\eta}{q_i}\right)}\right) + \sum_{k=1}^{r_i}\log\left(\frac{\Gamma\left(N_{ijk}+\frac{\eta}{r_i q_i}\right)}{\Gamma\left(\frac{\eta}{r_i q_i}\right)}\right)\right]\right] \tag{4.20}$$

For information theory based functions, the log-likelihood of the DAG G is maximised given the data set D, formulated in Equation 4.21 [77]. However, Log-likelihood has a high probability to overfit. Therefore, a number of scoring functions have been developed which add a penalising term in the log-likelihood expression. This is shown in Equation 4.23, where the penalising term a function of the number of samples and the complexity of the found network $|G|$, which can be calculated using Equation 4.24 [77]. One of the most commonly used scoring functions is Bayesian Scoring criterion (BIC) scoring function, also known as the Minimum Description Length (MDL), shown in Equation 4.22 [77, 80].

$$\mathrm{LL}(G \mid D) = \sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i} N_{ijk}\log\left(\frac{N_{ijk}}{N_{ij}}\right) \tag{4.21}$$

$$\mathrm{BIC}(G \mid D) = \mathrm{LL}(G \mid D) - \frac{1}{2}\log(N)|G| \tag{4.22}$$

$$\phi(G \mid D) = \mathrm{LL}(G \mid D) - f(N)|G| \tag{4.23}$$

$$|G| = \sum_{i=1}^{n}\left(r_i - 1\right) q_i \tag{4.24}$$

CONSTRAINT/DEPENDENCY BASED

Dependency based or constraint based methods work with conditional independence tests on the variables in the input data set. The found network relies thus on the effectiveness of the used conditional independence test. Cheng *et al.* [75] and Natori *et al.* [81] mention three possible conditional independence tests:

- Conditional mutual information test
- Likelihood ratio chi-squared

- Pearson chi-squared

The conditional independence between variables is often expressed in terms of d-separation, which is often used when using graphical network structures or DAGs. D-separation determines whether a path between two variables is closed, meaning that these two variables are conditional independent. Applying this to the example Bayesian Network shown in Figure 4.2, it can be said that variable x_5 is d-separated from variables x_2 and x_3, given variable x_4. It can thus be stated that variable x_5 is conditionally independent from variables x_2 and x_3, given variable x_4. Mathematically, this can be formulated as follows: $x_5 \perp\!\!\!\perp x_2, x_3 \mid x_4$.

One of the most commonly used constrained based algorithms is the PC algorithm [49, 80, 82]. In this method, the structure is learned by starting from a complete, undirected DAG G, where all variables or nodes are connected by edges. The algorithm then performs a conditional independence test on the variables, and in case variables are d-separated, the link between them is removed [82]. The tests for conditional independence are done as follows. First, the complete independence of two variables, x and y is tested, and repeated for all combinations of variables in the data set. If the two variables are independent, as shown in Equation 4.25, the link between them is removed [82]. The following step is to test the conditional independence of variables which were not independent, given a neighbouring variable. If the two variables are found to be conditionally independent or d-separated, as shown in Equation 4.26, where z is a neighbouring node of x, the link between them is removed [82].

Another popular constraint based algorithm is the Fast Causal Inference Algorithm (FCI) [49], which can also detect unmeasured or confounding variables.

$$x \perp\!\!\!\perp y \mid \emptyset \qquad (4.25) \qquad\qquad x \perp\!\!\!\perp y \mid z \qquad (4.26)$$

COMPARISON

Several scientific studies have focused on researching the performance of the different algorithms for learning Bayesian networks [74, 81]. According to Natori *et al.* [81], score based methods are computationally more demanding, which makes them only applicable to relatively smaller networks. However, they are also seen as more accurate [81]. de Campos [76] agrees that score based methods are more accurate, as the results of dependency based methods can be unreliable due to the conditional independence tests. However, de Campos also recognises the computational complexity of the dependency based method as an important drawback of the method, which is in contrast with what was stated by Natori *et al.* [81]. Additionally, Scutari *et al.* [74] made an extensive comparison of the two classes of methods, by using both real-world and simulated data sets. Scutari *et al.* found that for small networks, dependency based algorithms are actually more accurate. Additionally, score based functions using tabu search outperformed the other methods in computational time, for most of the cases [74]. Additionally, Acid *et al.* [80] studied several learning algorithms on a single data set. They found that, in contrast to other work, scoring algorithms did not perform better than dependency based ones, in fact the latter had a better performance.

It can thus be concluded that there is no real consensus on the performance and computational load differences between the two methods. Overall, score based methods are said to have the best performance, but this is not universally confirmed. Therefore, this can be potentially incorporated in the research questions of this project.

4.3.3. BAYESIAN NETWORKS FOR CAUSAL ANALYSIS OF FLIGHT DELAYS

When established, Bayesian networks can be used in many ways [32, 73]. One of them is to make predictions about the resulting variable. This is called forward inference, and can be achieved by setting the value of the parent or input nodes [32]. Additionally, the network can also be used to understand the main causes or influencing factors of the dependent variable in a child node [73]. This can be done by using backward inference, which is the opposite of forward inference, and thus requires the fixation of the variable in the child node [32].

Bayesian networks have been widely applied in the research area of flight delay, especially for delay propagation throughout a network of airports or flights [7, 8]. Additionally, BN have also been used by Rodriguez-Sanz *et al.* [32] to perform a causal analysis of flight delays, where the attributes in the data set represented the nodes and the links represented causal inference, or conditional dependence. The latter application of Bayesian networks is by far the most interesting in this research project, as it also aims to discover complex causal relationships between the variables in a big data context.

Wu and Law [7] acknowledged that Bayesian networks are less fit to be used in a big data or machine learning environment, as the increased complexity and variable numbers can make the computational cost very high or even impracticable. Especially for an increase in the number of variables or attributes, the search space of possible Bayesian Network structures grows exponentially [82]. However, as the relationships between the variables or nodes need to be statistically significant, Bayesian networks still require a large number of data tuples or records, as investigated by Zuk *et al.* [83].

In the study by Wu and Law [7], the authors have applied the methodology to a small number of flights, which were very limited in number of variables and states. However, Rodriguez-Sanz *et al.* has used 34,000 flights at Madrid airport to construct a Bayesian Network, which represented two busy Summer months in 2016. In the Bayesian Network developed by Xu *et al.* [8], also three months of airport flight data was used, with a total of 18 nodes or variables.

4.4. MACHINE LEARNING

In past research, Machine Learning (ML) methods have been used for flight delay prediction, and in some cases also to find the influencing factors of delay [60, 84]. The most commonly used models have been determined from literature and are discussed in this section. In order to extract causal knowledge from these machine learning models, Explainable Artificial Intelligence (xAI) is introduced and assessed for application in this research.

4.4.1. MODELS

Many past scientific studies have researched different machine learning techniques and their performance in the context of flight delays. From literature, the most commonly used machine learning models could be identified and are listed below. These models vary from linear, relatively simple models such as linear regression, to highly non-linear and complex methods, of which neural networks and random forests are good examples.

1. Multiple Linear Regression [85]
2. Logistic Regression [32]
3. (Deep) Neural Networks[60, 84, 85]
4. Support Vector Machines [31, 84]
5. Random Forest [84, 86–88]
6. Decision tree [31, 88]

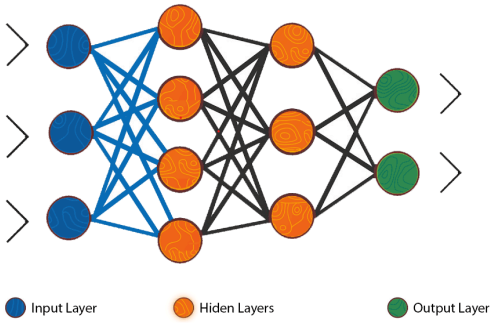


Figure 4.4 – Illustration of a feedforward NN [89].

In a study by Sridhar *et al.* [85], the linear regression model was compared to Neural Networks (NN). It was found that the non-linear neural networks performed better in flight delay prediction than a (multiple) linear regression model. Kulkarni *et al.* [31] used both Support Vector Machines (SVM) and decision tree classifiers for flight delay prediction, but found no significant differences in the performance of the models. In a study by Belcastro *et al.* [86], the Random Forest (RF) model was used as this method outperformed SVM, logistic regression and other models in the preliminary results. Rebollo and Balakrishnan [87]also used random forest, as it had the best performance out of a range of models including several regression models, decision trees and neural networks. Fernandes *et al.* [84] followed the reasoning of Rebollo and Balakrishnan, but also added neural networks and support vector machines due to their high performance in other studies. However, also in the work of Fernandes *et al.*, RF outperformed NN and SVM. The authors of this study however acknowledged that using a deep neural network with multiple hidden layers could improve the performance of the neural network model, and possibly outperform RF [84].

From all these studies, it can be concluded that both random forest and neural network models are capable of detecting highly non-linear patterns in the data. These models are discussed in more detail below.

An illustration of the most straightforward neural network format is shown in Figure 4.4 [89]. A neural network consists of multiple (hidden) layers, which each consist out of a number of nodes. The NN always consist out

of an input layer, with the data set variables as input nodes, and an output layer, containing the predicted value of the neural network. In between are a number of hidden layers, which consist out of a number of nodes [90]. Both the number of hidden layers and associated nodes are to be set by the user. The nodes of subsequent layers are connected with each other, and each link is characterised by a weight. The output of the node is multiplied by the weight, and fed to a non-linear activation function at the other node [90]. This process is illustrated in Equation 4.27, which shows how the outputs of the nodes of the previous layer are summed together, and the output is fed to a non-linear activation function h, as shown in Equation 4.28 [90].

$$a_j = \sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)} \qquad (4.27) \qquad\qquad z_j = h\left(a_j\right) \qquad (4.28)$$

The most simple form of NN is the feedforward NN, also known as the Multilayer Perceptron [90]. Additionally, there are also forms as Recursive Neural Networks and Deep Belief Networks [60, 89]. The most commonly used training method of neural networks is gradient descent and backpropagation. Here, the gradient of the error function in function of the weights, $E\left(\mathbf{w}^{(\tau)}\right)$, needs to be determined [90]. Then, gradient descent can be used, such that the weights in the network are updated in the direction of a smaller error, formulated in Equation 4.29, where η denotes the learning rate, a user set parameter, and τ the learning iteration or epoch [90].

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \eta \nabla E\left(\mathbf{w}^{(\tau)}\right) \qquad (4.29)$$

Random forest is an ensemble model, which combines multiple decision trees. A decision tree is a top-down model in which each internal node corresponds to a condition on an input variable [31]. Each of these nodes then further splits into other nodes, as seen in the decision trees displayed on Figure 4.5 [91]. Decision trees can be used for both classification as regression models. In case of classification, the 'Gini impurity' or 'Information Gain' measures are used to assess the quality of a split in the decision tree [31]. For regression purposes, an error metric such as the mean squared error or mean absolute error is used.

As the random forest model combines multiple decision trees, it makes the model less likely to overfit on the data, compared to neural networks [92]. This is because the method learns the decision trees at the same time, which later on combines the learned patterns and theories [86]. The working principle of this model is illustrated below in Figure 4.5 [91].

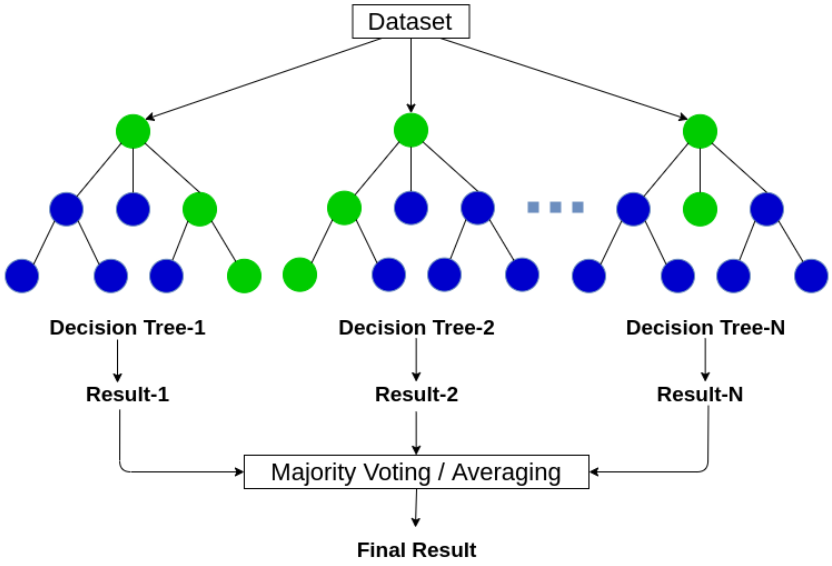


Figure 4.5 – Illustration of the random forest model [91].

4.4.2. EXPLAINABLE ARTIFICIAL INTELLIGENCE

Explainable Artificial Intelligence(xAI) is used in many fields in order to make black box machine learning models and its decisions more transparent and understandable [93]. As acknowledged in a review by Adadi and Berrada [94], there are several reasons to use xAI on machine learning algorithms. One of them is explanation of the model to discover new insights in the data, which is in line with the aim of this research.

Therefore, xAI can be used is in causality research, and understanding the causal influences of the different variables [93].

In general, machine learning models can only detect and learn correlations among the variables in the training data set. However, explainable models and methods can help to validate whether the correlation also entails a cause effect relation [93]. In this section, two particular xAI methods are discussed, namely rule extraction and sensitivity analysis. The former can be classified as a model simplification method, whereas sensitivity analysis is a feature relevance explanation method [93].

RULE EXTRACTION

Neural networks are great machine learning models, as they have the ability to capture complex and non-linear patterns in the data sets. However, because of this feature, they are often considered as black-box models and it is very hard to explain the reasoning of the model in an understandable manner. Therefore, many rule extraction algorithms have been developed over the years. Rule extraction is a method applied to black box machine learning models, with the aim to better understand how the neural network has come to its predictions. In rule extraction , there are three main algorithm categories[95]:

- Decompositional
- Pedagogical
- Eclectic

The decompositional algorithms go at a node level in the neural network, and analyse each individual node and weight of the network. Pedagogical algorithms do not depend on the internal structure of the neural network, but also treats the learned network as a black box and determines the rules based on the in- and output [95]. Lastly, the eclectic method is a combination of the decompositional and pedagogical approaches [95].

Additional to the different types of algorithms used for discovery of the rules, there are also different types of rules that can be extracted, which are shown below [95].

- If-then rules
- M-of-N rules
- Decision tree

The if-then rules are conditional, and have a format which specifies a certain condition on a variable, leading to another variable condition [95]. M-of-N rules are extracted from boolean conditions, where M out of the N sets need to be fulfilled [95]. Lastly is the decision tree also a sort of rule, as it is a white box model. The manner in which it makes predictions can be retrieved and understood based on input data conditions at the splits, as illustrated in the previous subsection.

Although many different rule extraction algorithms have been developed, some of these are restricted to be used on pure discrete data sets [96]. Another concern about rule extraction is that it does belongs to the model simplification class of xAI, meaning that it inherently simplifies the black-box neural network. By doing so, the extracted rules may not represent the original model in an accurate manner, and thus important information might be lost [97].

As the decompositional approach uses the information contained by the nodes and the weights, this method is in general more transparent than the pedagogical approach. However, the latter has lower computational complexity and loads because of that same reason [95]. Additionally, the pedagogical method is more flexible in the architecture of the used neural network, as it does not analyse the layers and nodes between the in- and outputs [95]. However, as the decompositional and eclectic algorithms take into account the neural networks internal format, they can achieve a higher accuracy, but this comes at a cost of a higher complexity [96].

In the application of rule extraction for causal relationship discovery, the pedagogical approach has little to no contribution. As the pedagogical approach draws rules based on the in and output of the neural network, the use of a neural network in this application becomes redundant, as it is not aimed to make predictions about the occurrence of ATC delays. Therefore, this method has little value for determination of causal relationships, and is therefore excluded in this research and the used methodology.

SENSITIVITY ANALYSIS

As mentioned before, sensitivity analysis is a method that falls under the feature relevance area in xAI. In order to determine feature importance, many methods have been researched. In Yu *et al.* [60], the feature relevance was found by leaving one feature out of the data set and quantifying the (negative) impact on the accuracy of the model. This method is very computationally heavy, as the model needs to be retrained for the importance analysis of every feature, and the outcome is very dependent on the performance of the prediction method [60]. If the model has a bad performance, the found feature relevance values will have low confidence as well.

Cortez and Embrechts [98] proposed a number of data-driven sensitivity analysis techniques, which can be universally applied to supervised learning methods, in contrast to the rule extraction method, which has been designed for neural networks. In sensitivity analysis (SA), the input variables of the model are varied across their range of values, and the impact on the output of the model is quantified [97]. This method has been used for feature selection, but also for explainability of the model [97].

The sensitivity analysis is done by letting the trained model predict the target outputs, originally $\hat{y}$, again, but now with modified input attributes x_a, resulting in predictions $\hat{y}_a$. Each of these attributes are discretized into a number of levels, denoted as L. Thus when a variable x_a for examples ranges between 0 and 4, and $L = 4$, then $x_{a_j} \in \{0,1,2,3,4\}$, where j denotes the level of x_a [98].

The sensitivity analysis techniques identified and developed in Cortez and Embrechts [98] are listed below, of which the symbols are defined in Table 4.4.

- **One Dimensional SA (1DSA)** was developed by Kewley *et al.* [99],and is seen as a computationally efficient method[97], with a complexity of Order $O(MxLxP)$ [98]. This method varies the input of only one variable, while the other attributes are kept at their average value [97]. Therefore, this method cannot detect the influence of features upon each other [98].

- **Global SA (GSA)** has been developed by Cortez and Embrechts [97]. It is a computationally demanding method, but can detect more than 1D sensitivity analysis, as it varies multiple input features at the same time. However, this comes at a high computational cost, with an order in function of the number of features varied at the same time $O(L^{\#F}xP)$ [98].

- **Data-based SA (DSA)** was proposed by Cortez and Embrechts [98]. This method has the benefits of Global SA, but has a lower computational load, with an order of $O(MxLxN_sxP)$ [98]. This is achieved by sampling the training data from the data set, with length N_s.

- **Monte-Carlo SA (MSA)** is used where the data required for training is not available, but the fitted model is, which can be the case due to privacy issues [84]. The only difference with the DSA is that the data for sensitivity analysis is taken from a uniform distribution, in stead of sampled from the actual data input. Therefore, it has the same computational load as DSA [98].

- **Cluster-based SA (CSA)** is a method in which the variables are first clustered using a very fast method [98]. CSA has a lower computational load as DSA, as the order is namely $max(O(NxP), O(MxL))$ [98], but also a reduced performance compared to GSA,DSA and MSA, as validated by Cortez and Embrechts [98].

It can be seen that for sensitivity analysis, the number of features M, data samples N and computational load of the data mining model P are important factors in the computational load of the sensitivity analysis. As these techniques are applied to a data mining method, characterised by a high number of attributes and data samples, this should be taken into account while selecting the best applicable method. However, DSA and MSA, developed by Cortez and Embrechts [98], have achieved great reduction of the computational load. Cortez and Embrechts even report that both methods perform well if only 1% of the data set is used in the sample.

In order to quantify the sensitivity, several sensitivity metrics have been defined such as shown below [97, 98]. It should be noted that these metrics are only applicable to regression problems,and a higher value indicates a higher importance of the input feature [97].

Table 4.4 – Explanation of the variables used in sensitivity analysis, retrieved from the expressions used and defined by Cortez and Embrechts [98].

Variable	Meaning
P	Data mining method to predict y based on input x
M	Number of attributes or features
N	Number of data samples
#F	Number of features varied
x	Input data set
x_a	Variable in input data set
$\hat{y}_a$	Set of sensitivity responses
L	Number of levels in the range of values of variable x_a
x_{a_j}	Level j of the variable x_a
s	Any of the defined sensitivity measures

- range S_r
- variance S_v
- gradient S_g
- average absolute deviation from median (AAD)

$$S_r = \max\left(\hat{y}_{a_j} : j \in \{1,\ldots,L\}\right) - \min\left(\hat{y}_{a_j} : j \in \{1,\ldots,L\}\right) \tag{4.30}$$

$$S_g = \sum_{j=2}^{L} \left|\hat{y}_{a_j} - \hat{y}_{a_{j-1}}\right| / (L-1) \tag{4.31}$$

$$S_v = \sum_{j=1}^{L} \left(\hat{y}_{a_j} - \bar{y}_a\right)^2 / (L-1) \tag{4.32}$$

$$S_d = \sum_{j=1}^{L} \left|\hat{y}_{a_j} - \tilde{y}_a\right| / L \tag{4.33}$$

Finally, the above sensitivity measures can be translated into a measure of relative importance, shown in Equation 4.34 [100]. Here, s represents any of the sensitivity measures above [97], as the higher the variance, gradient or range of the sensitivity responses, the higher the impact of the attribute on the output $\hat{y}$. Originally, the relative importance R_a was computed using the variance metric, S_v [100].

$$R_a = s_a / \sum_{i=1}^{M} s_i \times 100(\%) \tag{4.34}$$

Additional to the sensitivity measures, Cortez and Embrechts [97] also proposed a number of visualisation methods for opening the black box models. One of them is to plot the output of Equation 4.34 in the form of a bar plot, to visually illustrate the importance of the input features. Additionally, the Variable Effect Characteristic curve (VEC) is proposed, which uses the results from the sensitivity analysis to visually represent the influence of the input variable on the output [97, 98]. This is achieved by plotting the used input variables x_{a_j} and outcomes $\hat{y}_{a_j}$ on the x- and y-axis respectively.

The work of Cortez and Embrechts [98] has been applied to flight delays by Fernandes *et al.* [84]. In this study, three data mining techniques were used to perform prediction of flight delays, namely neural networks, support vector machine and random forest model. The sensitivity analysis was executed using the RF model, as this gave the best prediction performance. A disadvantage of this method is that only the relevance and impact of one feature on the target variable is assessed. Therefore, sensitivity analysis cannot capture the possible complex relationships between various attributes and the influence on the target variable.

4.4.3. NEURAL NETWORKS FOR CAUSAL MODEL LEARNING

Additional to making predictions, neural networks can be used in other forms and ways to retrieve (causal) relationships in data, by rule extraction, as discussed before, and for causal inference, studied by Goudet *et al.* [50].

In Goudet *et al.* [101], neural networks are used to find the causal model or DAG representing the data. As discussed earlier, these Bayesian networks are often learnt using score based functions. However, these functions are often not differentiable and thus cannot be used to train a neural network [50]. Therefore, Goudet *et al.* proposed and developed a neural network that can be used for causal model discovery, the

Causal Generative Neural Networks(CGNN). The main novelty of this method is that it uses neural networks to model the probability distribution of variable X_i, given its parent set O_i, denoted as $P(X_i|O_i)$ [101].This model has the following properties [101]:

- The edges in the DAG all have a confidence score.
- Multivariate dependencies between in- and output can be identified.
- The methods makes no assumptions on the prior distribution of the data or the generative model, and it is therefore non-parametric.

The CGNN is in fact a Causal Functional Model, which can be simply expressed using Equation 4.35 [101]. The model is graphically illustrated in Figure 4.6, which shows the application of Equation 4.35, worked out in Equation 4.36 [101]. As shown in Figure 4.6, the functions f_i denote in fact a generative neural network with 1 hidden layer [101]. A generative neural network is a network which is trained to generate new data from an existing input data set. The neural networks used in this application are also generative, as the noise vector can be used to develop a new data set as $\hat{x}_1, \hat{x}_2, ..., \hat{x}_n$.

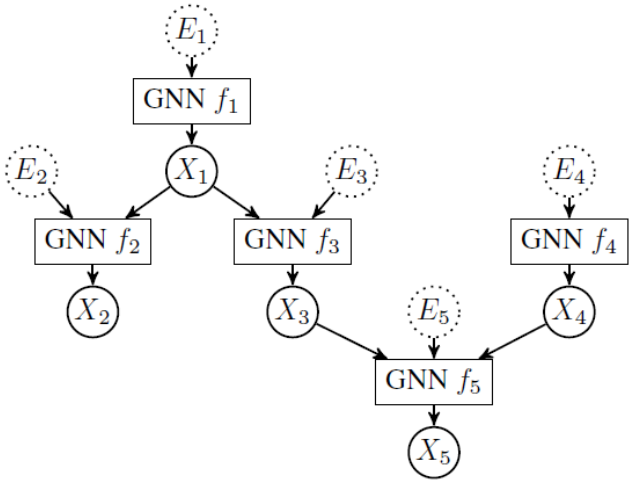


Figure 4.6 – Illustration of the CGNN model, where functions f_i each are a GNN with a hidden layer [101].

$$X_i = f_i\left(O_i(G), E_i\right) \tag{4.35}$$

$$\begin{cases} X_1 = f_1\left(E_1\right) \\ X_2 = f_2\left(X_1, E_2\right) \\ X_3 = f_3\left(X_1, E_3\right) \\ X_4 = f_4\left(E_4\right) \\ X_5 = f_5\left(X_3, X_4, E_5\right) \end{cases} \tag{4.36}$$

Like all other models, CGNN needs to be optimised as well. The optimisation process consists out of two problems, one where the neural networks are optimised, and one where the found structure, the DAG, is optimised.

The first is done using the Maximum Mean Discrepancy, of which the formulation is shown in Equation 4.37, where k denotes a kernel, which is most often the Gaussian Kernel [50]. A kernel is a function that computes the dot product of its inputs [90]. D is the original data set, and $\hat{D}$ is the data set sampled from the causal structure analysed. The MDD will go to zero if the actual data x_i and the data from the analysed structure $\hat{x}_j$ is drawn from the same distribution, meaning that the found structure represents the data set perfectly for an infinite number of data samples [101]. Therefore, this objective function should be minimised to be optimal. The advantage of using this function in combination with a kernel, is that is becomes differentiable. Therefore, the principle of gradient descent and backpropagation can be used to learn the networks $\hat{f}_i$, as discussed earlier in this section [101].

$$\hat{\text{MMD}} = \frac{1}{n^2}\sum_{i,j=1}^{n} k\left(x_i, x_j\right) + \frac{1}{n^2}\sum_{i,j=1}^{n} k\left(\hat{x}_i, \hat{x}_j\right) - \frac{2}{n^2}\sum_{i,j=1}^{n} k\left(x_i, \hat{x}_j\right) \tag{4.37}$$

The optimisation of the found DAG is complex as the possible structures for a problem with m variables and thus m nodes has an exponential number of possibilities of degree m [101]. Thus for big data approaches, the search space for the number of possible DAGs is enormous. Without an efficient searching algorithm, the computational load would make the problem unsolvable. Therefore, an initial structure can be formed using feature selection methods, combined with a greedy search method, as proposed by Goudet *et al.* [101].
This combination of Bayesian networks and neural networks has been recently proposed by Goudet *et al.* [101]. The concern with this method is that it might become very computationally heavy. Additional to finding the DAG, which is already a computational challenge for a data set with a large number of features, multiple neural networks need to be trained. Therefore, this method is seen as out of scope for the research objective and size of the data set.

4.5. COMPARISON OF CAUSAL METHODS

In this section, the previous discussed methods are compared. This is done in terms of performance and applicability to the research objective and question, and in their scalability with respect to data set size.

4.5.1. PERFORMANCE FOR CAUSAL ANALYSIS

In this chapter, several causal models have been presented. For the state-of-the-art method, two large methods can be distinguished, machine learning with xAI and Bayesian networks.

Bayesian networks have excellent properties for causal analysis, as they can be constructed based on a data set and represent the conditional dependencies between the variables. Additionally, through backward inference in the network, the main causes or drivers of a variable can be found and quantified [32, 73]. According to Rodriguez-Sanz *et al.* [32], Bayesian networks are a good method to analyse airport saturation or flight delays, for various reasons. First of all, they are white boxes, meaning that they can be interpreted in a straightforward manner, and can be used for causal analysis. Additionally, the analysis is done from a probabilistic perspective, which is very suitable for a stochastic process such as flight delays. Lastly, Bayesian networks support multiple control variables, such that complex relationships and interactions among the variables can be detected [32].

Although a lot of research has been performed in the field of learning the structure of a Bayesian Network[7, 8, 32, 49, 75, 76], a few challenges in the application of this method remain present [75]:

1. The need of node ordering for some algorithms.
2. The computational load & complexity.
3. The lack of public accessible algorithms & applicability to data mining applications.

The discussed machine learning models are actually models which are used for prediction purposes. However, this is not the goal of the research, but by using xAI, causal relationships can be discovered [93]. Therefore, rule extraction and sensitivity analysis techniques have been investigated [93]. Rule extraction methods have the disadvantage of simplifying the machine learning model, and therefore might not capture complex relationships [98]. The sensitivity analysis on the other hand is rather computationally expensive especially for high dimensional data sets, but has shown to give good results [84]. However, Adadi and Berrada [94] state that sensitivity analysis is not often used as a pure explanation method, but more as a method to validate the model and the learned patterns.

A drawback of using machine learning followed by an xAI method is that the knowledge discovered from explaining the black box model is dependent on the performance of the machine learning model. If the model does not have a good performance, and thus cannot capture the real complex patterns between the variables, the explanation of this method will be of little value to expose the causes of the analysed dependent variable. However, the same can be said about Bayesian networks. If the discovered BN does not fit the underlying data well, the extracted knowledge will not represent reality.

4.5.2. REQUIRED AMOUNT OF DATA & DATA SCALABILITY

The scalability of (machine learning) methods is the influence that an increasing size of the data set has on the computational performance of that method [102]. The size of a data set can increase in two dimensions, the number of data records, and the number of attributes or features. The computational performance of a method consists of the accuracy, required training or running time and the needed memory, and optimising

a model often consists out of a trade-off between these parameters [102]. Therefore, evaluating the scalability of a method requires the use of measures relating to accuracy/error, time and memory usage [102].

The scalability of a method is important to evaluate when the method wants to be used on a high dimensional data set, in order to guarantee that the method can handle the amount of data. A data set is seen as high dimensional when there is a very large amount of data samples, a large amount of input features, and/or classification groups, according to Bolón-Canedo *et al.* [103]. Additionally, a data set is particularly high dimensional when the number of features is larger than the number of data samples [103].

For machine learning methods, the rule of thumb is that the more training data is used, the better the prediction of the model, as illustrated in Figure 4.7 [104]. However, not only the number of samples in the data set determines the performance of the prediction model, but also the number of attributes or features. As shown on Figure 4.8 [104], the prediction of the model can deteriorate when the dimensionality becomes too large in comparison with the size of the data set. This phenomena is also better known as 'The curse of dimensionality', resulting in the fact that high dimensionality can often lead to overfitting of the data and thus to a reduced model performance [40, 104].

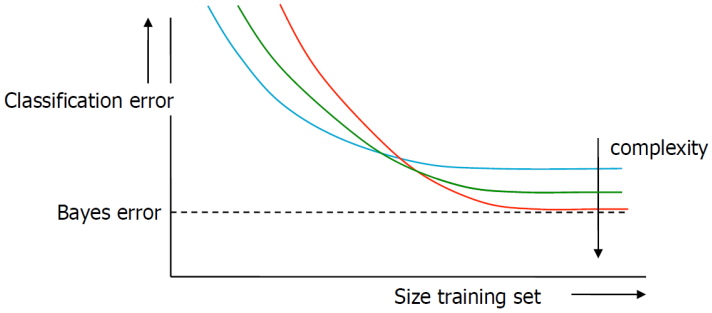


Figure 4.7 – The relation between error and size of data set used for training for ML models [104].

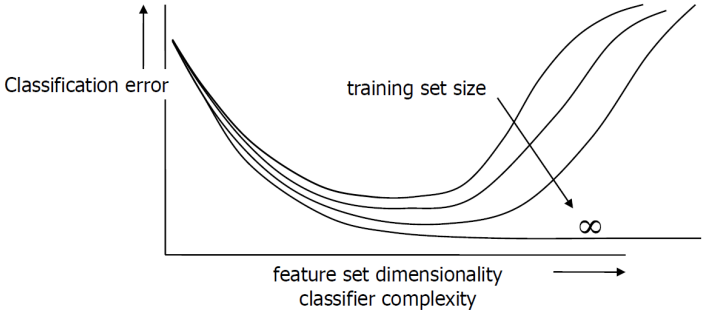


Figure 4.8 – The relation between error and dimensionality for ML models [104].

The random forest model requires less data than neural networks, and are less prone to overfitting [92]. Li *et al.* [105] have developed a scalable RF model by making use of cloud computing. This model can handle up to 110 GB of data, with about 10 million records and 1000 features. This is out of the scope of this research project, and it is therefore deemed that random forests are scalable for the application in this research.

In the area of deep learning is the most common used method Deep Belief Networks, which is a combination of multiple Restricted Boltzmann Machines. In general, this method outperforms classic machine learning ensemble trees, such as random forest, but they also require a lot of data to train the model, and are computationally heavy [106]. Additionally, these methods also reduce the need for feature selection and reduction, as proven by Yu *et al.* [60] [106].

Support Vector Machines are generally outperformed by other non-linear models, as discussed earlier in this chapter [84]. Additionally, SVM models do not have good scalability characteristics, due to their high computational time and memory usage [106]. Also multivariate linear regression models suffer from these characteristics, as they have a high training time and are not guaranteed to perform well on large data sets [106].

In machine learning models, the data set also needs to be divided into a training and test data set. The most common ratios have been investigated using past studies. Rebollo and Balakrishnan [87] used a test-training data set ratio of 75%-25%. Yu *et al.* [60] used 60% of the flights for training, 20% for testing and the remaining 20% for validation. Rodriguez-Sanz *et al.* [32] used a ratio of 90% and 10% to construct and test the Bayesian Network.

As discussed before, the search space of finding the optimal Bayesian Network structure from data grows exponentially with the number of attributes in the data set [7, 82]. However, as the relationships between the variables or nodes need to be statistically significant, Bayesian networks still require a large number of data tuples or records. In Zuk *et al.* [83], the number of samples needed to find the real structure of the data in Bayesian networks is researched. Here, a distinction is made between the probability to underfit, thus not capture the actual patterns in the data set, and the probability to overfit, where the structure is too adjusted

to the input data set. It was found that the probability for underfitting errors decades exponentially with an increase in the number of samples. For overfitting, this is not the case, and when using a data set with a large number of samples, overfit errors become more likely than underfitting errors.

In contrast to what is discussed above, Rodriguez-Sanz *et al.* [32] states that BNs can efficiently perform inference, even for a data set with a high number of variables. Also Kalisch and Buhlmann [107] acknowledges that the exponential growth of the search space is a problem, also in the algorithm that they studied, the PC algorithm. However, Kalisch and Buhlmann also claim that when a Bayesian Network structure is sparse, the PC algorithm search space no longer grows exponential, making the computational load feasible. In order to say that a BN is sparse, the number of possible node connections or parent nodes are limited per variable, and one can assume that the number of nodes are approximately equal to the number of links [14].

In Table 4.5, the past studies which are somewhat comparable to this research project, their methodology and used data is summarised. It can be seen that the amount of data samples used in the studies vary significantly. For example, both Fernandes *et al.* [84] and Yu *et al.* [60] used (deep) neural networks and support vector machines to make flight delay predictions, but Fernandes *et al.* only used a data set with a size of less than 2% compared to Yu *et al.* [60]. However, Yu *et al.* used deep neural networks, which needs a lot more training data than conventional and supervised machine learning methods [106]. Rebollo and Balakrishnan [87] used random forests and used as little as 3000 data samples to train the model. Additionally, both Fernandes *et al.* [84] and Sridhar *et al.* [85] used feedforward neural networks where the input data was limited to 1000-5000 data samples, which both attained an accuracy of around 70%.

Both Rodriguez-Sanz *et al.* [32] and Truong [73] used Bayesian networks to find causal relationships for flight delay. It can be seen from Table 4.5 that these studies actual use the highest number of attributes out of all listed studies, although Bayesian networks do not scale well with the dimensionality of the data set. Additionally, the difference in used data records of the two studies is large. In Truong [73], the author claims that the number of data samples should be 100 times as large as the number of attributes, but does not use this rule of thumb in the performed research.

Sternberg *et al.* [3] used almost 3,000,000 data samples for frequent pattern mining, using the Apriori algorithm. However, in frequent pattern mining, the computational time is mostly determined by the number of attributes and their range possible values. In Sternberg *et al.* [3], only 16 attributes were present in the data set, but it is expected that the number of features of the used data set in this research will be larger due to the large amount of data sources, as discussed in chapter 5.

Table 4.5 – Summary of previous research work and data used.

Reference	Methodology	Application	Nr. Data samples	Nr. Attributes
Rodriguez-Sanz *et al.* [32]	Score based Bayesian Network	Madrid Airport	34,000	51
Truong [73]	Bayesian Network	US airport delay	1058	43
Fernandes *et al.* [84]	Data-based sensitivity analysis of NN, SVM & RF	EU charter flights	5484	33
Yu *et al.* [60]	Deep belief network model	Peking airport	528,471 total 317,082 training	16
Rebollo and Balakrishnan [87]	Random Forest	US delayed airports	3000 training 1000 test	..
Sridhar *et al.* [85]	Multiple linear regression &NN	US flight delays due to weather	Varying between 730-1293	...
Sternberg *et al.* [3]	Frequent pattern data mining Apriori algorithm	Brazilian domestic flights	2,818,898	16

As introduced in the next chapter, the number of data records available for this research is in the order of 30,000. Comparing this to earlier studies, this might not be sufficient to use deep neural networks as they require a lot of data to be trained correctly, as Yu *et al.* [60] used almost 10 times as many data samples. Additionally, compared to the data samples used in earlier studies for random forest or feedforward neural networks, as can be seen from Table 4.5, there is sufficient data available to use these methods.

Compared to the previous studies on Bayesian networks and the data set size, it can be said that there is

sufficient data available to obtain meaningful results. The number of attributes is here the most constraining factor, as many data sources from the different involved stakeholders are combined together.

4.5.3. CONCLUSION

Taking into account the presented performance, limitations and data scalability of the proposed methods, the most suitable methods for this application can be selected.

For the baseline method, statistical analysis and frequent pattern mining have been discussed. It has been shown that the statistical methods, such as correlation, regression and Granger causality can either not capture complex dynamics between the variables, or they are not suitable due to the limitation to time series analysis. Therefore, frequent pattern mining will be used, as it has been shown to discover underlying complex relations between the variables in a study by Sternberg *et al.* [3]. Although the Apriori method is most commonly applied and used by Sternberg *et al.*, the FP-growth method has better scalability features and has not yet been applied in the context of flight delays [68]. Therefore, this algorithm will be used for association rule mining. This method remains a baseline method as it measures the statistical significance and correlation between variables, but does this at the level of the value of each variable, and for multiple variables at the same time. Frequent pattern mining has only been applied once to find the causes of flight delays, and not to the causes of specific ATC delays, to the best of the author's knowledge.

As a state-of-the-art method, Bayesian networks will be used. This method has a number of advantages to be used in causal analysis, as listed earlier in this section. Compared to machine learning with xAI methods, this method can better capture complex dynamics between multiple features and their impact on the dependent variable [32]. The major concern of this method is the scalability with the size of the data set, however, the work of Rodriguez-Sanz *et al.* [32] has proven that this is acceptable for the size of this problem, which is comparable to the data size used in this study.

To the best of the author's knowledge, these methods have not been compared for the causal analysis of ATC delays. Therefore, this work will add to the body of knowledge by assessing and quantifying the performance of both methods.

5

DATA SOURCES & PROCESSING

Due to the introduction of data recording systems, data warehouses and lakes, big data volumes have become available in recent years for the aviation industry [108, 109]. This chapter discusses which data and its sources are used in this research project, and which processing techniques exist to transform the raw data into a data set which can be directly used in the developed models. According to Moreira *et al.* [109], the five data processing steps for data mining models are the following: data integration & cleaning, transformation, reduction and data balancing.

5.1. DATA SOURCES

The data of interest is the year 2019. Before the relapse of air traffic in the year 2020, air traffic had been experiencing an average yearly growth of 5.5% [2]. Therefore, the airspace was at its busiest level in 2019, which makes it good data to analyse the Air Traffic Control delays, due to their high occurrence in the saturated airspace.

5.1.1. FLIGHT DATA

KLM flight data was accessed through the data warehouse of the industry partner, KLM Royal Dutch Airlines. The data warehouse contains all flight records from KLM and KLM Cityhopper. The queried flight data was filtered on the following conditions:

- The scheduled departure and arrival time in the year 2019.
- The actual arrival or departure airport is Amsterdam Airport Schiphol.
- The airline code is 'KL', indicating KLM or KLM Cityhopper as operating airline.
- The flight number is < 2000, indicating a commercial flight.

In total, 249,517 flight records are present in the 2019 data set, which is an average of 683 flights a day. This is lower than the average number of flights reported by KLM, which is a result of the filtering conditions above. Out of all flights, 84.57% of all flights was a flight in Europe, with a total of 211,005 flights in total. Some general statistics on the data set are shown in Table 5.1.

The format of this data set is horizontal, meaning that the one data sample represents one flown flight. This must be transformed into a format where each data sample corresponds with a turnaround procedure at Amsterdam Airport, as this research project looks at flights from a cruise-to-cruise point of view, and analyses the turnaround process of one aircraft. This means that the in- and outbound flight of the same aircraft needs to be coupled to form one data sample.

As a dependent variable, the last known ATC delay of a KLM flight is used. This variable is available for 24.2% of the outbound flights, for a total of 30,224 flights in 2019. A box plot of this delay information is presented in Figure 5.1. Here, the samples with delay values above 300 minutes or 6 hours have been removed, which were a total of 6 flights or less than 0.1% of the data set. The shown attributes are, last known ATC delay, minimum ATC delay, maximum ATC delay, pure ATC delay and local ATC delay. The pure ATC delay is the delay issued by the CFMU, and the local ATC delay is the delay which is caused by congestion at the departure airport, in this case AMS.

The distribution of arrival and departure delay for KLM flights at AMS is shown in Figure 5.2. It can be observed that the distribution of departure delays is shifted to the right compared to the arrival delay, and the majority of the flights have positive flight delay. This indicates that many flight delays are incurred at the airline's hub.

Additionally, the average number of KLM flights and delays in function of time of the day are plotted in Figure 5.3 and 5.4. The time of the day is expressed in Coordinated Universal Time (UTC). The plots have been split for EUR and ICA flights, and it can be observed that the flight and delay distribution shows different trends. For EUR flights, the departure delay shows an increasing pattern over the course of the day, which cannot be seen for ICA flights. This is the result of the high fleet utilisation for European flights, with shorter flight and turnaround times.

Table 5.1 – Statistics on KLM flights at AMS from 2019.

Metric (%)	Inbound	Outbound
Delayed	43.17	77.46
EUR	84.57	84.57
ICA	15.43	15.43
EUR & Delayed	45.27	77.09
ICA & Delayed	31.66	79.51
Regulated	53.50	20.68
Multiple regulations	13.88	5.83

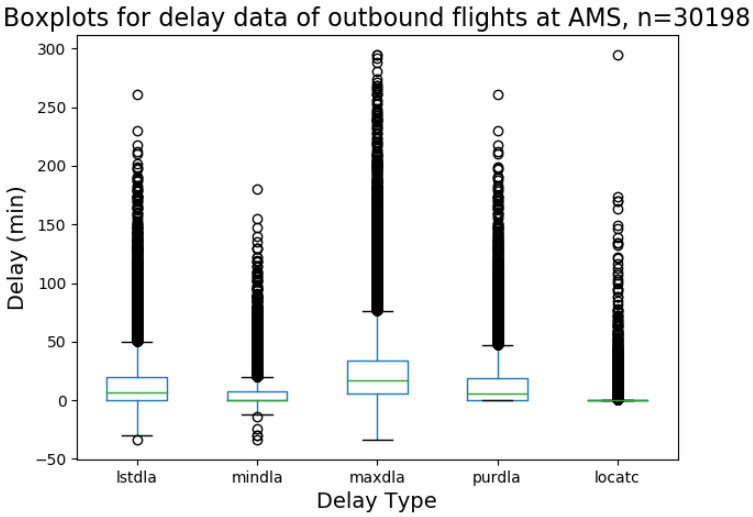


Figure 5.1 – Boxplot of the ATC delay of outbound KLM flights in 2019.

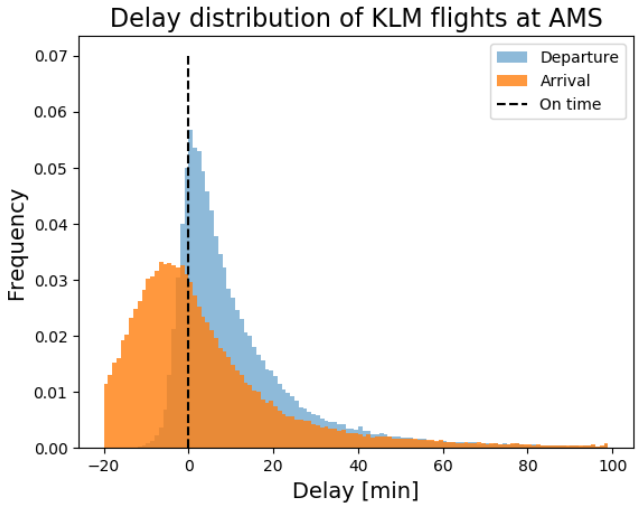


Figure 5.2 – Distribution of the delay of KLM flights at AMS in 2019.

From that same data warehouse, the flight plan data or route data of all KLM flights is available. As the spatial scope of this research is limited to the Dutch airspace, all the waypoints which are situated outside of this area have been eliminated.

5.1.2. A-CDM DATA

Data from the A-CDM system at Amsterdam Airport Schiphol could be retrieved from the airport's data base, which is accessible through KLM. Here, the access is restricted to flights which were handled by KLM ground handling services. This includes operations from KLM, KLM Cityhopper, Transavia, Alitalia,..

The retrievable A-CDM data is different for inbound and outbound flights. For all flights, variables such as flight number, call sign, runway usage, aircraft type and registration and the CDM flight state (refer to chapter 2) are present. Additionally, the data set for inbound flights contains the following information: SIBT, ELDT, ALDT, EIBT, ramp & gate, the consecutive flight number and day of departure of the following flight.

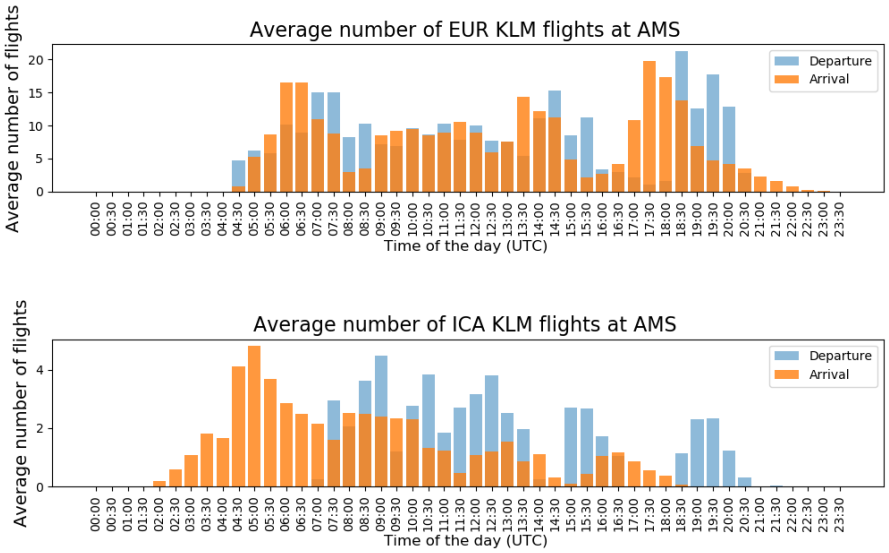


Figure 5.3 – Time distribution of the average number of KLM flights at AMS in 2019.

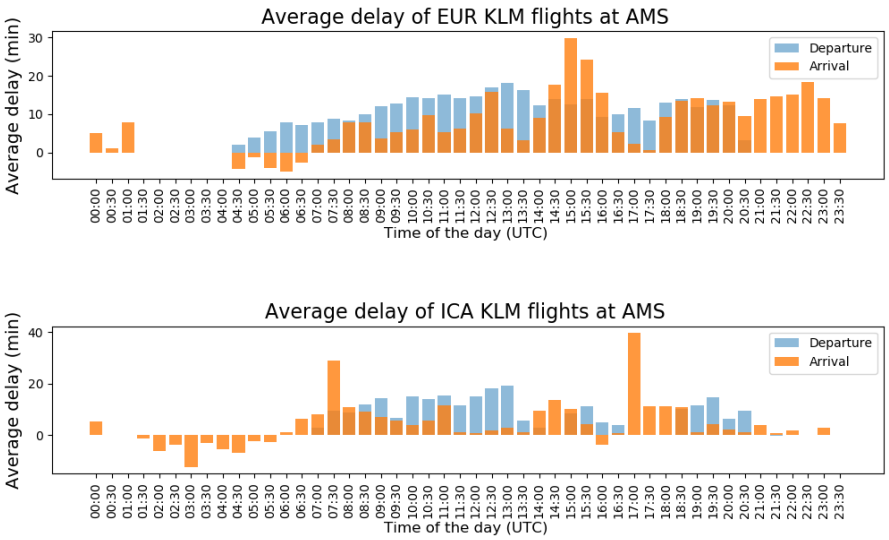


Figure 5.4 – Time distribution of the average delay of KLM flights at AMS in 2019.

The data set for outbound flights contains other CDM data and milestones: SOBT, TOBT, EOBT, TSAT, ASRT, EXOT, AOBT, TTOT, CTOT, the used SID.

The A-CDM data is a vertical data set, which means that every update in the A-CDM process of a flight is registered as a new data sample. As all other flight and route data is a horizontal data set, this data will need to be aggregated to an individual sample per flight by extracting new features from the data set.

5.1.3. OPERATIONAL DATA

Operational data in the Amsterdam FIR and at Amsterdam Airport is received from LVNL, the Air Navigational Service Provider in the Dutch airspace. The following data has been made accessible to be used in this research project:

- Runway usage & capacity declarations
- (Anonymous) demand data on departure/arrival at AMS and the IAFs in the AMS FIR per 20 min.

- Outbound traffic data containing most penalising regulation, ETOT and ATOT.
- Inbound traffic data containing most penalising regulation, ETA and ATA.
- Regulation data including timing, reduced capacity rate, point of application and reason for regulation.

5.1.4. AIRSPACE INFORMATION

Information on the location of the different waypoints in the Dutch airspace could be retrieved from open source navigation website [110]. The information regarding the sectors and the IAFs is found from the (electronical) Aeronautical Information Package [13]. The information of the CTA sectors was retrieved as the coordinates defining the sector edges, which could be transformed into an area defined by these edge points.

5.1.5. WEATHER DATA

Hourly weather data on different weather stations across the Netherlands is available from the public database of the weather institute of the Netherlands, the KNMI [111]. The locations of these weather stations are plotted on the Dutch airspace division, as seen in Figure 5.5. The data includes wind direction and average speed, maximum wind gust, temperature, dew point temperature, precipitation duration and hourly amount, air pressure, horizontal visibility, cloud coverage, relative humidity, weather code, and the presence of fog, rainfall, snow, thunderstorms and ice formation.

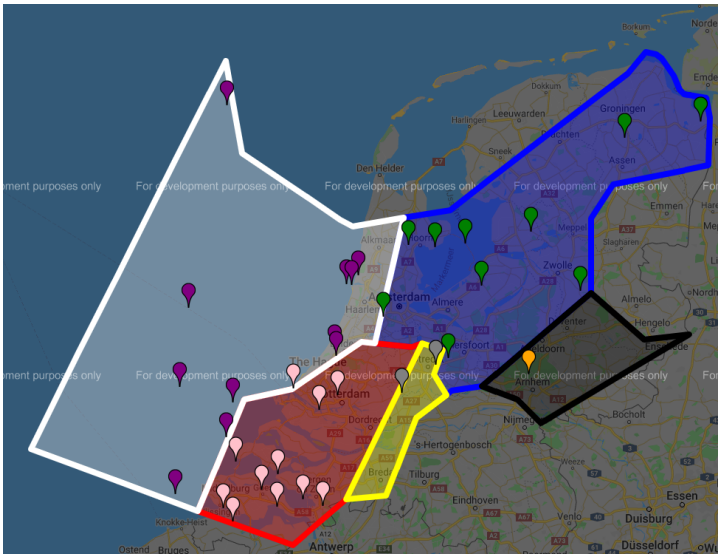


Figure 5.5 – Plot of the available KNMI weather stations and their location per CTA sector.

5.2. INTEGRATION & CLEANING

This section discusses the different techniques available for data integration and cleaning. The data pipeline and integration procedure of the above mentioned data sources and tables is presented.

5.2.1. INTEGRATION

Data integration is the process of combining the data records from different sources [67]. Additionally, this also consist of the detection of discrepancies between the different data sets, such as varying values for the same attribute. As discussed under section 5.1, the used data in this research project originates from different data sources, and therefore comes in different data formats and layouts.

Figure 5.6 presents the data pipeline and integration procedures applied in this project. As it can be seen, the integration process consists of multiple sequential merging procedures, due to the large number of sources and associated data tables. It can be seen that most data needs to be pre-processed, which then leads to an initial data table of the data available from that source. After this, several merging procedures have to be done. It is important to split up the merging of the data into several sub processes, such that data tables can be

integrated on a common column. For each integration procedure, the column(s) or key(s) on which the data tables will be matched and merged are written next to the respective arrow. When all merging procedures are done, the final data set can be reformatted, such that the in- and outbound flight of one aircraft are linked with each other to form a single data sample.

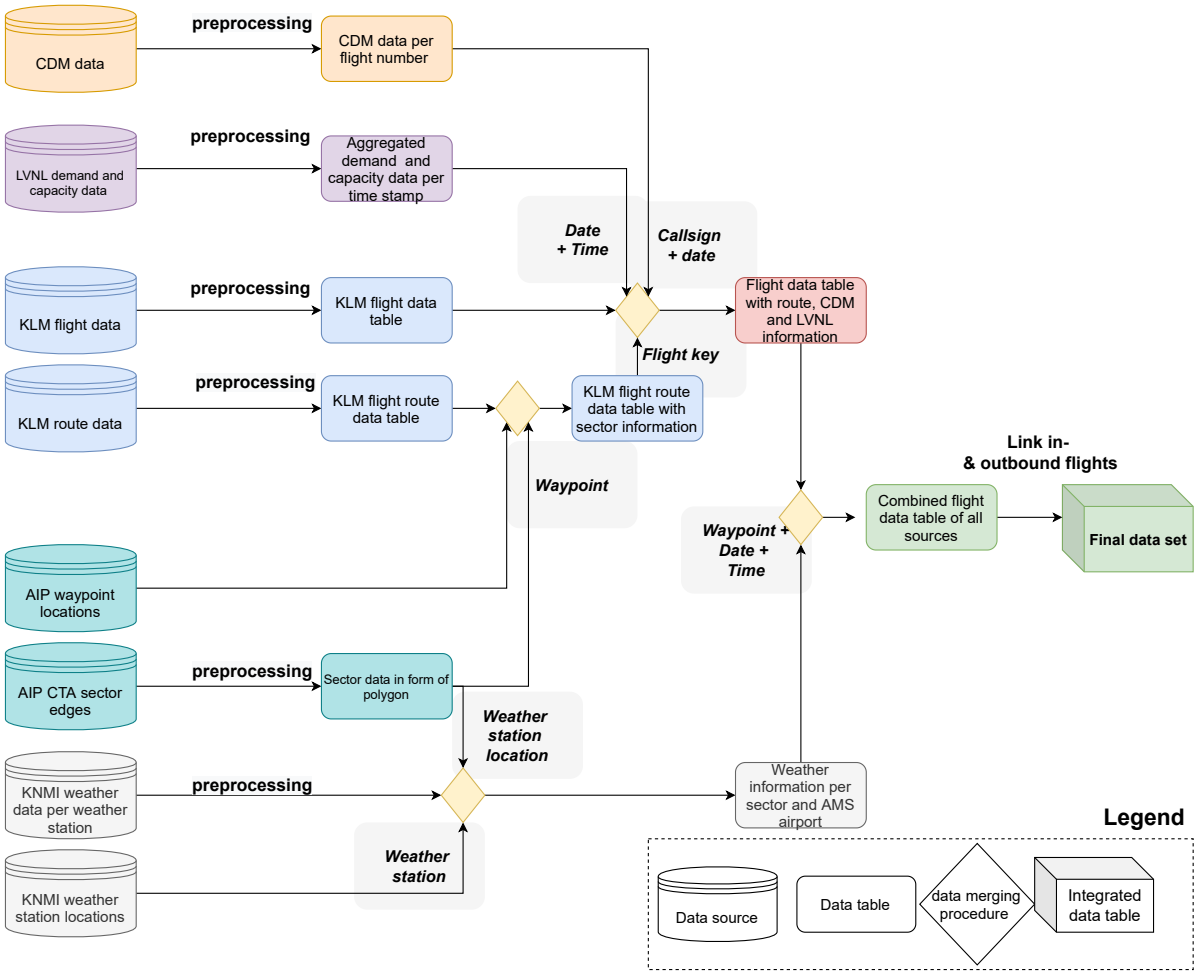


Figure 5.6 – Data integration pipeline.

5.2.2. CLEANING

Data cleaning consist out of the following steps: outlier removal, noise smoothing and handling missing data values [67]. Each of them are discussed below.

Missing Values

for missing data values, there are several options to consider them in the data set [67]:

1. Ignore the record with missing attribute(s).
2. Fill in the missing attributes manually.
3. Use a global constant such as NaN, 'Not known' or ∞.
4. Use a global constant of the attribute such as mean or median.
5. Use a global constant, such as mean or median, of the attribute over the class of that data record.
6. Estimate the value of the missing attribute by using regression or decision tree models.

The first option is the simplest, however it lacks efficiency. It may be a good option in case the data record has many or multiple missing attributes [67]. The second option is not often used as it requires too much

manual work, and the missing value may not be known. Options three to six are good options to retain the data record in the data set, but add bias in the data as the filled in value can be incorrect [67].

As discussed in the previous section are outbound flights with a missing value for the ATC delay omitted from the data set, as it is crucial that the dependent variable is known in order to analyse its causes from observational data. However, for other attributes, the other purposed methods can be used. It should be noted that all the gathered data is data retrieved from actual observations, and it is therefore likely that values will be missing in a large part of the data sets. This should however not always be a reason to discard the observation.

Noise smoothing

Noise in the data can be removed by binning, where the bins each consist out of the same number of observations, and then the bin can be smoothed by replacing all values in the bin by the average of all observations of the bin, or by replacing all values by the nearest edge of the bin [67]. These methods are denotes as smoothing by bin means and bin boundaries respectively. Additionally, also regression can be used for data smoothing [67].

Outlier removal

Outliers or anomalies are data points that are different than the other data samples, and thus by definition have a low frequency in the data set [67]. The removal of outliers is important in order to eliminate the samples which are inaccurate and contain bias, and thus to prevent this bias to be retrieved in the found results [32, 67]. Han *et al.* [67] stated the following: *"Outliers are interesting because they are suspected of not being generated by the same mechanisms as the rest of the data (Han* et al.*,2012, p 544)"*. In the context of this research, the delay outliers might have different causal mechanisms, which can influence the found results. However, Rodriguez-Sanz *et al.* [32] has been careful with outlier removal on a particular set of variables. According to Wu [112], the dynamics for short and long delays are very different. Therefore large delay values should not immediately be regarded as outlier, as the data records can still hold valuable information regarding the causes of long flight delays. Therefore, the choice of removing or including outliers in the data set can have large implications on the obtained results, and a distinction can be made between *error* and *interesting* outliers [113].

Outlier detection can be done in various ways [113]. One of the best known methods is the use of a box plot [113], of which an example is shown in Figure 5.1. It can be seen that for the ATC delays, many outliers are detected using this method. Outlier detection can also be done by approximating or assuming a normal distribution, where outliers are standard identified as values which deviate more than three times the standard deviation from the mean [67]. However, this method has to make an assumption on the distribution of the data. In order to overcome this, histograms can be used, which are non-parametric [67]. The basic idea in this method is to identify all values that do not fall in a bin of the histogram as outliers, however the specification of the bin width remains a user set parameter, which is hard to optimise [67]. Additionally, outlier detection can be done by performing clustering, where the data records that fall outside of the identified clusters are considered as outliers [67]. The details on clustering techniques are discussed in more detail in the next section.

5.3. TRANSFORMATION

The goal of data transformation is to represent the data in a better way, such that the model(s) can better understand the data, and identify patterns in it [67].

5.3.1. NORMALISATION

Normalisation is a technique where the attribute's values are mapped onto a new range, typically [0,1] or [-1,1]. The aim of this is to reduce the extra weight an attribute might have due to its high values, which is standardised by applying normalisation. Again, there are multiple methods for normalisation, such as min-max normalisation and z-score normalisation, shown in Equation 5.1 and 5.2 for the attribute x [67].

$$x'_i = \frac{x_i - \min_x}{\max_x - \min_x} (\text{new-max}_x - \text{ new-min }_x) + \text{ new-min }_x \qquad (5.1) \qquad x'_i = \frac{x_i - \bar{x}}{\sigma_x} \qquad (5.2)$$

Normalisation is necessary for methods where distance measurements or linear combinations of the input are used, such as for neural networks and clustering algorithms [67]. The aforementioned normalisation procedures are not necessary when using frequent pattern mining or Bayesian networks, but are needed when the data is clustered as part of the data processing.

5.3.2. DISCRETIZATION

Discretization is the process where continuous variables are mapped into discrete values, meaning that the number of values the attribute can take is finite. In order to discretize attributes, different methods can be used, such as conceptual hierarchy, binning and clustering analysis [3, 67].

Concept Hierarchy In this method, attributes are mapped to higher-level concepts [3, 67]. For example, the capacity of a certain sector can be mapped to a higher order attribute such as 'maximum capacity' or 'reduced'. This is a method that should be done manually in the data set.

Binning As discussed before, binning can be used as a smoothing technique, but also as a discretization method. When determining the bins, two methods can be used, namely equal bin width and equal bin frequency [67]. The latter means that each bin should contain an equal number of data observations, but will most likely result in varying bin widths. Determining the number of bins for discretization is important as this determines the quality of the resulting attribute [3].

Additionally, to actually discretize the data, again two methods can be used, bin by mean/median or bin by bin edge [67]. Binning can also be used to generate the hierarchical concepts discussed above, when performed recursively. It should be noted that the method of binning is sensitive to the user selected parameters, such as the bin width, and to outliers [3].

Cluster Analysis Next to the binning technique, data can also be grouped using cluster methods, where the bins are replaced by the identified clusters or groups [67]. As clustering methods use the distance or closeness measures between the values, it performs well for discretization purposes. Additionally to cluster analysis, decision trees can also be used for discretization [67]. This is a supervised learning method, and thus requires training or data of which the discretization is known. This is not applicable in this application, as there is no training data with labels available.

K-means clustering might be the most well known and used clustering algorithm. In this method, it is required that the user determines the number of clusters beforehand [67]. When the number of clusters is not known, density-based clustering algorithms can be used [106]. The best known algorithm in this area is DBSCAN (Density Based Spatial Clustering of Applications in Noise), which separates clusters by finding and distinguishing high and low density areas [106]. However, the computational load of DBSCAN is a disadvantage when working with large data sets, both in the number of samples and the number of attributes [106].

5.3.3. ONE-HOT ENCODING

One-hot encoding is a data processing method which maps categorical attributes into binary and numerical attributes, also known as categorical mapping [109]. This method is necessary for models to be able to interpret and handle the categorical data, which cannot be done in text format. Frequent pattern mining does not require this mapping, as it searches for patterns without using or interpreting the data. However, for Bayesian network learning, a lot of formulas are used, which require numerical features. A downside of one-hot encoding is that the number of features or attributed grows for each attribute that needs to be encoded.

This method can be used to represent the aircraft type for example. This attribute has several possible values, such as 'B737, B772,B787,A330,...'. The most important characteristic to be able to apply one-hot encoding is that the number of categories is finite and limited in size. The aircraft type attribute can then be mapped to binary attributes of the different categories. An example is illustrated below in Figure 5.7 [114].

Tier
None
Silver
Silver
Gold
Platinum
None
Ivory

None	Ivory	Silver	Gold	Platinum
1	0	0	0	0
0	0	1	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1
1	0	0	0	0
0	1	0	0	0

Figure 5.7 – Example of one-hot encoding [114].

5.4. DATA REDUCTION

Data reduction is a general term, which consist of several processes to reduce the data set, namely dimensionality reduction, numerosity reduction and data compression [67]. However, as only dimensionality reduction is relevant for this project, this is the only method discussed.

Dimensionality reduction is a process in which the number of attributes or features in the data set are reduced [67]. Dimensionality reduction can also be seen as attribute selection, as only the most important attributes or features should be kept in the data set [109]. This should be done as irrelevant features increase the probability of model overfitting and therefore reduce the model's performance [40]. To reduce the number of features, two main methods can be distinguished, listed below [67].

1. **Feature Selection** In this method, the number of features is reduced by selecting the most important features from the original set.

2. **Feature Extraction** This type of method extracts new features out of the given data set, with a lower dimensionality.

One of the most used methods for dimensionality reduction is Principle Component Analysis (PCA). This is a feature extraction method, which extracts principle components from the data by performing an eigenvalue and eigenvector analysis of the data set [67]. If the data set consists of m features, these orthogonal vectors will have a length of m. An illustration of this is shown in Figure 5.8 [115]. This example is for a two dimensional data set, and the two orthogonal principle components are denoted in blue and red. In this illustration, it can be seen that the principle component vectors are oriented in the direction of the highest variance in the data. The method can extract as many principle components as there are attributes or features in the data set. However, the principle components are ranked, and can thus be filtered to only keep l vectors, the ones with the highest variance, where $l \leq m$ [67]. PCA is performed as follows [67, 109]:

1. The data is normalised.

2. The principle components or eigenvectors are computed.

3. The principle components are ranked according to variance.

4. The weakest principle components are eliminated, such that the l eigenvectors with the highest variance remain.

5. The original data is projected onto the kept eigenvectors, and thus will only have l features.

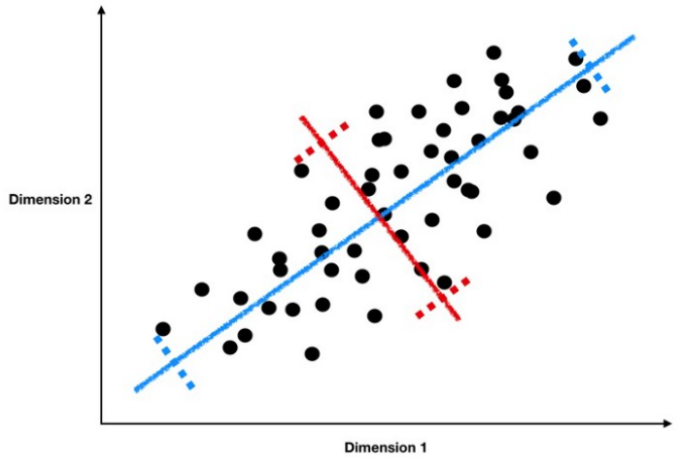


Figure 5.8 – Illustration of the PCA method on a 2-dimensional data set[115].

A drawback of the PCA method is that the features or extracted principle components become less interpretable, as the original features have been mapped onto another feature space. Additionally are the principle

components orthogonal to each other, meaning that they are uncorrelated by definition. As finding complex relationships between the observational data is the objective of this research, PCA might be a less suitable method for dimensionality reduction, due to the loss of interpretability of the features.

Additionally, deep neural networks can also be used as feature reduction techniques, as done in a study by Yu *et al.* [60]. In particular, a Deep Belief Network can be used, which consists out of multiple unsupervised neural networks, such as Restricted Boltzmann [60]. In this method, the number of linked neural networks each extract features from the input, which then again serve as input for the following neural network. However, this method also extracts new features from the attributes, which makes them not interpretable or linked to the actual attributes.

Next to the PCA method, attribute (subset) selection can also be performed to reduce the number of features of the input data set. The goal of this process is to remove irrelevant attributes in order to reduce the dimensionality of the data set, and it is thus a feature selection method [67]. As the number of subsets of attributes is an exponential search space, often heuristic methods are used to determine which attributes are of no or little importance [116]. Often, "greedy" heuristic methods are used, which means that choices are made based on what is the best solution at that point in time. In order to find the most suitable attributes, forward and backward selection can be used [67, 116]. Forward selection implies that the subset is initially empty, and the best attributes are added step by step. Backward selection implies that the subset initially contains all attributes, which are then step by step removed if found to be the least relevant [67]. The assessment whether a feature is kept in the subset or not, is often based on statistical tests, capturing independence between the attributes.

Additionally, *wrapper and filter* approaches can be segregated from each other [116]. A filter method is a method where the feature selection is performed before the actual model is applied, whereas the wrapper approach incorporates them into each other [116]. For filter approaches, the feature selection method is independent of the learning model, but with wrapper methods, the feature selection is combined and evaluated using the learning method [117]. The wrapper method is computationally expensive, as the learning process needs to be repeated for every evaluation of a feature subset [118]. However, an advantage of this approach is that the feature subsets are fully evaluated in their performance, and thus wrapper can achieve an optimal solution in theory, although this is not guaranteed due to computational limits [118].

Drugan and Wiering [116] proposed a feature selection algorithm, which uses the Minimum Description Length (MDL) scoring function which is used to learn a Bayesian network. However, Drugan and Wiering now propose to use the MDL to evaluate the features for the purpose of feature selection. As discussed before is the MDL, also known as BIC, a measure of the log-likelihood of the Bayesian network on the data set. This principle is used as a feature selection method by assessing the log-likelihood of the Bayesian network build using the full data set, and the one built using the subset of features. The difference between the two is then defined as the MDL-FS, the Minimum Description Length - Feature Selection [116]. The stopping criterion for this feature selection algorithm is that when the next found Bayesian network does not outperform the previously found one in terms of MDL-FS value, the algorithm is stopped and the solution is deemed as optimal [116]. Additionally, Ozçift and Gulten [119] have proposed a genetic algorithm in combination with a Bayesian network to perform feature selection. This is again a wrapper approach, as the genetic algorithm is combined with the BN. A genetic algorithm is a heuristic method, which evaluates the fitness of certain feature sets, and then adapts the feature set based on the best performing feature sets of the previous evaluation round.

Combining a wrapper approach with a Bayesian network learning method will be very computationally demanding, as the task of Bayesian network learning is already computationally complex when the set of features is fixed. Therefore, a wrapper approach for feature selection is deemed to be too computationally heavy for this application.

5.5. DATA BALANCING

For classifiers, an uneven distribution of the different classes in the data set leads to the inability to evaluate the performance of classifiers [109]. A classifier can namely only be truly evaluated in its performance when trained and tested on a balanced data set [87]. Over-sampling from minority data is a method that can be used to ensure that the minority classes are well represented in the data and the data set becomes balanced. The problem of data imbalances has mostly been researched and addressed for classification problems, however, also for regression models data imbalance can have a negative impact on the model's performance [120].

One of the most straightforward methods to balance the data is sampling. It can also be used as a data

reduction method, as it can represent data in smaller subsets. Data sampling thus reduces the data set in length, not in the number of attributes or features. The most two common methods for data sampling are random sampling and stratified sampling [109]. Random sampling is the process where each data tuple or record has the same probability to be part of the subset, whereas in stratified sampling the data set is split into several sets according to an attribute, and one sample is drawn from each of these attribute sets. For data imbalance, a popular method is the SMOTE, Synthetic Minority Oversampling Technique [109, 120]. This method generates 'synthetic' or new data records which are similar to the data samples in the minority group, and thus 'oversamples' this group in the data [67, 109]. For classification problems, the class with less data records is then oversampled, in order to achieve a balanced data sample [109]. In regression problems, the target variable is continuous, and it is the goal to oversample values which are less frequent or rare [120].

In this research and the used methodology, data balancing is not needed. As concluded in chapter 4, frequent pattern mining and Bayesian networks will be used as causal models. Both these methods work with the frequency of values in the data set to mine relationships from the data set. Therefore, adjusting the data set to obtain a balanced data set will change the frequency in the data set and therefore negatively impact the results.

6

RESEARCH APPROACH

The research questions presented in chapter 1 need to be answered following a clear methodology, for which the different models and methods have been discussed and selected in chapter 4 and 5. This chapter therefore presents the taken approach to apply the purposed methodology to the case study of ATC delay for KLM flights at Amsterdam Airport. First of all, the work is divided into several work packages, each consisting out of specific tasks. Additionally, the sequential timeline and dependencies between these work packages are illustrated using a work flow diagram.

6.1. APPROACH

The required tasks have been grouped together and divided into several work packages. The first two work packages, and the largest ones, are the processing of the data and the implementation of the decided methodology into a software environment.

In order to guarantee that all steps are correctly performed, a thorough verification process needs to be implemented. Verification is needed both in the data processing task, as well as for the implementation of the causal models in a programming environment. Verification will be done by using simple test cases of which the result is known beforehand. Additionally, it is done continuously throughout the data processing and methodology development, as it needs to be incorporated in every step of the work packages before the next step can be initiated. This continuous verification should prevent the discovery of mistakes at a late stage, which leads to an increased workload and complexity to correct the made mistake.

Once the results of both causal models have been obtained, the validation process will start. This is the process used to confirm that the obtained results represent reality. The validation strategy in this research is to keep validation data separate from the data set used to generate the results. This allows to validate the obtained results by comparing the results from the actual and the validation data set, which should be comparable and thus show that the method actually performs in a consistent manner.

As described above, the experimental set-up of this research project consists out of the creation of one or more causal analysis methods and its implementation into a computer programme. The hardware used in this research project is the laptop received from KLM to perform the research project, a Lenovo ThinkPad, model DEPLOY W10 STD V1909.01 / Node W10 STD US V1909.01. The processor on this machine is the Intel(R) Core(TM) i5-6300U CPU @ 2.4GHz 2.5 GHz, with an installed memory of 8 GB RAM. For compatibility reasons, it is highly preferred that the used software for both the data processing and causal analysis are the same. Therefore, Python, version 3.7, will be used throughout the research project, as its has a large availability of compatible modules and libraries, suitable for both the data extraction and processing, as well as for the causal model. Additionally, Python offers a wide range of libraries compatible with SQL (Structured Query Language), which is necessary in the project to access and extract data from the KLM databases.

6.2. WORK FLOW DIAGRAM

The developed work flow diagram for the next stage in this research project is presented in Figure 6.1. It can be seen that the data processing and methodology development work packages are connected with the verification work package with a recurrent arrow. This indicates a continuous or iterative process, meaning that verification will be executed at each stage of the data processing and methodology development.

The planned duration of each of these work packages is shown in Table 6.1. The data processing and methodology development are by far the largest work packages, and they are the first step in the remainder

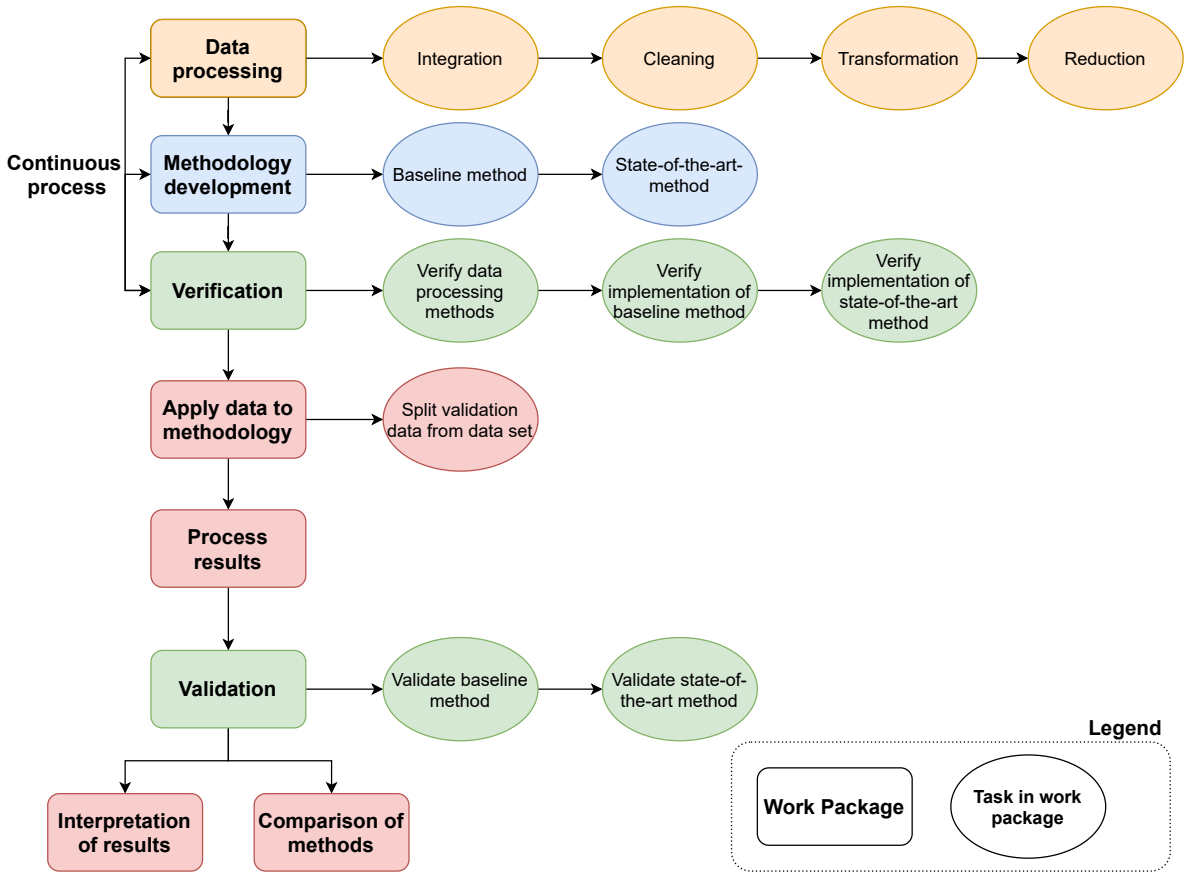


Figure 6.1 – Work flow diagram for the remainder of this research project.

of this project. For each of the work packages, time to document the steps taken and the made decisions has been included into the planned duration. Additionally, time has been reserved to allow for some contingency in the planning, as well as for holidays. As seen in Table 6.1, the completion of all work packages will take approximately 24 weeks. After this period, the green light review is planned, which is followed by completing the thesis report and finally the defence. According to the planning below, the green light review would be planned in the week of the 16th of August 2021. A detailed overview of each task duration and planning of the project can be found in Appendix A, in the form of a Gantt chart.

Table 6.1 – Planned duration of the work packages.

Work Package	**Duration (working weeks)**
Data processing	8
Methodology development	6
Verification	1
Apply methods to case study	1
Process results	2
Validation	1
Comparison of methods	1
Interpretation of results	1
Contingency	1
Holidays	2
Total	24

7

CONCLUSIONS

This chapter presents the conclusions that can be drawn from the performed literature review on the subject of the root causes of the Air Traffic Control delays for KLM flights at Amsterdam Airport Schiphol. The main research objective of this master thesis is to expose the drivers of the ATC delays encountered by KLM flights at Schiphol Airport, by performing a root cause analysis of these ATC delays and their impact on the KLM network.

ATC delays consist out of two types, ATFM delays and start-up delays. The former are a direct result of regulations in the airspace, such that the aircraft is kept on the ground. Start-up delays are delays assigned to a flight before departure, but the flight is not necessarily regulated.

It has been found that the system of Airport-Collaborative Decision Making captures the turnaround and pre-departure dynamics of a flight at Schiphol, which can contain valuable information on the causes for ATC delays. Additionally, regulations play an important role in ATC delays, for inbound flights at Amsterdam Airport, which also affects the departure process and its on time performance. Additionally, application of regulations can lead to air traffic bunching, resulting in unexpected arrival peaks. In order to prevent this, the capacity in regulations can be declared to be lower, in order to reduce this risk of exceeding the capacity.

Causal relationships are traditionally investigated using experiments. However, in some applications, the execution of experiments is very impractical or even impossible, as for this research. Therefore, a causal analysis model has to be used to discover the causes of ATC delays from pure observational data. Two models will be used, a baseline model and a state-of-the-art model, such that their performance and results can be compared.

For the baseline model, several statistical methods have been investigated. This includes the well known and used correlation coefficients, regression analysis and Granger causality, which is a statistical test that can find causal relationships between time series. Additionally, frequent pattern mining has been researched, a method which originates from the retail industry. This method mines the data set to find frequent patterns in the data, using statistical significance and correlation measures. In past research, several algorithms have been developed to perform frequent pattern mining, of which the most important ones are Apriori, FP-growth and Eclat. The former two can be applied to horizontal data sets, and are thus applicable in this research. The Apriori algorithm is computationally heavier than the FP-growth algorithm, as the former first generates all possible rules from the data set. FP-growth uses a tree structure, which eliminates the need for the generation of all possible rules. Additionally, FP-growth has never been used to study flight delays and its drivers. Therefore, the FP-growth algorithm has been selected to perform frequent pattern mining as baseline method.

For the state-of-the art model, both Bayesian networks and machine learning with an explainable artificial intelligence method have been explored. Bayesian networks are very suitable for the analysis of causal relationships, due to their probabilistic characteristics and ability to use multiple attributes together. Bayesian networks can be constructed using expert knowledge on the subject, but this is not suitable for big data applications with complex relations. However, Bayesian networks can also be learnt from observational data. This can be done using multiple approaches. The main distinction can be made between score based and constraint based methods. Bayesian networks have the disadvantage that the search space of possible network structures grows exponentially with the number of attributes, and therefore has low scalability for high dimensional data sets. For machine learning models, it was found from previous studies on flight delay pre-

diction that random forests and neural networks had the best performance. For xAI methods, rule extraction and sensitivity analysis have been investigated. Rule extraction has the disadvantage that it simplifies the original model, causing important information to be lost. For sensitivity analysis, the results only show the relevance and impact of a single feature on the target variable. This has the disadvantage that it can not fully capture the relationships between various attributes and the dependent variable, and can still be computationally heavy in combination with the training time of the machine learning model. Therefore, it has been decided to proceed with a Bayesian network approach to expose the drivers of ATC delays, as this method was found to have the best characteristics for causal analysis of complex relationships.

The data used in this research originates from many different sources, due to the different stakeholders and aspects of ATC delays. The sources include KLM flight data, operational LVNL data, weather and airspace layout information and A-CDM data. The required data tables all come in different formats due to the variety in sources, and therefore the data requires multiple integration processes. Once this is completed, the raw data needs to be processed such that it can be used in the developed causal models. The most important data processing methods for the used methodology are data cleaning, discretization and feature selection.

Once the methodology has been determined, the next steps in this research project had to be planned. The required tasks have been divided into several work packages, which are shown in Table 7.1. The data processing and methodology implementation in a software environment are by far the largest work packages. The software that will be used is Python 3.7, as it has extensive libraries for data processing as well as the causal analysis methods.

The duration of each of these work packages has been estimated and is presented in Table 7.1. In total, the next phase of this project will take up about 24 working weeks, taking into account some contingency and holidays. After this, the green light review will be planned, which is followed by the complete documentation of the project, and finally the thesis defence.

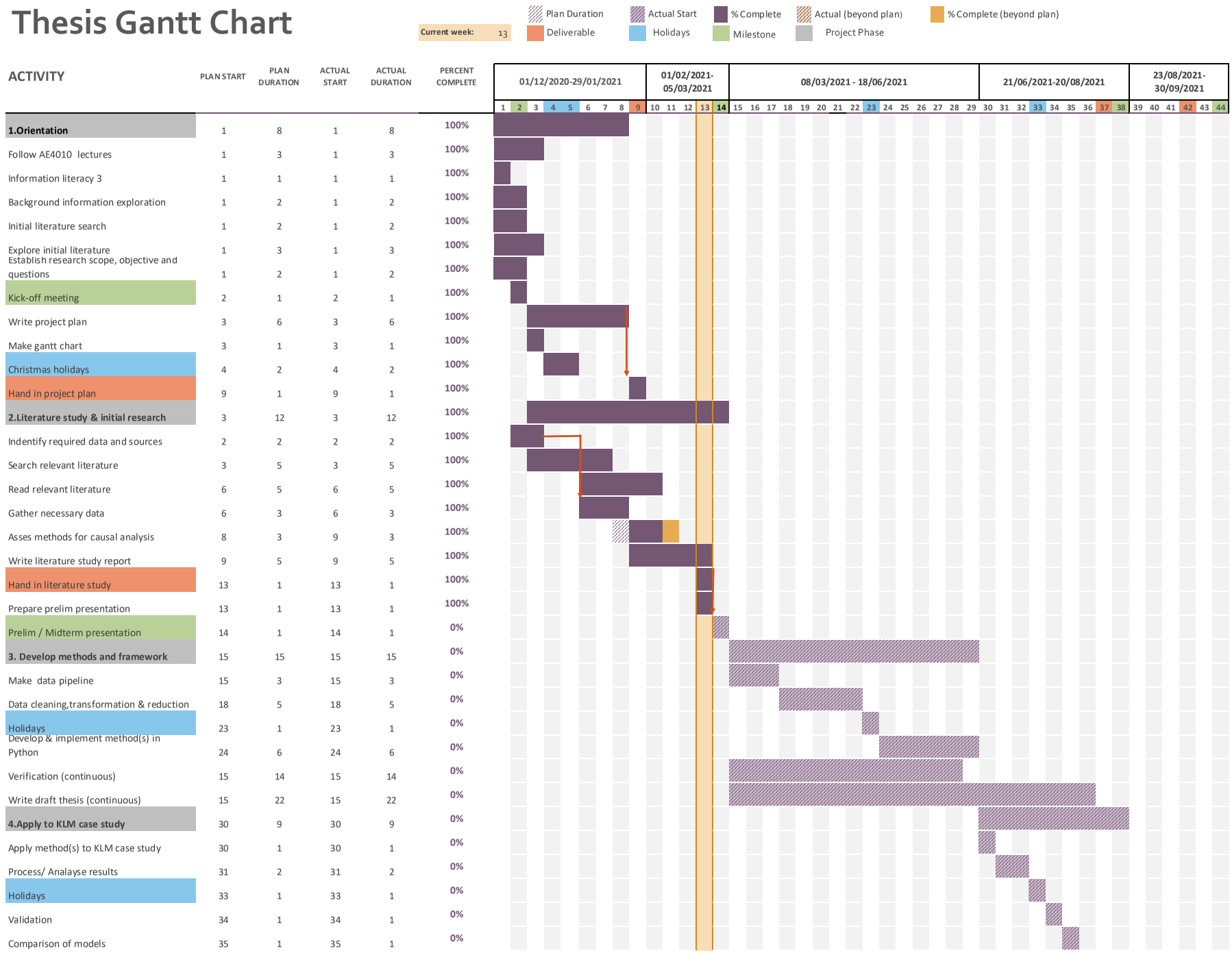
Table 7.1 – Work packages and their planned duration for the next phase of the research.

Work Package	**Duration (working weeks)**
Data processing	8
Methodology development	6
Verification	1
Apply methods to case study	1
Process results	2
Validation	1
Comparison of methods	1
Interpretation of results	1
Contingency	1
Holidays	2
Total	24

A

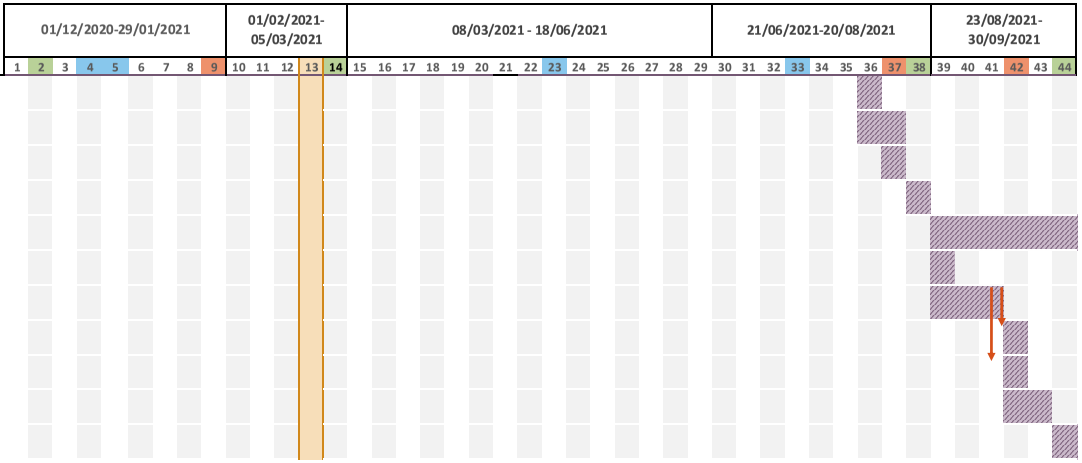
GANTT CHART

Thesis Gantt Chart



ACTIVITY	PLAN START	PLAN DURATION	ACTUAL START	ACTUAL DURATION	PERCENT COMPLETE
1.Orientation	1	8	1	8	100%
Follow AE4010 lectures	1	3	1	3	100%
Information literacy 3	1	1	1	1	100%
Background information exploration	1	2	1	2	100%
Initial literature search	1	2	1	2	100%
Explore initial literature	1	3	1	3	100%
Establish research scope, objective and questions	1	2	1	2	100%
Kick-off meeting	2	1	2	1	100%
Write project plan	3	6	3	6	100%
Make gantt chart	3	1	3	1	100%
Christmas holidays	4	2	4	2	100%
Hand in project plan	9	1	9	1	100%
2.Literature study & initial research	3	12	3	12	100%
Indentify required data and sources	2	2	2	2	100%
Search relevant literature	3	5	3	5	100%
Read relevant literature	6	5	6	5	100%
Gather necessary data	6	3	6	3	100%
Asses methods for causal analysis	8	3	9	3	100%
Write literature study report	9	5	9	5	100%
Hand in literature study	13	1	13	1	100%
Prepare prelim presentation	13	1	13	1	100%
Prelim / Midterm presentation	14	1	14	1	0%
3. Develop methods and framework	15	15	15	15	0%
Make data pipeline	15	3	15	3	0%
Data cleaning,transformation & reduction	18	5	18	5	0%
Holidays	23	1	23	1	0%
Develop & implement method(s) in Python	24	6	24	6	0%
Verification (continuous)	15	14	15	14	0%
Write draft thesis (continuous)	15	22	15	22	0%
4.Apply to KLM case study	30	9	30	9	0%
Apply method(s) to KLM case study	30	1	30	1	0%
Process/ Analayse results	31	2	31	2	0%
Holidays	33	1	33	1	0%
Validation	34	1	34	1	0%
Comparison of models	35	1	35	1	0%

ACTIVITY	PLAN START	PLAN DURATION	ACTUAL START	ACTUAL DURATION	PERCENT COMPLETE
Interpretation of results	36	1	36	1	0%
Prepare green light meeting	36	2	36	2	0%
Hand in draft thesis	37	1	37	1	0%
Green Light Meeting	38	1	38	1	0%
5. Finalize & defend	39	6	39	6	0%
Implement green light feedback	39	1	39	1	0%
Write thesis final version	39	3	39	3	0%
Hand in final thesis	42	1	42	1	0%
Upload final thesis on TU Delft repository	42	1	42	1	0%
Prepare final presentation & defense	42	2	42	2	0%
Thesis defense	44	1	44	1	0%



BIBLIOGRAPHY

[1] LVNL, *Beeldbank | LVNL,* (), accessed on 08-01-2021.

[2] International Civil Aviation Organisation, *Facts and Figures :World Aviation and the World Economy,* Accessed on 06-01-2021.

[3] A. Sternberg, D. Carvalho, L. Murta, J. Soares, and E. Ogasawara, *An analysis of Brazilian flight delays based on frequent patterns,* Transportation Research Part E: Logistics and Transportation Review **95**, 282 (2016).

[4] Y. Wang, Y. Cao, C. Zhu, F. Wu, M. Hu, V. V. Duong, M. Watkins, B. Barzel, and H. E. Stanley, *Universal patterns in passenger flight departure delays,* Scientific Reports **10** (2020), 10.1038/s41598-020-62871-6.

[5] I. Vlachos and Z. Lin, *Drivers of airline loyalty: Evidence from the business travelers in China,* Transportation Research Part E: Logistics and Transportation Review **71**, 1 (2014).

[6] A. Sadiq, F. Ahmad, S. a. Khan, J. Valverde, T. Naz, and M. Anwar, *Modeling and analysis of departure routine in air traffic control based on Petri nets,* Neural Computing and Applications **25**, 1099 (2014).

[7] C.-L. Wu and K. Law, *Modelling the delay propagation effects of multiple resource connections in an airline network using a Bayesian network model,* Transportation Research Part E: Logistics and Transportation Review **122**, 62 (2019).

[8] N. Xu, G. Donohue, K. B. Laskey, and C.-H. Chen, *Estimation of delay propagation in the national aviation system using bayesian networks,* in *Proceedings of the 6th USA/ Europe Air Traffic Management Research and Development Seminar* (Baltimore, MD, USA, 2005) p. 11.

[9] N. Pyrgiotis, K. M. Malone, and A. Odoni, *Modelling delay propagation within an airport network,* Transportation Research Part C: Emerging Technologies Selected papers from the Seventh Triennial Symposium on Transportation Analysis (TRISTAN VII), **27**, 60 (2013).

[10] B. Campanelli, J. J. Ramasco, P. Fleurquin, V. M. Eguíluz, I. Etxebarria, and A. Arranz, *Data-driven modelling of the tree of reactionary delays,* in *International Conference on Research in Air Transportation* (Istanbul Technical University, Turkey, 2014) p. 8.

[11] Eurocontrol, *Annual Network Operations Report 2019,* (2019).

[12] J. Hoekstra and J. Ellerbroek, *Lecture slides of AE4321 Air Traffic Management,* Delft University of Technology (2020).

[13] LVNL, *Aeronautical Information Package for The Netherlands,* (), accessed on 29-01-2021.

[14] G. Gurtner, S. Vitali, M. Cipolla, F. Lillo, R. Mantegna, S. Micciche, and S. Pozzi, *Multi-Scale Analysis of the European Airspace Using Network Community Detection,* PloS one **9** (2014), 10.1371/journal.pone.0094414.

[15] S. Stolz and P. Ky, *Reducing Traffic bunching through a more Flexible Air Traffic Flow Management,* in *Proceedings of the USA/FAA Air Traffic Management R&D Seminar 2001* (Sante Fe, New mexico USA, 2001).

[16] J. Bronsvoort, P. Zissermann, S. Barry, and G. McDonald, *A Framework for Assessing and Managing the Impact of ANSP Actions on Flight Efficiency,* Air Traffic Control Quarterly **23**, 29 (2015).

[17] T. Lehouillier, F. Soumis, J. Omer, and C. Allignol, *Measuring the interactions between air traffic control and flow management using a simulation-based framework,* Computers & Industrial Engineering **99**, 269 (2016).

[18] S. Ruiz, H. Kadour, and P. Choroba, *A novel air traffic flow management model to optimise network delay,* in *Air Traffic Management Research & Development Thirteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2019)Seminar* (Vienna, Austria, 2019) p. 10.

[19] S. Carlier, I. de Lépinay, J.-C. Hustache, and F. Jelinek, *Environmental impact of air traffic flow management delays,* in *Proceedings of the USA/FAA Air Traffic Management R&D Seminar 2007* (Barcelona, Spain, 2007).

[20] D. A. Pamplona and C. J. P. Alves, *An overview of air delay: A case study of the Brazilian scenario,* Transportation Research Interdisciplinary Perspectives **7**, 100189 (2020).

[21] Royal Schiphol Group, *Schiphol Airport CDM Operations Manual,* (2019).

[22] Eurocontrol, *Airport Collaborative Decision Making (A-CDM) implementation Manual,* (2017).

[23] A. J. Reynolds-Feighan and K. J. Button, *An assessment of the capacity and congestion levels at European airports,* Journal of Air Transport Management **5**, 113 (1999).

[24] Eurocontrol, *Public Airport Corner, Amsterdam Schiphol AMS / EHAM airport information,* (), accessed on 11-01-2021.

[25] M. Liang, D. Delahaye, and P. Marechal, *Conflict-free arrival and departure trajectory planning for parallel runway with advanced point-merge system,* Transportation Research Part C: Emerging Technologies **95**, 207 (2018).

[26] Schiphol Airport, *Start- en landingsbanen,* `https://www.schiphol24.nl/plattegrond-banen/`, accessed on 02-10-2020.

[27] G. Santos and M. Robin, *Determinants of delays at European airports,* Transportation Research Part B: Methodological Economic Analysis of Airport Congestion, **44**, 392 (2010).

[28] ACNL, *Airport Coordination Netherlands,* Accessed on 09-02-2021.

[29] KLM, *KLM Newsapp: What is NEW in our winter 2018 schedule?* (2018), accessed on 29-01-2021,Internal document KLM.

[30] LVNL, *Preferentievolgorde,* (), accessed on 08-02-2021.

[31] D. Kulkarni, Y. Wang, and B. Sridhar, *Analysis of Airport Ground Delay Program Decisions Using Data Mining Techniques,* in *14th AIAA Aviation Technology, Integration, and Operations Conference*, AIAA AVIATION Forum (American Institute of Aeronautics and Astronautics, 2014).

[32] A. Rodriguez-Sanz, F. G. Comendador, R. A. Valdes, and J. A. Perez-Castan, *Characterization and prediction of the airport operational saturation,* Journal of Air Transport Management **69**, 147 (2018).

[33] *Performance Review Report :An Assessment of Air Traffic Management in Europe during the Calendar Year 2019*, Tech. Rep. (Eurocontrol, 2019).

[34] H. Fricke and M. Schultz, *Delay Impacts onto Turnaround Performance,* in *USA/Europe Air Traffic Management Research and Development Seminar (ATM2009)* (2009) p. 10.

[35] J. Rakas, *Defining and Measuring Aircraft Delay and Airport Capacity Thresholds* (Airport Cooperative Research Program, 2014).

[36] Xiao-Bing Hu and Wen-Hua Chen, *Receding horizon control for aircraft arrival sequencing and scheduling,* IEEE Transactions on Intelligent Transportation Systems **6**, 189 (2005), conference Name: IEEE Transactions on Intelligent Transportation Systems.

[37] T. Bolić, L. Castelli, L. Corolli, and D. Rigonat, *Reducing ATFM delays through strategic flight planning,* Transportation Research Part E: Logistics and Transportation Review **98**, 42 (2017).

[38] F. Lillo, S. Miccic, R. Mantegna, V. Beato, and S. Pozzi, *Toward a complex network approach to ATM delays analysis,* SIDs 2011 - Proceedings of the SESAR Innovation Days (2011).

[39] Eurocontrol, *Standard IATA Delay Codes-Airport Handling Manual,* ().

[40] A. Sternberg, J. Soares, D. Carvalho, and E. Ogasawara, *A Review on Flight Delay Prediction*, Preprint (2017).

[41] B. Baspinar, N. K. Ure, E. Koyuncu, and G. Inalhan, *Analysis of Delay Characteristics of European Air Traffic through a Data-Driven Airport-Centric Queuing Network Model,* IFAC-PapersOnLine 14th IFAC Symposium on Control in Transportation SystemsCTS 2016, **49**, 359 (2016).

[42] M. Hansen, *Micro-level analysis of airport delay externalities using deterministic queuing models: a case study,* Journal of Air Transport Management **8**, 73 (2002).

[43] K. Gopalakrishnan, H. Balakrishnan, and R. Jordan, *Deconstructing Delay Dynamics,* in *Proceedings of the International Conference on Research in Air Transportation* (Drexel University,Philadelphia, United States of America, 2016) p. 8.

[44] S. AhmadBeygi, A. Cohn, Y. Guan, and P. Belobaba, *Analysis of the potential for delay propagation in passenger airline networks,* Journal of Air Transport Management **14**, 221 (2008).

[45] S. Belkoura and S. Cristobal, *New insights on non-linear delay causation network for passengers and flights in Europe,* in *International Conference on Research in Air Transportation* (2018) p. 9.

[46] J. Clausen, A. Larsen, J. Larsen, and N. J. Rezanova, *Disruption management in the airline industry—Concepts, models and methods,* Computers & Operations Research Disruption Management, **37**, 809 (2010).

[47] B. F. Santos, M. M. E. C. Wormer, T. A. O. Achola, and R. Curran, *Airline delay management problem with airport capacity constraints and priority decisions,* Journal of Air Transport Management **63**, 34 (2017).

[48] G. Lulli and A. Odoni, *The European Air Traffic Flow Management Problem,* Transportation Science **41**, 431 (2007), publisher: INFORMS.

[49] C. Glymour, K. Zhang, and P. Spirtes, *Review of Causal Discovery Methods Based on Graphical Models,* Frontiers in Genetics **10** (2019), 10.3389/fgene.2019.00524, publisher: Frontiers.

[50] O. Goudet, D. Kalainathan, P. Caillou, I. Guyon, D. Lopez-Paz, and M. Sebag, *Causal Generative Neural Networks,* arXiv:1711.08936 [stat] (2018), arXiv: 1711.08936.

[51] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Elsevier Science & Technology, 1988).

[52] A. P. King and R. J. Eckersley, *Chapter 2 - Descriptive Statistics II: Bivariate and Multivariate Statistics,* in *Statistics for Biomedical Engineers and Scientists*, edited by A. P. King and R. J. Eckersley (Academic Press, 2019) pp. 23–56.

[53] A. Sfetsos and D. Vlachogiannis, *A new approach to discovering the causal relationship between meteorological patterns and PM10 exceedances,* Atmospheric Research International Conference on Nucleation and Atmospheric Aerosols (Part 1), **98**, 500 (2010).

[54] S. Vardoulakis and P. Kassomenos, *Sources and factors affecting PM10 levels in two European cities: Implications for local air quality management,* Atmospheric Environment Fifth International Conference on Urban Air Quality, **42**, 3949 (2008).

[55] C. Mayer and T. Sinai, *Network Effects, Congestion Externalities, and Air Traffic Delays: Or Why Not All Delays Are Evil,* The American Economic Review **93**, 1194 (2003), publisher: American Economic Association.

[56] R. Aydemir, D. T. Seymour, A. Buyukdagli, and B. Guloglu, *An empirical analysis of delays in the Turkish Airlines network,* Journal of Air Transport Management **65**, 76 (2017).

[57] Y. Liu, M. Hansen, D. J. Lovell, C. Chuang, M. O. Ball, and J. M. Gulding, *Causal Analysis of En Route Flight Inefficiency – the US Experience,* in *Air Traffic Management Research & Development Seminar* (2017) p. 9.

[58] I. Mohammadian, B. Abbasi, A. Abareshi, and M. Goh, *Antecedents of flight delays in the Australian domestic aviation market,* Transportation Research Interdisciplinary Perspectives **1**, 100007 (2019).

[59] M. Abdel-Aty, C. Lee, Y. Bai, X. Li, and M. Michalak, *Detecting periodic patterns of arrival delay,* Journal of Air Transport Management **13**, 355 (2007).

[60] B. Yu, Z. Guo, S. Asian, H. Wang, and G. Chen, *Flight delay prediction for commercial air transport: A deep learning approach,* Transportation Research Part E: Logistics and Transportation Review **125**, 203 (2019).

[61] C. W. J. Granger, *Some recent development in a concept of causality,* Journal of Econometrics **39**, 199 (1988).

[62] C. W. J. Granger, *Causality, cointegration, and control,* Journal of Economic Dynamics and Control **12**, 551 (1988).

[63] S. Belkoura and M. Zanin, *Phase changes in delay propagation networks,* in *International Conference on Research in Air Transportation* (Drexel University,Philadelphia, United States of America, 2016) p. 8.

[64] W.-B. Du, M.-Y. Zhang, Y. Zhang, X.-B. Cao, and J. Zhang, *Delay causality network in air transport systems,* Transportation Research Part E: Logistics and Transportation Review **118**, 466 (2018).

[65] P. Mazzarisi, S. Zaoli, F. Lillo, L. Delgado, and G. Gurtner, *New centrality and causality metrics assessing air traffic network interactions,* Journal of Air Transport Management **85**, 101801 (2020).

[66] J. Hipp, U. Güntzer, and G. Nakhaeizadeh, *Algorithms for association rule mining - a general survey and comparison,* Association for Computing Machinery SIGKDD Explorations Newsletter **2**, 58 (2000).

[67] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. (Morgan Kaufmann, 2012).

[68] J. Han, H. Cheng, D. Xin, and X. Yan, *Frequent pattern mining: Current status and future directions,* Data Min. Knowl. Discov. **15**, 55 (2007).

[69] R. Agrawal, T. Imielinski, and A. Swami, *Mining Association Rules Between Sets of Items in Large Databases,* ACM Sigmod Record **22**, 207 (1993), journal Abbreviation: Sigmod Record.

[70] M. Zaki, S. Parthasarathy, M. Ogihara, and W. li, *New Algorithms for Fast Discovery of Association Rules.* in *Conference on Knowlegde Discovery and Data Mining* (1997) pp. 283–286.

[71] R. Agrawal and R. Srikant, *Fast Algorithms for Mining Association Rules in Large Databases,* in *Proceedings of the 20th International Conference on Very Large Data Bases*, VLDB '94 (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1994) pp. 487–499.

[72] J. Han, J. Pei, and Y. Yin, *Mining frequent patterns without candidate generation,* ACM SIGMOD Record **29**, 1 (2000).

[73] D. Truong, *Using causal machine learning for predicting the risk of flight delays in air transportation,* Journal of Air Transport Management **91**, 101993 (2021).

[74] M. Scutari, C. Graafland, and J. Gutiérrez, *Who Learns Better Bayesian Network Structures: Constraint-Based, Score-based or Hybrid Algorithms?* in *Proceedings of Machine Learning Research*, Vol. 72 (2018) pp. 416–427.

[75] J. Cheng, R. Greiner, J. Kelly, D. Bell, and W. Liu, *Learning Bayesian networks from data: An information-theory based approach,* Artificial Intelligence **137**, 43 (2002).

[76] L. de Campos, *A Scoring Function for Learning Bayesian Networks based on Mutual Information and Conditional Independence Tests,* Journal of Machine Learning Research **7**, 2149 (2006).

[77] A. M. Carvalho, *Scoring functions for learning bayesian networks,* Inesc-id Tec. Rep **12** (2009).

[78] F. S. Hillier and G. J. Lieberman, *Introduction to Operations Research*, Tenth edition, (McGraw-Hill Education, 2015).

[79] N. Dojer, *Learning Bayesian networks from datasets joining continuous and discrete variables,* International Journal of Approximate Reasoning **78**, 116 (2016).

[80] S. Acid, L. M. de Campos, J. M. Fernandez-Luna, S. Rodriguez, J. Maria Rodriguez, and J. Luis Salcedo, *A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service,* Artificial Intelligence in Medicine Bayesian Networks in Biomedicince and Health-Care, **30**, 215 (2004).

[81] K. Natori, M. Uto, Y. Nishiyama, S. Kawano, and M. Ueno, *Constraint-Based Learning Bayesian Networks Using Bayes Factor,* in *Advanced Methodologies for Bayesian Networks*, Lecture Notes in Computer Science, edited by J. Suzuki and M. Ueno (Springer International Publishing, Cham, 2015) pp. 15–31.

[82] D. Barber, *Bayesian Reasoning and Machine Learning* (Cambridge University Press, UK, 2012).

[83] O. Zuk, S. Margel, and E. Domany, *On the Number of Samples Needed to Learn the Correct Structure of a Bayesian Network,* in *Uncertainty in Artificial Intelligence* (2016) pp. 560–567, arXiv: 1206.6862.

[84] N. Fernandes, S. Moro, C. J. Costa, and M. Aparício, *Factors influencing charter flight departure delay,* Research in Transportation Business & Management Data analytics for international transportation management, **34**, 100413 (2020).

[85] B. Sridhar, Y. Wang, and A. Klein, *Modeling Flight Delays and Cancellations at the National, Regional and Airport Levels in the United States,* in *Eighth USA/Europe Air Traffic Management Research and Development Seminar* (Napa, California, USA, 2009).

[86] L. Belcastro, F. Marozzo, D. Talia, and P. Trunfio, *Using Scalable Data Mining for Predicting Flight Delays,* ACM Transactions on Intelligent Systems and Technology **8**, 5:1 (2016).

[87] J. J. Rebollo and H. Balakrishnan, *Characterization and prediction of air traffic delays,* Transportation Research Part C: Emerging Technologies **44**, 231 (2014).

[88] P. Monmousseau, D. Delahaye, A. Marzuoli, and E. Feron, *Predicting and Analyzing US Air Traffic Delays using Passenger-centric Data-sources,* in *Thirteenth USA/Europe Air Traffic Management Research and Development SeminarManagement Research & Development Seminar* (2019) p. 10.

[89] A. Pai, *ANN vs CNN vs RNN | Types of Neural Networks,* (2020), accessed on 04-02-2021.

[90] C. Bischop, *Pattern Recognition and Machine Learning* (Springer Science + Business Media, 2006).

[91] A. Sharma, *Decision Tree vs. Random Forest - Which Algorithm Should you Use?* (2020), accessed on 10-02-2021.

[92] S. Lu, Q. Li, L. Bai, and R. Wang, *Performance predictions of ground source heat pump system based on random forest and back propagation neural network models,* Energy Conversion and Management **197**, 111864 (2019).

[93] A. Barredo Arrieta, N. Diaz Rodriguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado González, S. Garcia, S. Gil-Lopez, D. Molina, V. R. Benjamins, R. Chatila, and F. Herrera, *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI,* Information Fusion **58** (2019), 10.1016/j.inffus.2019.12.012.

[94] A. Adadi and M. Berrada, *Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI),* IEEE Access **6**, 52138 (2018), conference Name: IEEE Access.

[95] T. Hailesilassie, *Rule Extraction Algorithm for Deep Neural Networks: A Review,* (IJCSIS) International Journal of Computer Science and Information Security **14** (2016), arXiv: 1610.05267.

[96] M. G. Augasta and T. Kathirvalavakumar, *Rule extraction from neural networks — A comparative study,* in *International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)* (2012) pp. 404–408.

[97] P. Cortez and M. J. Embrechts, *Opening black box Data Mining models using Sensitivity Analysis,* in *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)* (2011) pp. 341–348.

[98] P. Cortez and M. J. Embrechts, *Using sensitivity analysis and visualization techniques to open black box data mining models,* Information Sciences **225**, 1 (2013).

[99] R. H. Kewley, M. J. Embrechts, and C. Breneman, *Data strip mining for the virtual design of pharmaceuticals with neural networks,* IEEE transactions on neural networks **11**, 668 (2000).

[100] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, *Modeling wine preferences by data mining from physicochemical properties,* Decision Support Systems Smart Business Networks: Concepts and Empirical Evidence, **47**, 547 (2009).

[101] O. Goudet, D. Kalainathan, P. Caillou, D. Lopez-Paz, I. Guyon, M. Sebag, A. Tritas, and P. Tubaro, *Learning Functional Causal Models with Generative Neural Networks,* in *Explainable and Interpretable Models in Computer Vision and Machine Learning* (2018) pp. 39–80.

[102] D. Peteiro-Barral and B. Guijarro-Berdiñas, *A Study on the Scalability of Artificial Neural Networks Training Algorithms Using Multiple-Criteria Decision-Making Methods,* in *Artificial Intelligence and Soft Computing*, Lecture Notes in Computer Science, edited by L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada (Springer, Berlin, Heidelberg, 2013) pp. 162–173.

[103] V. Bolón-Canedo, D. Rego-Fernández, D. Peteiro-Barral, A. Alonso-Betanzos, B. Guijarro-Berdiñas, and N. Sánchez-Maroño, *On the scalability of feature selection methods on high-dimensional data,* Knowledge and Information Systems **56**, 395 (2018).

[104] D. Tax, M. Loog, and G. Migut, *Lecture Slides of CS4220 Machine Learning 1,* Delft University of Technology (2019).

[105] B. Li, X. Chen, M. J. Li, J. Huang, and S. Feng, *Scalable Random Forests for Massive Data,* in *Proceedings of the 16th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining - Volume Part I* (2012) pages: 146.

[106] M. P. Skënduli, M. Biba, and M. Ceci, *Implementing Scalable Machine Learning Algorithms for Mining Big Data: A State-of-the-Art Survey,* in *Big Data in Engineering Applications*, Studies in Big Data, edited by S. S. Roy, P. Samui, R. Deo, and S. Ntalampiras (Springer, Singapore, 2018) pp. 65–81.

[107] M. Kalisch and P. Buhlmann, *Estimating High-Dimensional Directed Acyclic Graphs with the PC-Algorithm,* Journal of Machine Learning Research **8**, 613 (2007).

[108] M. Z. Li and M. S. Ryerson, *Reviewing the DATAS of aviation research data: Diversity, availability, tractability, applicability, and sources,* Journal of Air Transport Management **75**, 111 (2019).

[109] L. Moreira, C. Dantas, L. Oliveira, J. Soares, and E. Ogasawara, *On Evaluating Data Preprocessing Methods for Machine Learning Models for Flight Delays,* in *2018 International Joint Conference on Neural Networks (IJCNN)* (2018) pp. 1–8, iSSN: 2161-4407.

[110] OPENNAV, *Waypoints in Netherlands : OpenNav aviation database,* Accessed on 05-02-2021.

[111] KNMI, *KNMI - Uurgegevens van het weer in Nederland - Download,* .

[112] C.-L. Wu, *Inherent delays and operational reliability of airline schedules,* Journal of Air Transport Management **11**, 273 (2005).

[113] H. Aguinis, R. K. Gottfredson, and H. Joo, *Best-Practice Recommendations for Defining, Identifying, and Handling Outliers,* Organizational Research Methods **16**, 270 (2013), publisher: SAGE Publications Inc.

[114] Data Science Community KLM, *Feature Engineering Workshop: Presentation Slides,* (2017), kLM Internal Document.

[115] T. Yiu, *Understanding PCA,* (2019), accessed on 03-02-2021.

[116] M. M. Drugan and M. A. Wiering, *Feature selection for Bayesian network classifiers using the MDL-FS score,* International Journal of Approximate Reasoning **51**, 695 (2010).

[117] A. L. Blum and P. Langley, *Selection of relevant features and examples in machine learning,* Artificial Intelligence Relevance, **97**, 245 (1997).

[118] I. Tsamardinos and C. Aliferis, *Towards Principled Feature Selection: Relevancy, Filters and Wrappers,* in *in Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics* (2002).

[119] A. Ozçift and A. Gulten, *Genetic algorithm wrapped Bayesian network feature selection applied to differential diagnosis of erythemato-squamous diseases,* Digital Signal Processing **23**, 230 (2013).

[120] L. Torgo, R. Ribeiro, B. Pfahringer, and P. Branco, *SMOTE for Regression,* in *Progress in Artificial Intelligence. EPIA 2013.*, Vol. 8154 (Springer, 2013) pp. 378–389.