Quality analysis OPS plan predictions

Comparative assessment of the traffic demand and runway predictions

Bachelor thesis

centre of excellence

knowledge & development centre Mainport Schiphol

KDC Mainport Schiphol – Centre of Excellence

A collaboration with the Aviation Academy, Amsterdam University of Applied Sciences

Janjaap Wijnker

June 18, 2021, Amsterdam

Page intentionally left blank

Quality analysis OPS plan predictions

Comparative assessment of the traffic demand and runway capacity predictions

Bachelor thesis

Author[1]		
Name	Responsibility	
Janjaap Wijnker	Research student KDC Mainport	
Student number	E-mail	Graduation Track
500661433	janjaap.wijnker@hva.nl	Flight Operations Engineering

Reviewers[2]	
Name	Responsibility

Acceptance by client[3]			
Name	Responsibility	Signature	Date
	Air Traffic Control the Netherlands (LVNL)		
	Knowledge & Development Centre Mainport Schiphol (KDC)		

[1] The author agrees with the content, did take notice of the review feedback, and applied it to his/her own insight.
[2] The reviewers declare that they reviewed the report based on their expertise and provided the author with feedback.
[3] The client uses or applies the result of the report, only applicable clients sign.

Acknowledgments

This research study is the result of an incredible chance given to me by the Knowledge & Development Centre. Throughout executing this research study, I have received a great deal of personal and professional support.

I would like to thank Dr. Catya Zuñiga Alcaraz for her guidance and feedback as my university advisor and Yoram Obbens, Ilse Megens, Emmelie Janse and Huy Vû for their continuous encouragement and collaboration as my company supervisors.

I am thankful for the tips and feedback I have received from Pauline Poot, Geert Boosten, Sashiko Shirai Reyna and Evert Westerveld that helped me improve my work, and all the colleagues from the Strategy & Capacity Management and Operations departments for letting me join several meetings that helped me learn more about the processes and projects related to the topic of my study.

Since the graduation phase of my study took place during the COVID-19 pandemic, many things had to be arranged in a different fashion. I would like to thank everyone from the Aviation Academy of the Amsterdam University of Applied Sciences and Luchtverkeersleiding Nederland that made this research position possible.

I would also like to thank my parents and siblings for their love and support throughout my entire life as I know I would not have been able to get this far without them. Thank you to my girlfriend for proofreading my thesis many times, helping me understand difficult subjects and pouring me wine when work took longer than expected. Furthermore, I could not have finished this dissertation without the support and insights from my friends and family, and finally I would like to thank my furry roommates, who have provided me with unconditional love and support.

June 18, 2021

Abstract

In 2020 the OPS plan was introduced at LVNL as a pre-tactical capacity management procedure with the purpose to better align traffic demand with the available runway capacity, and therefore tackle operational bottlenecks such as delays. An OPS plan is a collection of information and predictions that aim to help several stakeholders to prepare for the day of operations. A post-operations feedback mechanism was yet to be developed to assess the accuracy and precision of the predictions and provided improvements if necessary. This research study aims to develop such a feedback mechanism by analyzing three of the predictions: traffic demand, runway capacity and runway use.

The predictability of the predictions is assessed in terms of accuracy, the mean difference between the prediction and the realizations, and precision, the standard error of the mean. In addition, for the runway predictions, the most impacting factors have been identified.

The dataset was divided in three periods as traffic demand was considerably different in each of them. The dataset was divided in three periods as traffic demand was considerably different in each of them. During period two the average accuracy is 1.24 movements per day, with a precision of 0.63 movements per day. During period one the average accuracy is 6.80 with a precision of 0.59 movements per day, and during period three the average accuracy is 5.17 with a precision of 0.57 movements per day. When looking at the traffic demand per hour 20 minutes rolling, during period two the average accuracy is 0.05 movements with a precision of 0.05 movements per hour. During period one the average accuracy is 0.28 with a precision of 0.03 movements per hour, and during period three the average accuracy is 0.21 with a precision of 0.04 movements per hour.

The runway capacity and usage predictions were equal to the realization on approximately 90 percent of the time analyzed. Human error was the main factor that caused the differences in runway capacity, while changing weather conditions mainly caused differences in runway usage. The period analyzed was severely impacted by the COVID-19 pandemic and therefore does not represent ordinary operations. Nevertheless, some findings are unrelated to this impact and this study allows to be used as a framework for future analyses.

Table of contents

List of tables

List of figures

List of abbreviations

ACC	Area Control Center
ADP	ATFCM Daily Plan
ANSP	Air Navigation Service Provider
ASM	Air Space Management
ATC	Air Traffic Controller
ATFCM	Air Traffic Flow & Capacity Management
ATFM	Air Traffic Flow Management
ATM	Air Traffic Management
ATS	Air Traffic Services
BZO	Beperkt Zicht Omstandigheden (Limited Visibility Conditions)
CB	Cumulonimbus clouds
CDM	Collaborative Decision-Making
CoE	Center of Excellence
COVID-19	Coronavirus Disease
CSV	Comma-Separated Values
D0	Day of operations
D-1	Day before operations
DDR	Demand Data Repository
ECAC	European Civil Aviation Conference
EHAA	ICAO code for Amsterdam FIR
EHAM	ICAO code for Amsterdam Airport Schiphol
EHFIRAM	Traffic volume indicating EHAM traffic within EHAA
FIR	Flight Information Region
GC	Ground Controller
IAF	Initial Approach Fix
ICAO	International Civil Aviation Organization
KDC	Knowledge & Development Center
KLM	Koninklijke Luchtvaart Maatschappij (Royal Dutch Airlines)
KNMI	Koninklijk Nederlands Meteorologisch Instituut (Royal Dutch Weather Institute)
LVNL	Lucht Verkeersleiding Nederland (Air Traffic Control The Netherlands)
MUAC	Maastricht Upper Area Control center
NATS	National Air Traffic Services
NEST	Network and Strategy Tool
NM	Network Manager
NMOC	Network Manager Operations Center
NOP	Network Operations Plan
NOTAM	Notice to Airman
OHD	Operational Helpdesk
OPS	Operations
QRC	Quick Reference Card
RVR	Runway Visual Range
SKV	Schiphol Kans Verwachting (Schiphol Chance Prediction)
TMA	Terminal Maneuvering Area
UDP	Uniform Daylight Period
UIR	Upper Information Region
UTC	Coordinated Universal Time

Summary

As Schiphol's traffic demand has been increasing up to 500,000 movements per year, the need for improving the utilization of the available capacity intensifies. Several research studies were conducted at LVNL to investigate the possibilities of enhancing the capacity management process, and it was concluded that an operational planning (OPS plan) would help LVNL, as well as its stakeholders to optimize the efficiency of their operations. The OPS plan's objective is to align the traffic demand with the available capacity, prior to the day of operations. This is done by gathering all sorts of information that is relevant to or might affect the operations. In addition, several predictions are made regarding the expected traffic demand and the available runway capacity. This study aims to provide a post-operations analysis, where the predictability of the inbound traffic demand predictions for EHFIRAM, as well as the runway capacity and usage predictions for the first inbound peak are examined.

The predictability of the predictions is determined in terms of accuracy and precision, where accuracy is calculated as the mean difference between the prediction in the OPS plan and a reference dataset, and the precision is defined as the spread of the accuracy. For calculating accuracy, the traffic demand prediction in the OPS plan was compared to the reference datasets of realization and last filed flight plans. Additionally, the runway usage was compared to the realization and the capacity forecast form, whereas the runway capacity could only be compared to the capacity forecast form, as the realized runway capacity data was not available. The data that was analyzed was gathered between April 1st, and November 14th of 2020.

This research study was conducted during the COVID-19 pandemic. The data that was available for analyzing was affected by that same pandemic, or restrictions following it. It needs to be stressed that this data does not represent the operations that the air traffic control organization is used to. Furthermore, the situation may also have affected the ability to make accurate predictions. Some of the findings of this study can thus not be applied to situations where the industry has recovered from the impact of the pandemic.

The average accuracy of the inbound traffic demand predictions for EHFIRAM during period two compared to the realization is 1.24 movements per day, with a precision of 0.63 movements per day. During period one the average accuracy is 6.80 with a precision of 0.59 movements per day, and during period three the average accuracy is 5.17 with a precision of 0.57 movements per day. When looking at the traffic demand per hour rolling, during period two the average accuracy is 0.05 movements with a precision of 0.05 movements per hour rolling. During period one the average accuracy is 0.28 with a precision of 0.03 movements per hour, and during period three the average accuracy is 0.21 with a precision of 0.04 movements per hour. The traffic demand predictions were most inaccurate when traffic demand was at its lowest, that being ± ten percent compared to the year before, while the predictions seemed to get more accurate as traffic demand increased, that being ± fifty percent compared to the year before. Throughout the day, the predictions seem to be more accurate and precise around inbound peak periods, implying that predictions are more reliable when traffic demand is high. Through testing for statistical significance, it can be concluded that the predicted traffic demand was not significantly different from the realization for the period with highest traffic demand (that being the period from July 1st until September 1st of 2020), whereas the predictions in the periods with lower traffic demand were significantly higher than the realization.

An inbound peak is defined as a period where two runways are in used simultaneously for handling inbound traffic. During normal operations there are multiple inbound peaks, however, due to the declined traffic demand during the period of analysis, a single inbound peak was not recorded nor predicted until July 1st. The OPS plan contains a primary and alternative runway capacity prediction for the first inbound peak of the day. For instances where the primary runway capacity prediction differed from that of the forecast form, the alternative prediction was considered. When solely comparing the primary prediction to the realization, and when comparing both primary and alternative prediction, there was no significant difference between the

OPS plan prediction and the forecast form. On 85 of the 94 days where an inbound peak was anticipated, the predicted runway capacity in the OPS plan (that being the primary or alternative prediction) was equal to the forecast form. The main cause for differences is human error, as in those cases information was used incorrectly, or available information was not used. The predicted runway usage was equal to the realization and equal to the capacity forecast form on 91 of the 97 days. While the main cause for these differences is changing wind direction, the difference is not statistically significant.

As the OPS plan was introduced to make optimal use of the available resources, it is recommended to adjust the predictions according to the accuracy calculated. Since it is expected that the aviation industry will recover from the COVID-19 pandemic, the predictability of the predictions should be reassessed when traffic demand has reached the 2019's level. Based on the findings of this study it is expected that the traffic demand predictions will be more accurate when demand increases.

As human error, particularly the use of information, is the main cause for differences in runway capacity, it is recommended that the gathering of information should be automated to a certain extent. An algorithm for the selection of runway combinations should be developed, which would allow for automatic consideration of weather conditions and availability, therefore reducing the human errors observed. The final predication would then still be made based on expert judgement, but now only using the most suitable possibilities. Another key recommendation is that a framework should be developed so runway combination can be analyzed in the future. The runway predictions were extracted from the OPS plans manually, thus the resulting dataset could be used as a template for the framework. Finally, steps should be taken to gather the realized runway capacity as it is yet to be done. It would be beneficial if the realized runway combinations were combined with weather conditions and the information from QRC09 to determine the realized capacity.

1 Introduction

This research study was conducted on behalf of Air Traffic Control the Netherlands (LVNL) at the Centre of Excellence (CoE), a part of the Knowledge & Development Centre (KDC). KDC was founded by Amsterdam Airport Schiphol, Royal Dutch Airlines (KLM) and LVNL with the aim to support sustainable development and growth of Schiphol by defining and executing target orientated projects in close collaboration with both the government and the aviation industry. Through the CoE, KDC joins forces with its research partners Aviation Academy at Amsterdam University of Applied Sciences and the Technical University Delft. This partnership gives students the chance to develop their knowledge and talents while contributing to the efficiency and sustainability of Schiphol Mainport.

This first chapter describes the background of this study, the justification, objective, research questions, hypotheses, scope and finally, the structure of the rest this report is detailed.

1.1 Background

In 2016, Schiphol became the busiest airport in the European network (Eurocontrol, 2017). For several years, the airport has been closing in on its political limit of 500,000 flight movements a year (Zuidberg & Vinkx, 2018), as well as the operational limits, such as runway or airspace capacity (LVNL, 2020). An indicator that the capacity lags regarding traffic demand, is the growing number of delays (Eurocontrol, 2019). Within the European network, the amount of Schiphol related delays has been rising steadily, and it was estimated that this cost airlines around 100 million euros in 2017 (van Delden, Westerveld, Lieshout, Obbens, & Devilee, 2017).

The increase of delays was one of multiple factors that highlighted the need for improvements in capacity management at LVNL. In 2017, multiple research studies were conducted to investigate the possibilities of enhancing the capacity management process. The studies concluded that a well-structured planning for the day of operations, referred to as an operations plan (OPS plan), would help LVNL to optimize the efficiency of its operations.

The aim of an OPS plan is to align the traffic demand with the available capacity prior to the day of operations (D-1). This is thought to increase the predictability, reliability, and stability of the operations of the following day (D0).

The pre-tact unit, a specialized team of experts, gather input information on D-1 that could have an impact on the operational capacity. For example, the predicted traffic demand, weather predictions, planned maintenance, and other relevant events are detailed. Based on this information, recommendations are formulated (i.e., output information) which would lead to an increase in safety and efficiency because the operations get more predictable, and the use of capacity is optimized (Lieshout, Devilee, & Obbens, 2017). Examples of recommendations that inform the OPS plan are runway usage, sector grouping, traffic regulation scenarios and expected required runway capacity. When the aim of the OPS plan is met, it provides several benefits.

- A clear, complete, and transparent insight into traffic demand, capacity and availability of air space, weather, and irregularities on the day before operations.
- Predictable operations and services that could help stakeholders improve their processes.
- Collaborative decisions made with adjacent air navigation service providers regarding the distribution of capacity.
- The optimal utilization of the available air space in collaboration with military air traffic control.
- Reduction in workload for supervisors and traffic controllers on the day of operations.
- Schedules that are better aligned with the demand for personnel on the day of operations to increase efficiency on the deployment of staff.

The development of the OPS plan was initiated in 2018, and in January 2019 a trial begun wherein approximately three times a week a planning was created a day in advance. In the end of March 2020, LVNL started creating a plan for every operational day.

1.2 Problem statement

The accuracy and precision of the traffic demand and runway capacity predictions made by the OPS plan have yet to be assessed.

1.3 Objective

In order to address this problem, the objective of this study is to conduct a post-operational comparative analysis to assess to what extent the realized traffic demand and runway capacity numbers match the predicted ones in the OPS plan, by means of accuracy and precision.

1.4 Research questions

To meet the objective mentioned above, the following research question will need to be answered.

> *'How predictable are the traffic demand and runway capacity predicted in the OPS plan in terms of accuracy and precision, and what are the key factors that affect the accuracy of the runway capacity predictions?'*

This research question can be divided into smaller sub questions which will be answered in different parts of the study.

'Why is the OPS plan necessary for LVNL and the stakeholders?'

'What information is included in the OPS plan?'

'What is the quantitative difference between the predicted and realized traffic demand?'

'What is the quantitative difference between the predicted and realized runway capacity, and what factors are the cause for these differences?'

1.5 Research design

This research study was conducted in three steps: desk research, exploratory research, and data analysis, as displayed in Figure 1.

Figure 1 – Process diagram of the three steps of the research study

Desk research

Qualitative research in the form of a literature review was conducted in order to answer the first sub question stated in section 1.4. During this step of the research two subjects are reviewed: air traffic management, and capacity management.

Air traffic management is the system in which multiple stakeholders contribute to safe and efficient air traffic. The desk research focusses on the relationship between two of the stakeholders within air traffic management: Eurocontrol and LVNL. One aspect of air traffic management (ATM) is capacity management, which is an effort by the network manager and LVNL to make optimal use of the available airspace capacity. The desk research aims to identify the gap in the existing literature that will be closed by the findings of this research study.

The literature that was reviewed is mainly sourcing from studies conducted at or for LVNL and operational manuals drafted by Eurocontrol and LVNL.

Exploratory research

The second step of the research aims to clarify the process of making the OPS plan, the types of information used during that process, and the predictions that are made in the plan. The qualitative research conducted during this step of the research study is in the form of a content analysis and aims to answer the second sub question stated in section 1.4. During this step, a distinction was made between information used as input for, and information that is output of the OPS plan. The output information is used to identify what aspects of the OPS plan could and should be analyzed during this research study.

Data analysis

The last step of the research study is the analysis of the data, where the OPS plan predictions are compared to a reference in order to determine the accuracy and precision of those predictions in relation to the reference. Wherein, accuracy is conceptualized as the quantitative difference between the predicted and reference datasets. Precision is defined by the spread of the mean accuracy, that being the standard error of the mean. The purpose of looking at precision is to determine if the accuracy of the predictions is consistent. For example, if the traffic is consistently overestimated by X movements per unit of time (where consistency describes precision and X describes accuracy), then it could be recommended that the prediction should be adjusted accounting for X. However, if the accuracy is not consistent, then accounting for X in the prediction would maybe improve some predictions while worsening others. The quantitative data analysis aims to answer the last two research questions discussed in section 1.4. The methods, phases of the analysis, and the data that was used for this step are elaborated in the methodology in chapter 4.

1.6 Hypotheses

In order to validate the findings of the data analysis, the following hypotheses were posed.

Accuracy of traffic demand prediction

$Null\ hypothesis\ (H_0)$: $\mu_1 = \mu_2$, which translates to; the mean number of movements predicted is equal to the mean number of movements realized.

$Alternative\ hypothesis\ (H_a)$: $\mu_1 \neq \mu_2$, which translates to; the mean number of movements predicted is not equal to the mean number of movements realized.

Accuracy of runway use prediction

$Null\ hypothesis\ (H_0)$: $p_1 = p_2$, which translates to; the predicted runway usage (proportion) in OPS plan is equal to the runway use realized.

$Alternative\ hypothesis\ (H_a)$: $p_1 \neq p_2$, which translates to; the predicted runway usage (proportion) in OPS plan is not equal to the runway use realized.

Accuracy of runway capacity prediction

$Null\ hypothesis\ (H_0)$: $\mu_1 = \mu_2$, which translates to; the mean capacity predicted for the first inbound peak is equal to the mean capacity in the forecast form.

$Alternative\ hypothesis\ (H_a)$: $\mu_1 \neq \mu_2$, which translates to; the mean capacity predicted for the first inbound peak is not equal to the mean capacity in the forecast form.

1.7 Limitations and assumptions

Limitations

From March 2020 onwards, the coronavirus disease (COVID-19) had a severe impact on the worldwide aviation industry. To minimize the risk of spreading the virus, international travel restrictions were introduced, and consequently, airlines suspended parts of their operations. Consequently, this affected the traffic demand at Schiphol, and therefore the data analyzed during this study. The data analyzed is not an accurate representation of the operations that the Dutch air traffic system is designed for, and thus the findings of this study might not be generalizable to times when traffic demand increases back to the original level.

The realized runway capacity is not available for analysis as the capacity is dependent on various external factors like wind speed and visibility. To determine the accuracy of the runway capacity prediction for the first inbound peak, the prediction was compared to the data in the capacity forecast form, which is a forecast of the runway capacity made a few hours before the first inbound peak.

Assumptions

The runway use, and capacity predictions were not yet available as a dataset and thus were extracted from each individual OPS plan. As the OPS plan usually contains a primary and secondary prediction, both were extracted and used for the analysis. In cases where more than two predictions were provided, it was chosen to only extract the most immediate predictions (i.e., the primary and secondary) as minimal OPS plans contain more than two predictions. This also takes a step towards standardizing the OPS plan format by only accounting for a primary and secondary prediction.

The runway use describes whether the landing direction is either north or south. As none of the datasets mention landing direction specifically, it was determined by the first of the two runways in use. If for example runway 18R and 27 were in use, the runway usage is assumed to be south, as the first runway, 18R, lays in the southern direction.

1.8 Scope of research

This study analyzes data from April 1st until November 14th of 2020. Days where no predictions were available were excluded from analysis. The number of days excluded thus varied based on which prediction dataset was analyzed.

The OPS plan has many different types of users, each with their own reason for utilizing it, and thus a different definition of accuracy. As the pre-tact unit is considered the main user for this research study, accuracy will be conceptualized from their perspective (accuracy and precision are elaborated in section 4.2).

This research study analyzed the traffic demand prediction for EHFIRAM inbound, which entails traffic inbound for EHAM, measured at FIR entry. The runway use, and capacity predictions are analyzed for the first inbound peak.

1.9 Thesis structure

The remainder of this document contains the following sections.

- Desk research – The Dutch air traffic management context
- Exploratory research – Breakdown of the OPS plan
- Methodology
- Research findings
- Discussion
- Conclusion
- Recommendations
- List of references
- Appendences

2 The Dutch air traffic management context

As the demand for transportation of persons and cargo grows, and the impact of flying on the environment becomes clearer, the need for smarter and more sustainable planning increases (Schiphol, 2020). Several studies were conducted to improve the utilization of the available airspace and its capacity while enabling industry growth in a safe and sustainable manner (e.g., Westerveld, 2017; van Delden, Westerveld, Lieshout, Obbens, & Devilee, 2017). Relevant literature will be reviewed in this chapter to describe the tools and methods that are in place now, and further, how there is a gap in the literature. It will become clear that this study contributes to the steps necessary to improve operational planning for LVNL.

Air traffic management is a broad concept that inherently refers to air space management, air traffic services and air traffic flow & capacity management (ICAO, 2007). Figure 2 displays the three roles within ATM, which are then explained in the following sub-sections.

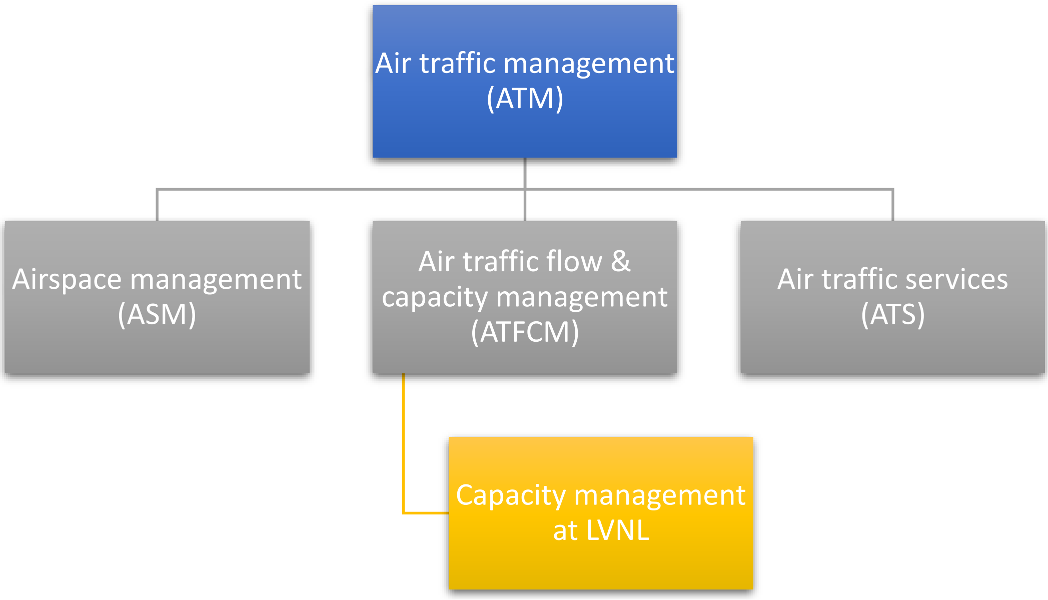


Figure 2 – Division of the ATM roles

2.1 Airspace management

Eurocontrol is the network manager (NM) for the European Civil Aviation Conference (ECAC), which consists of 41 European countries and two states outside of Europe. As the NM, Eurocontrol aims for a seamless airspace by connecting the elements of the European air traffic system (see also Figure 2). By means of collaborative decision making, data sourcing from all over the network is being used to create safe and cost-efficient operations for all members, while at the same time, minimizing environmental impact, and using capacity to its full potential (Eurocontrol, 2020).

Airspace is a valuable resource for Eurocontrol and its members, especially when it is subject to high traffic demand. Since airspace has a fixed volume, ASM is of great importance. ASM involves planning, defining, and managing airspace to ensure the users can utilize it in the safest and most efficient manner possible. Traffic demand, however, is fluctuating over time, which means that the available air traffic control capacity requires a mechanism to avoid overloads and to ensure safety and efficiency. Such a mechanism is made available by the NM Operations Center (NMOC) and provides dynamic ATFCM (Eurocontrol, 2020).

2.2 Air traffic flow & capacity management

Capacity is defined as the maximum number of aircraft existing in a land- or airside segment during a given period (Flynn, 2014). However, airspace capacity is also dependent on a large range of influencing factors. One might therefore conclude that the capacity is defined by multiple factors that all play a different and dynamic role (Mujica Mota, Boosten, & Zuniga, 2017).

The second role of ATM is ATFCM, (see Figure 2), which endeavours to align traffic demand with aerodrome and airspace capacity. When traffic demand threatens to exceed the available capacity, the demand will be adjusted so it meets the available capacity again. These adjustments might involve allocation of individual flight departure times to avoid bottlenecks and maintain safe operations throughout the network. The ATFCM process requires continuous exchange of information and communication with all air navigation service providers (ANSPs) and airspace users. The objective of ATFCM is to ensure that the network's capacity is utilized to its maximum possible extent. This service is provided by the NMOC to airspace users throughout the ECAC (Eurocontrol, 2020), and relies heavily on collaborative decision-making (CDM).

CDM is a philosophy aiming to improve efficiency and flexibility of airport operations by optimizing the use of resources and improving the predictability of the operations. CDM encourages the whole ATM system, and its users, to exchange relevant and accurate data. This may not only optimize the operations at a local level, but also across the network.

All the stakeholders of the airport's operations will benefit from CDM. Airlines can optimize their planning, which leads to fuel savings, reduction of delays, and optimized fleet utilization. Ground handlers would benefit by efficient allocation of resources. The local ATC organization will be able to create a clear and accurate operational planning horizon, and the airport itself benefits from more improved airside utilization (gates, parking, runway, etc.) and more available capacity. Aside from efficiency, safety is also an aspect the would be improved because of CDM (Eurocontrol, 2017). Even though the local airports partners could benefit directly from the CDM principles, Eurocontrol plays a vital role in guiding the implementation of it.

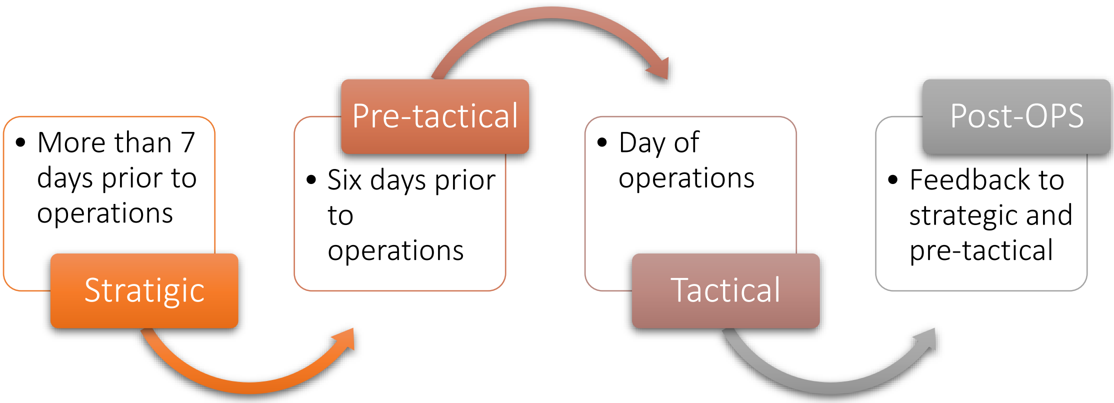


Figure 3 – The four ATFCM phases

As can be seen in Figure 3, the ATFCM is divided into four phases: strategic flow management, pre-tactical flow management, tactical flow management and post operational analysis.

Strategic Flow Management

Strategic flow management is the phase from twelve months, up to seven days prior to the day of operations. This process involves research, planning and coordinating through CDM. During this phase, the NMOC collects and reviews data with the aim of early identification of major demand/capacity discrepancies. After identification, the NM coordinates and executes the strategic ATFCM planning to optimize the utilization of available capacity. The result of this phase is the Network Operations Plan (NOP) (Eurocontrol, 2020).

Pre-Tactical Flow Management

After the strategic phase, during the six days prior to the day of operations, the NM starts to utilize the data collected prior. During pre-tactical planning, the expected demand is compared to the predicted maximum available capacity. Adjustments are made to the NOP if necessary. The pre-tactical phase aims to balance demand and capacity by effectively organizing resources (e.g., sector configuration or demand regulations) and implementing appropriate ATFCM measures. The pre-tactical planning involves CDM between the NM, ANSPs and aircraft operators. The output of this phase is the ATFCM Daily Plan (ADP) (Eurocontrol, 2020).

Tactical Flow Management

On the day of operations, tactical flow management involves real time analysis of events that are possibly affecting the ADP, and if necessary, adjustments are made. The aim of this phase is to ensure that any measures taken in the NOP or ADP are sufficient for solving demand /capacity imbalances. Ideally, adjustments needed on this day are caused by disturbances that were not foreseen during the previous phases such as significant meteorological phenomena, crises, staffing issues, etc. The presence of reliable information is of vital importance during this phase because this allows for short-term forecasts and maximizing available capacity (Eurocontrol, 2020).

Post Operational Analysis

The final phase of ATFCM is the post operational analysis. During this phase, all processes relevant to the ATFCM service are assessed, examined, and reported. All stakeholders of the ATFCM service ought to provide feedback on the accuracy of the ADP (e.g., ATFCM measures, delays or implementation of predefined regulations). This allows for comparing the anticipated operations with the actual measured operations, usually in terms of delays. The result of this phase is the development of best practices and/or lessons learnt allowing for future improvements on operational processes and activities as well as increasing the reliability of predictions and planning's (Eurocontrol, 2020).

2.3 Air traffic services

The last role of ATM is air traffic services (ATS) (see Figure 2). The purpose of ATS is to maintain safe distance between aircraft and objects and to maintain orderly and efficient flow of traffic by providing appropriate services to the airspace users. Within ATS, the ATFCM principles discussed above are implemented on a local level. An Air Navigation Service Provider (ANSP) is responsible for providing the air traffic control services in flight information region (FIR) (NATS, 2020). Dutch FIR is the airspace between the geographic borders of the Netherlands and a large part of the airspace above the North Sea, as can be seen in Figure 4. Dutch airspace can also be referred to as Amsterdam FIR.

According to Eurocontrol (2016), a FIR is "a three-dimensional area in which aircraft are under control of usually a single authority (i.e., ANSP)". A FIR could be divided into several smaller areas, in both longitudinal and lateral directions. Smaller countries may have one FIR in the airspace above them while larger countries may have several. Sometimes one or more FIRs have a combined upper area control. Dutch airspace is divided into two layers: the lower FIR called Amsterdam ACC (EHAA[4]) and the upper information region (UIR) called Maastricht ACC. LVNL is the ANSP responsible for EHAA, the airspace ranging from sea level to 24,500 feet which covers the complete longitudinal area of the airspace as seen in Figure 4. Maastricht ACC starts at an altitude of 24,500 feet and is controlled by Maastricht Upper Area Control center (MUAC). It covers airspace over Belgium, the Netherlands, Luxembourg, and north-west Germany, one of Europe's busiest and most complex airspace areas (Eurocontrol, 2016).

[4] The International Civil Aviation Organization (ICAO) has defined codes based on four letters to designate airports and other locations around the world. The first letter of the code represents the area of the world, the second letter is unique to a country, and the last two letters are used to indicate a location such as an airport. Schiphol for example has ICAO airport code EHAM, where 'E' stands for Northern Europe, 'H' for the Netherlands and 'AM' for Schiphol Amsterdam Airport. Aside from airports, the 'EH' prefix is also used for other locations like weather station De Bild (EHDB) and the Dutch FIR (EHAA). In the rest of this document the ICAO prefixes will be used when referring to Schiphol (EHAM) and Dutch FIR (EHAA).

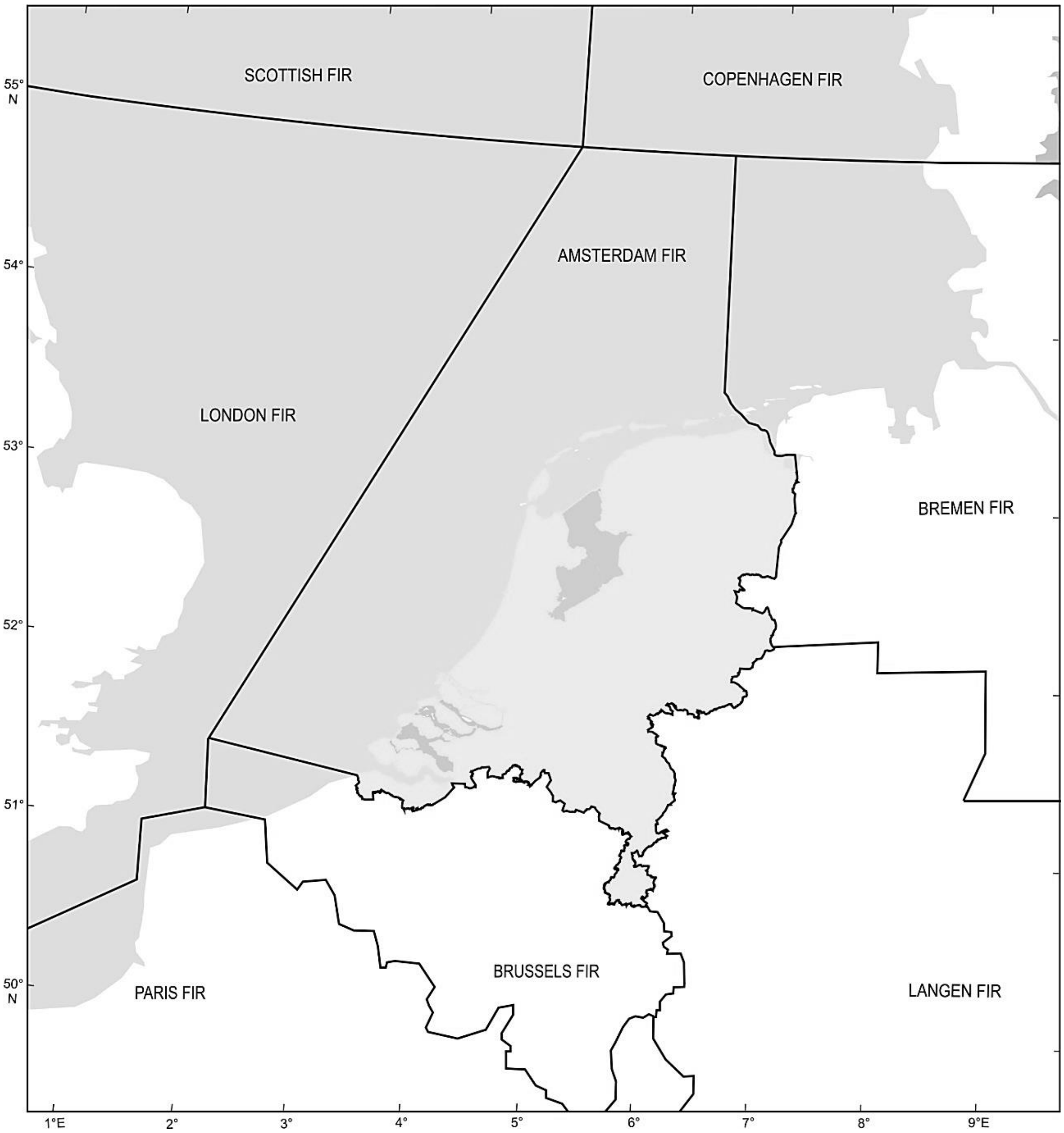


Figure 4 – Chart illustrating the borders of EHAA

Amsterdam ACC is a control area which contains civil and military parts. The civil part is divided in 5 sectors that lay around the EHAM terminal maneuvering area (TMA[5]). In and outbound traffic for EHAM must cross at least one of the sectors, as can be seen in Figure 5. EHAM inbound traffic must enter the TMA at one of the three initial approach fixes (IAF, or also referred to as holding stacks): ARTIP, RIVER or SUGOL. Displayed are the five sectors, with sector 1 in the top right and sector 2, 3, 4 and 5 laying clockwise around the EHAM TMA. In addition, the three holding stacks are shown on the TMA border.

[5] "A TMA is a designated area of controlled airspace surrounding a major airport where there is a high volume of traffic." (Eurocontrol, 2020)

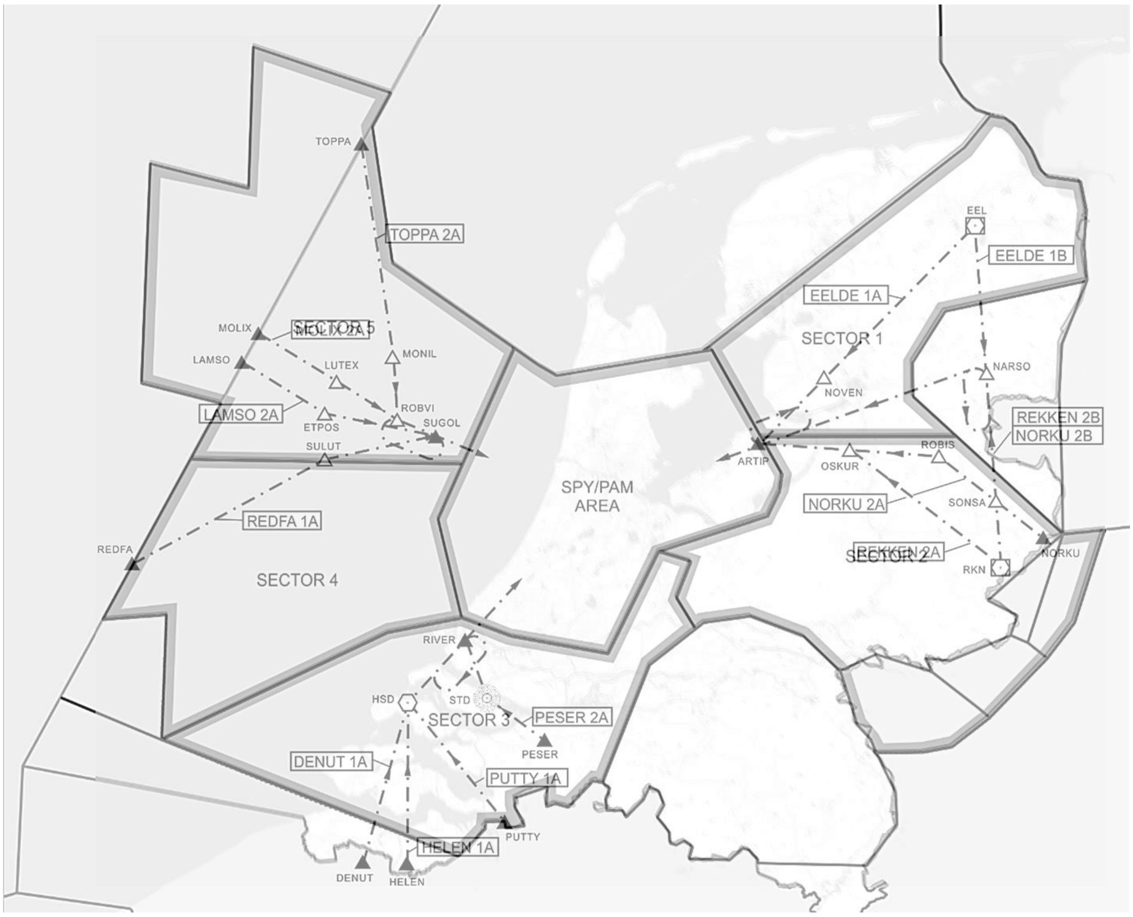


Figure 5 – Sector 1-5 clockwise around EHAM TMA, plus the holding stacks on the TMA border

2.4 ATFCM phases at LVNL

The four ATFCM planning stages that Eurocontrol uses for managing traffic flow, as displayed in Figure 3, were also adopted by LVNL for capacity management on the ATC level. At LVNL, activities conducted during the strategic phase entail among others research and planning in collaboration with the NM or other ANSPs. During this phase big differences between demand and capacity can be identified and resolved. During the pre-tactical phase, the pre-tact unit creates a short-term planning for the days and hours before operations, also known as the OPS plan. The OPS plan aims to align demand and capacity based on the latest operational conditions (e.g., weather, availability of staff, special events). The post-operations phase is the moment where the operations are assessed. This allows for a feedback mechanism that provides insights to improve future planning in the strategic, pre-tactical and tactical phases. This step in ATFCM is extensively executed at LVNL. However, the specific evaluation of the OPS plan is, at the moment of conducting this research, not yet fully established.

2.5 Capacity management at LVNL

In the summer of 2016, air traffic controllers at LVNL reported high workloads at ACC sector 3. This led to the pre-study ‘Solving bottlenecks Sector 3’ (Westerveld, 2017) which investigated the possibilities of improving capacity management (see Figure 6). The foundation for this study was that Schiphol risked facing a large problem, as it was expected to reach its maximum limit of 500,000 movements per year by 2017 or 2018. Short term solutions were expected to solve the high workloads for the following years. However, by 2021 the government may allow Schiphol to expand, so long-term solutions for improving capacity management were also required. Most recommendations of this initial study were considered not feasible in the short term, because the data necessary for making decisions related to capacity management was not accurate enough.

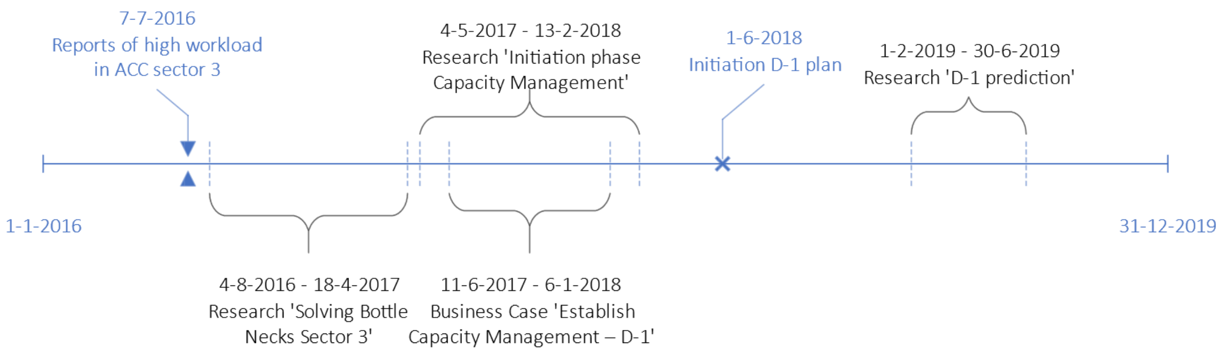


Figure 6 – Timeline of research studies towards capacity management

In July 2017 (see Figure 6), LVNL prioritized the development of establishing capacity management, which lead to the study 'Initiation Phase Capacity Management' (van Delden, Westerveld, Lieshout, Obbens, & Devilee, 2017). This study aimed to develop a plan of approach for designing capacity management at LVNL. The necessity of capacity management was supported by the following issues.

- Growth of the aviation industry leads to shifting and/or creating capacity bottlenecks.
- Increasing number of delays in the European network.
- Further increasing workloads for the air traffic controllers.
- Dynamic traffic demand (peak hours) of Schiphol requires a system to adapt dynamically.
- Insufficient collaborative decision making among stakeholders hinders increasing capacity.
- Inaccurate predictions of traffic demand could lead to more nuisance to the environment (e.g., noise and fuel emissions).

This was echoed by findings of the 2017 study called 'Capacity Management: design D-1' (Lieshout, Devilee, & Obbens, 2017) (see Figure 6). This study researched the implementation of an OPS planning as improvement to capacity management. The following flaws that impede improvements on capacity planning were identified.

- Within the operation, information related to capacity impact (e.g., NOTAMs, weather forecast, Daily Operating Plan, Airspace Use Plan etc.) is scattered.
- LVNL has limited sight on the expected traffic demand of the next day. This might cause residual capacity to not be used.
- Personnel planning in alignment with the demand on the day of operation is challenging due to lack of (technical) support.
- Deployment of capacity limiting regulations is avoided as possible negative consequences are unknown.
- Traffic demand in Dutch airspace is not aligned with the capacity, causing increasing delays when demand expands.

These problems have many practical implications as regulating traffic in Dutch airspace caused on average 600,000 minutes of delay for aircraft operators between 2012 and 2017 (van Delden et al., 2017), as can be seen in Figure 7.

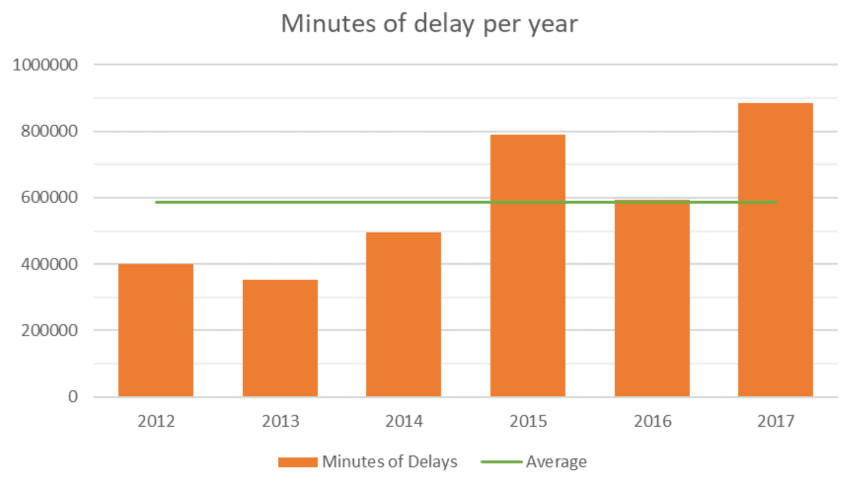


Figure 7 – LVNL related delay for airlines from 2012 until 2017

The cost of one minute of delay was in 2015 estimated to be €50 which translates to costing airlines roughly €30,000,000 per year (Lieshout, Devilee, & Obbens, 2017). The OPS plan is expected to reduce the amount of minutes of delay by approximately 1 to 5%, which is a cost reduction of €300,000 to €1,500,000 per year (Lieshout, Devilee, & Obbens, 2017). Airlines will also benefit from the increased predictability of operations provided by the OPS plan, as it will allow airlines to optimize their processes and better anticipate on irregularities.

Van Delden et al., (2017) compared the capacity management procedures at LVNL with those of NATS and SkyGuide, two other European ANSP's. To improve their capacity management procedures, both these ANSPs had implemented a method for creating an operational planning for every day. The comparison showed several deficiencies in the four capacity management phases at LVNL.

Van Delden et al., (2017) found that during the strategic phase, capacity management rarely happens, even though the knowledge and tools to do so are available. Additionally, when it does occur, the process is time consuming, and results are difficult to interpret. The processes and tools used during pre-tactical planning were considered to lack consistency, making it hard to anticipate the implications of disruptions. Additionally, only in cases of extreme weather was a planning briefed to local stakeholders. In the tactical phase, the biggest concern was that decisions about regulating the traffic flow were being made using unreliable information. Thereafter, in the post-operational phase, very few improvements for the strategical and pre-tactical phases were proposed.

In addressing the shortcomings identified by van Delden et al., (2017), Lieshout, Devilee, & Obbens (2017) suggested the implementation of an OPS plan. An OPS plan is a planning for the day of operations, created the day before. The overall aim of an OPS plan is to align the traffic demand with the available capacity, prior to of the day of operations (Lieshout, Devilee, & Obbens, 2017). This would lead to an increase in safety and efficiency because the operations get more predictable, and the use of capacity is optimized. When the aim of the D-1 plan is met, it provides several benefits.

- A clear, complete, and transparent insight into traffic demand, capacity and availability of airspace, weather, and irregularities the day before operations.
- Predictable operations and services that could help stakeholders improve their processes.
- Collaborative decisions made with adjacent ANSPs regarding the distribution of capacity.
- Optimal utilization of the available airspace in collaboration with military air traffic control.
- Reduction in workload for supervisor and traffic controller on the day of operations.
- Schedules that are better aligned with the demand for personnel on the day of operations to increase efficiency on the deployment of staff.

Lieshout, Devilee, & Obbens (2017) found that the implementation of an OPS plan could be done with various degrees of detail. Each additional step is retrospective, including information of the previous version, resulting in more comprehensive planning each time. The most basic version gathers already existing, but scattered information. The second level was to include traffic demand predictions (provided by the NM) and schedules for air traffic controllers. Next, it is possible to use the planning of adjacent centers to create scenarios for the day of operations. Finally, the most detailed version adds the possibility to create a planning in such a way that both the deployment of air traffic controllers and usage of runways are dynamic, changing to meet the needs of a given moment.

Following the introduction of OPS plan concepts, research was conducted that assessed the reliability of the traffic demand predictions made in the early OPS plan concept by comparing them to the actual demand that occurred on D0 (Wassenberg, 2019). In comparing the datasets, a difference in the recorded times of entry was found, in addition to the number of entries measured. This was the case across several airspace segments described in the OPS plan, such as: traffic entering sector 2 and 3, traffic passing the three stacks (ARTIP, SUGOL and RIVER), traffic in- and outbound at EHAM, inbound traffic to EHFIRAM and peak hour traffic.

The study found significant differences between entry times for traffic within the various airspace segments, which experts supported by considering the initial OPS plan predictions inaccurate. The conclusion of the study, however, is that the prediction data might still be useful, if broken down. For example, the predictions for the first peak hour seam acceptable, where the accuracy for the rest of the day variates.

Furthermore, the study claims that the accuracy of the predictions could be increased by implementing more input and by improving collaboration between the network manager and ANSP's. The implementation of an OPS plan does show potential, and further research was deemed compulsory. Evaluation of the accuracy should be more thorough and the causes for the inaccuracy should be identified.

Wassenberg (2019) also researched the relevance of an OPS plan at KLM, the main Dutch airline which, together with its codeshare partners (SkyTeam) are responsible for approximately 64% of the total air traffic movements at Schiphol (LVNL, 2020). This study found that KLM already creates an operational plan for internal use and is therefore dependent on the information that LVNL provides on the day of operations. If a more accurate capacity planning can be provided by LVNL on the day before operations, KLM would be able to make better predictions for their own and SkyTeam's operations. As supported by the finding of Lieshout et al., (2017), this provides opportunities for avoiding cancellations and reducing delays. Risks would consequently be smaller, and the financial impact of delays reduced. Consistent and transparent capacity planning allows for better anticipation on disruptions, and improved usage of the available capacity, which is beneficial for all stakeholders.

2.6 Concluding remarks

A growing industry causes new – and shifting of existing – capacity bottlenecks for LVNL and its network. The bottlenecks entail increasing workloads for air traffic controllers, higher delays in the network, and inefficient use of the airspace. By means of air traffic flow and capacity management, LVNL and several other stakeholders try, through collaborative decision-making, to ensure optimal usage of the available air- and landside capacity. During the pre-tactical planning phase of capacity management, an OPS plan is created which enables LVNL to share information regarding the operations throughout the organization and with partners and stakeholders. Since the OPS plan was first conceptualized in 2018, the information, methods and conditions have changed. A research study in 2019 concluded that the traffic demand predictions included in the OPS plan were not accurate and more research was needed to improve the predictability of the plan. The analyses conducted in this study are part of the post-operations phase and aim to provide feedback that would help improve the predictions made in future OPS plans.

3 Breakdown of the OPS plan

The OPS plan implemented since 2020 is a document that is created by the pre-tact unit on all business days, for the next operational day. It is a collection of existing information (i.e., input information) that could have an impact on the operations, like the weather forecast, scheduled maintenance, military events, or relevant information from adjacent ANSPs. In addition, some of this information is used to make predictions (i.e., output information) that aims to help LVNL to prepare and adjust the operations.

3.1 Input of the OPS plan

The main information analyzed by the pre-tact unit to produce the OPS plan can be divided into four types: traffic forecast, weather information, special events and preferred runway combinations.

3.1.1 Traffic forecast

Eurocontrol provides operational stakeholders with information of historical and future traffic demand throughout the network. This information is maintained and made available through the Demand Data Repository (DDR). Traffic trends are forecasted based on past traffic flows and information shared by network members. New data is collected and analyzed daily to support strategic and pre-tactical planning (Eurocontrol, 2021). Pre-tactical planning is supported by the PREDICT system which combines historical data with expected demand (Eurocontrol, 2020). The PREDICT data is used as an input parameter by LVNL to predict the traffic demand for Amsterdam ACC, and several smaller segments of it. The PREDICT dataset contains expected trajectories for individual flights.

The PREDICT data from Eurocontrol is processed in NEST (Network and Strategy Tool), a tool created by Eurocontrol, used by the pre-tact unit to analyze, and process the DDR data, for the purpose of facilitating airspace design and capacity planning. This allows for determining the optimal use of resources (e.g., runway capacity). Based on the flight trajectories in PREDICT, NEST calculates the corresponding entry and exit times of the sectors, stacks and as well as take-offs and landings at different airports, as explained in sub-section 2.3. The predictions obtained from NEST are reviewed and might be adjusted based on expert judgement and the latest meteorological information. NEST is also used for simulating the effect of regulations on the operation (see sub-section 3.2.4) (LVNL, 2020).

3.1.2 Weather information

The Koninklijk Nederlands Meteorologisch Instituut (KNMI, the royal Dutch meteorological institute) provides a Schiphol Chance Prediction (SKV, Schiphol Kans Verwachting), displayed in Figure 8. A SKV is a table with weather predictions specifically concentrated on the conditions at EHAM. The table consists of six types of information (horizontally), and the predictions are made per three-hour intervals for the day of operations (vertically). The initial SKV provided on the day before operations gives a forecast up to 18:00. An updated prediction for the remaining part of that day is provided at approximately 11:00.

	00	03	06	09	12	15	18
Zicht < 5 km en/of wolkenbasis < 2000 ft (%)	10	15	15	10	10	10	5
Zicht < 5 km en/of wolkenbasis < 1000 ft (%)	5	10	10	5	0	0	0
RVR < 1500 m en/of wolkenbasis < 300 ft (%)	0	0	0	0	0	0	0
RVR < 550 m en/of wolkenbasis < 200 ft (%)	0	0	0	0	0	0	0
RVR < 350 m (%)	0	0	0	0	0	0	0
Windrichting (deg)	050	060	050	050	050	040	040
Windsnelheid (kt)	8	9	8	8	10	9	7
Windstoten (kt)		14				15	12
Standaarddeviatie windrichting (deg)	15	15	20	20	20	25	35
Standaarddeviatie windsnelheid (kt)	2	2	2	2	2	3	3
CB (%)	0	0	0	0	0	5	5
Onweer (%)	0	0	0	0	0	0	0
Aircraft induced lightning	N	N	N	N	N	N	N
Temperatuur (C)	-5	-5	-6	-5	-2	-2	-3
Dauwpunt (C)	-6	-6	-7	-6	-5	-5	-6
Relatieve vochtigheid (%)	93	93	93	93	80	80	80
Gevoelstemperatuur (C)	-11	-11	-12	-11	-8	-7	-8
Sneeuw (%)	0	0	0	0	0	5	5
Matige/zware sneeuw (%)	0	0	0	0	0	5	5
Onderkoelde neerslag (%)	0	0	0	0	0	0	0
Luchtdruk (hPa)	1009	1008	1009	1011	1013	1015	1018

Figure 8 – Schiphol weather prediction provided by KNMI

The first type of information in the SKV is related to visibility. The visibility is separated into five levels, each having a predicted cumulative probability of occurrence per time interval. The first two levels entail a visibility of less than 5000m and/or a ceiling, that being the altitude of the bottom side of the clouds, of less than 2000 or 1000ft. These conditions are considered to produce marginal visibility, making it so that visual approaches are prohibited, and some runway combinations cannot be used. The other three levels are considered Limited Visibility Conditions (BZO, Beperkt Zicht Omstandigheden). There are four BZO categories: A, B, C and D. BZO A can be described as reduced visibility and occurs when the runway visual range (RVR) is less than 1500m and/or a ceiling of less than 300ft. A reduced visibility only has an impact on ground operations regarding departing traffic. BZO B, C and D are considered low visibility. BZO B occurs when RVR is less than 550m and/or a ceiling of less than 200ft, C when RVR is less than 350m, and D when RVR is less than 200m. No probability is predicted for BZO D conditions in the SKV (LVNL, 2020).

The second part of the SKV provides information regarding wind: the expected direction and speed, gusts and an expected deviation for the direction and speed (KNMI, 2011). Then the probability predictions for Cumulonimbus clouds (CB) and thunderstorms are provided. CB are clouds that could have a low ceiling yet could be as tall as 15km above sea-level. These clouds can produce severe weather conditions such as thunder, rain, and hailstorms. Furthermore, it is provided whether there is a chance of aircraft induced lightning, a phenomenon where the airplane itself produces lightning (Chesapeake Aviation Training, 2015).

The last three types of information are: the forecasted temperature (including dewpoint temperature, humidity, and apparent temperature), probability of snow and freezing precipitation, and the predicted air pressure.

In addition to this prediction, KNMI also gives a short briefing every morning. The briefing is intended to further explain the SKV as well as to provide additional, more up to date information. The predictions, and when applicable the remarks discussed in the briefing, are included in the OPS plan.

3.1.3 Special events

The OPS plan also intends to inform operational personnel and stakeholders about events and disruptions that might impact operations. The events entail: OHD events, NOP, scheduled maintenance, and Notices to Airman (NOTAMs).

OHD events are non-standard aviation related events that require special attention. Examples of these events are test/survey flights, parachute jumps from aircraft and temporary or reoccurring air space closings. These events are evaluated by experts who can judge what the impact of these situations will be on the capacity or operations. The information regarding the events contains the time and location at which the event takes place, the callsign if appropriate, and the impact it is expected to have on the operations (Westerveld, 2017).

The Network Operations Portal (NOP) is an online portal hosted by Eurocontrol. The network manager provides information that could have an impact on one or more ANSPs in the network (Eurocontrol, 2020). This information could for example be international military operations, organized strikes in adjacent airspaces or other disruptions impacting the operations anywhere in the network.

Scheduled maintenance to systems that may have an impact on the capacity, or the operational process are also included in the OPS plan. A distinction is made between maintenance to systems that fall under LVNL's responsibility, like surveillance systems, and systems of other organizations such as runways, or flight plan processing systems at adjacent centers.

A NOTAM is an announcement that contains information essential to airspace users concerned with flight operations. The information provided through NOTAMs are usually not known far enough in advance to be announced in through the channels mentioned above (FAA, 2021).

3.1.4 Preferred runway combinations

The preferred runway combinations and the related capacity are mainly defined in so-called Quick Reference Charts (QRCs), documents containing useful information used internally by LVNL for quick reference. The two main QRCs used for making runway predictions are QRC09 and QRC25. QRC25 contains the most preferred options for the runway combinations used during a given peak moment, as well as scheduled runway inspections.

Traffic demand at EHAM is not constant but is distributed over multiple 'waves' (also referred to as peaks) throughout the day. There are three different types of peaks: inbound peak, outbound peak and off-peak. Inbound and outbound peaks are usually characterized using respectively two runways for landing, one for take-off (2+1) and one runway for landing, two for take-off (1+2). During the off-peak one runway is used for landing and one for take-off (1+1). Additionally, a night period is defined where also a 1+1 runway combination is applied. In some cases, when traffic demand does not comply to the defined peak periods, a 2+1+1 runway combination could be used (LVNL, 2021).

The table in Figure 9 is an example of the information found in QRC25. It describes the four most preferred runway combinations for the morning inbound peak and their visual requirements, as determined in the Nieuwe Normen Handhavingsstelsel. The table also mentions the least preferred usage and constrains regarding the use of a 2+2 combination or other combinations.

INBOUND PEAK				
0620 – 0740 UTC				
Pref	LND-1	LND-2	STR	Vereist
1	06	36R	36L	VIS ≥ 5 km en CEIL > 2000 ft
2 *	18R	18C	24	RVR ≥ 200 m
3	36R	36C	36L	RVR ≥ 200 m
4	18R	18C	18L	RVR ≥ 550 m en CEIL ≥ 200 ft
5	Overige combinaties			

- 2+2 baangebruik tot 1430 UTC toegestaan.
- Minst preferent: LND 27, LND 22, STR 09.

* Als 550 m ≤ VIS < 5 km of 200 ft ≤ CEIL ≤ 2000 ft, preferentie 4 i.p.v. 2 toepassen.
Als VIS ≥ 5 km en CEIL > 2000 ft, en inbound peak valt buiten UDP, mag TWR-SUP zelf bepalen of preferentie 2 of 4 wordt gebruikt.

Figure 9 – Part of QRC25 containing the preferred runway combinations for the inbound peak

QRC09 contains most of the possible runway combinations and their corresponding runway capacity. The runway capacity is defined as the maximum number of movements per hour, given a nominal traffic mix. The capacity is provided for the three peaks and the night period. The table in Figure 10 is an example of the information found in QRC09. The figure shows the first twelve runway combinations for the inbound peak. The capacity is provided for multiple visibility conditions: good visibility (during or outside of Uniform Daylight Period (UDP)), marginal visibility and BZO A (with two or three ground controllers (GC's)), B and C conditions.

INBOUND peak		Capaciteit						
Landen / Starten		Goed zicht [1]		Marg. zicht	BZO A		BZO B	BZO C
		UDP	Not UDP		2 GC's	3 GC's		
06+09	09	35+20 / 20						
06+18C	09	34+34 / 37	32+35 / 32					
06+18C	18L	34+34 / 37						
06+18R	09	34+34 / 37	32+35 / 32	32+35 / 32	28+28 / 24	30+30 / 30	22+22 / 22	17+17 / 23
06+18R	18L	34+34 / 37	32+35 / 32	32+35 / 32	28+28 / 24	30+30 / 30		
06+27	36C	34+34 / 37	35+32 / 32					
06+27	36L	34+34 / 40	34+34 / 40					
06+36R	09	34+34 / 32						
06+36R	36C	34+34 / 40	34+34 / 40					
06+36R	36L	34+34 / 40	34+34 / 40					
18C+22	18L	35+30 / 30	35+30 / 30	35+20 / 30	30+20 / 20 [2]	30+20 / 20 [2]		
18C+22	24	35+30 / 37	32+30 / 32	32+30 / 32	28+28 / 24 [2]	30+30 / 30 [2]		

Figure 10 – Part of QRC09 containing the capacity for some of the inbound peak runway combinations

Detailed information about the preferred runway combination and its related declared capacity can be found in internal documents from LVNL. Based on QRC25 and QRC09 and the other information mentioned in this chapter, an appropriate runway combination is selected and proposed for usage on the day of operations. The main factors that affect the different combinations in QRC09 are runway availability, wind direction and speed, visibility, and traffic demand. Wind direction is important as it is preferred to land and take-off with head wind. The traffic demand affects the amount of capacity required and the visibility might cause limitation on the use of specific runways or combinations.

3.2 Output of the OPS plan

The information discussed in section 3 is used to make four predictions for the day of operations: traffic demand, runway combinations, sector configuration and (when applicable) regulation scenarios.

3.2.1 Traffic demand

Based on the PREDICT data and NEST tool from the NM, four predictions are made for different airspace segments: EHFIRAM[6] inbound, the three inbound holding stacks, the five sectors and the number of landings and arrivals at EHAM. The prediction for EHFIRAM includes traffic inbound to EHAM, which is measured at the ACC entry. The demand is predicted per 60 minutes 20 min rolling and per 20 minutes, as can be seen in Figure 12.

Per 60 minutes 20 min rolling (hereafter also referred to as per hour rolling) means that for each of the twenty-minute intervals the movements of the next two intervals are added. This principle is illustrated in Figure 11. Consequently, a flight (e.g., with entry time 11:19:55) is counted in three rolling hour intervals. This allows for a continuous account for upcoming movements; as soon as a 20-minute interval passes (e.g., 10:20:00-10:40:00), the planning horizon extends so that the movements for the next 40-60 minutes are always considered. This gives a better understanding of the amount of traffic to be expected in contrast to only looking at fixed 20-minute intervals.

[6] EHFIRAM is a term used for a traffic volume that indicates EHAM air traffic movements within EHAA (To70 & NLR/FerWay, 2020).

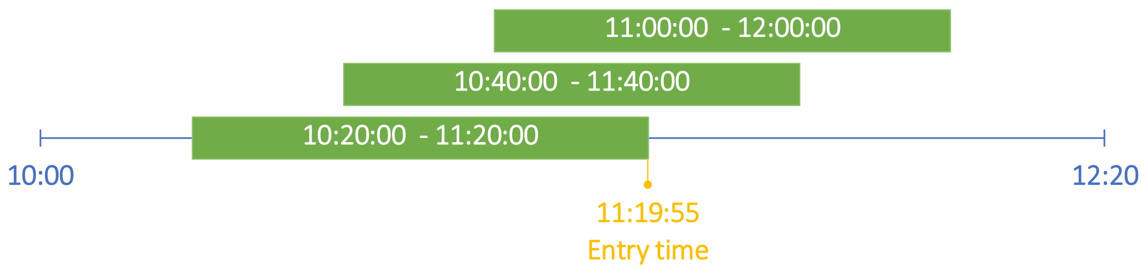


Figure 11 – Illustration of a rolling hour window

The graphs in Figure 12 include the expected traffic demand and the nominal (i.e., unregulated) runway capacity (sub-section 3.2.2). If regulations (sub-section 3.2.4) are expected, the reduced capacity is also displayed. As can be seen when comparing 60 min rolling 20 min to inbound per 20 min, there is a more uniform distribution as a flight is accounted for in three 20-minute intervals, instead of one.

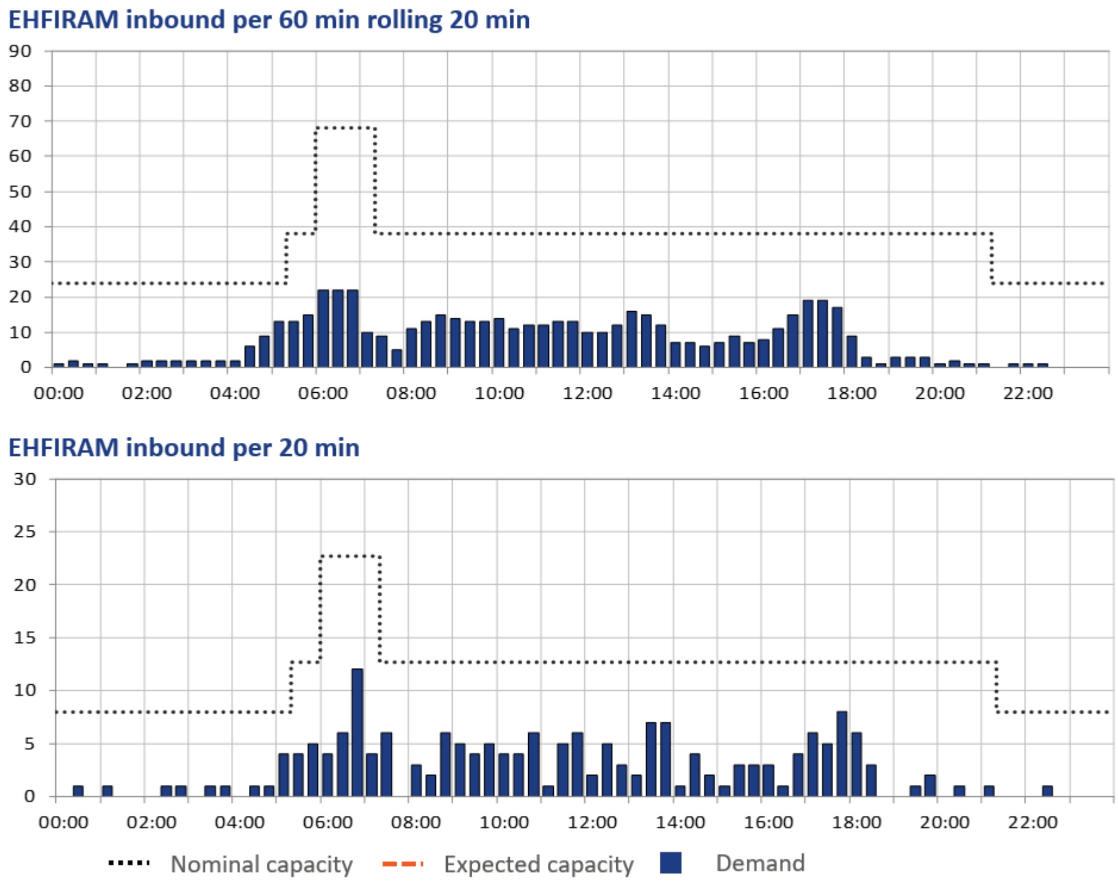


Figure 12 – EHFIRAM inbound traffic per 60 minutes 20 min rolling (top) and per 20 minutes (bottom)

The predictions for the three holding stacks contain information regarding EHAM inbound traffic demand measured at the three inbound holding stacks: ARTIP, RIVER and SUGOL. The demand is predicted for each of the stacks per 20 minutes. The example in Figure 13 shows the nominal (i.e., unregulated) capacity of the ARTIP stack, as there is a maximum number of aircraft that could hold at each stack per unit of time. If the occurrence of bunching[7] is expected, it is also presented in the graph.

[7] Bunching is the phenomenon where traffic demand is not distributed evenly but appears in bunches. In that case the hourly capacity might not be exceeded yet bunching could cause overload of the ATS system (LVNL, 2019).

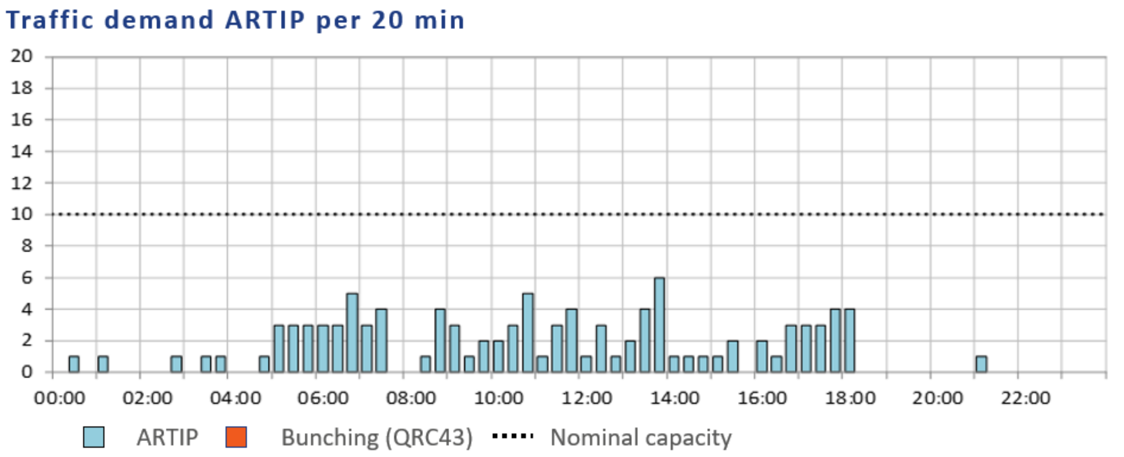


Figure 13 – EHFIRAM inbound traffic measured per 20 minutes at ARTIP, one of the three holding stacks

The demand is also predicted for the three stacks combined per 20 minutes, as seen in Figure 14. Since the predicted traffic is mostly EHAM inbound, this graph contains the nominal runway capacity as a reference.

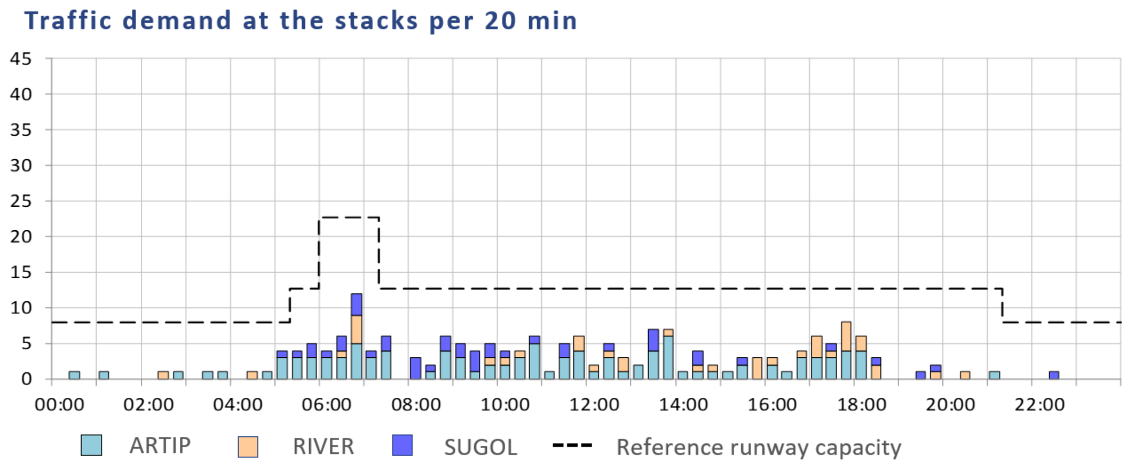


Figure 14 – EHFIRAM inbound traffic measured at the three holding stacks per 20 minutes

The traffic demand for EHFIRAM is predicted per 20 minutes for the five sectors. This data represents the in- and outbound EHAM movements, transits, and regional traffic to and from the other Dutch airports and is presented in three separate groups: sector 1+2 (shown in Figure 15), sector 3 and sector 4+5. The traffic prediction is split in three groups to illustrate how the traffic is distributed through the ACC sectors.

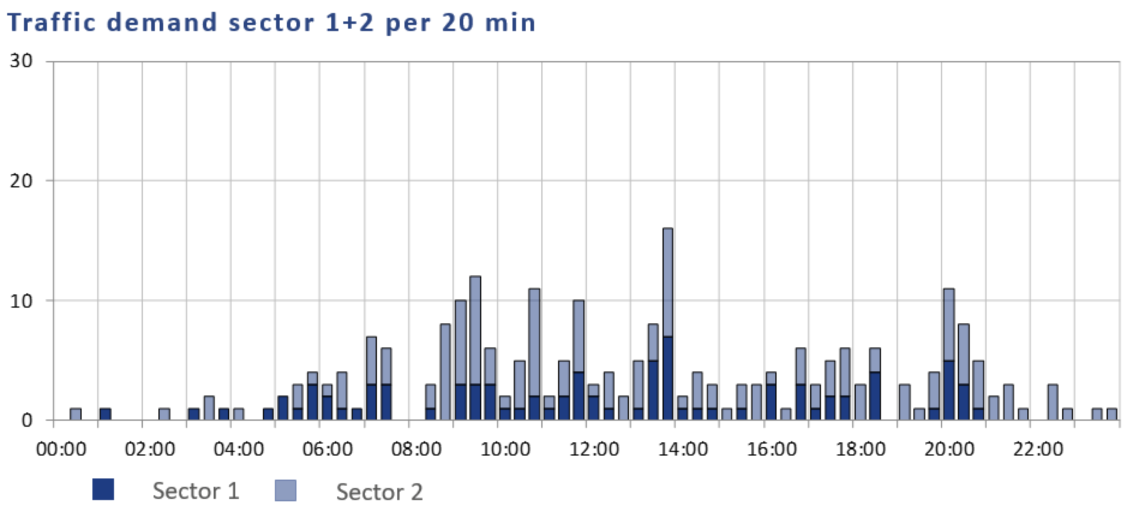


Figure 15 – EHFIRAM traffic measured at sector 1 and 2 per 20 minutes

Finally, the take-offs and landings per 60 minutes (per 20 min rolling) at EHAM are predicted, as can be seen in the graph in Figure 16 with the take-off in the upper part of the graph, and the landings in the lower. In the graph the nominal runway capacity is also displayed. The runway capacity fluctuates based on the number of runways in use during different peak periods. If the capacity is reduced, for example due to weather influences, the expected decrease will be displayed in the graph as well. In case of high traffic demand (i.e., when the demand exceeds the nominal runway capacity) regulations might be introduced to spread the demand over a longer period of time.

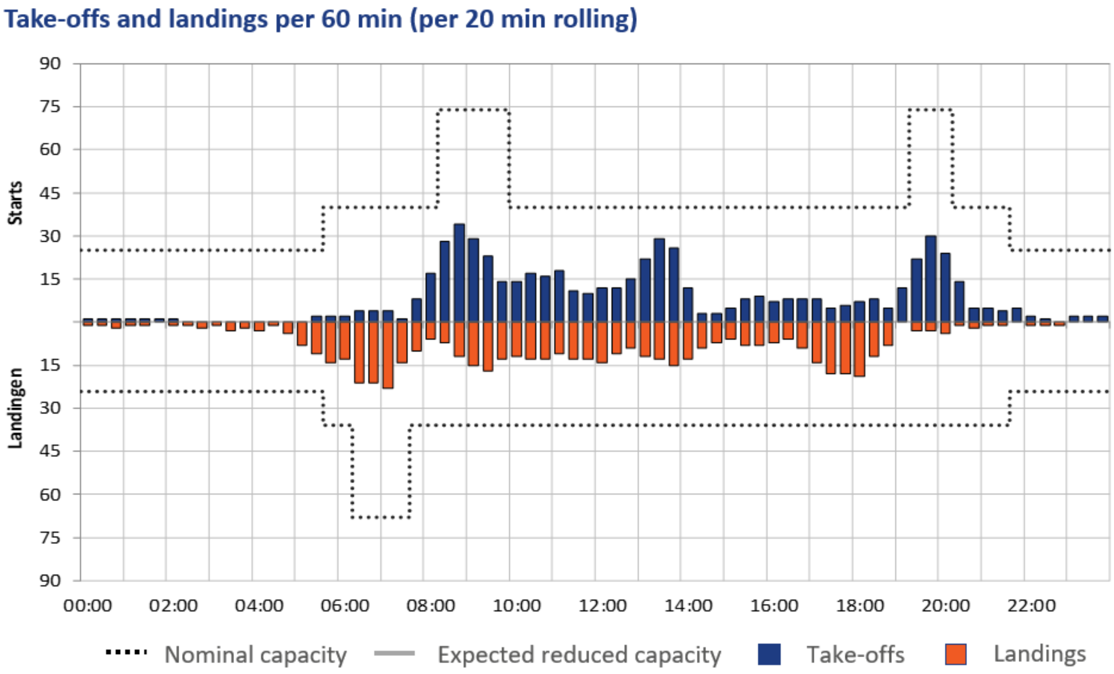


Figure 16 – Take-offs and landings at EHAM measured at the runway per 60 minutes per 20 min rolling

3.2.2 Runway combinations and capacity

Based on the KNMI SKV, QRC25 (preferred runway combinations), QRC09 (runway combinations and capacities) and the expected traffic demand, a runway combination is proposed for the day of operations. Operational expertise is required to describe an appropriate runway combination based on the available data. The pre-tact unit specifies runway combinations for the first in- and outbound peaks, and if possible, predictions for the rest of the day are provided (LVNL, 2020). A runway prediction entails a primary and alternative combination, where an alternative is usually proposed if there is a chance of changing weather conditions on the day of operations, during the peak period. The combinations are presented in the OPS plan as can be seen in Figure 17. The runway combinations are usually also justified by comments that explain why certain combinations are proposed. In some cases, additional capacities are mentioned (in parentheses shown in Figure 17). These are additional expected capacities if, for example, there is a chance of reduced visibility on the day of operations.

Period	Possible runway use				Possible alternative runway use				Comments
	Landing		Take-off		Landing		Take-off		
Inbound peak Capacity	**18R** 34	**27** 34	**24** 40		**18R** 35	**22** 30	**24** 40 (35)		- Alternative runway use if visibility < 5km and/or CB <2000 - (35) if marginal visibility
Outbound peak Capacity		**18R** 38	**24** 37	**18L** 37					
Off-peak Capacity		**18R** 38	**24** 40						

Figure 17 – Proposed runway use and the corresponding available capacity for the different peak periods

3.2.3 Sector configuration

The sectors can be combined or split up during specific moments of the day, as seen in Figure 18. The amount of sector groups (i.e., 1, 2, 3 or 4) can relate to the number of controllers needed. The decision to split the sectors in separate groups is calculated based on the amount of traffic demand that is expected.

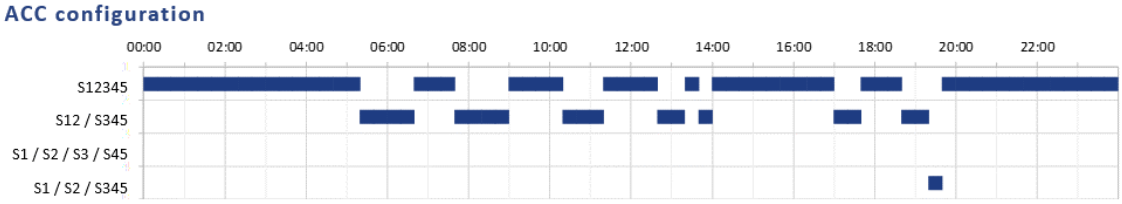


Figure 18 – Proposed sector configuration

3.2.4 Regulation scenarios

A regulation scenario is introduced to decrease traffic demand at the FIR entry. NEST is used to simulate the regulations and determine if the regulation is effective (LVNL, 2020). Possible regulation scenarios are based on all the information discussed above. They describe a scenario where the demand is higher than the available capacity. Regulations are introduced when the weather conditions limit the runway capacity and when demand is just higher than the nominal runway capacity.

3.3 Concluding remarks

The OPS plan contains several types of information and predictions that could impact the traffic demand or the available runway capacity on the day of operations. The information required for the OPS plan is retrieved from several locations and provided by many stakeholders. The information plays a big role in the development of the predictions. Three of the predictions are going to be analyzed in this study: the predictions for traffic demand, runway capacity and runway combinations.

The traffic demand is predicted for four segments of the ACC, as explained in section 3.2.1. It was chosen to analyze predicted inbound traffic demand at EHFIRAM for this research study because high accuracy of this type of traffic demand is important as it describes two possible bottlenecks of the ATS system: the capacity of the holding stacks and the demand to the runway capacity. Outbound traffic demand is predicted in the OPS plan but will not be analyzed.

Runway combinations describe their usage and the capacity, where the usage means the direction in which the runway is positioned. The analysis in this study is focused on determining the accuracy of the usage and capacity predictions for the first inbound peak of each day.

4 Research methodology

This research study analyzes two aspects of the OPS plan: the inbound traffic demand predictions for EHFIRAM and the runway combination and capacity predictions during the first peak hour of the day. The flowchart in Figure 19 displays the research methodology used for this study.

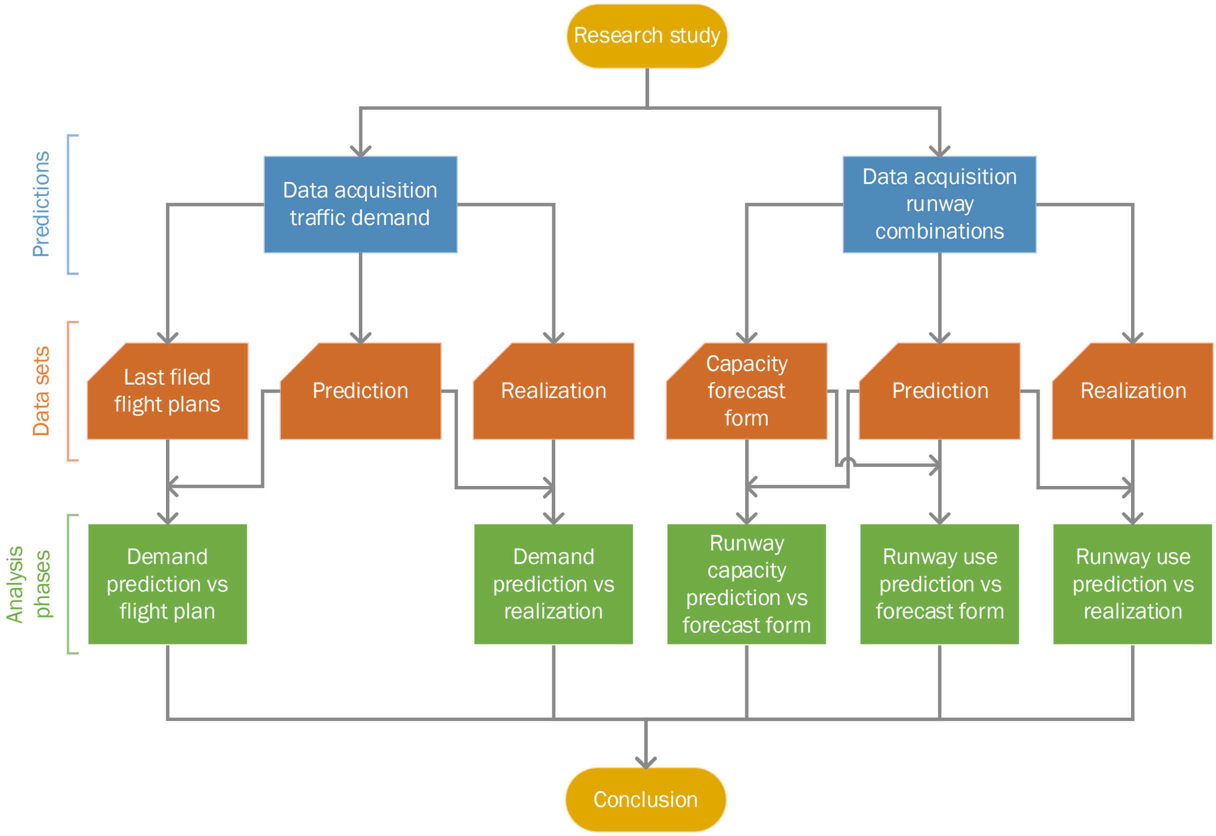


Figure 19 – Flow chart of the analysis framework

As can be seen in Figure 19, the traffic demand prediction made the day before operations by the OPS plan was compared with the last filed flight plans and the realization. The realization is the actual amount of traffic measured at D0, which was then compared to the OPS plan traffic demand prediction. Next, the OPS plan prediction was compared to the flight plans submitted until approximately 09:00 on the day of operations. While flight plans are continuously submitted by airspace users, the last filed flight plans are retrieved at 09:00 on D0 and are thus expected to be more accurate than the predictions made in the OPS plan.

Similar methodology was used for the data analysis of runway combination. The OPS plan runway combination prediction was compared with the latest capacity forecast, a prediction regarding runway use, that is created closer in time to the realization than the prediction made by the OPS plan. The runway combination prediction by the OPS plan was then also compared to the realization. Figure 20 depicts the approximate times at which the predictions described above are made.

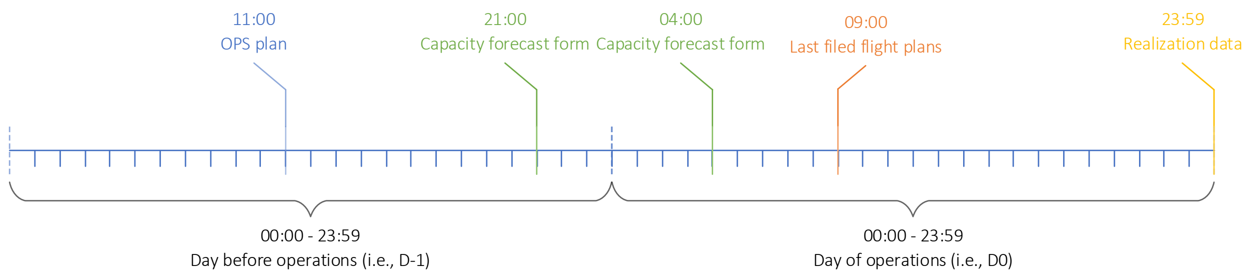


Figure 20 – Approximate timeline of activities related to traffic demand and capacity predictions in local time

The analysis was conducted in the programming language Python with the Pandas library. The Plotly libraries were used for creating most of the graphs, and the statistical tests required the SciPy library.

4.1 Data sample and collection

For this study, data was available for the timeframe between April 1st and November 14th of 2020. During this period, traffic demand at Schiphol was heavily impacted by uncertainties following the COVID-19 pandemic. Based on distinctive patterns in the traffic demand, the data was divided into three sections which were analyzed individually.

Most available data was extracted from databases as .csv files, however, the runway prediction data was not yet available as a dataset and thus had to be extracted manually from each OPS plan (see Appendix I). After extraction, the data was cleaned to only include essential columns from the exported data files, as some of the exports contain multiple columns of records that are not used for the analysis. Since the analysis is concentrated on EHFIRAM inbound traffic, only flight information that has location EHFIRAM and destination EHAM was selected (see Appendix II).

The datasets concerning the runway combination contain information for the whole day, yet only the first inbound peak was to be analyzed. The inbound peak times mentioned in the prediction were used as the baseline, and the same times were selected in the forecast form dataset. The peak times in the realization dataset were conceptualized as the first time in which two runways were in use, as the exact predicted peak times made in the OPS plan were not available in the realization dataset (see also inbound peak time chart in sub-section 5.7.4).

The times presented in the graphs hereafter are expressed in Coordinated Universal Time (UTC), unless otherwise stated. This results in some graphs displaying a shift in observations on October 25th, the day where daylight saving time ended and standard time started.

4.2 Analyzing the datasets

The predictions were analyzed in terms of accuracy and precision. As illustrated in Figure 21, accuracy is defined by how close the prediction is compared to a reference, while precision is defined by how close the predictions are compared to each other. Accuracy and precision are independent of each other, so observations can be accurate while not being precise, precise, and not accurate, both or neither (Forecast, 2021). The accuracy can be described by the mean difference (i.e., the sum of the observations divided by the total number of observations) between the prediction and the reference dataset. The precision can be described by the standard error (i.e., the standard deviation of the mean accuracy of a dataset) (Bhandari, 2020).

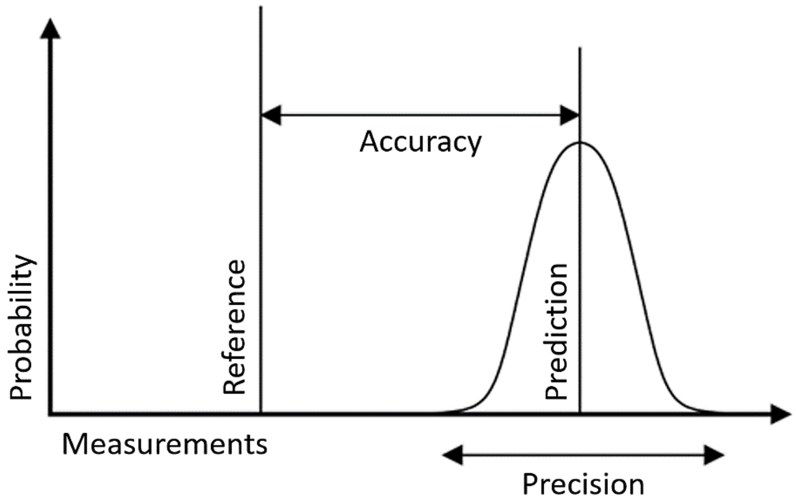


Figure 21 – Illustration of accuracy and precision

Two datasets were used as reference for the demand prediction: the realization and the last filed flight plans. For the runway combination analysis, the two reference datasets are the capacity forecast forms and the realization.

4.3 Testing for significance

The predictions are compared to reference datasets with statistical tests to determine if there is a statistically significant difference. The outcome of the tests is a p-value, or probability value, which indicates how likely it is that the data could have occurred under the null hypothesis (as stated in sub-section 1.6). The p-value is a proportion, so a p-value of 0.05 means that 5% of the time the test statistic would be at least as extreme as the one calculated if the null hypothesis were true. If the p-value is bigger than the alpha threshold of $\alpha = 0.05$, the null hypothesis cannot be rejected, and it could be assumed that there is no significant difference between what was predicted and the reference dataset. If, however, the p-value is smaller than 0.05, it could be assumed that the difference is statistically significant (Bhandari, 2020).

5 Traffic demand and runway analyses

To analyze the predictability of the traffic demand and runway usage predictions made in the OPS plan, each of the predictions is compared to two reference datasets which are composed on different moments, as described in section 4. The analysis is divided in five phases, where in each phase the prediction is contrasted to another reference dataset.

5.1 Framework

Figure 22 compares the number of flights at EHAM in 2020 compared to 2019. The graph shows that in April, the starting point of the period analyzed in this study, the traffic demand decreased with approximately 90% compared to that of 2019. The figure also shows the variation per day between 2019 and 2020 and the weekly moving average variation. The weekly moving average is a short-term indicator of how traffic demand is moving on average over seven days. The decrease in traffic demand might have an impact on the accuracy of the predictions made in the OPS plan.

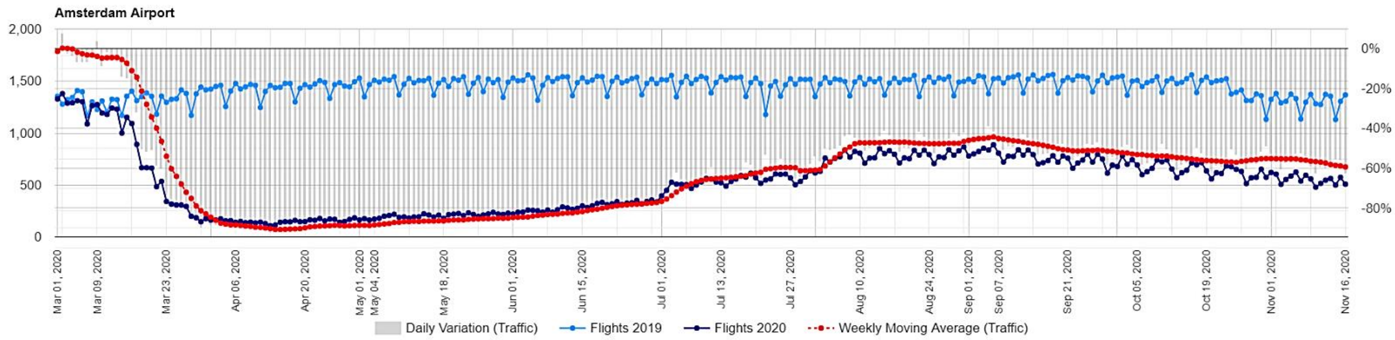


Figure 22 – EHAM traffic per day in 2020 compared to 2019. Source: Eurocontrol

5.1.1 Phase one and two – Inbound traffic demand EHFIRAM

Traffic demand translates to the amount of air traffic movements per unit of time. The analysis will determine the accuracy and precision per day and per hour rolling twenty minutes (see also subsection 3.2.1). However, the prediction and realization datasets used for this analysis contain timestamps for flights entering the EHFIRAM. The timescale in the datasets is therefore resampled to days and twenty minutes, meaning the number of flights is counted per day or twenty-minute intervals (e.g., a flight with entry time 01:33:00 gets resampled to 01:20:00 – 01:40:00). The flight plan data is already sampled to twenty minutes. On some days, the first flight of the day may be predicted on 01:33:00. When the dataset gets resampled to twenty-minute intervals, it will contain no intervals before 01:33:00 (i.e., 00:00:00 – 01:20:00). When that dataset is then used for analyzing, no accuracy can be calculated for the first few intervals, while the prediction of zero aircraft during that time may have been correct. To tackle this issue, the empty intervals are assigned the value ‘0’.

On September 22nd, the operations were heavily impacted due to weather conditions, and on April 5th, 24th, 26th, 27th and 28th, July 18th and November 1st and 2nd no prediction was made. These days are excluded from the analyses, resulting in a total of 219 days to analyze.

Determining the accuracy and precision per day will show whether the total number of EHFIRAM inbound flights predicted is equal to what is realized. Differences would mainly be caused by canceled flights or incorrect flight plans. The predictions are also analyzed per 60 minutes, 20 minutes rolling (see also subsection 3.2.1). The accuracy of the predictions per hour rolling would indicate whether movements took place too late or too early (i.e., delays or early arrivals). Of both the daily and hourly rolling data, the 28-day moving average is calculated to illustrate how the accuracy of the prediction changes over time. For the moving average, a 28-day period was chosen as it is a period long enough to smooth out weekly fluctuations, yet it still allows for showing fluctuations within the three periods.

The accuracy of the traffic demand prediction is expressed in two ways: as the difference and as a ratio. The two different representations allow for underlining different behaviors. The difference is calculated by subtracting the reference value (that being the last filed flight plans or realization) from the prediction, as can be seen in Equation 1. This equation can result in a positive or negative value or zero (i.e., no difference), where a positive result indicates that more traffic was predicted (i.e., overestimation) and negative indicates less was predicted (i.e., underestimation). The accuracy as a difference represents the magnitude of the accuracy.

Equation 1

$$accuracy\ = prediction - reference\ value$$

The ratio (hereafter also referred to as relative accuracy) is calculated by dividing the prediction by the reference value, as can be seen in Equation 2. The ratio expresses underestimation as a value smaller than one, overestimation as a value bigger than one and no difference as value one. Calculating the accuracy as a ratio is useful as it indicates how big the accuracy is in relation to the reference value.

Equation 2

$$accuracy\ = \frac{prediction}{reference\ value}$$

Expressing the accuracy as a ratio has a disadvantage when the prediction for a specified interval is bigger than zero, while the reference value is equal to zero. The denominator of Equation 2 becomes zero, resulting in an infinite value. Instances where the reference is zero can therefore not be used to determine the accuracy as a ratio. In Figure 23 the twenty-minute intervals where the realization is zero and the prediction is bigger than zero are presented in yellow to illustrate their position, which was four percent of all the observations. These instances are not used for the analyses. It can be seen that the realization is usually only zero during the night period where the traffic demand is typically low. Hence, instances where the prediction and the reference value are both zero are still considered equal and therefor result in a ratio of one.

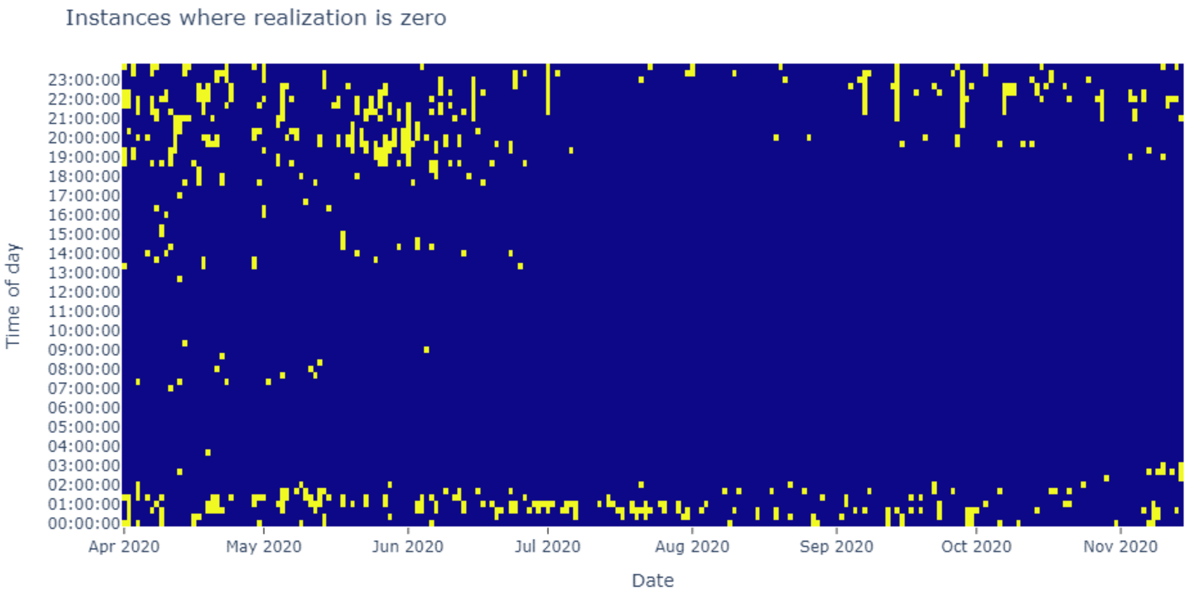


Figure 23 – Instances (yellow) where the realization is zero

The precision can be described by the standard error of the mean accuracy, and it assesses how far the sample mean is likely to fall from the population mean. The precision could be determined by dividing the standard deviation of the accuracy of the sample by the square root of the number of observations in the sample, as displayed in Equation 3.

Equation 3

$$precision = \frac{standard\ deviation}{\sqrt{number\ of\ observations}}$$

A statistical test is used to compare the means of two samples and determine if that difference is statistically significant. As the predicted traffic is paired by a time interval in the realization, a dependent sample test will be used. A T-test assumes the data to be normally distributed, or parametric. In Appendix IV the distributions of the datasets are presented, and none of the datasets show a normal distribution. However, the central limit theorem states that with a sufficiently large sample size (i.e., more than 30 observations), the sampling distribution starts to approximate a normal distribution (Frost, 2018). Because the smallest sample used for this analysis contains 62 observations, the data is assumed to be normally distributed, and a dependent samples T-test is used to determine if the traffic demand prediction is significantly different from the realization or the last filed flight plans.

5.1.2 Phase three and four – Runway use during the first peak hour

Runway use describes whether the landing direction is either north or south. As the datasets do not mention this specifically, the usage was justified based on the runway direction. Schiphol has six runways, all of which could be used in both directions. For example, runway 27 is the same physical runway as 09, but used in the opposite direction, as seen in Figure 24. If the first of the two runways in the prediction, forecast form, or realization is 04, 06, 09, 36L (left), 36C (center) or 36R (right), the landing direction is considered north. In any other case the landing direction is considered south.

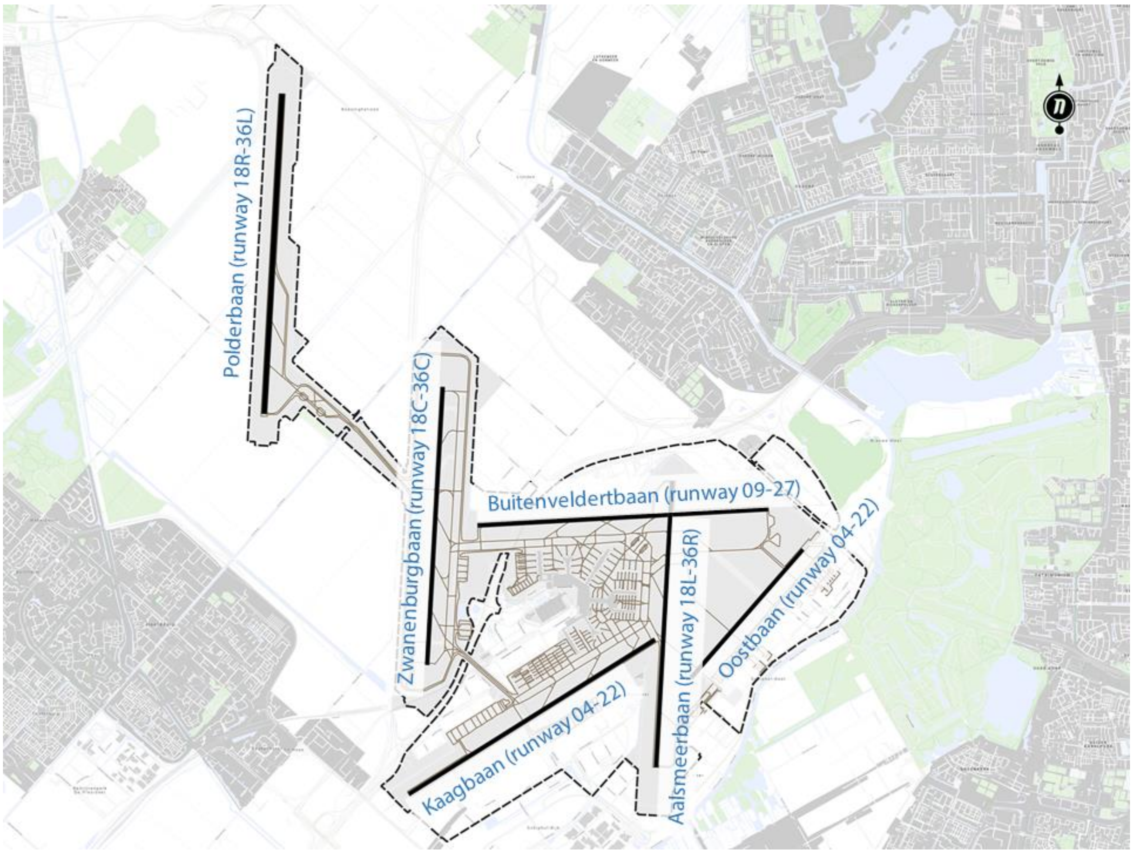


Figure 24 – Placement and direction of the six runways at Schiphol. Source: to70

As can be seen in Figure 24, the direction of most runways is clear, however, the direction of runway 27 (or 09) could be considered both north and south as it is oriented form east to west. Therefor the realization dataset was used to determine with what other runway it was used most often during the first inbound peaks. Runway 09 was not used for landing in this time range, however, runway 27 was used eighteen times, of which seventeen times in combination with runway 18R or 18C and once without any other runway. Runway 27 will therefore be considered southern, and 09 northern. In a few cases, a combination of runways was predicted or realized which could not be categorized north or south (e.g., 27+36C). These instances are considered outliers, and not used for the analysis.

The first inbound peak is the first moment of the day where two runways are used for landing. Due to the decreased traffic demand caused by COVID-19 restrictions, only one runway was predicted and used from April until July. The runway usage analysis is therefore concentrated on the period between July 1st and November 14th of 2020, which equals to 97 days.

The runway usage analysis counts the number of times the direction predicted was the same as that in the reference dataset. The two reference datasets available for the runway usage are the capacity forecast form (hereafter also referred to as 'forecast form') and realized datasets.

The runway usage is also tested to determine if there is a significant difference between the prediction and realization. As runway usage can only be north or south, a related samples McNemar test is used to determine if the proportions of runway use (i.e., north/south) in the two related datasets are significantly different from each other.

5.1.3 Phase five – Runway capacity during the first inbound peak

Capacity can be defined by the maximum number of aircraft that can utilize a runway in a given period of time (see also sub section 2.2) and is thus deeply intertwined with traffic demand. Therefore, the accuracy of the capacity predicted in the OPS plan is examined. The combined runway capacity of the primary prediction in the OPS plan is compared with the forecast form, as capacity data cannot be available for the realization. In case a difference is observed, the alternative prediction (if available) is used for the comparison. The accuracy of the runway capacity prediction is calculated with Equation 4, that subtracts the forecast form value from the prediction.

Equation 4

$$accuracy = prediction - forecast\ form$$

The precision of the prediction is calculated by dividing the standard deviation of the accuracy by the square root of the number of observations with Equation 5.

Equation 5

$$precision = \frac{standard\ deviation}{\sqrt{number\ of\ observations}}$$

No predictions were made for all Sundays and Mondays. Instances for which no data was available were not included in the analyses. Furthermore, in one case, the times for the first inbound peak in the OPS plan did not align with the times in the capacity forecast form and is considered an outlier and was thus excluded from analysis. On three days, a difference of more than 20 movements per hour was registered. As these differences are vastly larger than the other differences measured, these three days are considered extreme outliers, and therefore were also not used for this analysis.

The runway capacity predictions are related to the same time intervals in the capacity forecast form. Therefore, the same related samples test method is used to test for significance as discussed in sub-section 5.1.1.

5.2 Phase one – Traffic demand OPS plan vs realization

During the first phase, the accuracy and precision of the traffic demand predictions are compared to the realization on two levels of detail: per day and per hour rolling. Furthermore, a distinction is made between the data of the three separate periods. Figure 25 shows the inbound traffic demand for EHFIRAM predicted and realized for the analyzed period. The graph shows three periods with different demand behavior. During the first period, from April 1st until

June 30th, the traffic demand slightly increases from 75 movements to 150 movements per day. During period two, from July 1st until August 31st, the demand increases from around 200 to 400 movements per day. During period three, from September 1st until November 14th, the demand gradually decreases from 400 to 300 movements per day. As presented in Figure 22, the traffic demand on September 1st was approximately 50% of that of 2019.

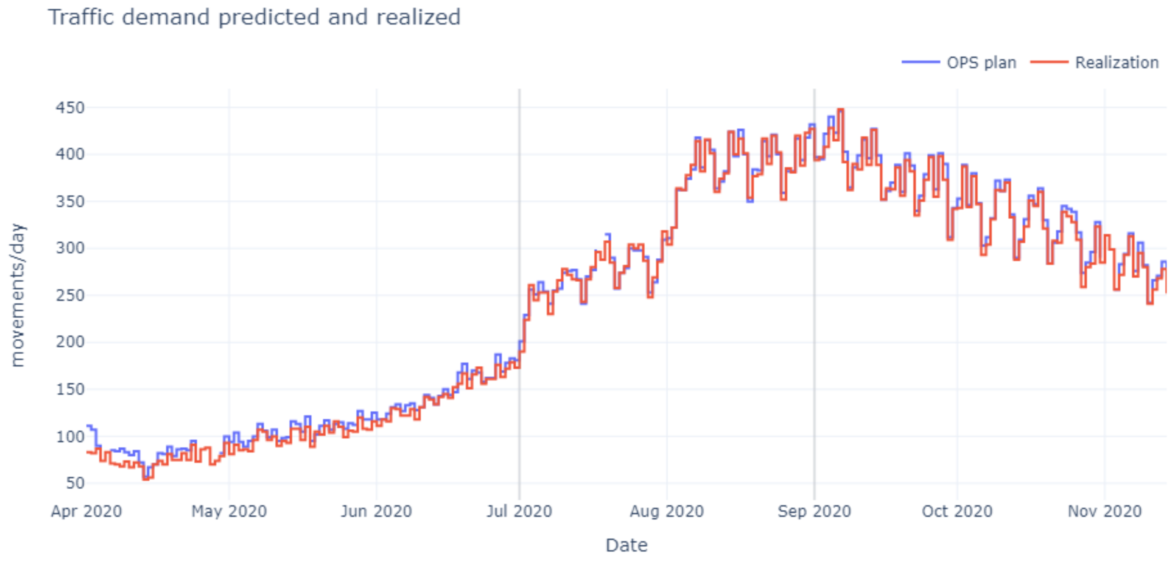


Figure 25 – Predicted and realized traffic per day at EHFIRAM

5.2.1 Accuracy and precision per day

Figure 26 displays the accuracy of the OPS plan predictions compared to the realization, following Equation 1. The mean accuracy of the prediction is 4.69, the standard deviation is 5.56, the mode is 3 and the accuracy ranges between -9 and 28 with a median of 4 movements per day (see also Table 11 in Appendix III). In Figure 26 it can be seen that on 213 of the 219 days (i.e., 97% of the time) the accuracy lays between -5 and 15 movements. 85% of the observations are bigger than zero, which means that 85% of the time the traffic demand was overestimated.

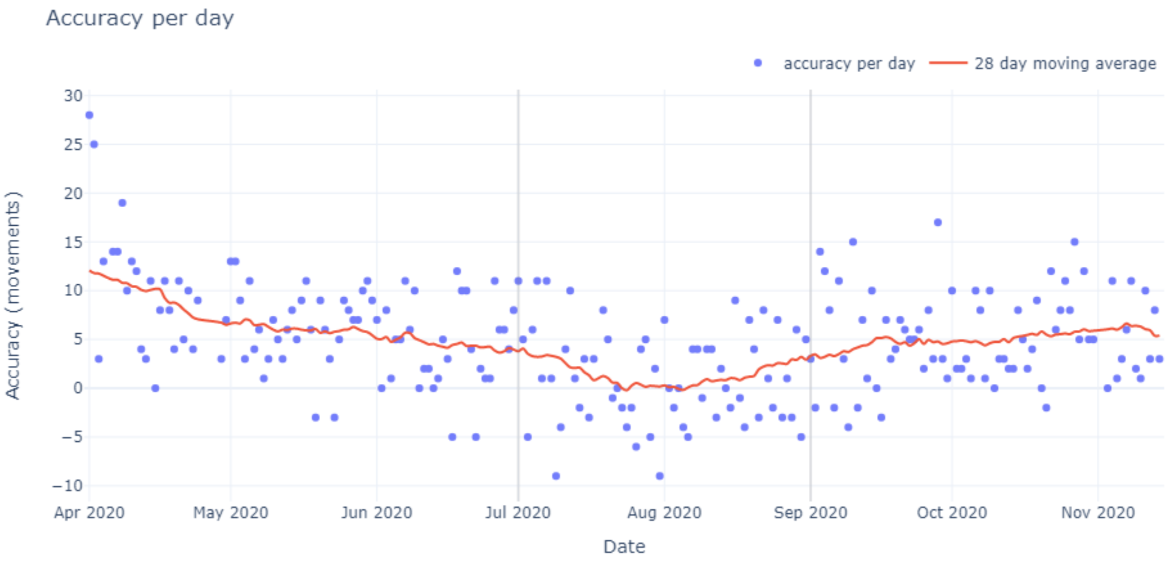


Figure 26 – Accuracy per day

The red line in Figure 26 represents the 28-day moving average of the accuracy per day. A 28-day (four week) moving average is used to smooth out short-term fluctuation and to emphasize long-term trends. During period one the moving average accuracy increases from 12 to 5 movements per day. During period two, the average accuracy improves further to zero movements per day. In period three the average accuracy worsens again to just over 5 movements per day.

When representing the accuracy of the OPS plan as a ratio of predicted to realized, see also Equation 2, the prediction could be considered accurate when it is equal to one, which means that the prediction and realization were equal. Figure 27 presents the accuracy per day. It can be noticed that the predictions per day relative to the realization get more accurate and more precise over time, as the observations lay closer to each other. For the whole period, 84% of the observations indicate overestimation, and on average, the accuracy is 1.04, which translates to 4%, the standard deviation is 0.06, the mode 1.00, and the accuracy ranges between 0.97 and 1.34 with a median of 1.02, as can be seen in Table 11 in Appendix III.

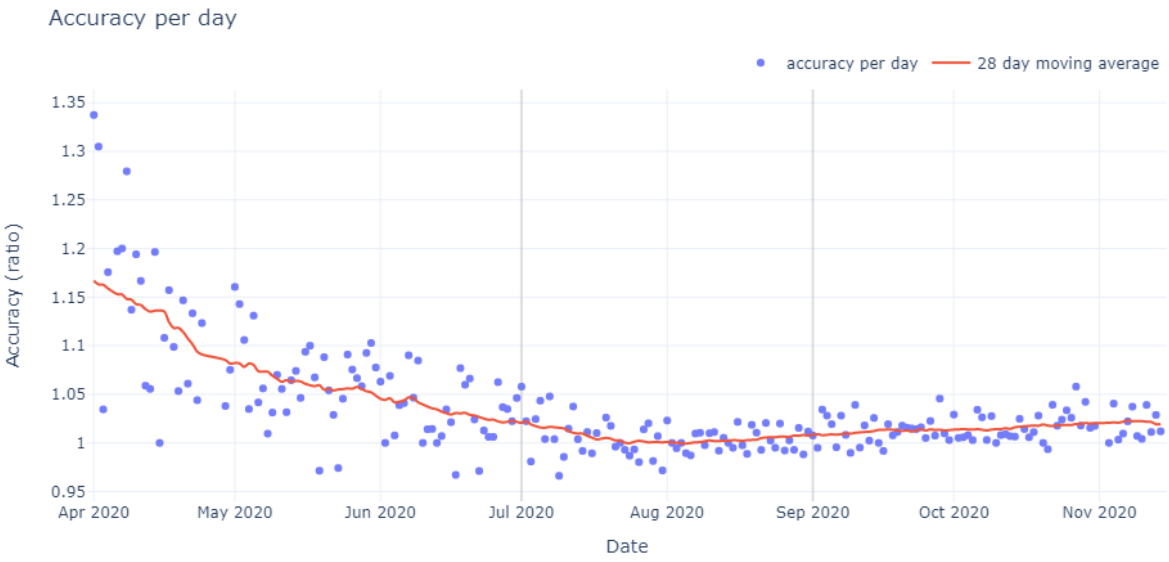


Figure 27 – Accuracy per day (ratio)

The 28-day moving average of the relative difference per day displayed as the red line in the figure shows that the accuracy in period one and two improves to around 1.00, and then worsens again to approximately 1.02 in November.

In Figure 25 it was shown that the traffic demand per day was vastly different for each of the three periods. In Figure 26 and Figure 27 it could be seen that the accuracy and precision behave differently during the three periods as well.

During period one, the mean accuracy of the prediction is 6.80, the standard deviation 5.48, the mode 11 and the accuracy ranges between -5 and 28 with a median of 6 movements per day (see also Table 12 in Appendix III). When assessing the accuracy as a ratio, the mean accuracy is 1.07, the standard deviation 0.07, the mode 1.00 and the accuracy ranges between 0.97 and 1.34 with a median of 1.06.

During period two, the mean accuracy of the prediction is 1.24, the standard deviation 4.93, the mode 4 and the accuracy ranges between -9 and 11 with a median of 6 movements per day (see also Table 13 in Appendix III). When assessing the accuracy as a ratio, the mean accuracy is 1.01, the standard deviation 0.02, the mode 1.00 and the accuracy ranges between 0.97 and 1.06 with a median of 1.01.

Figure During period three, the mean accuracy of the prediction is 5.17, the standard deviation 4.72, the mode 3 and the accuracy ranges between -4 and 17 with a median of 6 movements per day (see also Table 14 in Appendix III). When assessing the accuracy as a ratio, the mean accuracy is 1.02, the standard deviation 0.01, the mode 1.00 and the accuracy ranges between 0.99 and 1.06 with a median of 1.01.

Since the daily traffic demand differs greatly between the three periods, it would not be representative to test for significance for the entire period. To validate if the differences between the OPS plan prediction and the realization per period are significant, the following hypotheses were proposed.

$$Null\ hypothesis\ (H_0): \mu_1 = \mu_2$$

$$Alternative\ hypothesis\ (H_a): \mu_1 \neq \mu_2$$

Where:

μ_1 is the mean predicted traffic per day

μ_2 is the mean realized traffic per day

A two-sided related samples T-test conducted at a significance of $\alpha = 0.05$ compared the number of movements per day predicted in the OPS plan to the number of movements realized for each of the three periods.

Table 1 – Summary results of statistical tests for inbound EHFIRAM predictions vs realization per day

period analyzed	mean difference	standard deviation	standard error of mean	p-value	null hypothesis
Period one	6.80	5.48	0.59	< .001	Reject
Period two	1.24	4.93	0.63	.0517	Cannot reject
Period three	5.17	4.72	0.57	< .001	Reject

As can be seen in Table 1, the null hypothesis ($\mu_1 = \mu_2$) can be rejected for periods one and three and one can conclude that for those periods the traffic demand per day predicted was significantly higher than the traffic demand realized. For period two one can conclude that there is no significant difference between predicted and realized. It can be highlighted that the standard error of the mean accuracy is highest (i.e., precision is lowest) in period two, while not being drastically higher than period one and three.

5.2.2 Accuracy and precision per hour rolling

The analysis of accuracy of the predictions per hour rolling gives a more detailed comprehension of how accurate the predictions are during the day. Figure 28 depicts the accuracy of the OPS plan demand predicted per hour rolling. 95% of the 15,768 observations lay between -5 and 5 movements per hour rolling. In period two and three the observations are more spread compared to the period before, as more markers lay between ± 5 and ± 12 movements per hour rolling. This increase in spread indicates that the predictions became less precise.

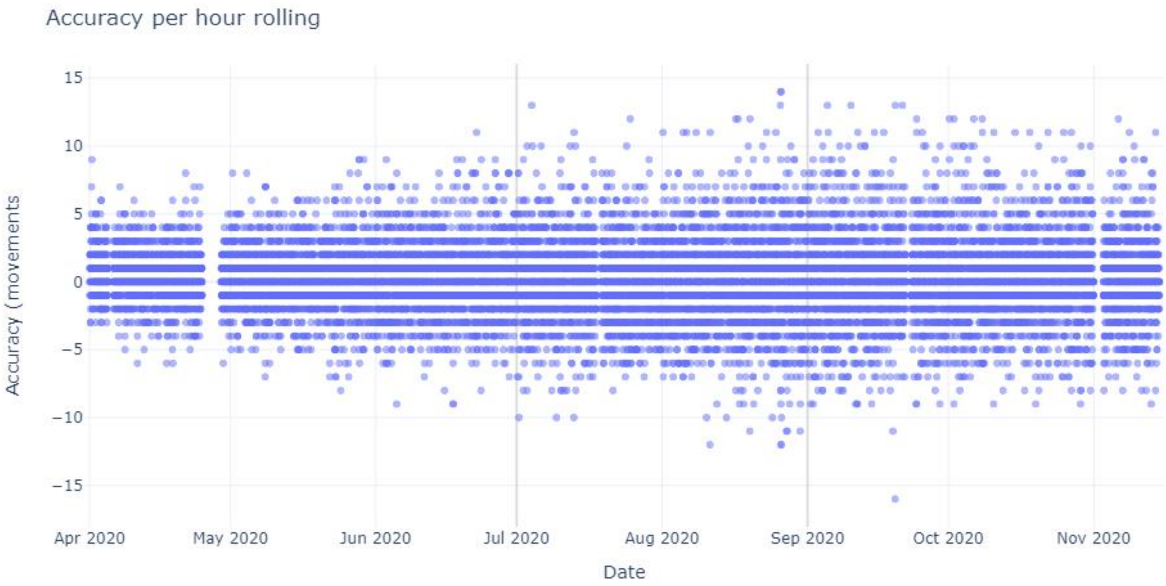


Figure 28 – Accuracy per hour rolling

The overall average accuracy is 0.20 movements per hour rolling, the standard deviation is 2.72, the mode is 0 and the accuracy ranges between -16 and 14, with a median accuracy of 0 movements per hour rolling (see also Table 15 in Appendix III). Overall, 42% of the observations are bigger than zero, thus overestimated, the predictions were equal to the realization on 22% of the observations, and 36% of the observations are smaller than zero, thus underestimated.

The 28-day moving average of the accuracy per hour rolling, displayed in Figure 29, emphasizes that the traffic demand is usually overestimated. The graph below also shows that the accuracy is close to 0 in period two, while during that period the traffic demand was highest of the period analyzed.

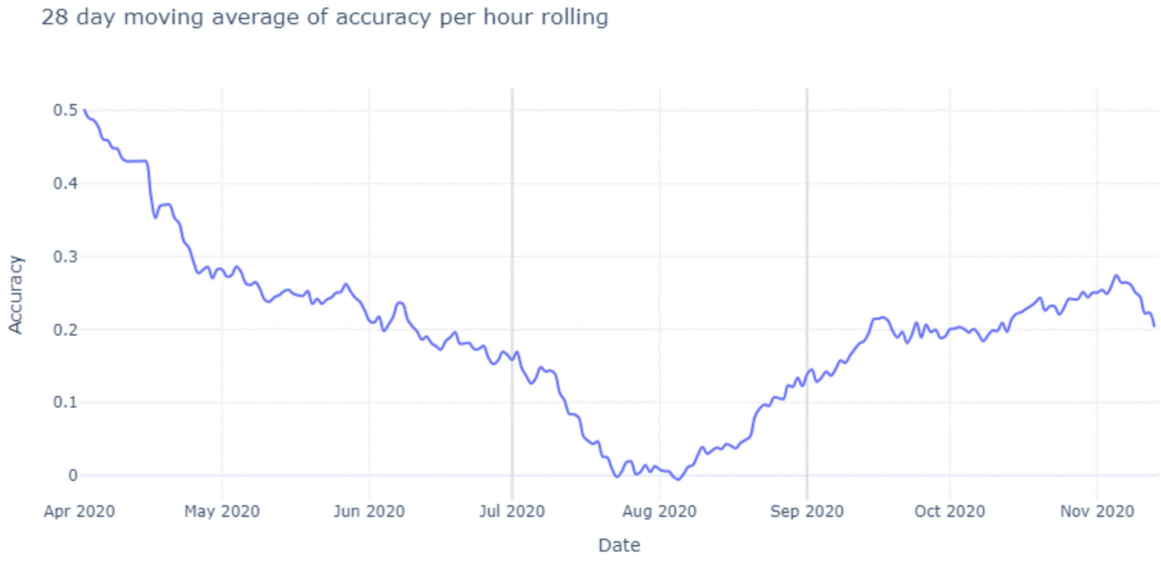


Figure 29 – 28-day moving average of the accuracy per hour rolling

Figure 30 depicts that the predictions are more accurate during the night, which could be caused by low traffic demand during that period. During the day, the observations are more often yellow or blue, compared to night period, indicating respectively over- and underestimation. In period two and three, between 04:00 and 05:00, observations are often underestimated, while between 05:00 and 07:00 there are more observations that are overestimated. This indicates that during this period of the day, traffic arrives earlier than predicted by NM.

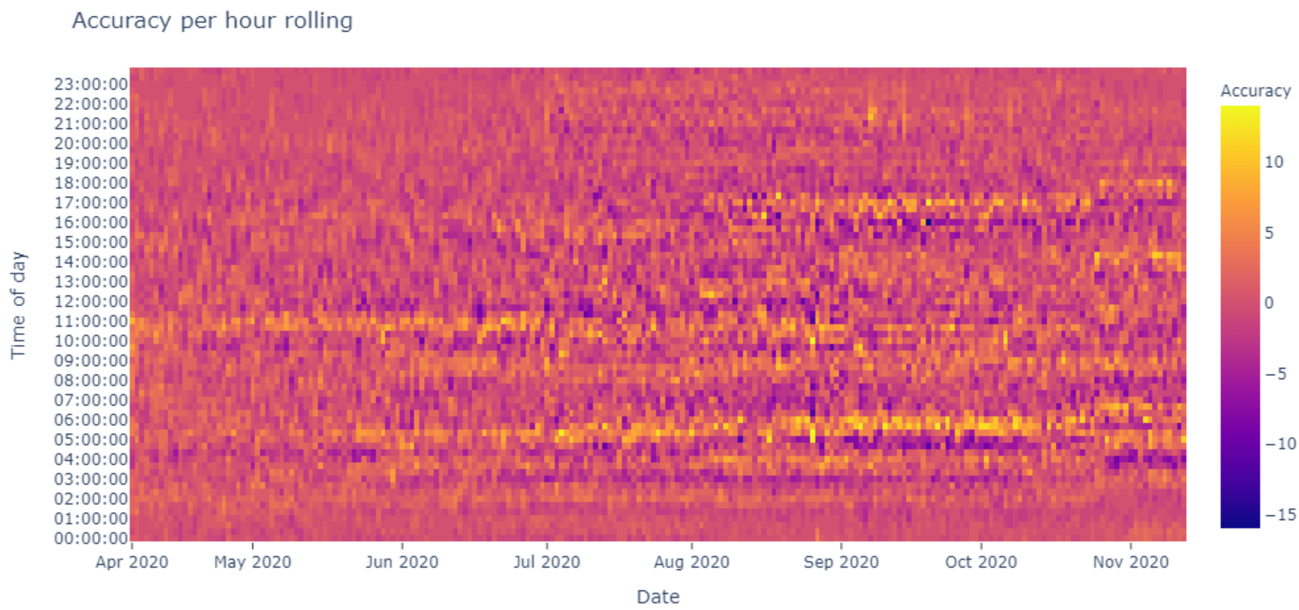


Figure 30 – Accuracy per hour rolling, per day

The accuracy per hour rolling relative to the realization, shown in Figure 31, demonstrates that the observations are more spread in period one, while in period two and three the accuracy is less spread (i.e., higher precision) and closer to one (i.e., higher accuracy). The average accuracy is 1.07, the standard deviation is 0.58, the mode is 1.00 and the accuracy ranges between 0.00 and 7.00, with a median of 1.00 (see also Table 15 in Appendix III).

Furthermore, in period one, there are more observations with a ratio between 2.00 and 7.00 compared to the other two periods, indicating bigger differences. Overall, 89% of the observations lay between zero and two. 37% of the observations are bigger than one, and thus overestimated, while 23% is exactly equal to one.

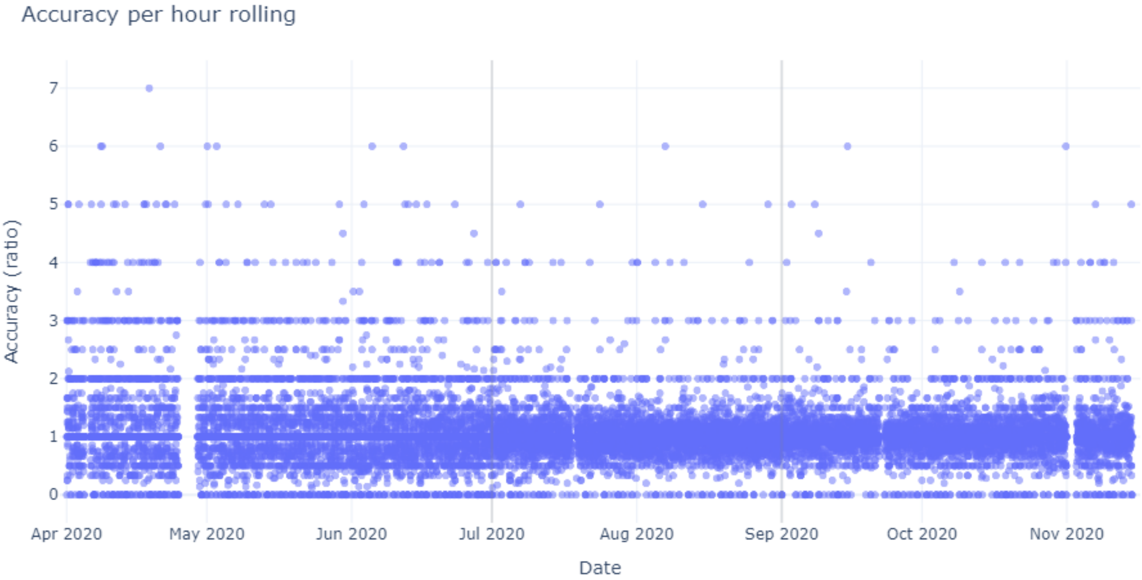


Figure 31 – Accuracy per hour rolling (ratio)

Figure 32 shows that the 28-day moving average of the accuracy per hour rolling is increasing during period one and two to a ratio of approximately 1.00 whereas it decreases to 1.05 again in period three.

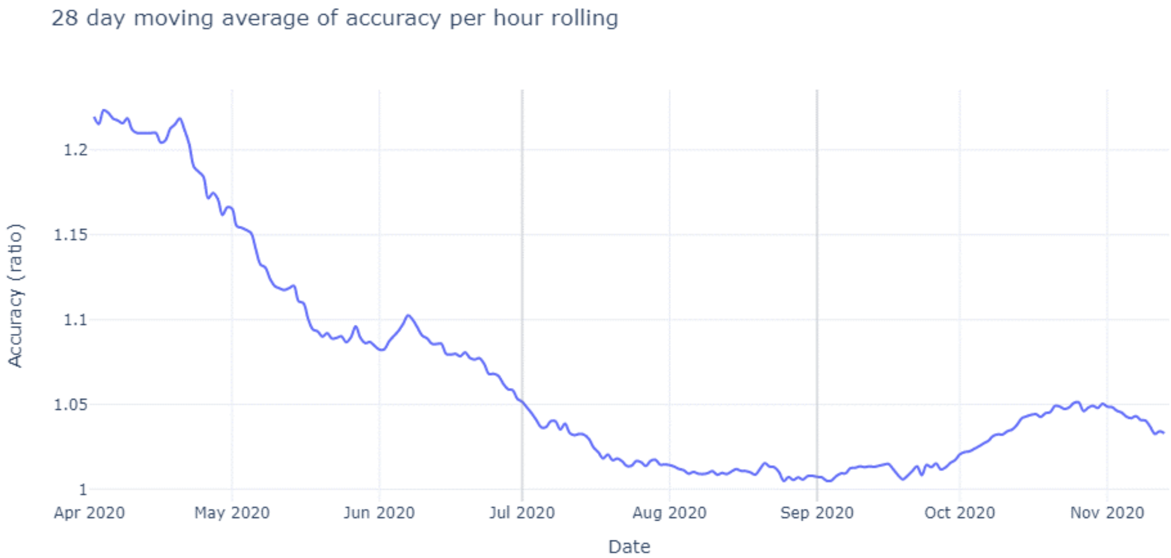


Figure 32 – 28-day moving average of the accuracy per hour rolling (ratio)

Figure 33 shows that in period one the predictions are less accurate and precise compared to the periods after. In periods two and three the predictions during the day seem more accurate than the ones during the night. Overall, during the peak moments with higher demand there is less difference in color, which indicates higher accuracy and precision. Around midnight a line of underestimation can be seen, followed by a line of overestimation from 01:00 to 02:00. In period three the accuracy of the evening period (18:00 – 00:00) shows decreased accuracy and precision.

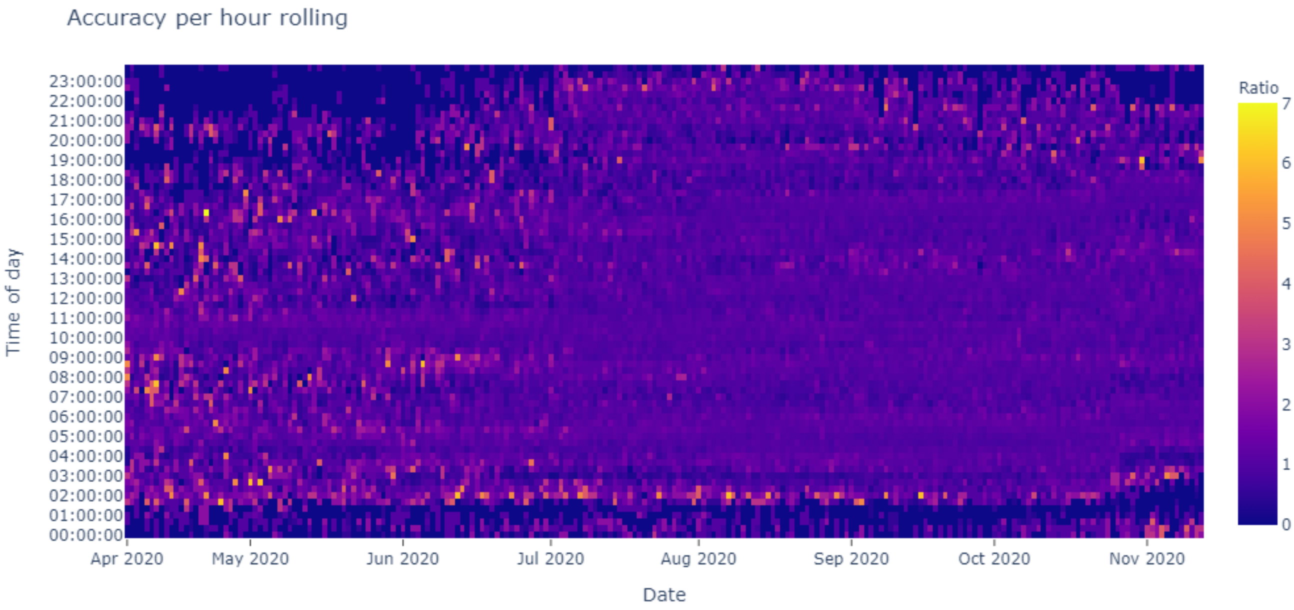


Figure 33 – Accuracy per hour rolling, per day (ratio)

Figure 33 shows low relative accuracy at nighttime which is usually not an issue as demand is low on those moments. During daytime, the relative accuracy seems better, as a small difference in movements does not impact the relative accuracy when there is a higher amount of traffic. However, one movement difference could have a much bigger impact on the operations during the day as the demand might already be close to the available capacity. The accuracy represented as a ratio does not reflect this possible impact.

In Figure 28 and Figure 29 it could be seen that the accuracy and precision do not behave in the same way throughout the three periods.

During period one, the mean accuracy of the prediction is 0.28, the standard deviation 2.11, the mode 0 and the accuracy ranges between -10 and 11 with a median of 0 movements per day (see also Table 16 in Appendix III). When assessing the accuracy as a ratio, the mean accuracy is 1.13, the standard deviation 0.73, the mode 1.00 and the accuracy ranges between 0.00 and 7.00 with a median of 0.00 per hour rolling.

During period two, the mean accuracy of the prediction is 0.05, the standard deviation 2.99, the mode 0 and the accuracy ranges between -12 and 14 with a median of 0 movements per day (see also Table 17 in Appendix III). When assessing the accuracy as a ratio, the mean accuracy is 1.03, the standard deviation 0.45, the mode 1.00 and the accuracy ranges between 0.00 and 6.00 with a median of 1.00 per hour rolling.

During period three, the mean accuracy of the prediction is 0.21, the standard deviation 3.11, the mode 0 and the accuracy ranges between -16 and 13 with a median of 0 movements per day (see also Table 18 in Appendix III). When assessing the accuracy as a ratio, the mean accuracy is 1.04, the standard deviation 0.46, the mode 1.00 and the accuracy ranges between 0.00 and 6.00 with a median of 1.00 per hour rolling.

To validate if the differences between the OPS plan prediction and the realization per period are significant, the following hypotheses were proposed.

$$Null\ hypothesis\ (H_0): \mu_1 = \mu_2$$

$$Alternative\ hypothesis\ (H_a): \mu_1 \neq \mu_2$$

Where:

μ_1 is the mean predicted traffic per hour rolling

μ_2 is the mean realized traffic per hour rolling

A two-sided related samples T-test conducted at a significance of $\alpha = 0.05$ compared the number of movements per hour rolling predicted in the OPS plan to the number of movements realized for the whole period, as well as for each of the three periods.

Table 2 – Summary results of statistical tests for inbound EHFIRAM predictions vs realization per hour rolling

period analyzed	mean difference	standard deviation	standard error of mean	p-value	null hypothesis
period one	0.28	2.11	0.03	< .001	Reject
period two	0.05	2.99	0.05	.2389	Cannot reject
period three	0.21	3.11	0.04	< .001	Reject

As can be seen in Table 2, the null hypothesis ($\mu_1 = \mu_2$) can be rejected for period one and three, and it can be concluded that the mean traffic demand predicted per hour rolling is significantly higher than the mean realization in those periods. For period two the null hypothesis cannot be rejected, and one can conclude that there is no significant difference between the prediction and realization. It can be highlighted that the precision of the predictions is highest in period one while the accuracy in that period is worst.

5.3 Phase two – Traffic demand OPS plan vs flight plans

In addition to calculating the accuracy of the predictions compared to the realization, the accuracy of the prediction is also determined by comparing it to the last filed flight plans. The last filed flight plan data is a dataset which is retrieved on the day of operations, thus after the OPS plan is created, and is therefore expected to be more accurate than the OPS plan predictions. The results and patterns of this analysis were found to be similar to those of the analysis in section 5.2 and are therefore presented with fewer graphs. In subsection 5.7.2 a comparison is made between the accuracy of the prediction regarding the realization and the predictions regarding the last filed flight plans.

5.3.1 Accuracy and precision per day

In Figure 34 the accuracy of the OPS plan predictions compared to the last filed flight plans is depicted. The mean accuracy of the prediction is 4.1, the standard deviation 5.5, the mode 3 and the accuracy ranges between -12 and 28 with a median of 3 movements per day (see also Table 19 in Appendix III).

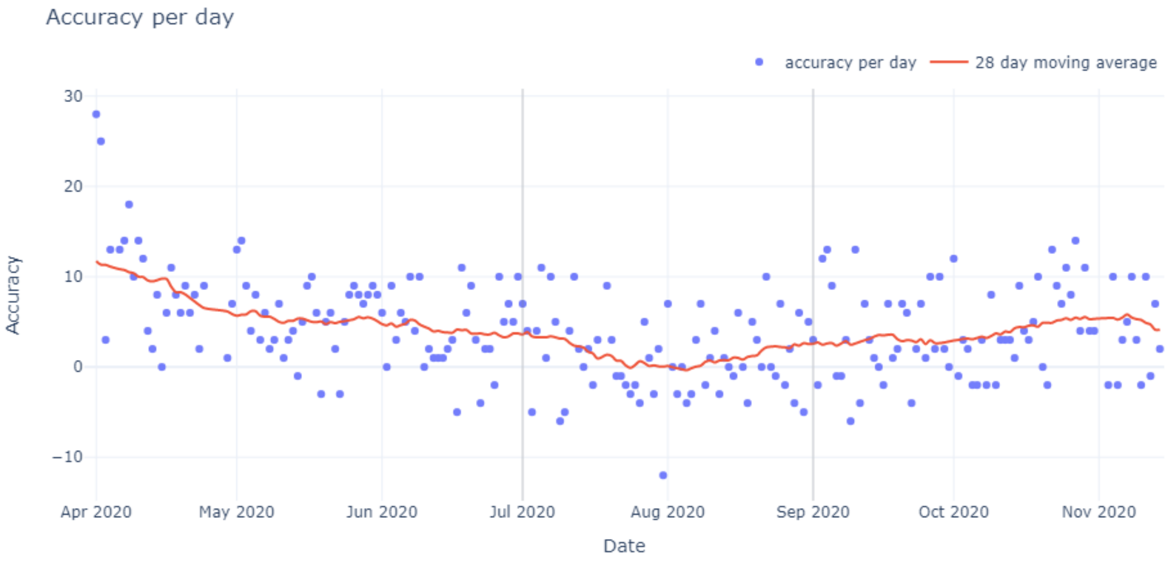


Figure 34 – Accuracy per day

Since the daily traffic demand differs greatly between the three periods, it would not be representative to test for significance for the entire period. To validate if the differences between the OPS plan prediction and the last filed flight plans per period are significant, the following hypotheses were proposed.

$$Null\ hypothesis\ (H_0): \mu_1 = \mu_2$$
$$Alternative\ hypothesis\ (H_a): \mu_1 \neq \mu_2$$

Where:

μ_1 is the mean predicted traffic per day

μ_2 is the mean traffic per day in the flight plans

A two-sided related samples T-test conducted at a significance of $\alpha = 0.05$ compared the number of movements per day predicted in the OPS plan to the number of flight plans filed for each of the three periods.

Table 3 – Summary results of statistical tests for inbound EHFIRAM predictions vs flight plans per day

period analyzed	mean difference	standard deviation	standard error of mean	p-value	null hypothesis
period one	6.22	5.45	0.58	< .001	Reject
period two	1.21	4.61	0.59	.0431	Reject
period three	4.04	4.89	0.57	< .001	Reject

As can be seen in Table 3, the null hypothesis ($\mu_1 = \mu_2$) can be rejected for each of the three period, and it can be concluded that the traffic demand per day predicted is significantly higher than the traffic demand in the last flight plans filed. It can be highlighted that the predictions for period one were least accurate and period three least precise compared to the predictions of period two and three.

5.3.2 Accuracy and precision per hour rolling

Figure 35 depicts the accuracy of the prediction per hour rolling. The average accuracy is 0.17 movements per hour rolling, the standard deviation is 2.32, the mode is 0 and the accuracy ranges between -13 and 14, with a median of 0 movements per hour rolling (see also Table 20 in Appendix III).

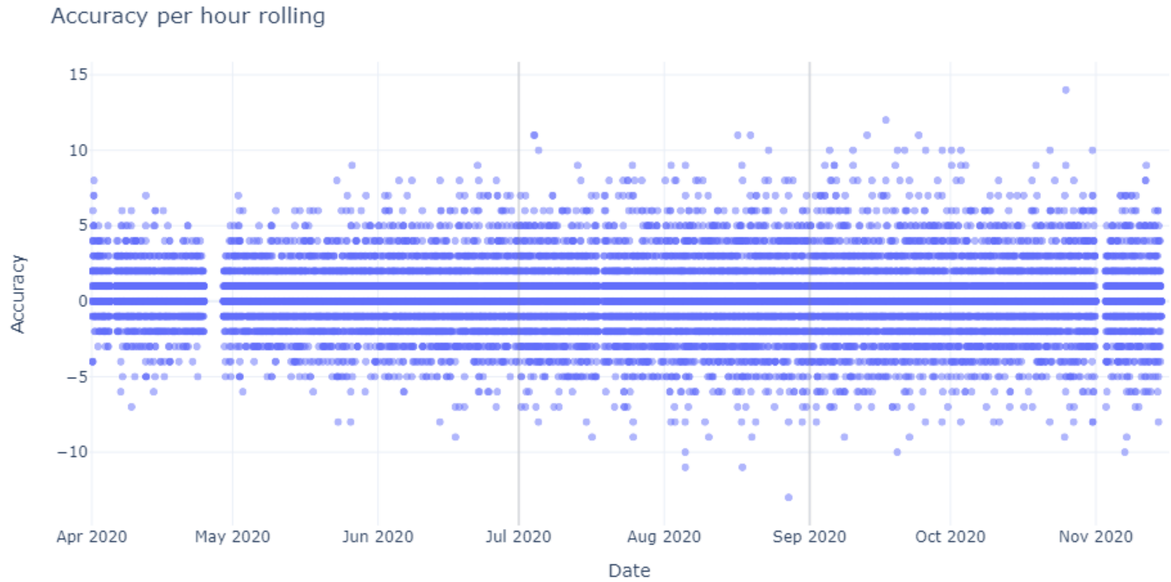


Figure 35 – Accuracy per hour rolling

To validate if the differences between the OPS plan prediction and the last filed flight plans per period are significant, the following hypotheses were proposed.

$$Null\ hypothesis\ (H_0): \mu_1 = \mu_2$$

$$Alternative\ hypothesis\ (H_a): \mu_1 \neq \mu_2$$

Where:

μ_1 is the mean predicted traffic per hour rolling

μ_2 is the mean traffic per hour rolling in the last filed flight plans

A two-sided related samples T-test conducted at a significance of $\alpha = 0.05$ compared the number of movements per hour rolling predicted in the OPS plan to the flight plans filed for each of the three periods.

Table 4 – Summary results of statistical tests for inbound EHFIRAM predictions vs flight plans per hour rolling

period analyzed	mean difference	standard deviation	standard error of mean	p-value	null hypothesis
period one	0.26	1.93	0.02	< .001	Reject
period two	0.05	2.54	0.04	.1742	Cannot reject
period three	0.17	2.52	0.04	< .001	Reject

As can be seen in Table 4, the null hypothesis ($\mu_1 = \mu_2$) cannot be rejected for period two, and one can conclude that there is no significant difference between the traffic demand predicted and the last filed flight plans. However, for period one and three the null hypothesis cannot be rejected, and one can conclude that the prediction was significantly higher than the last filed flight plans.

5.4 Phase three – Runway use OPS plan vs realization

Table 5 describes how many times northern or southern runway use was predicted and realized. The primary prediction was equal to the realization on 88 of the 97 occasions, whereas it was not equal on nine days.

Table 5 – Runway use during first inbound peak predicted (primary) vs realized

Runway usage OPS plan	Runway usage realized	Frequency of observations
North	North	20
South	South	68
		88
North	South	3
South	North	6
		9

Of the nine days where the primary prediction was not equal to the realization, no alternative prediction was provided on four days. For the five days where the primary prediction was not equal to the realization, the alternative prediction was compared to the realization, as presented in Table 6 . As can be seen, when considering the alternative prediction as well, the OPS plan prediction was equal to the realization on 91 of the 97 days. Together with the four days on which no alternative prediction was provided, the runway capacity prediction (primary and alternative) was not equal to the realization on six occasions.

Table 6 – Runway use during first inbound peak predicted (primary and alternative) vs realized

Runway usage OPS plan	Runway usage realized	Frequency of observations
North	North	22
South	South	69
		91
North	South	2
South	North	4
		6

To determine if the primary runway use predictions are significantly different from the realization, the following hypotheses were proposed.

$$Null\ hypothesis\ (H_0): p_1 = p_2$$

Which translates to:

There is no difference in runway use (proportion) between the prediction and realization

$$Alternative\ hypothesis\ (H_a): p_1 \neq p_2$$

Which translates to:

There is a difference in runway use (proportion) between the prediction and realization

A related samples McNemar test conducted at a significance of $\alpha = 0.05$ compared the primary and alternative runway usage predictions to the realized runway use.

Table 7 – Summary results of statistical tests runway usage predictions vs realization

prediction analyzed	p-value	null hypothesis
primary	.5078	Cannot reject
primary and alternative	.6875	Cannot reject

As can be seen in Table 7, the null hypothesis ($p_1 = p_2$) cannot be rejected when comparing either the primary or primary and alternative prediction with the realization, and one can conclude that there is no significant difference between the runway use prediction and the realization. In Table 23 in Appendix V an overview is provided for the six days where the primary and alternative prediction in the OPS plan were not equal to the realization. Runway use is mainly dependent on wind direction, and the comments in the table indicate that changing wind direction is often mentioned in the OPS plan and forecast form.

5.5 Phase four – Runway use OPS plan vs forecast form

As a reference, the predicted runway direction was also compared to the capacity forecast form, as this dataset is assumed to be more accurate since it was made available closer to the realization. In Table 8 the frequency of northern or southern runway use is presented for the prediction in the OPS plan and the forecast form.

Table 8 – Runway use during first inbound peak predicted (primary) vs forecast form

Runway usage OPS plan	Runway usage forecast form	Frequency of observations
North	North	19
South	South	67
		86
North	South	4
South	North	7
		11

The data shows that the runway usage predicted was equal to the forecast form on 86 of the 97 occasions, whereas the predictions were equal to the realization on 88 occasions. Of the 11 days where the OPS plan prediction was not equal to the forecast form, no alternative was provided on four days. For the seven remaining days where the primary prediction was not equal to the forecast form, the alternative prediction was compared to the forecast form instead, as presented in Table 9. As can be seen, when considering the alternative prediction as well, the OPS plan prediction was equal to the forecast form on 91 of the 97 days. Including the four days on which no alternative prediction was provided, the runway capacity prediction (primary and alternative) was not equal to the forecast form on six occasions.

Table 9 – Runway use during first inbound peak predicted (primary and alternative) vs forecast form

Runway usage OPS plan	Runway usage forecast form	Frequency of observations
North	North	22
South	South	69
		91
North	South	2
South	North	4
		6

In order to determine if the primary runway use predictions are significantly different from the forecast form, the following hypotheses were proposed.

$$Null\ hypothesis\ (H_0): p_1 = p_2$$

Which translates to:

There is no difference in runway use (proportion) between the prediction and forecast form

$$Alternative\ hypothesis\ (H_a): p_1 \neq p_2$$

Which translates to:

There is a difference in runway use (proportion) between the prediction and forecast form

A related samples McNemar test conducted at a significance of $\alpha = 0.05$ compared the primary and alternative runway usage predictions to the runway use in the forecast form.

Table 10 – Summary results of statistical tests runway usage predictions vs forecast form

prediction analyzed	p-value	null hypothesis
primary	.5488	Cannot reject
primary and alternative	.6875	Cannot reject

As can be seen in Table 10, the null hypothesis ($p_1 = p_2$) cannot be rejected when comparing either the primary or primary and alternative prediction with the forecast form, and one can conclude that there is no significant difference between the runway use prediction and the forecast form. In Table 24 in Appendix V an overview is provided for the six days where the primary and alternative predictions were not equal to the forecast form.

5.6 Phase five – Runway capacity OPS plan vs forecast form

For the runway capacity analysis, no data was available for realization. Therefore, the runway capacity predictions are only compared to the capacity forecast forms. The OPS plan can contain two predictions, the primary and an alternative prediction. If the primary prediction turned out to not be correct, the alternative (if available) was used for the analysis. The runway usage analysis is conducted to see if the landing direction (north or south) was correctly predicted with regards to the realization and forecast form. As the last forecast form is usually drafted around midnight each day, just a few hours before the first inbound peak (see also chapter 4), the forecast form is considered a reliable representation of the realization. This analysis aims to examine the available runway capacity during the first inbound peak. Therefore, the start and end times of every first inbound peak in the predictions are compared to the same timeframes in the forecast forms.

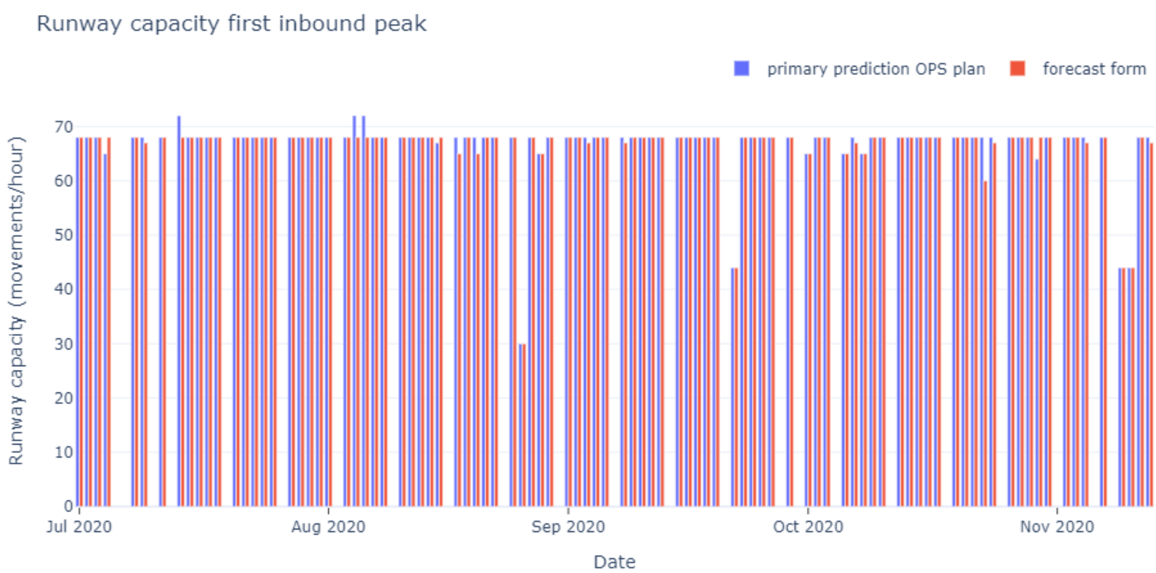


Figure 36 – Runway capacity predicted in the OPS plan (primary) and the forecast form

Figure 36 shows that three peaks have a capacity of 72 movements per hour, whereas the other days have a maximum capacity of 68. On 26th September one runway was predicted for the first inbound peak in both the OPS plan and forecast form. On the other days two runways were predicted.

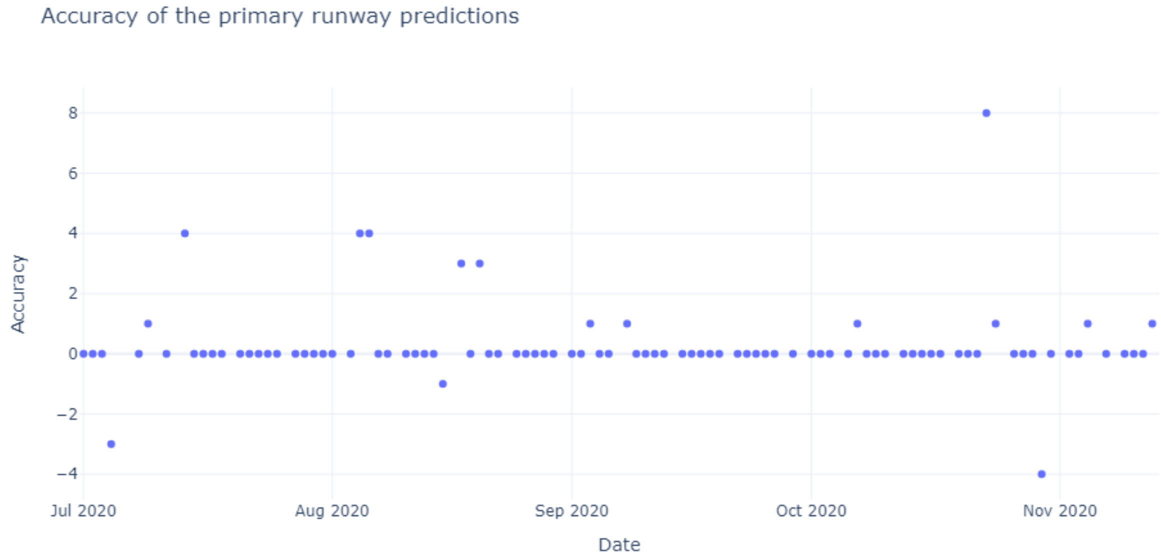


Figure 37 – Accuracy of the primary predictions

The graph in Figure 37 displays the accuracy of the primary prediction in the OPS plan compared to the forecast form. On 78 of the 94 instances the prediction was equal to the forecast form. On 16 occasions the prediction was not equal to the forecast form. Overall, the mean accuracy is 0.27 movements per hour, the standard deviation is 1.30, the mode is 0 and the accuracy ranges between -4 and 8 with a median accuracy of 0 movements per hour (see Table 21 in Appendix III).

Of the 16 instances where the prediction was not equal to the forecast form, no alternative prediction was provided for five days. For the 11 instances where an alternative was provided, the alternative prediction from the OPS plan was compared to the forecast form and are displayed in Figure 38.

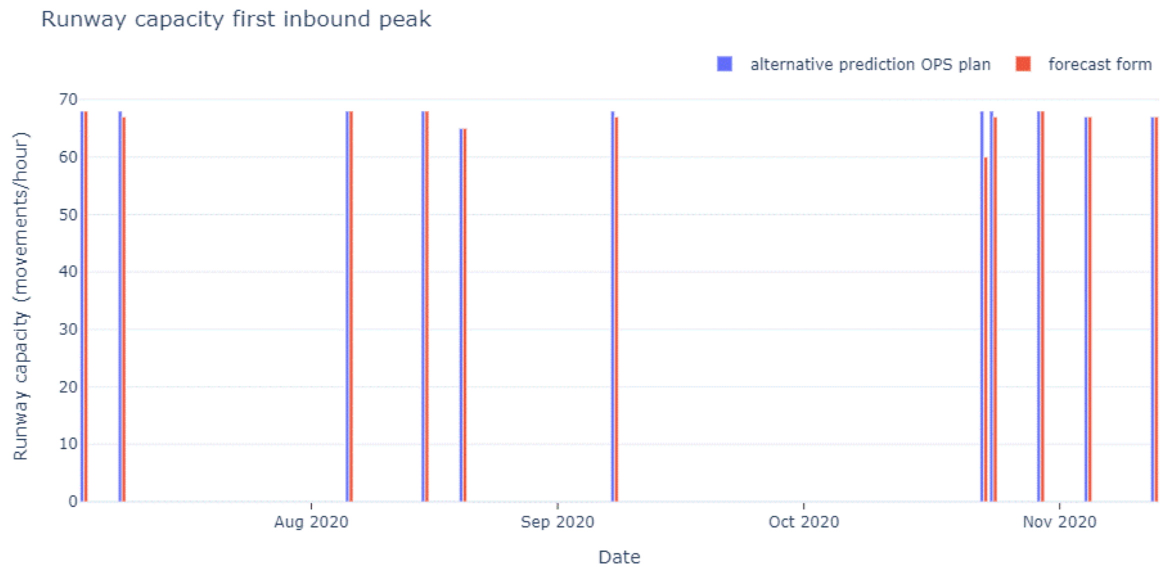


Figure 38 – Runway capacity predicted in OPS plan (alternative) and forecast form

Figure 39 displays the accuracy of the alternative predictions where the primary prediction was not equal to the forecast form. The alternative was not equal to the forecast form on four instances. Together with the five days where no alternative prediction was made, the OPS plan prediction (primary or alternative) was not equal to the forecast form on nine of the 94 days.

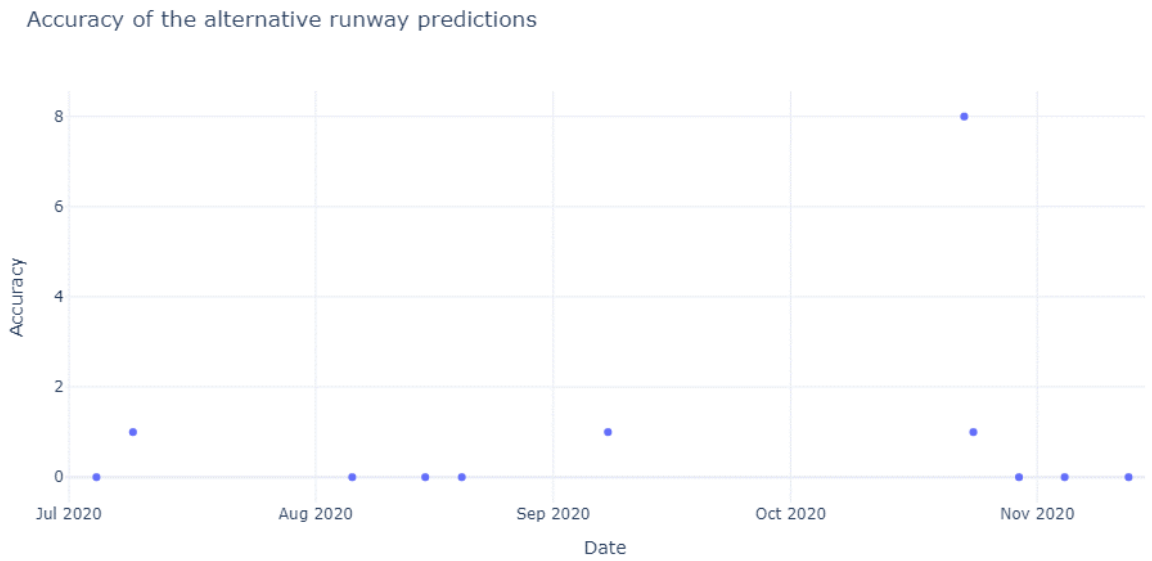


Figure 39 – Accuracy of the alternative predictions

For the instances where an alternative prediction was provided, the prediction and the forecast forms are either equal, or the difference is smaller than the primary prediction. Overall, the mean accuracy of the primary and alternative prediction is 0.26 movements per hour, with a standard deviation of 1.01, the mode is 0 and the accuracy ranges between 0 and 8 with a median of 0 movements per hour (see Table 22 in Appendix III).

To validate if the accuracy of the runway capacity prediction is significant, the following hypotheses were proposed.

$$Null\ hypothesis\ (H_0): \mu_1 = \mu_2$$

$$Alternative\ hypothesis\ (H_a): \mu_1 \neq \mu_2$$

Where:

μ_1 is the mean capacity per hour predicted

μ_2 is the mean capacity per hour in the forecast form

A two-sided related samples T-test conducted at a significance of $\alpha = 0.05$ compared the runway capacity predicted to the runway capacity in the capacity forecast form.

Table 11 – Summary results of statistical tests runway capacity predictions vs realization

prediction analyzed	mean difference	standard deviation	standard error of mean	p-value	null hypothesis
primary	0.27	1.30	0.13	.0497	Reject
primary and alternative	0.25	1.06	0.11	.1913	Cannot reject

As can be seen in Table 11, when only the primary prediction is considered, the null hypothesis can be rejected, and it can be concluded that the mean runway capacity predicted (66.74) is significantly higher than the mean runway capacity in the forecast form (66.48). However, the null hypothesis ($\mu_1 = \mu_2$) cannot be rejected when considering both the primary and alternative prediction provided in the OPS plan, and it can therefore be concluded that there is no significant difference between both. Additionally, when accounting for the primary and alternative prediction, the prediction is more precise compared to when only accounting for the primary prediction.

Table 25 in Appendix V provides an overview of the days where the primary and alternative capacity predictions were not equal to the forecast form, or no alternative prediction was available. The table includes the capacity form the OPS plan, and the forecast form. In addition, possible causes for the differences are discussed. Based on the information in the table, the reasons for a difference between the prediction and the forecast form could be categorized into two groups: human error and changing weather conditions. Three instances were categorized human error, whereas five instances categorized as changing weather conditions.

5.7 Subsequent analyses

5.7.1 Traffic demand – Accuracy and precision per day of the week

Figure 40 depicts the accuracy and precision per day of the week for the whole period analysed. Wednesdays are most accurate with a mean accuracy of 3.88 movements per day. Thursdays are least accurate with a mean accuracy of 5.48 movements per day. It could also be seen that Wednesdays are least precise with a standard error of 1.25, whereas Saturdays are most precise with a standard error of 0.77 movements per day (see also Table 12).

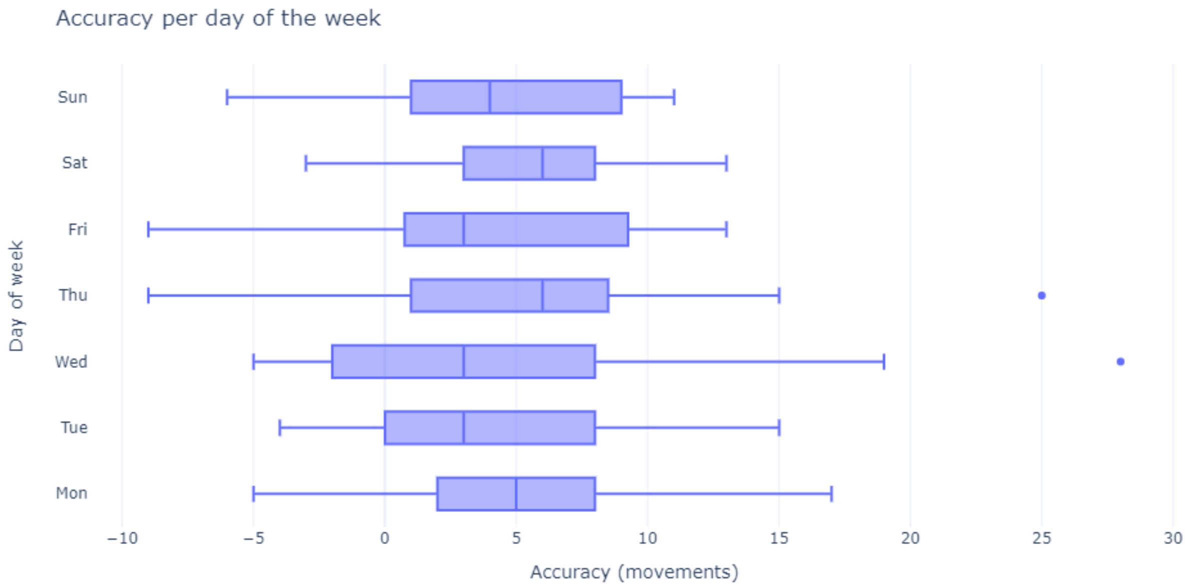


Figure 40 – Accuracy per day of the week

Table 12 – Descriptive statistics per day of the week

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
count	30	31	33	33	33	31	29
mean	5.00	4.32	3.88	5.48	3.91	5.52	4.79
mode	5	0, 3	-3, 0, 1	1	3	8	11
std	4.83	5.02	7.16	6.46	5.69	4.26	4.87
sem	0.88	0.90	1.25	1.12	0.99	0.77	0.90
min	-5	-4	-5	-9	-9	-3	-6
25%	2.25	0.00	2.00	1.00	1.00	3.00	1.00
median	5.00	3.00	3.00	6.00	3.00	6.00	4.00
75%	7.75	8.00	8.00	8.00	9.00	8.00	9.00
max	17	15	28	25	13	13	11

When the accuracy is presented as a ratio of the prediction to the realization as in Figure 41, Mondays, Tuesdays, Fridays, and Sunday are most accurate, with a mean accuracy of 1.03.

Wednesdays, Thursdays, and Saturdays are slightly less accurate with a mean accuracy of 1.04 (see also Table 13).

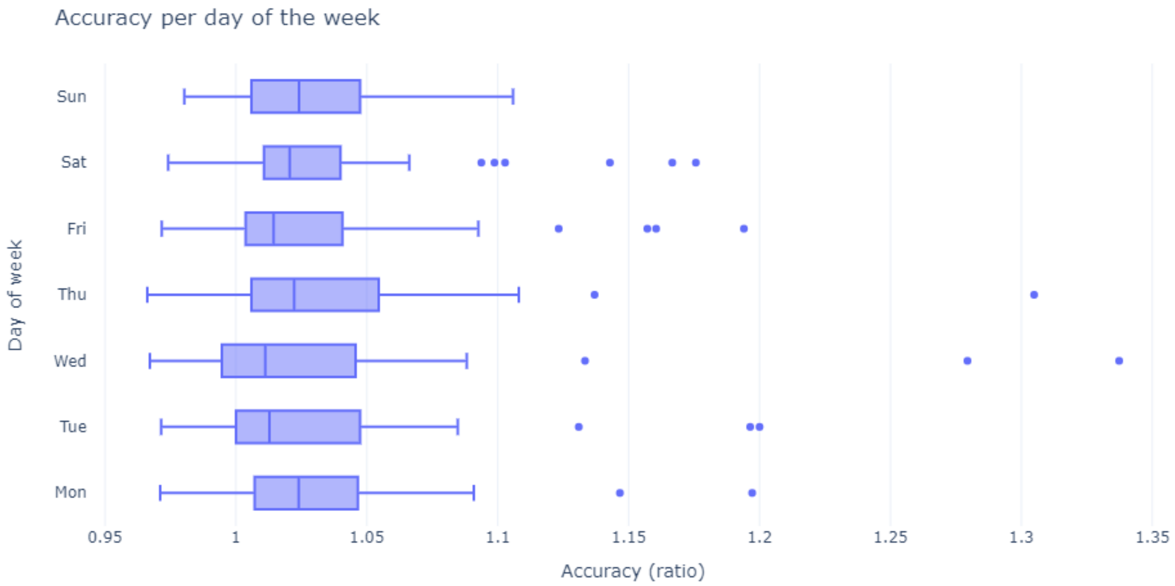


Figure 41 – Accuracy per day of the week (ratio)

Table 13 – Descriptive statistics per day of the week (ratio)

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
count	30	31	33	33	33	31	29
mean	1.03	1.03	1.04	1.04	1.03	1.04	1.03
mode	1.03	1.01	0.99	1.01	1.01	1.02	1.01
std	0.05	0.05	0.08	0.06	0.05	0.05	0.03
min	0.97	0.97	0.97	0.97	0.97	0.97	0.98
25%	1.01	1.00	0.99	1.01	1.01	1.01	1.01
median	1.02	1.01	1.01	1.02	1.01	1.02	1.02
75%	1.05	1.05	1.04	1.05	1.04	1.04	1.05
max	1.20	1.20	1.34	1.30	1.19	1.18	1.11

5.7.2 Traffic demand – Comparing OPS plan vs realization with OPS plan vs flight plans

As the accuracy of the OPS plan prediction found in section 5.3 seemed similar to the accuracy found in section 5.2, in this subsection both results are compared. In this subsequent analysis, only the 28-day moving average accuracies per day will be analyzed.

When comparing the two accuracies as line graphs, as shown in Figure 42, instances where the lines are equal, inaccuracy in the predictions is typically caused by factors that occur between when the predictions are made in the OPS plan and when the flight plans are filed (see also Figure 20). A difference between the two lines indicates inaccuracies are usually caused by factors that occur after the last flight plan is filed.

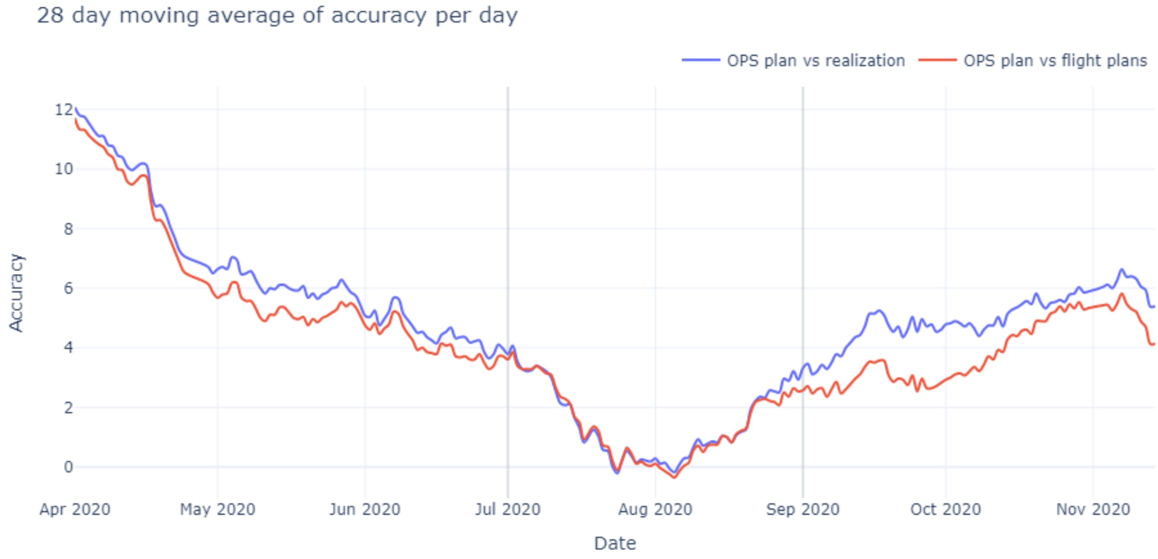


Figure 42 – 28 day moving average of accuracy per day

The mean accuracy of OPS plan vs realized is 4.69 movements per day, the standard deviation is 5.55, the mode is 3 and the accuracy ranges between -9 and 28 with a median of 4 movements per day. The mean accuracy of OPS plan vs flight plans is 4.06, the standard deviation is 5.45, the mode is 3 and the accuracy ranges between -12 and 28 with a median of 3 movements per day.

Table 14 – Descriptive statistics per day

	OPS plan	realized	accuracy	OPS plan	flight plans	accuracy
count	219	219	219	219	219	219
mean	251.35	246.66	4.69	251.35	247.30	4.06
mode			3			3
std	119.34	121.08	5.55	119.34	121.13	5.45
sem			0.38			0.37
min	57	54	-9	57	55	-12
25%	120.25	116.00	1.00	120.25	116.75	0.00
median	276.00	272.00	4.00	276.00	272.50	3.00
75%	361.25	360.00	8.00	361.25	359.25	8.00
max	446	448	28	446	447	28

5.7.3 Traffic demand – Day vs night period

Figure 31 shows a high frequency of observations with a ratio of either two or zero. This could be due to low traffic demand during the night, where for example one movement was predicted and two realized (that being a ratio of two). The instances where the ratio is zero or two are displayed in yellow in Figure 43, and most of the observations lay in periods where traffic demand has been low (i.e., the night period). This example could be used for future research where the accuracy and precision of only high demand periods (i.e., daytime) is analyzed.

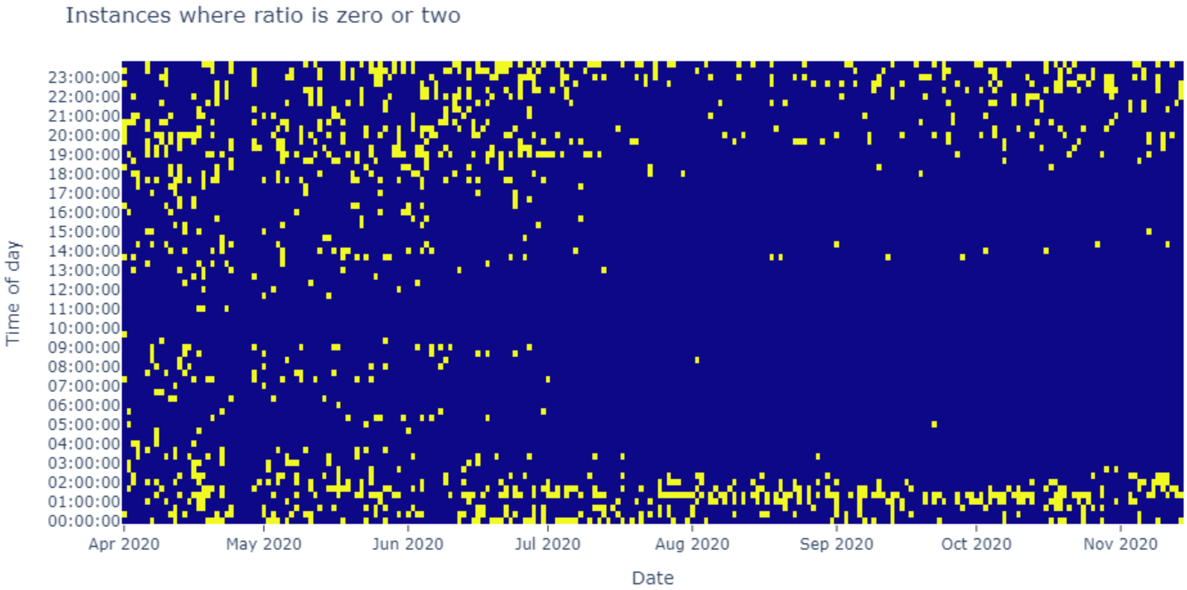


Figure 43 – Instances where the accuracy per hour rolling is equal to zero or two

5.7.4 Inbound peak times

Figure 44 displays the start and end times for the first inbound peak as predicted in the OPS plan, and the realized peak times (i.e., the start and end time of the period where two runways were in use). Note that the timescale of the Y axis is not continuous. In the month of July, the first inbound peak occasionally took place from later than predicted. This short analysis was done as an indication and could be used for future research, as elaborated in Chapter 6.

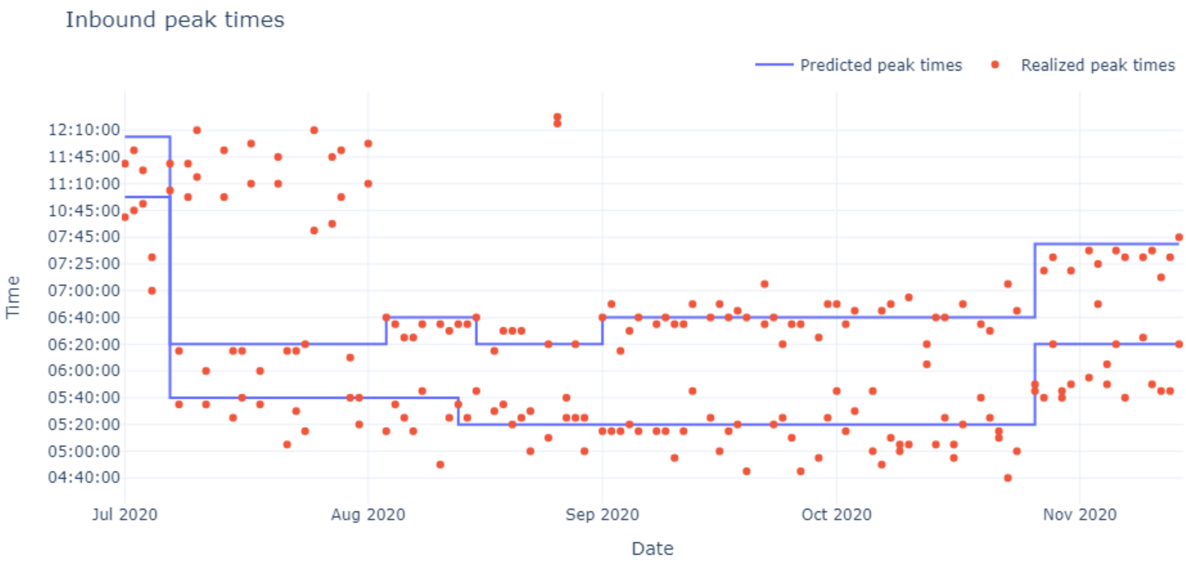


Figure 44 – Difference between inbound peak times predicted in OPS plan and realization

6 Discussion

This study researched the predictability of the inbound traffic demand for EHFIRAM and the runway combinations during the first inbound peak per day, predicted in the OPS plan. The OPS plan predictions were compared to two moments in time. The traffic demand prediction is compared to the realization and the flight plan data. The runway combination predictions are compared to the realization and the latests capacity forecast form. The flight plan data and the latest capacity forecast form are both issued on the day of operations and are therefore expected to be more accurate than the predictions. The key finding from this study can be summarized as:

- When compared to the realization, the traffic demand predictions (per day and per hour rolling) are significantly higher in period one and three, and not significantly different in period two.
- When compared to the last filed flight plans, the traffic demand predictions (per day and per hour rolling) are significantly higher in period one and three, and not significantly different in period two.
- There is a slight difference between OPS plan vs realization and OPS plan vs flight plans, yet the data shows similar patterns.
- Runway usage prediction in OPS plan is not equal to realization on 6 of the 97 days. Inequality mainly caused by changing wind direction.
- Runway usage prediction in OPS plan is not equal to forecast form on 6 of the 97 days, however, these are not the six exact days as mentioned above. Inequality mainly caused by changing wind direction.
- Runway capacity prediction is not equal to forecast form on 10% of the days. Inequality mainly caused by human error (five of ten instances) and changing weather conditions (four of ten instances).

The average accuracy of the inbound traffic demand predictions for EHFIRAM during period two compared to the realization is 1.24 movements per day, with a precision of 0.63 movements per day. This means that on 95% of the time the average accuracy of the traffic demand prediction in the OPS plan fluctuates between -0.02 and 2.50 movements per day. A statistical test found with an interval of confidence of 95% that for period two there is no significant difference between the mean number of movements predicted and the mean number of movements realized, whereas for periods one and three the mean number of movements predicted was significantly higher than the realization. During period one the average accuracy is 6.80 with a precision of 0.59 movements per day, and during period three the average accuracy is 5.17 with a precision of 0.57 movements per day. It can be highlighted that even though the accuracy was highest in period two, the precision was lowest.

When looking at the traffic demand per hour rolling, during period two the average accuracy is 0.05 movements with a precision of 0.05 movements per hour rolling. A statistical test found with an interval of confidence of 95% that for period two there is no significant difference between the mean number of movements predicted and the mean number of movements realized, whereas for periods one and three the mean number of movements predicted was significantly higher than the realization. During period one the average accuracy is 0.28 with a precision of 0.03 movements per hour, and during period three the average accuracy is 0.21 with a precision of 0.04 movements per hour. It can be highlighted that even though the accuracy was highest in period two, the precision was lowest.

Since the runway's capacity is a fixed resource, it is safer to overestimate traffic demand, rather than underestimate. However, this safety net should be as small as possible as overestimation could be extremely costly for stakeholders, as explained in section 2.5. Moreover, one of the objectives of making an operational planning is to align traffic demand and the available capacity to the most optimal extent possible. Thus, predictions should focus on reflecting the realization (i.e., making the most accurate predictions possible). Then, the person(s) responsible for the tactical capacity management (that being on the day of operations) should use the predictions to anticipate the expected demand and decide what capacity margin should be implemented, allowing for a reduction of overestimation.

The average accuracy of the predictions compared to the realization is 0.6 movements per day higher (less accurate) then the average accuracy of the prediction compared to the last filed flight plans. When plotted in a graph, the accuracies are equal at some points and vary at others. Points of equality indicate that inaccuracy is often caused by factors that occur in the time between the creation of the OPS plan and the moment the last flight plans are filed. When there is a difference between the accuracy of the OPS plan prediction vs flight plan and OPS plan vs realization, it indicates that inaccuracy is often caused by a factor that occurred after the last flight plans are filed.

The predicted runway usage for the first inbound peak was equal to the realization on 91 of the 97 days. On six of the days where it was not, a changing wind direction was the main cause for the difference. In two instances, runway usage could not be clearly classified by means of directionality. The predicted runway use was equal to the capacity forecast form on 91 of the days as well. A changing wind direction caused a difference between the prediction and forecast form in six instances. Despite the negative impact wind direction had on accuracy, the results of this study indicate that the runway usage predictions are not significantly different from the runway usage realized nor from those indicated on the forecast from.

When looking at the accuracy of only the primary runway capacity prediction in the OPS plan, the average accuracy is 0.03 movements with a precision of 0.13 movements per hour. When considering the alternative predictions for the instances where the primary runway capacity prediction was not equal to the forecast form, the average accuracy is 0.11 movements with a precision of 0.11 movements per hour. A statistical test, with an interval of confidence of 95%, found that there is no significant difference between the mean capacity predicted (primary and alternative) and the mean capacity in the forecast form.

The predicted runway capacity was not equal to the forecast form on nine days. The difference can be explained by two main factors: changing weather conditions and human error. While the predictions were not equal to the forecast form, on five days the predictions could be considered correct regarding the weather information available at the time the OPS plan was created. Thus, the differences in capacity observed were caused by changing weather conditions that occurred after the creation of the OPS plan.

On three days, the inaccuracy can be attributed to human error. Once, the OPS plan provided eight different possible capacities for different weather conditions. The analysis only considers the first two options, which differed from the capacity mentioned in the forecast form. As providing eight different capacities is not the conventional way, this instance is considered a human error in the form of a formatting mistake. Twice, a total capacity of 72 (38+34) movements per hour was predicted. However, this is not possible as the maximum available runway capacity with two runways is 68 (34+34).

On the remaining day, runway combination 18R+27 was proposed in the prediction. The forecast form proposed 18R+22. The runway combination proposed in the OPS plan offers a slightly higher capacity, while the combination in the forecast form is a safer option in case of possible go-arounds. The runway combination of the OPS plan was the combination realized. The reasoning for the change of runway combinations in the forecast form was not clear, however, it was not due to human error or changing weather.

The period that has been analyzed in this study, from April 1st 2020, until November 14th 2020, is heavily affected by the COVID-19 pandemic. Travel restrictions and other measures that aim to limit the spread of the virus have caused Schiphol's traffic demand to drop drastically, which has a major impact on the planning procedures and execution of the operations. During the first three months of the period that was analyzed, the traffic demand had declined 90% compared to the year before. During this period, the accuracy and precision of the traffic demand predictions were particularly low. As the COVID-19 pandemic had just started, the inaccuracy could be explained by uncertainties related the new measures and restrictions. Other factors that could contribute to the inaccuracy of the prediction – regardless of if it is compared to the last filed flight plans or realization dataset – were found in the traffic prognoses provided by LVNL. According to these prognoses, it was hard to forecast which flights the airlines would cancel due to declining passenger numbers. Also, the recovery of the European air traffic was becoming more influenced by regional measures regarding the spread of the COVID-19 pandemic. For instance, in October 2020, KLM cancelled over 200 flights per week due to travel restrictions in Europe. Lower traffic demand could also be explained by the start of the winter planning which started in the beginning of October.

The data analyzed in this study is not representative of the traffic that Schiphol, and therefore LVNL, are designed to control. In future studies, when the industry has recovered from the COVID-19 pandemic, new data should be analyzed again to answer the same research question. Instead of analyzing traffic demand per day, future studies may assess the periods of the day that are subject to high traffic demand (e.g., daytime, or peak moments). Furthermore, future research might prove that there is a relationship between high traffic demand and higher accuracy. As this study examined the OPS plan predictions made for traffic demand and capacity, more research is needed to identify and analyze other key aspects of the OPS plan.

The times for the peaks in the OPS plan and forecast form are almost identical, however, these times are not always equal to the realized peaks. Additional research should further examine this aspect, as it could help improve the predictability of the operations further. If the predictions do not align with the realizations for peak times, then the capacity could be consistently impacted. This future research would require the analyzation of the realized runway capacity which was not available at the time of this study.

7 Conclusion

The aim of this study was to examine how predictable the predictions included in the OPS plan are in terms of accuracy and precision. The study analyzed three predictions: predicted inbound traffic demand for EHFIRAM, predicted runway capacity during the first inbound peak and the predicted runway use during the first inbound peak. The accuracy and precision of the traffic demand prediction were assessed on two levels of time: per day and per hour rolling. The runway capacity predictions, and the runway use were analyzed for the first inbound peak moment.

Based on the outcomes of the statistical tests and the data that was available for analysis, it can be concluded that the inbound traffic demand predictions for EHFIRAM per day and per hour rolling are accurate during the period with highest traffic demand, while the periods with lowest demand were found to be least accurate. Whereas the accuracy changes with traffic demand, the precision of the predictions is similar in each of the three periods. The runway usage predictions are accurate regarding the realization and the forecast form. When only considering the primary runway capacity predictions for the first inbound peak, the predictions are not significantly different from the forecast form. When considering the alternative predictions for the instances where the primary was not equal to the forecast form, the predictions are also not significantly different form the forecast form. It was revealed that human errors are the most impacting factor for differences in runway capacity. For runway use, changing weather conditions are the most impacting factor.

Based on how predictable the OPS plan predictions have been during the COVID-19 pandemic, it can be concluded that the predictions are accurate in all aspects researched. While it seems as though the predictions are more accurate than they are precise, the accuracy of the predictions generally improves as traffic demand increases.

8 Recommendations

The OPS plan was introduced while the aviation industry was in the middle of a unique situation where traffic demand had dropped over 90% compared to the year before. Aside from the fact that the demand was so much lower than it should have been, this new situation also caused uncertainties. Especially during these times, the OPS plan has suggested its added value, yet its predictability could continue to be researched. Even though the predictions seem more accurate when traffic demand increased, it is impossible to know if that would still be true when traffic demand reaches the same level as 2019 again. It is therefore recommended to research the accuracy of the predictions again with data that better represents 'normal' operation. In addition, it is recommended to implement a tool that enables day-to-day analyses of the operations. The dashboard developed during this study could be implemented in such a way that it provides post-ops analysis of the predictions.

On average, inbound traffic demand for EHFIRAM is 4.7 movements overestimated per day, and 0.2 movements per hour rolling, which equals to 4% overestimated per day, and 7% per hour rolling, relative to the realization. Even though overestimation is safer than underestimation, the purpose of the OPS plan is to align the expected traffic demand with the available runway capacity. It is therefore recommended to develop a framework that allows for adjusting the traffic demand predictions according to the accuracy. Before this could be done, it is recommended to determine what level of accuracy and precision is acceptable for the stakeholders of the OPS plan. This requires the accuracy and precision of the predictions (that being the spread of the accuracy) to be analyzed for a longer period (e.g., a year). Hence, the data that would be analyzed should be more representative to normal operations. Furthermore, research should point out if a distinguish should be made between periods with fluctuating demand, that being night period, peak, and off-peak moments.

The differences in runway capacity are caused five out of the ten times by human factors like formatting or style of the predictions and using the wrong information. It is therefore recommended to investigate possibilities of automating parts of the development of the OPS plan. That being, for instance, that the possible runway combinations are automatically gathered based on weather information and availability (e.g., due to maintenance). Future research is required to determine how often the realized runway combination is equal to the most preferred combinations (as in QRC25), based on pre-described conditions (i.e., wind direction, visibility, etc.) The final prediction is then still drafted following expert judgement, yet the chances of making mistakes might decrease.

A framework should be developed that enables analyzing runway capacity in a similar fashion as what is possible regarding analyzing traffic demand. The runway capacity prediction was manually extracted from the individual OPS plans. This dataset could be used as a template for this future framework. The realized runway capacity was also not available. Since the runway combinations are recorded, this could be combined with the weather conditions on the day of operations, and information found in QRC25 and QRC09 to determine the realized capacity. Future research may then analyze the predictability of the runway capacity predictions compared to the realization. Furthermore, the start and end times of the peaks in the OPS plan and forecast form are almost identical, however, these times are not always equal to the realized peaks. Additional research should further examine this aspect, as it could help improve the predictability of the operations further. If the predictions do not align with the realizations for peak times, then the capacity could be consistently impacted.

List of references

Bhandari, P. (2020, July 9). *An introduction to descriptive statistics*. Retrieved from Scribber: https://www.scribbr.com/statistics/descriptive-statistics/

Chesapeake Aviation Training. (2015, March 17). *Lightning strikes to aircraft while in flight*. Retrieved from AvWxWorkshops.com: https://avwxworkshops.com/etips/08-24-09.html#:~:text=Most%20of%20the%20aircraft%2Dinduced,close%20to%20the%20melting%20level.&text=The%20conditions%2C%20however%2C%20are%20not,by%20aircraft%20intercepting%20the%20lightning.

Eurocontrol. (2016). *Eurocontrol ATM Lexicon*. Retrieved from Eurocontrol: https://ext.eurocontrol.int/lexicon/index.php/Main_Page

Eurocontrol. (2017). *Airport CDM Implementation Manual.* Eurocontrol.

Eurocontrol. (2017). *Eurocontrol Anual Report 2016.* Eurocontrol.

Eurocontrol. (2019). *The Netherlands - Local Single Sky Implementation.* Eurocontrol.

Eurocontrol. (2020, June 23). *Eurocontrol ATFCM Operations Manual.* Retrieved from Eurocontrol: https://www.eurocontrol.int/sites/default/files/2020-05/eurocontrol-atfcm-operations-manual-22052020.pdf

Eurocontrol. (2020). *Optimised operational performance*. Retrieved from Eurocontrol: https://www.eurocontrol.int/network-manager

Eurocontrol. (2020, November 11). *TMA - Terminal Manoeuvring Area*. Retrieved from Eurocontrol: https://ansperformance.eu/acronym/tma/

Eurocontrol. (2021). *Daily Traffic Variation - ACCs (2020)*. Retrieved from Eurocontrol: https://www.eurocontrol.int/Economics/2020-DailyTrafficVariation-ACCs.html

Eurocontrol. (2021, February 10). *Demand data repository*. Retrieved from Eurocontrol: https://www.eurocontrol.int/ddr

FAA. (2021, February 8). *What is a NOTAM?* Retrieved from Federal Aviation Administration: https://www.faa.gov/about/initiatives/notam/what_is_a_notam/

Flynn, B. (2014). *Defining and Managing capacities.* Retrieved from ICAO: https://www.icao.int/SAM/Documents/2014-ATFM8/DAY_TWO_5_Cancun%20ATFM%20defining%20managing%20capacities_day2%2013h30_B%20Flynn.pdf

Forecast. (2021, Januari). *What is the Difference Between Accuracy & Precision in Scoping a Project?* Retrieved from Forecast: https://www.forecast.app/faqs/what-is-the-difference-between-accuracy-and-precision

Frost, J. (2018, October 28). *Central Limit Theorem Explained*. Retrieved from Statistics By Jim: https://statisticsbyjim.com/basics/central-limit-theorem/#comments

ICAO. (2007). *AIR TRAFFIC MANAGEMENT (Doc 4444).* INTERNATIONAL CIVIL AVIATION ORGANIZATION.

ITL. (2003, 6 1). *Engineering Statistics Handbook*. Retrieved from Chi-Square Test for the Variance: https://www.itl.nist.gov/div898/handbook/eda/section3/eda358.htm

KNMI. (2011). *Factsheet updatecriteria (AUTO) SPECIAL.* KNMI.

Lieshout, A., Devilee, J., & Obbens, Y. (2017). *Capaciteitsmanagement: Inrichten D-1.* LVNL.

LVNL. (2019). *De vierde baan regel en de Schiphol operatie.* Schiphol-Oost: LVNL.

LVNL. (2020). *Handleiding NEST ten behoeve van de Pre-Tact Unit.* Schiphol-Oost: LVNL.

LVNL. (2020). *Handleiding opstellen OPS plan.*

LVNL. (2020, July 12). *Jaarverslag 2019.* Retrieved from LVNL: https://www.lvnl.nl/media/3179/jaarverslag-2019-voor-publicatie-lr-1.pdf

LVNL. (2020, November 3). *Weersomstandigheden*. Retrieved from LVNL: https://www.lvnl.nl/veiligheid/veilige-uitvoering-in-de-praktijk/weersomstandigheden

LVNL. (2021, March 7). *Baangebruik Schiphol*. Retrieved from LVNL: https://www.lvnl.nl/omgeving/baangebruik

MedCalc Software Ltd. (2021). *Comparison of standard deviations (F-test)*. Retrieved from MedCalc: https://www.medcalc.org/manual/comparison_of_standard_deviations_f-test.php#:~:text=Comparison%20of%20two%20standard%20deviations,of%20two%20variances%20is%20calculated.&text=When%20the%20calculated%20P%20value,deviations%20are%20statistically%20significa

Mujica Mota, M., Boosten, G., & Zuniga, C. (2017). *Time to Sweat the Assets?* Querétaro: International Transport Forum.

NATS. (2020, December 3). *Introduction to Airspace*. Retrieved from NATS: https://www.nats.aero/ae-home/introduction-to-airspace/

Schiphol. (2020). *Monthly traffic data 1992-current.* Schiphol.

Schiphol. (2020). *Traffic review 2019.* Schiphol.

To70 & NLR/FerWay. (2020). *Multi-Airport Concept.* The Hague: Knowledge and Development Centre Mainport Schiphol.

van Delden, M., Westerveld, E., Lieshout, A., Obbens, Y., & Devilee, J. (2017). *Initiatieffase Capacity Management.* LVNL.

Wassenberg, F. (2019). *D-1 prediction.* Schiphol: KDC.

Westerveld, E. (2017). *Solving bottlenecks sector 3.* LVNL.

Zuidberg, J., & Vinkx, K. (2018). *Capacity demand at Schiphol Airport in 2023.* Amsterdam: SEO Amsterdam Economics.

Zuniga, C., & Boosten, G. (2020). *A Practical Approach to Monitor Capacity under the CDM Approach.* Amsterdam: Aerospace.

Appendix I Raw datasets

Column names

TIME	The exact day and time an aircraft enters the specified air space section. In some data frames also referred to as ENTRY, T_ENTRY or TIJD (Dutch for time).
CALLSIGN	A combination of letters and numbers used to identify a flight. A callsign is unique to an airline and the departure and destination airport, however, a callsign can be used several times a day.
FLIGHT_ID	In order to identify every individual flight, a unique id is given to each of them. This flight id is only used by LVNL, yet a flight id is not always used to identify the same aircraft in different data sources. A unique flight might therefore have multiple flight IDs in different data frames. Additionally, due to a technical error, a flight ID is sometimes used twice for different aircraft.
TV	In some data frames, the traffic volume represents the segment of airspace where the time is measured.
ADEP	The departure airport of the flight. This could be an airport in or outside of the Netherlands.
DEST	The destination airport of the flight. This could be an airport in or outside of the Netherlands.

Table 1 - EHFIRAM prediction

	CALLSIGN	FLIGHT_ID	ADEP	DEST
ENTRY				
2020-04-01 00:06:00	UAE9961	200389	KORD	EHAM
2020-04-01 01:33:00	KLM862	152891	RJAA	EHAM
2020-04-01 02:17:00	BCS21G	146244	EDDP	EHAM
2020-04-01 02:32:00	CSN451	134385	ZSPD	EHAM
2020-04-01 02:39:00	ABW325	195138	UUEE	EHAM
….	….	….	….	….
2020-11-14 19:44:00	TRA2K	651717	GCTS	EHAM
2020-11-14 19:51:00	ABW451	712278	UUEE	EHAM
2020-11-14 19:53:00	GTI8153	727539	UAAA	EHAM
2020-11-14 20:22:00	TFL398P	714870	TNCC	EHAM
2020-11-14 21:50:00	TRA3P	648724	LPMA	EHAM

[55,298 rows x 4 columns]

Table 2 - EHFIRAM realization

	CALLSIGN	ADEP	DEST	LOCATION_TYPE	...
ENTRY					
2020-04-01 00:33:00	ABW454	EHAM	EDDP	AIRFIELD	...
2020-04-01 00:34:00	ABW454	EHAM	EDDP	FIR	...
2020-04-01 00:44:00	BCS31L	EHAM	EGNX	AIRFIELD	...
2020-04-01 00:45:00	BCS31L	EHAM	EGNX	FIR	...
2020-04-01 01:16:00	KLM862	RJAA	EHAM	FIR	...
...	...	...	...	...	...
2020-11-14 23:05:00	QTR579L	OTHH	EHAM	AIRFIELD	...
2020-11-14 23:34:00	BBD315	BIKF	EBLG	FIR	...
2020-11-14 23:44:00	FDX5225	EDDK	KMEM	FIR	...
2020-11-14 23:47:00	ABW645	EHAM	UUDD	AIRFIELD	...
2020-11-14 23:48:00	ABW645	EHAM	UUDD	FIR	...

[247,438 rows x 8 columns]

Table 3 - EHFIRAM flight plans

	TV_ID	TR
TIJD		
2020-04-01 00:00:00	EH12345	1
2020-04-01 00:00:00	EH3+4+5	1
2020-04-01 00:00:00	EHACOD	1
2020-04-01 00:00:00	EHS3KLUX	1
2020-04-01 00:00:00	EHS3OUT	1
....	....	....
2020-11-15 23:20:00	EHTMAIN	1
2020-11-15 23:40:00	EHAM	1
2020-11-15 23:40:00	EHAMARR	1
2020-11-15 23:40:00	EHAMTMA	1
2020-11-15 23:40:00	EHTMAIN	1

[436,362 rows x 2 columns]

Table 4 - Runway prediction

	TIME_START	TIME_END	PEAK	LAND_1	LAND_2	...
DATE						
2020-04-01	2020-04-01 06:00:00	2020-04-01 23:40:00	off1	6	NaN	...
2020-04-02	2020-04-02 05:20:00	2020-04-02 23:40:00	off1	18R	NaN	...
2020-04-03	2020-04-03 04:20:00	2020-04-03 23:40:00	off1	36C	NaN	...
2020-04-04	2020-04-04 06:00:00	2020-04-04 23:40:00	off1	18R	NaN	...
2020-04-07	2020-04-07 06:00:00	2020-04-07 09:00:00	off1	6	NaN	...
...	...	...	...	...		...
2020-11-14	2020-11-14 10:00:00	2020-11-14 17:40:00	off3	18R	NaN	...
2020-11-14	2020-11-14 17:40:00	2020-11-14 19:00:00	in2	18R	18C	...
2020-11-14	2020-11-14 19:00:00	2020-11-14 19:20:00	off4	18R	NaN	...
2020-11-14	2020-11-14 19:20:00	2020-11-14 20:20:00	out2	18R	NaN	...
2020-11-14	2020-11-14 20:20:00	2020-11-14 21:30:00	off5	18R	NaN	...

[1045 rows x 22 columns]

Table 5 – Runway realization

	T_END	LANDING	TAKEOFF	LANDING_1	...
T_START					
2020-04-01 00:10:00	2020-04-01 01:45:00	18C	24	18C	...
2020-04-01 01:45:00	2020-04-01 06:35:00	06	36L	6	...
2020-04-01 06:35:00	2020-04-02 02:10:00	18R	24	18R	...
2020-04-02 02:10:00	2020-04-02 03:05:00	06	36L	6	...
2020-04-02 03:05:00	2020-04-03 06:15:00	18R	24	18R	...
...	...	...	...	...	...
2020-11-13 19:55:00	2020-11-14 06:20:00	18R	24	18R	...
2020-11-14 06:20:00	2020-11-14 07:45:00	18R-18C	24	18R	...
2020-11-14 07:45:00	2020-11-14 08:05:00	18R	24	18R	...
2020-11-14 08:05:00	2020-11-14 09:50:00	18R	24-18L	18R	...
2020-11-14 09:50:00	2020-11-15 05:50:00	18R	24	18R	...

[1838 rows x 8 columns]

Table 6 - Capacity forecast form

	DATUM_TIJD_PIEK_START	DATUM_TIJD_PIEK_EIND	PLR1	PLR2	PLC1	...
DATE_TIME						
2020-03-05 02:00:00	2020-03-05 05:30:00	2020-03-05 06:00:00	6	NaN	38.0	...
2020-03-05 02:00:00	2020-03-05 06:00:00	2020-03-05 06:40:00	6	NaN	38.0	...
2020-03-05 02:00:00	2020-03-05 06:40:00	2020-03-05 08:20:00	6	36R	34.0	...
2020-03-05 02:00:00	2020-03-05 08:20:00	2020-03-05 10:00:00	6	NaN	38.0	...
2020-03-05 08:15:00	2020-03-05 08:20:00	2020-03-05 10:00:00	6	NaN	38.0	...
...	...	...	...	...		...
2020-11-15 19:30:00	2020-11-15 20:20:00	2020-11-15 21:30:00	18R	NaN	38.0	...
2020-11-15 19:30:00	2020-11-16 06:20:00	2020-11-16 07:40:00	18R	18C	35.0	...
2020-11-15 19:30:00	2020-11-16 19:20:00	2020-11-16 20:20:00	18R	NaN	38.0	...
2020-11-15 19:30:00	2020-11-15 21:30:00	2020-11-16 05:30:00	18R	NaN	24.0	...
2020-11-15 19:30:00	2020-11-16 05:30:00	2020-11-16 06:20:00	18R	NaN	38.0	...

[3167 rows x 22 columns]

Appendix II Cleaned and resampled datasets

Table 7 - EHFIRAM per day

	Prediction	Flight plans	Realization
	MOVEMENTS	MOVEMENTS	MOVEMENTS
ENTRY			
2020-04-01	111	83	83
2020-04-02	107	82	82
2020-04-03	90	87	87
2020-04-04	87	74	74
2020-04-06	85	72	71
….	….	….	….
2020-11-10	242	244	241
2020-11-11	266	256	256
2020-11-12	271	272	268
2020-11-13	286	279	278
2020-11-14	255	253	252

[219 rows x 1 columns]

Table 8 - EHFIRAM prediction per hour rolling

	Prediction	Flight plans	Realization
	MOVEMENTS	MOVEMENTS	MOVEMENTS
ENTRY			
2020-04-01 00:00:00	1	0	0
2020-04-01 00:20:00	0	1	1
2020-04-01 00:40:00	1	1	1
2020-04-01 01:00:00	1	1	1
2020-04-01 01:20:00	2	0	0
….	….	….	….
2020-11-14 22:20:00	0	1	1
2020-11-14 22:40:00	0	0	1
2020-11-14 23:00:00	0	0	0
2020-11-14 23:20:00	0	0	0
2020-11-14 23:40:00	0	0	0

[15768 rows x 1 columns]

Table 9 - Runway use during first inbound peak

	Prediction	Forecast form	Realization
	DIR	DIR	DIR
TIME_START			
2020-07-01	South	South	South
2020-07-02	South	South	South
2020-07-03	South	South	South
2020-07-04	South	South	South
2020-07-08	South	South	South
….	….	….	….
2020-11-10	South	South	South
2020-11-11	South	South	South
2020-11-12	South	South	South
2020-11-13	South	South	South
2020-11-14	South	South	South

[97 rows x 1 columns]

Table 10 - Runway capacity during first inbound peak (primary)

	Prediction	Prediction	Forecast form
	PLC_TOT	PALC_TOT/ ALC_TOT	PLC_TOT
TIME_START			
2020-07-01	68	68	68
2020-07-02	68	68	68
2020-07-03	68	68	68
2020-07-04	65	68	68
2020-07-08	68	68	68
….	….	….	….
2020-11-07	68	68	68
2020-11-10	44	44	44
2020-11-11	44	44	44
2020-11-12	68	68	68
2020-11-13	68	67	67

[94 rows x 1 columns]

Appendix III Descriptive statistics

Table 11 – Descriptive statistics inbound traffic demand EHFIRAM per day

	OPS plan	Realized	Accuracy	Accuracy (ratio)
count	219	219	219	219
mean	251.35	246.66	4.69	1.04
mode			3	1.00
std	119.34	121.08	5.55	0.06
sem			0.38	
min	57	54	-9	0.97
25%	120.25	116.00	1.00	1.00
median	276.00	272.00	4.00	1.02
75%	361.25	360.00	8.00	1.05
max	446	448	28	1.34

Table 12 – Descriptive statistics inbound traffic demand EHFIRAM per day period one

	OPS plan	Realized	Accuracy	Accuracy (ratio)
count	87	87	87	87
mean	116.66	109.85	6.80	1.07
mode			11	1.00
std	32.00	33.44	5.48	0.07
sem			0.59	
min	57	54	-5	0.97
25%	92.00	82.50	3.00	1.03
median	112.00	105.00	6.00	1.06
75%	134.00	130.50	10.00	1.10
max	201	190	28	1.34

Table 13 – Descriptive statistics inbound traffic demand EHFIRAM per day period two

	OPS plan	Realized	Accuracy	Accuracy (ratio)
count	62	62	62	62
mean	331.29	330.05	1.24	1.01
mode			4	1.00
std	65.43	66.00	4.93	0.02
sem			0.63	
min	201	190	-9	0.97
25%	274.00	269.75	2.75	0.99
median	318.50	320.00	1.00	1.01
75%	392.00	388.75	4.75	1.01
max	432	427	11	1.06

Table 14 – Descriptive statistics inbound traffic demand EHFIRAM per day period three

	OPS plan	Realized	Accuracy	Accuracy (ratio)
count	72	72	72	72
mean	345.30	340.14	5.16	1.02
mode			3	1.00
std	49.27	49.45	4.68	0.01
sem			0.55	
min	242	241	-4	0.99
25%	306.00	304.00	2.00	1.01
median	347.00	344.00	5.00	1.01
75%	388.00	382.00	8.00	1.03
max	446	448	17	1.06

Table 15 – Descriptive statistics inbound traffic demand EHFIRAM per hour rolling

	OPS plan	Realized	Accuracy	Accuracy (ratio)
count	15768	15768	15768	15165*
mean	10.47	10.28	0.20	1.07
mode			0	1.00
std	11.33	11.07	2.72	0.58
sem			0.02	
min	0	0	-16	0.00
25%	2.00	2.00	-1.00	0.83
median	6.00	6.00	0.00	1.00
75%	15.00	15.00	2.00	1.20
max	66	67	14	7.00

* the count for accuracy is lower due to instances where the realization is zero

Table 16 – Descriptive statistics inbound traffic demand EHFIRAM per hour rolling period one

	OPS plan	Realized	Accuracy	Accuracy (ratio)
count	6264	6264	6264	5868*
mean	4.86	4.58	0.28	1.13
mode			0	1.00
std	5.43	5.12	2.11	0.73
sem			0.03	
min	0	0	-10	0.00
25%	1.00	1.00	-1.00	0.75
median	3.00	3.00	0.00	1.00
75%	6.00	6.00	1.00	1.33
max	42	40	11	7.00

*the count for accuracy is lower due to instances where the realization is zero

Table 17 – Descriptive statistics inbound traffic demand EHFIRAM per hour rolling period two

	OPS plan	Realized	Accuracy	Accuracy (ratio)
count	4464	4464	4464	4391*
mean	13.80	13.75	0.05	1.03
mode			0	1.00
std	11.98	11.58	2.99	0.45
sem			0.04	
min	0	0	-12	0.00
25%	5.00	5.00	2.00	0.83
median	10.00	10.00	0.00	1.00
75%	20.00	20.00	2.00	1.16
max	61	63	14	6.00

*the count for accuracy is lower due to instances where the realization is zero

Table 18 – Descriptive statistics inbound traffic demand EHFIRAM per hour rolling period three

	OPS plan	Realized	Accuracy	Accuracy (ratio)
count	5184	5184	5184	5040*
mean	14.39	14.17	0.21	1.04
mode			0	1.00
std	13.11	12.81	3.11	0.46
sem			0.04	
min	0	0	-16	0.00
25%	4.00	3.00	1.00	0.88
median	11.00	11.00	0.00	1.00
75%	22.00	22.00	2.00	1.15
max	66	67	13	6.00

*the count for accuracy is lower due to instances where the realization is zero

Table 19 – Descriptive statistics inbound traffic demand EHFIRAM vs flight plans per day

	OPS plan	Flight plans	Accuracy	Accuracy (ratio)
count	219	219	219	219
mean	251.35	247.30	4.06	1.03
mode			3	1.00
std	119.34	121.13	5.45	0.05
sem			0.37	
min	57	55	-12	0.96
25%	120.25	116.75	0.00	1.00
median	276.00	272.50	3.00	1.01
75%	361.25	359.25	8.00	1.04
max	446	447	28	1.34

Table 20 – Descriptive statistics inbound traffic demand EHFIRAM vs flight plans per hour rolling

	OPS plan	Flight plans	Accuracy	Accuracy (ratio)
count	15768	15768	15768	15238*
mean	10.47	10.30	0.17	1.06
mode			0	1.00
std	11.33	11.10	2.32	0.53
sem			0.02	
min	0	0	-13	0.00
25%	2.00	2.00	-1.00	0.86
median	6.00	6.00	0.00	1.00
75%	15.00	15.00	1.00	1.17
max	66	66	14	7.00

*the count for accuracy is lower due to instances where the realization is zero

Table 21 – Descriptive statistics runway capacity primary prediction

	OPS plan	Forecast	Accuracy
count	94	94	94
mean	66.74	66.48	0.27
mode			0
std	5.80	5.75	1.30
sem			0.13
min	30	30	-4
25%	68	68	0
median	68	68	0
75%	68	68	0
max	72	68	8

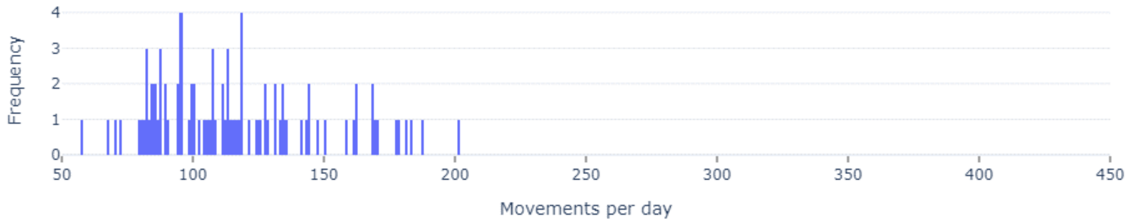
Table 22 – Descriptive statistics runway capacity primary and alternative prediction

	OPS plan	Forecast	Accuracy
count	94	94	94
mean	66.73	66.48	0.26
mode			0
std	5.77	5.75	1.06
sem			0.11
min	30	30	0
25%	68	68	0
median	68	68	0
75%	68	68	0
max	72	68	8

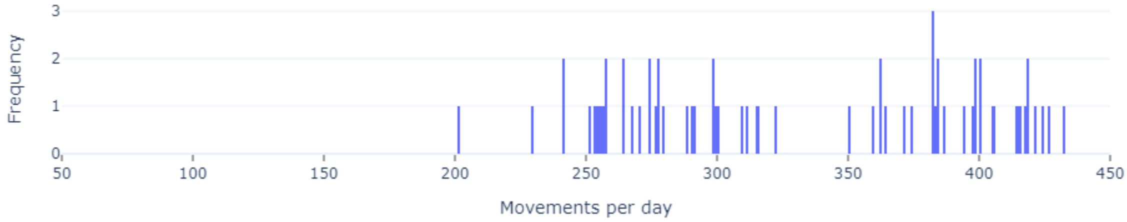
Appendix IV Distribution of datasets

Graph 1 – Data distribution predicted traffic demand EHFIRAM inbound per day

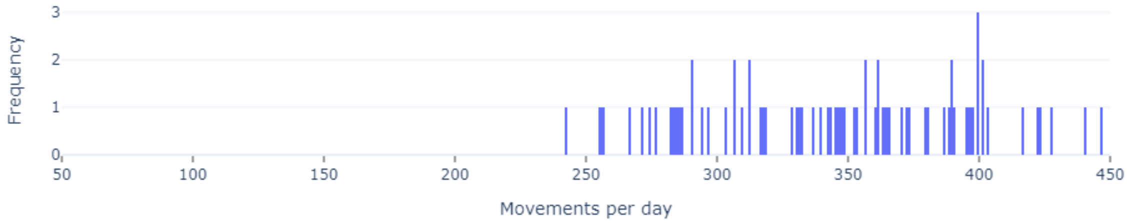
Distribution of traffic demand predicted period one



Distribution of traffic demand predicted period two

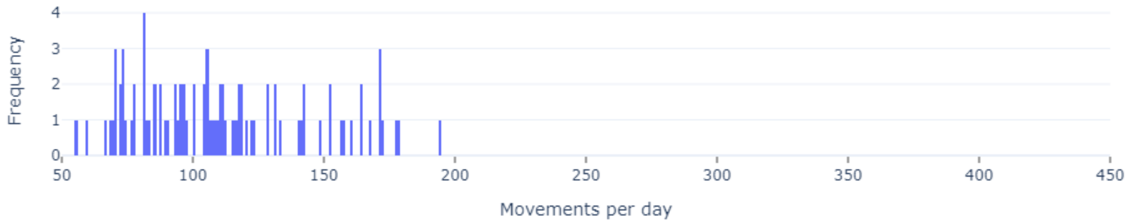


Distribution of traffic demand predicted period three

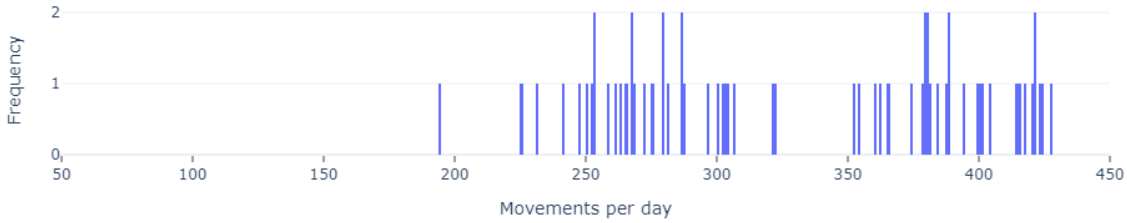


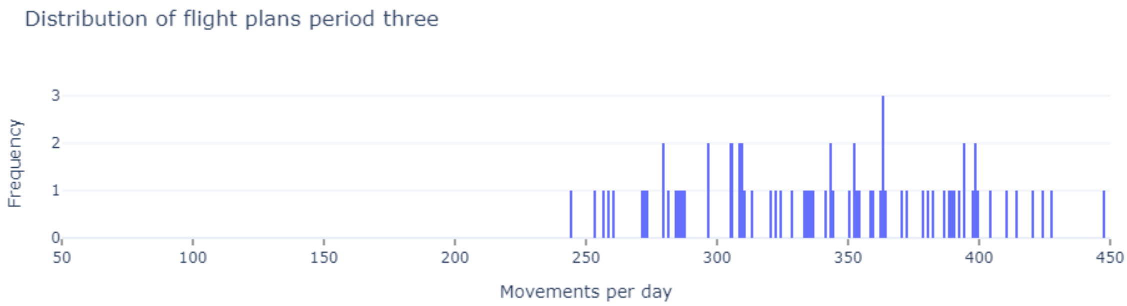
Graph 2 – Data distribution flight plans EHFIRAM inbound per day

Distribution of flight plans period one

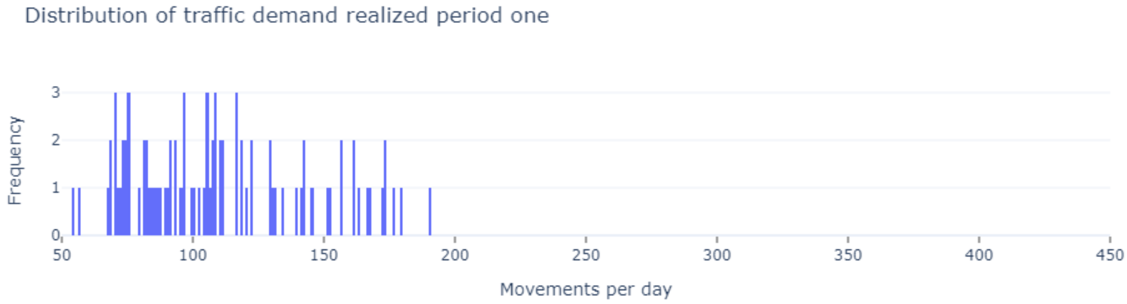


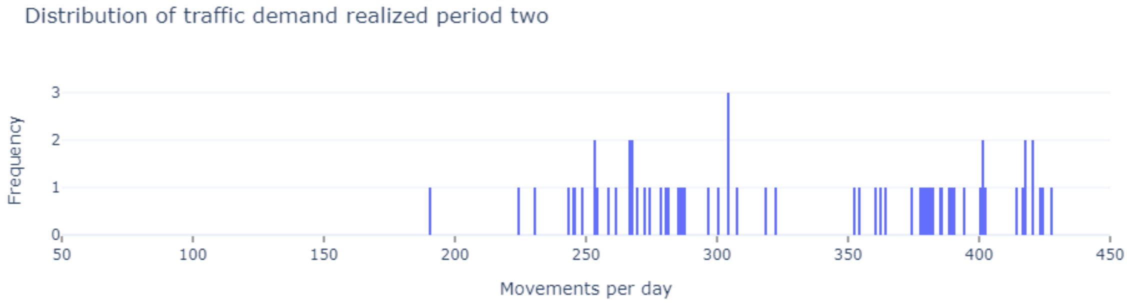
Distribution of flight plans period two

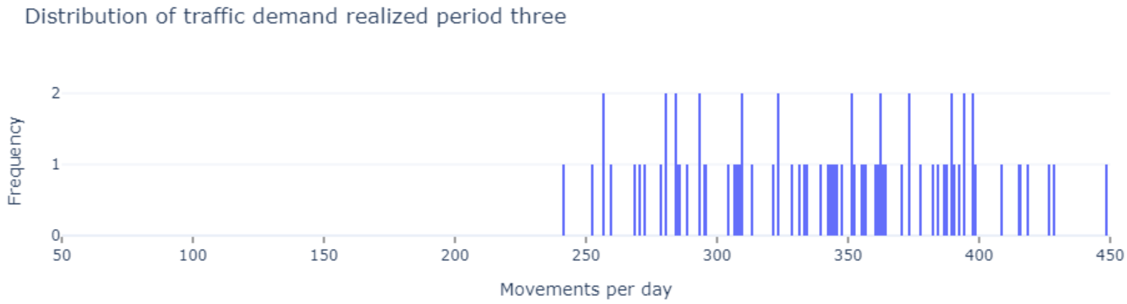




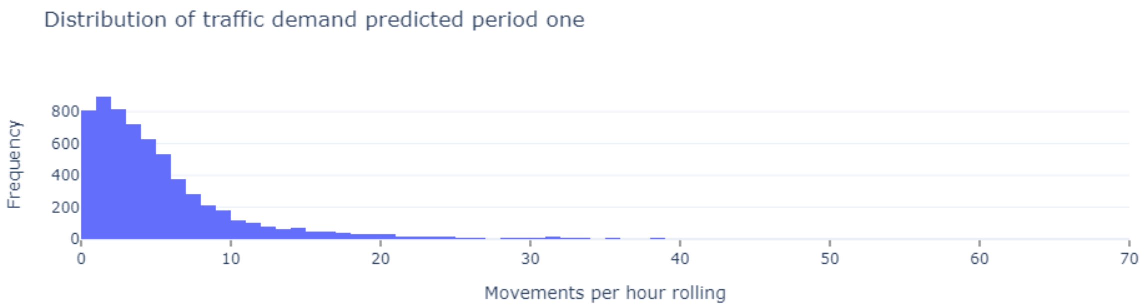
Graph 3 – Data distribution realized traffic demand EHFIRAM inbound per day

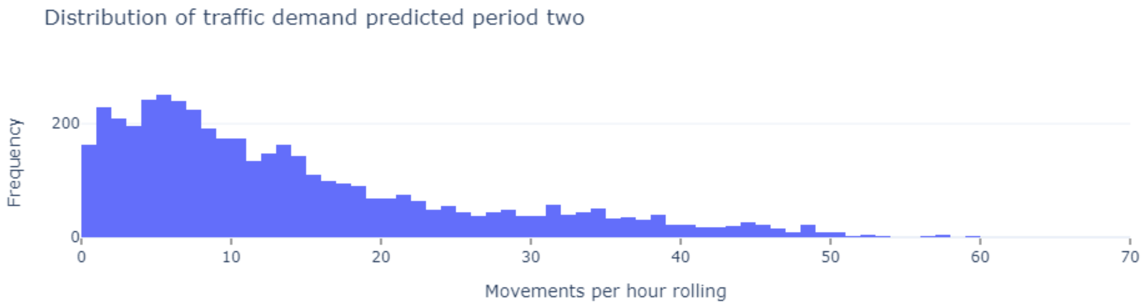


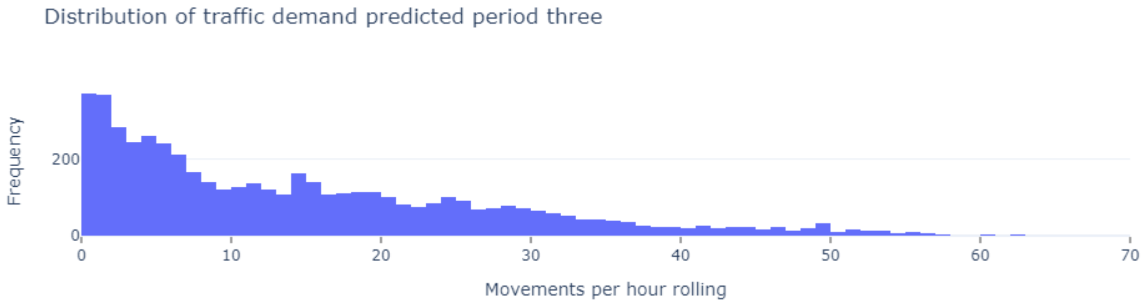




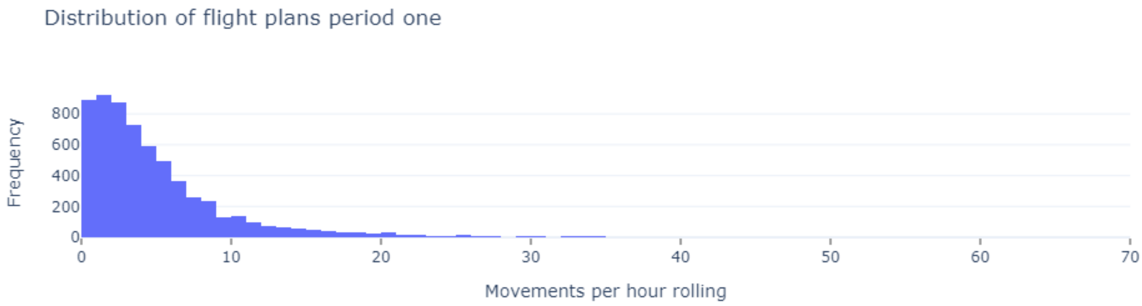
Graph 4 – Data distribution predicted traffic demand EHFIRAM inbound per hour rolling

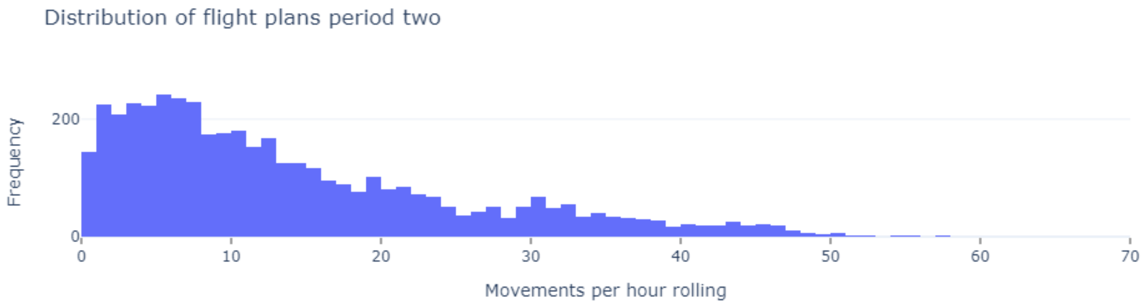


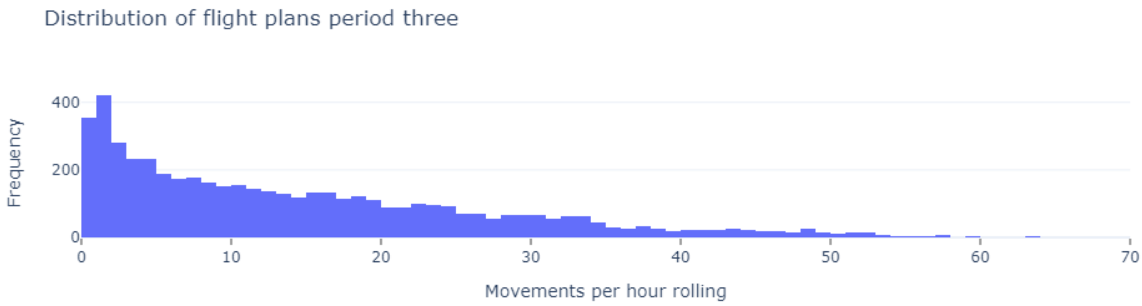




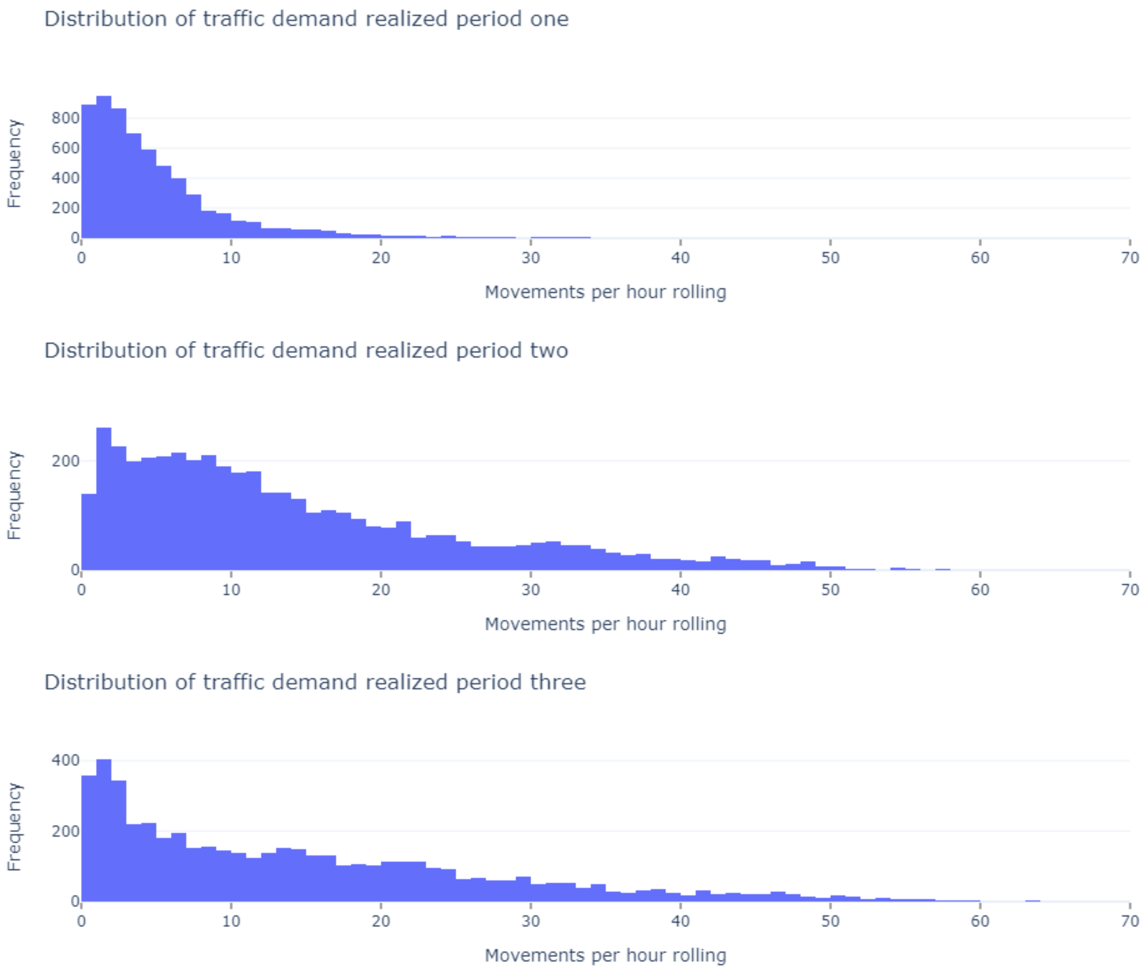
Graph 5 – Data distribution flight plans EHFIRAM inbound per hour rolling



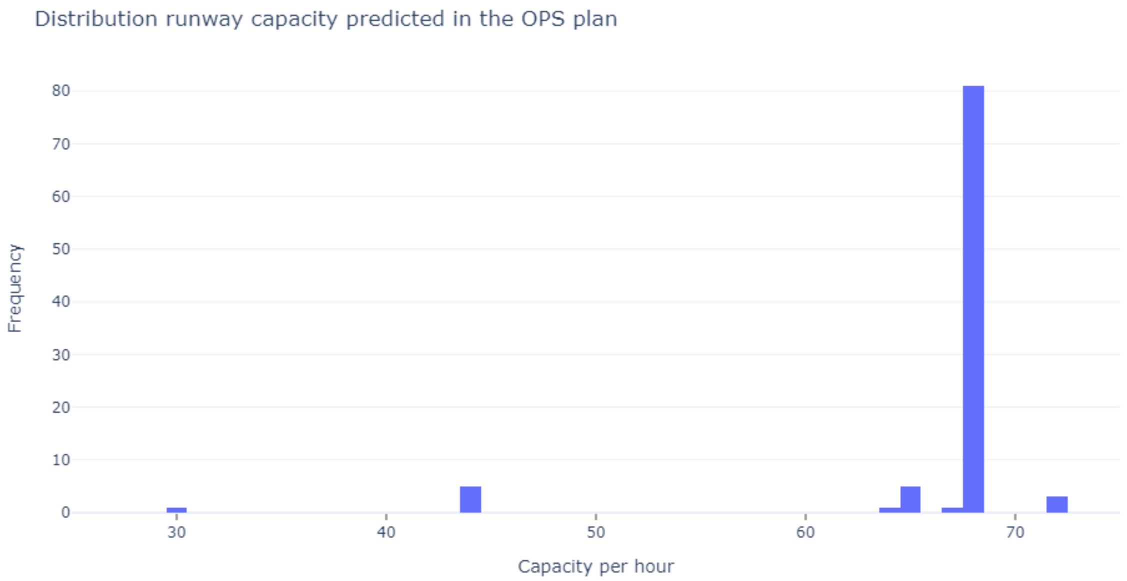




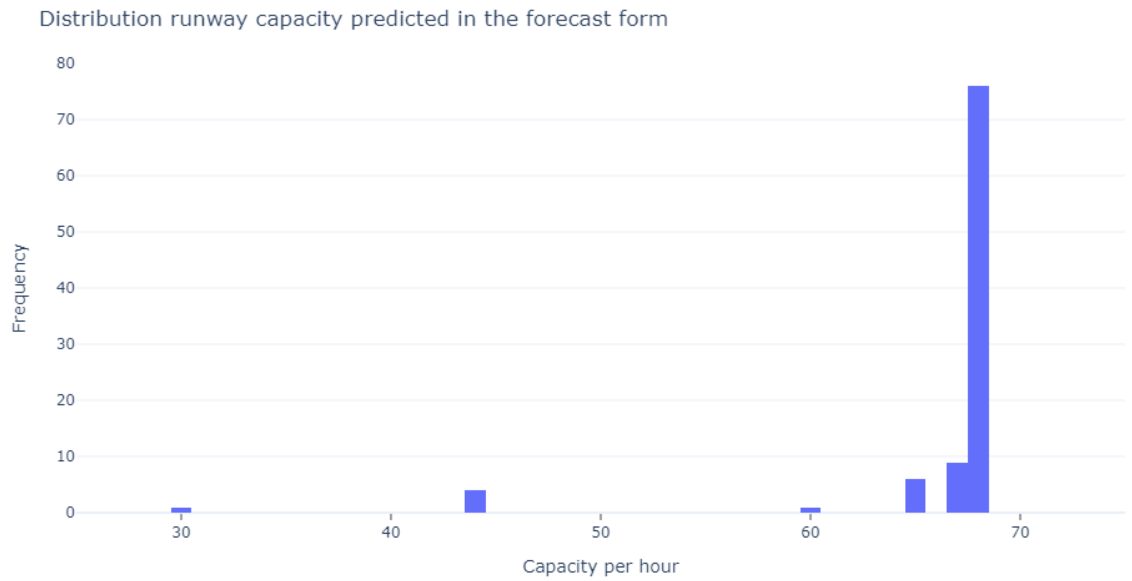
Graph 6 – Data distribution realized traffic demand EHFIRAM inbound per hour rolling



Graph 7 – Data distribution predicted runway capacity



Graph 8 – Data distribution runway capacity in forecast form


Distribution runway capacity predicted in the forecast form
Frequency
80
70
60
50
40
30
20
10
0
30
40
50
60
70
Capacity per hour

Appendix V Additional runway information

Table 23 – Runway combinations where OPS plan prediction was not equal to the realization

Date	OPS plan		Forecast form	Realization	Comments OPS plan	Comments forecast form
	Primary	Alternative				
Jul 11th	**06+36R** North	**36R+36C** North	**18R+18C** South	**18R+18C** South	Due to possible tail-wind 06	-
Sep 15th	**18R+18C** South	**18R+18C** South	**06+36R** North	**06+36R** North	Small chance on BZO during inbound peak / alt for BZO	Depending VIS
Jul 14th	**18R+18C** South	-	**18R+18C** South	**36R+36C** North	After 1400UTC north-ern rwy use	-
Aug 13th	**18R+18C** South	-	**06+36R** North	**06+36R** North	Rwy use depending development show-ers, rwy config based on stable situation in morning	Alt. depending on wind
Sep 11th	**18R+18C** South	-	**06+36R** North	**06+36R** North	-	-
Sep 29th	**36R+36C** North	-	**18R+18C** South	**18R+18C** South	-	Alt. due vis/ceiling.

Table 24 – Runway combinations where OPS plan prediction was not equal to the forecast form

Date	OPS plan		Forecast form	Realization	Comments OPS plan	Comments forecast form
	Primary	Alternative				
Jul 11th	**06+36R** North	**36R+36C** North	**18R+18C** South	**18R+18C** South	Due to possible tail-wind 06	-
Sep 15th	**18R+18C** South	**18R+18C** South	**06+36R** North	**06+36R** North	Small chance on BZO during inbound peak / alt for BZO	Depending VIS
Jul 17th	**18R+18C** South	-	**06+36R** North	**18R+18C** South	-	-
Aug 13th	**18R+18C** South	-	**06+36R** North	**06+36R** North	Rwy use depending development show-ers, rwy config based on stable situation in morning	Alt. depending on wind
Sep 11th	**18R+18C** South	-	**06+36R** North	**06+36R** North	-	-
Sep 29th	**36R+36C** North	-	**18R+18C** South	**18R+18C** South	-	Alt. due vis/ceiling.

Table 25 – Runway combinations where OPS plan prediction was not equal to forecast form

Date	OPS plan		Forecast form	Realization	Comments forecast form	Possible cause for difference*
	Primary	Alternative				
Jul 14th	**18R+18C** 38+34	-	**18R+18C** 34+34	**36R+36C**	-	**Human error** (capacity not possible)
Aug 5th	**18R+18C** 38+34	-	**18R+18C** 34+34	**18R+18C**	-	**Human error** (capacity not possible)
Aug 18th	**18R+27** 34+34	-	**18R+22** 35+30	**27+18R**	18C/36C NA due to WIP	Rwy 22 is safer due to possible go-arounds (convergent flightpaths), 27 higher capacity.
Sep 3rd	**18C+27** 34+34	-	**18C+27** 32+35	**27+18C**	18R NA due to WIP	**Changing weather conditions** (prediction correct according to weather (visibility))
Oct 7th	**27+18R** 34+34	-	**18R+27** 35+32	**27+36C**	-	**Changing weather conditions** (prediction correct according to weather (visibility))
Jul 9th	**18R+18C** 34+34	**18R+18C** 34+34	**18R+18C** 35+32	**18R+18C**	-	**Changing weather conditions** (prediction correct according to weather (visibility))
Sep 8th	**18R+18C** 34+34	**18R+18C** 34+34	**18R+18C** 35+32	**18R+18C**	-	**Changing weather conditions** (prediction correct according to weather (visibility))
Oct 23rd	**18R+18C** 34+34	**18R+18C** 34+34	**18R+18C** 22+22	**18R+18C**	BZO A, alt BZO B	**Human error** (formatting)
Oct 24th	**18R+18C** 34+34	**18R+18C** 34+34	**18R+18C** 35+32	**18R+18C**	ALT in case of CB's	**Changing weather conditions** (prediction correct according to weather (visibility))

* Causes elaborated in this table are the result of an Expert Session with members of the OPS plan development team.

Table 26 – Runway combinations where OPS plan prediction was not equal to forecast form and were considered outlier

Date	OPS plan		Forecast form	Realization	Comments forecast form	Possible cause for difference*
	Primary	Alternative				
Jul 10th	**36R+36C** 34+34	-	**06** 38	**06+36R**	-	**Analysis limitation** (peak times did not correspond in prediction and forecast form)
Nov 6th	**18R+18C** 22+22	-	**18R+18C** 35+32	**18R+18C**	-	**Changing weather conditions** (prediction correct according to weather (BZO B))
Sep 30th	**18R+27** 34+34	**18R+22** 35+30	**18R+18C** 22+22	**18R+22**	marg Wx/bzo A, reduced cap because twy Z clsd	**Human error** (information available, but not used (reduced cap due twy Z clsd)
Nov 14th	**18R+18C** 22+22	**18R+18C** 28+28	**18R+18C** 35+32	**18R+18C**	-	**Changing weather conditions** (prediction correct according to weather (BZO B))

* Causes elaborated in this table are the result of an Expert Session with members of the OPS plan development team.