Dynamic Booking Forecasting

A Kenya Airways Case Study

T.H.T. van Ostaijen

Delft University of Technology - Kenya Airways

Dynamic Booking Forecasting

A Kenya Airways Case Study

by

T.H.T. van Ostaijen

For obtaining the degree of Master of Science in Aerospace Engineering
at Delft University of Technology

May 8, 2017

Graduating professor:	Dr.ir. H.G. Visser,	Delft University of Technology
University supervisors:	Dr. ir. B.F. Santos,	Delft University of Technology
	Dr. M.A. Mitici,	Delft University of Technology
	Dr. ir. M. Pini	Delft University of Technology
Industry supervisors:	Thomas Omondi,	Kenya Airways, Director Strategy & Performance mgmt.
	Paul Mukunzi,	Kenya Airways, Head Commercial Analysis RM

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

Abbreviations

A/C	Aircraft
ARIMA	Autoregressive Integrated Moving Average
AU	Authorization Level
BH	Bookings Horizon
BOAC	British Overseas Airways Corporation
CDF	Cumulative Density Function
DBD	Days Before Departure
DOW	Day-Of-Week
HA	Historical Average
HOPM	Higher Order Polynomial Model
IATA	International Air Transport Association
JKIA	Jomo Kenyatta International Airport
KLM	Koninklijke Luchtvaart Maatschappij
KQ	Kenya Airways
LF	Load Factor
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MC	Markov Chain
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimation
MMFE	Martingale Method of Forecast Evolution
OD	Origin-Destination
PI	Prediction Interval
PIS	Prediction Interval Score
PLS	Penalized Least Squares
PU	Pick-Up
SAS	Scandinavian Airline System
SSE	Error Sum of Squares
STD	Standard Deviation
TDC	True Destination Code
TOC	True Origin Code

Nomenclature

K	The upper limit of the state space
N	The length of the forecast horizon
$P_{t,t-1}$	The transition matrix of the Markov Chain describing the transition from X_t to X_{t-1}
$S_\alpha(p_u, p_l, X_t)$	The interval score of a $(1-\alpha)\cdot 100$ prediction interval
T	The set of time indices for the forecast
$U_{t,0}$	The bookings expected to be picked-up from time t to the departure day
Z_t	The random variable of booking increments from time t to time *t-1*
π_t	The probability vector of the Markov Chain at time t
$\pi_t(i)$	The probability to have i net bookings in the system at time t
f	A flight in the bookings dataset
m	The number of forecasts made of a specific flight
n	The number of flights in the bookings dataset
p_l	The lower limit of the prediction interval
p_u	The upper limit of the prediction interval
t	The number of days remaining of days remaining until departure
$E[X_t]$	The expected net number of bookings at time t
$X_{0,f}$	The final bookings for the flight that departed at date f
X_t	The net number of bookings in the system at time t
α	The weighting factor balancing the empirical and parametric probabilities
$\hat{X}_0$	Forecast of the net number of bookings at the day of departure

Executive Summary

Airline revenue management aims to sell the seats on their planes to passengers at a price that is as close as possible to their maximum willingness to pay for a seat. In practice, airlines try to achieve this by creating a fare structure for the seats on their planes after which the seats are grouped under these different fares. Revenue management practice for airlines therefore consists of a pricing part and a seat allocation part. The fares can be updated, however, the fare structure is considered static compared to the seat allocation (also called allocation) control that takes place more dynamically in revenue management systems. Seat allocation control can therefore be considered as the key process within revenue management of an airline.

The booking forecasting model plays a central role in seat inventory control and therefore in the revenue management process. In theory, a 100% accurate forecast will result in a maximum revenue figure and inaccurate forecasts will result in a deviation from this maximum achievable revenue. Studies on the impact of forecasting accuracy on passenger revenues have demonstrated this by showing that a 10% improvement in forecasting accuracy can lead to a 0.5% to 3% increase in annual passenger revenue. For a major US airline with high demand flights this comes down to \$10 to \$60 million.

The quality of a revenue management forecast is of outmost importance for an airline. However, forecasting for revenue management has proven to be a challenging task because of the dynamics and complexity of the bookings process which is influenced by many factors such as pricing discounts, special events or defections of passengers from delayed or cancelled flights. As a result, the prediction for the future demand expressed as a single value is mostly wrong. In order to have an accurate forecasting model it is therefore important not only to forecast future demand close to the actual realized values, but also to provide the revenue management controller with information on the degree of uncertainty that is involved in the prediction. From there the research objective is formulated as follows,

"Increase the accuracy of bookings forecasts by contributing to the development of a dynamic forecasting model that includes the uncertainty in the forecast on a cabin class and flight level.""

In this study a new approach is proposed in which the dynamic booking forecast model includes a probability distribution representing the uncertainty of the future value using the transitional probabilities in a Markov Chain model. The proposed forecasting method is a novel approach to bookings forecasting which has not been proposed within airline revenue management before. The model is tested for several parameter settings in order to determine the best setting for the forecasting model in terms of average forecast error and uncertainty modeling using actual bookings data of Kenya Airways. Thereafter the model is compared with traditional forecasting methods to assess its relative performance.

The model is tested using two years of historical data of Kenya Airways for two different flights: KQ101 (LHR-NBO) and KQ512 (NBO-BKO). The model parameters are calibrated to achieve a minimal average forecasting error. For the parametric distribution setting of the model, fitting no distribution yielded the lowest error for both flights. Nevertheless, the lognormal distribution performed slightly better for short-term forecasts. A flight specific performance has been observed for the various weighting factors α. This weighting factor determines the relative weight attributed to the empirical and parametric distributions. The forecasts of KQ101 show the lowest error for an α of around 0.8 whereas a weighting factor of 0.5 is best for KQ512. Regarding the selective data that is chosen as an input to the forecasting model, the monthly seasonal data provided the best results for KQ101 and day-of-week data yielded the lowest error for KQ512.

The best parameter settings for the model to achieve the best representation of the forecast uncertainty are much similar to the parameter settings that lead to the lowest absolute forecast error. An exception to this is the weighting factor. Both for KQ101 and KQ512, a high value of 0.8 is found to yield the best interval scores. This does not correspond to the value of α that is found to result in the lowest forecast error for KQ512 as explained above. Therefore, it can be stated that the model settings that follow from the model calibration on forecast error do not necessarily also lead to an optimal uncertainty modeling. As a result, for new flights, the model should be calibrated on both performance measurements.

The mean absolute error of the new model with the best parameter settings is compared with the error resulting from traditional forecasting methods. It is found that the new model yields a lower mean forecasting error than the traditional methods for flight KQ101. Based on the Kruskal-Wallis significancy test it is found that a numerous amount of differences are significant enough to state that the Markov Chain method proposed in this study provides on average a lower error than the traditional methods. Especially for the medium-term forecasts of 100 to 39 days before departure the Markov Chain method is significantly more accurate than the traditional methods. For KQ512 the new model also yield lower forecasting errors, nevertheless, these differences are relatively small. From the Kruskal-Wallis test it followed that these differences for KQ512 are not significant enough.

The probability distributions are validated in order to determine if the model succeeds in modeling the uncertainty in the forecasts. For KQ512 it is found that the probability distributions represent the uncertainty in the future number of bookings well. It only captures slightly more values for lower prediction intervals and slightly more for the high prediction intervals, which could be improved by increasing the sharpness of the distribution around the center and increasing the trailing tails of the distribution. For KQ101 it is found that overall the distributions fail to display the uncertainty well. After a more detailed investigation it is found that the distributions for predictions from 70 days out of long-term forecasts with an horizon of more than 80 days are overconfident and capture a significantly lower number of bookings than indicated by the probability distributions. This asks for distributions that are wider in order to represent the high uncertainty. For the remaining horizons and days before departure the distributions are found to model the uncertainty well.

Altogether, it is found that the Markov Chain method proposed in this study showed to be more accurate for one of the test cases while for the other test case this could not be concluded. Nevertheless, it can definitely be stated that the Markov Chain method does not perform worse than the traditional methods. Furthermore, the method has shown to be able to display the uncertainty in the future number of bookings well. For only a part of the forecasts made for KQ101 that statement does not hold. However, the method has shown to yield a lower forecasting error for those forecasts. From there a conclusive answer can be formulated on the research objective. Overall, it can be stated that the research objective has been reached in this study following from the above mentioned ability of the model to achieve a high accuracy and provide a presentation of the future uncertainty as well.

Regarding the further development of the model the main next step would be to test the method using unconstrained data. Constrained data might be the reason why the Markov Chain method outperforms the traditional methods only for the high load factor flight (KQ101). Furthermore, the model can be tested in a different setting, such as another airline, or on another level; e.g. on fare class forecasting level, OD-level or flights aggregated over a period of a week or month. Finally, the model could be improved by considering bookings and cancellations separately, individual and group bookings separately in the transition matrices and by exploring the advantage of using sophisticated data pooling methods.

Acknowledgements

It was in the last week of September 2015 that I first stepped aboard the Boeing 787-8 of Kenya Airways in Amsterdam, with final destination Nairobi. I was quite excited for my first visit to Africa and not completely sure what to expect. It marked the start of my internship project with Kenya Airways and the beginning of an amazing journey. I am very grateful that I was given the opportunity to extend the journey with 9 months by working on this thesis in collaboration with the people that I by now consider being my colleagues of Kenya Airways. In the end I have had a great time in Nairobi and Kenya, a place that felt like home to me from the very beginning. This is all due to the people that I have met along the way who "Kenyanised" me to the fullest.

First of all, I would like to thank Thomas Omondi for the opportunity to work on an assignment at Kenya Airways and for providing me with all the help and information for the practical matters of my stay in Nairobi. Then, sincere thanks to my direct supervisor Paul Mukunzi as well as the other commercial analysts, Caroline Lodenyo, Joseph Ayieko and Catherine Waa for the the help in formulating the model, providing me with data and the feedback along the project.

Furthermore, I want to thank the colleagues at the revenue management department for the pleasant conversations over the coffee and lunch and the wonderful time I have had at the office. Special thanks for Sabina Bwire who has been there both during my internship and thesis project to help me out with all the minor and major issues that I have been facing while working in Nairobi.

My stay in Nairobi, Kenya could not be that successful without the other Delft students that have stayed in the apartment along Ndemi Rd. I am grateful for the unforgettable experience and memories we share from our stay in "Delft Office Nairobi".

Next to the people in Kenya, I would also like to thank the TU Delft for the support. I am especially grateful for the help and support I received from Bruno and Mihaela in the many discussions on the project and the assistance in the development of the model. Furthermore, I would like to thank Prof. dr. Curran and Dr. ir. Visser for providing feedback at the halfway point of the project.

Working and living in two countries can be challenging. A special word of thanks to my family, who supported me with everything during this project and the entire 6.5 years of my student time in Delft. My girlfriend, Sophie, I would like to thank for her support, patience during the times I was away from home and the warm welcomes back at home in The Hague. I am glad that you have been able to visit Kenya so we share a part of my amazing Kenyan experience. And to my friends, for accepting the absence at soccer matches, at social gatherings and for proofreading the report.

The journey has come to an end. It is time to take on new challenges and to start a new adventure. I am sure the knowledge and experiences I have gained during this project will help me in chasing my dreams.

T.H.T. van Ostaijen
Delft, May 2017

Contents

List of Figures

List of Tables

1 Introduction

This chapter presents an introduction to the study. Firstly, revenue management is introduced including an explanation of revenue management forecasting, complexity of forecasting and the practice of revenue management at Kenya Airways. Then, the importance of bookings forecasts is addressed. Thereafter, the introduction is finalized with a roadmap of this thesis.

1.1. Introduction to Revenue Management

The department of revenue management within an airline aims to sell the seats on their planes to passengers at a price that is as close as possible to their maximum willingness to pay for a seat. In practice, airlines try to achieve this by creating a fare structure for the seats on their planes and then the seats are grouped under these different fares. Revenue management practice for airlines therefore consists of a pricing part and a seat allocation part. The fares can be regularly updated, however, the fare structure is considered static compared to the seat allocation control that takes place more dynamically in revenue management systems (Usman, 2003). Seat allocation control can therefore be considered as the key process within airline revenue management. A typical seat inventory control system can be seen in Figure 1.1. This study focusses on booking forecasting for airline revenue management. Although the seat optimization model and the overbooking model are also important elements of the seat allocation control system in airline revenue management, these will therefore not be part of the scope of this study. The revenue management forecasting and its challenges are explained in the following sections.

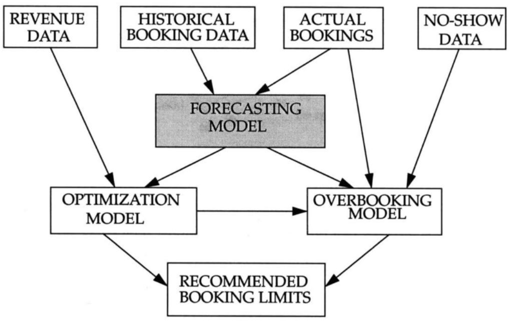


Figure 1.1: An automated booking limit system (Wickham, 1995)

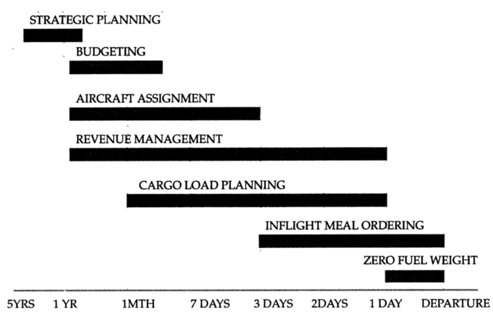


Figure 1.2: The applications of forecasting in the airline industry relative to the departure of the flight (Wickham, 1995)

1.1.1. Forecasting for Revenue Management

Booking forecasting for revenue management is one of the multiple ways in which forecasting is critical for airline management. Producing accurate forecasts for revenue management is a challenging task and can be considered more difficult than other forecasts critical for airlines. Examples are forecasts of airline demand in the short-term to facilitate tactical decisions such as load planning, catering and aircraft scheduling and in the long-term to determine strategic decisions such as route planning or the acquisition of new aircraft as can be seen in Figure 1.2.

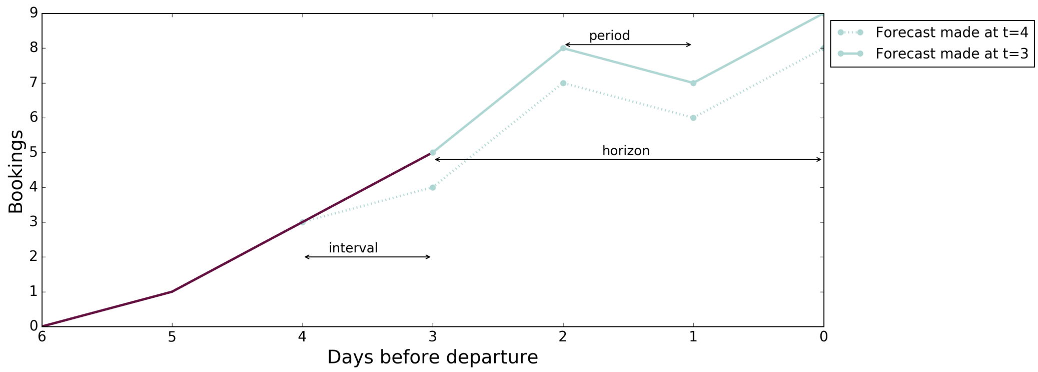


Figure 1.3: An example of a booking curve with a forecast and revised forecast indicating the forecast horizon, period and interval

In Figure 1.2 it can be seen that forecasting for revenue management spans a period of about a year. This makes it challenging to compute accurate predictions over the entire forecasting period. Another characteristic that makes forecasting for revenue management a challenging task is that instead of one, which is normally the case in forecasting problems[1], it involves two time variables. These two variables are the time the booking is made and the time the flight departs. In the discussion of a revenue management forecast, the forecast is mostly denoted by the departure date of the flight and the number of days before departure the forecast is created (Lee, 1990).

The timespan from the forecast computation day to the day of departure is called the *forecasting horizon*. Every forecast also has a period. The *forecasting period* is part of the forecasting horizon and is the unit of time over which the forecast is produced. Finally, every forecast has a *forecast interval*. This is the time after which the forecast is revised. In general the forecasting interval and period are equal and called the *booking interval*. This is displayed in Figure 1.3, where the forecast gives a prediction of the number of bookings for every day in the future. The initial forecast made 4 days before departure is revised one day later.

1.1.2. Complexity of bookings forecasting

The bookings process for airline reservations has a high complexity as a result of the interrelated processes that influence the demand. Each factor impacting the demand forecast presents a challenge to the forecaster. Firstly, the factors that influence the bookings an airline receives for its flights are explained. Secondly, approaches to deal with censored data, an important factor for revenue management, are explained.

Factors affecting demand

In order to forecast demand as accurately as possible, the goal is to create a model that is similar to reality. From this point of view it can be argued that the factors affecting the bookings numbers in real-life should be taken into account in the forecasting model. Guo et al. (2012) and McGill and Van Ryzin (1999) discussed a list of main factors that influence the bookings process and demand numbers which is shown below.

- Seasonality
  - Monthly
  - Yearly
  - Day-of-week
  - Time-of-day
- Special events
- Pricing actions
- Demand dependencies between fare classes
- Group bookings

[1] Traditional forecasting problems have only one time variable because there is no consumption date, i.e. based on the stock exchange rate of the last 10 years, the quote for the next 3 months can be predicted (Wickham, 1995)

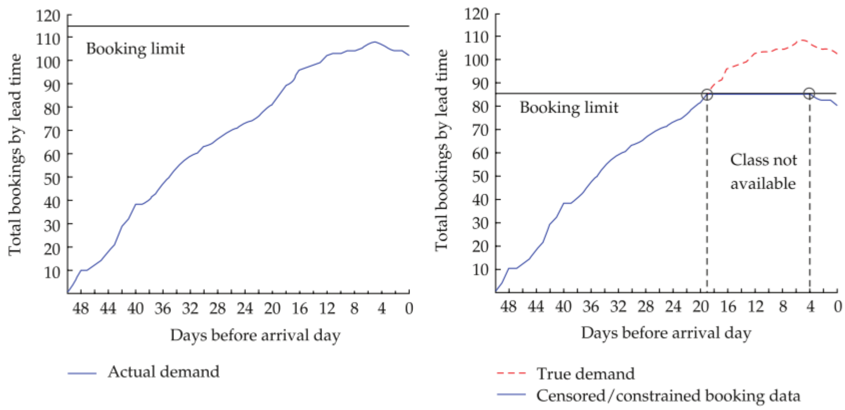


Figure 1.4: An example of an unconstrained booking curve (left) and a constrained curve (right) (Guo et al., 2012)

- Cancellations
- Censorship of historical demand data
- Defections from delayed flights
- No shows
- Recapture

Ideally, the forecasting model captures all these dynamic aspects, however, this will not be desirable because of the increasing complexity of the model, the increase in computational time and the amount of data needed to capture these aspects. Therefore, it is important to implement only factors that have a significant effect on the forecast.

Forecasting with censored data

The observed bookings data does in fact not have to represent the actual demand but the tickets sold per bookings class. The seats available can be constrained by bookings limits and other measures that an airline applies to fare classes which results in bookings data to be censored at a certain capacity level. The difference between a constrained and unconstrained booking curve is illustrated in Figure 1.4.

In the application of demand forecasting the models need unconstrained data to predict future demand. In the work of Skwarek (1996) it was concluded that the use of unconstrained data can result in a revenue gain of 3.5%. L. R. Weatherford and Pölt (2002) analyzed the impact of the unconstraining process in the revenue management process of a major US airline and found that a 2-12% revenue gain can be achieved. According to Guo et al. (2012) the unconstrained demand data is usually obtained in two steps. First, the parameters of unconstrained demand curves are derived based on historical bookings data that was not censored. Secondly, these parameters are applied when predicting future demand to end up with unconstrained demand curves.

Different approaches to obtain an unconstrained dataset and an unconstrained forecast exist. Ideally, one would observe the rejected bookings and would determine the uncensored data accordingly. Nevertheless, this is in most cases impracticable for an airline. Both naive methods where censored data is removed or replaced by uncensored data as well as sophisticated statistical methods have been proposed to forecast unconstrained demand (L. R. Weatherford & Pölt, 2002).

1.1.3. Revenue Management practice at Kenya Airways

This study is conducted in collaboration with Kenya Airways. Kenya Airways (KQ), the pride of Africa, is a leading African airline and the flag carrier of Kenya. KQ is operating a fleet of 30 aircraft consisting of Embraer 190, Boeing 737-800 and Boeing 787-8. It serves flights to numerous destinations across Africa, Europe, the Middle East, the Far East and Asia operating a hub and spoke network with Jomo Kenyatta International Airport (JKIA) in Nairobi as its main hub. In 2007 Kenya Airways became a full member of the SkyTeam alliance (SkyTeam, 2017).

Within the revenue management department of the airline, bookings forecasts are mostly used by flight demand analysts and commercial analysts. The flight demand analysts determine whether the seat allocation and pricing should be adapted for specific flights. Commercial analysts aim to specify the budgets for the airline by looking at the forecast of bookings for specific flights and on a higher level for an entire route or market. For the determination of these budgets also long-term forecasts (of 6 months and longer) are used and the findings shared with the network planning department.

The research context is centered around Kenya Airways by using data of the airline to construct, test and validate the forecasting model. Nevertheless, the aim of the research is to provide a new approach to forecasting demand for revenue management which can be used by other players in the airline industry and potentially other industries that are involved with revenue management such as the hotel industry.

1.2. Importance of bookings forecasts for Revenue Management

The booking forecasting model plays a central role in seat inventory control and therefore in the revenue management process. In theory a 100% accurate forecast will result in a maximum revenue figure and inaccurate forecasts will result in a deviation from this maximum achievable revenue and this deviation will become larger for increasing inaccuracies. Studies on the impact of forecasting accuracy on passenger revenues have demonstrated this by showing that a 10% improvement in forecasting accuracy can lead to a 0.5% to 3% increase in annual passenger revenue (Lee, 1990). For a major US airline with high demand flights this comes down to $10 to $60 million.

The quality of a revenue management forecast is therefore of outmost importance for an airline. However, forecasting for revenue management has proven to be a challenging task because of the dynamics and complexity of the bookings process which is influenced by many factors such as pricing discounts, special events or defections of passengers from delayed or cancelled flights. As a result, the prediction for the future demand expressed as a single value is mostly wrong. In order to have an accurate forecasting model it is therefore important not only to forecast future demand close to the actual realized values but also to determine the distribution of the bookings. This provides additional insight into the spread of the future bookings and, most importantly, enables the stochastic optimization of seat prices, seat allocation and booking limits, which, in turn, provides support for enhanced profitability.

This study will investigate the forecasting methods that have been applied both within and outside airline revenue management. From there a new approach is proposed in which the dynamic booking forecast model includes a probability distribution representing the uncertainty of the future value using the transitional probabilities in the matrix of a Markov Chain model. Variations of this model are considered and studied to find the best model configuration. In order to assess whether the accuracy of revenue management bookings forecasts are improved with the new model, the performance is compared with traditional forecasting methods. Furthermore, the performance of the probability distributions is analyzed for a variety of model configurations. The proposed forecasting method is a novel approach to bookings forecasting which has not been proposed within airline revenue management before.

The research is commissioned by and conducted in corporation with Kenya Airways and will therefore use the flights within the network of Kenya Airways as a case study. For these flights Kenya Airways aims to have a more accurate forecast of the bookings they can expect to have in their bookings system. This enables the airline to improve their seat allocation and budgeting within the revenue management department. As the current forecasts are volatile and yield forecasting accuracies that do not meet the requirements of Kenya Airways, the airline desires a model that is able to give a better representation of the future bookings situation of the flights in their network. The model is developed with a focus on forecasts on a cabin class (economy bookings) and on a flight leg level. This way it can be investigated whether the newly developed method is able to capture the trends that are present in the data. From there the research objective is established as follows:

"Increase the accuracy of bookings forecasts by contributing to the development of a dynamic forecasting model that includes the uncertainty in the forecast on a cabin class and flight level.""

1.3. Thesis roadmap

The remainder of this thesis is organized as follows. Firstly, Chapter 2 gives an overview and a critical review of the existing work on forecasting methods in the research field of revenue management with a focus on the airline industry. Secondly, Chapter 3 describes the research plan for this study including the research objective and the framework that is used as a guideline to achieve the research goal. Thirdly, the methodology of the research project including the model formulation, construction of the transition matrices and the measures used in the comparative study are addressed in Chapter 4. Subsequently, Chapter 5 elaborates on the experiment that is carried out for this research covering the experimental design, data selection and data analysis. Then, the results of the experiments consisting of three parts are explained in consecutive order in Chapter 6. After, the verification and validation of the proposed forecasting model is explained in Chapter 7. Finally, the conclusion of this study and the formulated recommendations are discussed in Chapter 8.

2
Literature study

The relevant literature for this study is addressed in this chapter. Firstly, the forecasting methods that have been applied in a revenue management context are discussed. The second part contains existing research that modeled a level of uncertainty in the demand forecast. The third part focusses on dynamic forecasting models. Finally, the chapter is ended with the main conclusions regarding the reviewed literature.

2.1. Revenue Management forecasting methods

In research, many different forecasting techniques and methods are used to address the problem of forecasting bookings within the airline industry. Lee (1990), Wickham (1995) and L. R. Weatherford and Kimes (2003) grouped the methods based on the dataset that is used to make the forecast. The same approach is applied in this study. The method can either make use of only historical data (complete booking curve data of departed flights), only the actual recent bookings at hand data or a combination of both to determine the future number of bookings. This is illustrated by the complete and incomplete bookings curves in Figure 2.1. Accordingly, three groups can be formed: historical bookings methods, advanced bookings methods and combined methods.

In this chapter, first the historical bookings forecasting methods that have been proposed for airline revenue management are discussed. This also includes methods introduced in the hotel industry of which the bookings process has similar characteristics as the airline bookings process. Secondly, the advanced bookings methods are explained and thereafter the combined bookings methods are discussed. Finally, an overview is given of the studies that compared multiple forecasting methods of which the results are summarized in Table 2.1.

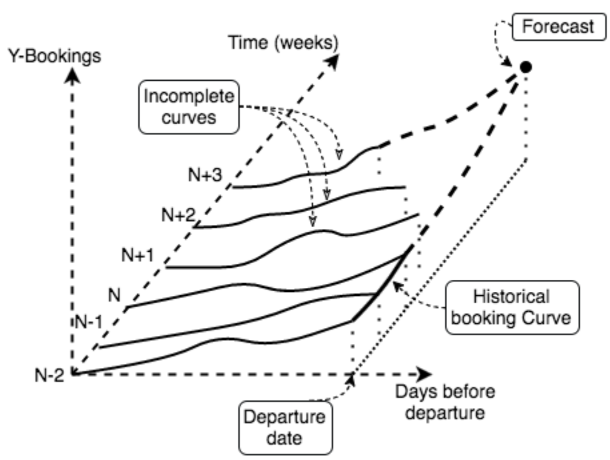


Figure 2.1: Historical and advanced bookings curves

2.1.1. Historical booking methods

The historical bookings methods only consider booking data of flights that already departed to forecast bookings for the flight of interest. The dataset therefore consists of complete bookings curves. This section covers the forecasting methods that have been introduced for the airline industry. Firstly, the first simple methods that have been introduced from the 70s until the early 80s are discussed. Thereafter, the more sophisticated methods that have been introduced from the late 80s are explained. Then, some interesting historical bookings methods that have been addressed for the hotel industry are presented. Furthermore, stochastic models make use of historical data and are often applied in airline revenue management practice. These are discussed in the final part of this section.

Methods introduced until early 80s

In the 70s the first work has been published on the characteristics of airline bookings and forecasting on a flight level for flights on specific dates. These papers focused on forecasting the total numbers of bookings at the day of departure. Littlewood (1972) proposed a simple forecasting model for flight bookings while working with British Overseas Airways Corporation (BOAC)/British Airways. This forecast is basically calculated by computing the average of the historical bookings as displayed in Equation 2.1.

$$\hat{X}_0 = \frac{1}{n}\sum_{f-1}^{f-n} X_{0,f} \tag{2.1}$$

where $\hat{X}_0$ represents the forecasted bookings at the day before departure, n the number of flights in the historical dataset and $X_{0,f}$ the final bookings for the flight in the dataset departing at day f.

Littlewood (1972) used the forecast of the number of bookings to determine the optimal bookings limits based on the probability distribution of the final number of bookings. This seat allocation control is nowadays well known as *Littlewood's Rule*.

While working at Scandinavian Airline System (SAS) Duncanson (1974) introduced a forecasting model that incorporated exponential smoothing for a forecasting horizon between 1 and 3 months. Simple exponential smoothing techniques remained typical applications for the airline industry in the 1980's and 1990's (L. Weatherford, 2016). Scandinavian Airline System (1978) primarily extends the work of Duncanson (1974) by addressing the quantity of historical datasets that is required to compute accurate forecasts. Furthermore, the work addressed the issue of outliers corresponding to non-recurrent events such as the Gulf war.

Some decision making areas within an airline where the passenger forecast can be used have been reviewed by Adams and Vodicka (1987) while working at Qantas Airways. Simple arithmetic means and subjective estimates from marketing experts were the methods that have been incorporated in the model. The performance of these methods has, however, not been addressed.

Methods introduced from the late 80s

From the 80s the subject of bookings forecasting has been of larger interest for the academe which resulted in the introduction of less simple approaches and multiple papers have been published on forecasting for airline revenue management. Sa (1987) performed forecasts using the Autoregressive Integrated Moving Average (ARIMA) time-series and regression models. The ARIMA models yielded discouraging results which led the author to abandon the analysis on time series and continue with the regression model which will be discussed in more detail in 2.1.2. The disappointing performance of the sophisticated ARIMA model is in line with the conclusion of the first Makridakis forecasting competition that more complex models in general do not outperform simple methods (Makridakis et al., 1982).

In L. R. Weatherford, Gentry, and Wilamowski (2003) a neural network approach has been proposed to forecast bookings for airline revenue management. The neural network technique is able to recognize patterns by learning from input data which is called *training*. This study used a single-layer feedforward network and a functional link network to determine the ability of neural networks to forecast airline bookings.

In Belobaba and Hopperstad (2004) a new forecasting method called Q-forecasting was proposed for airline revenue management. The method aims to forecast the demand in the lowest class (Q-class) only. Then it estimates the willingness-to-pay of passengers and closes lower fare classes to force passengers to sell-up to higher classes. It showed to be an effective technique for forecasting in a restriction-free fare structure. The downsides are that the passenger willingness-to-pay is hard to determine in most cases and that very rarely a fare structure is totally unrestricted which makes Q-forecasting more suitable to forecast a price-oriented demand.

From the early 00's the development of new revenue management forecasting methods for the airline industry stagnated. Historical bookings models are, however, studied in the hotel industry to forecast the number of reservations.

Methods introduced for the hotel industry The hotel industry shares some characteristics similar to the airline industry that are relevant for revenue management practice. Both these industries have a limited capacity, have a comparable forecasting horizon of in general multiple months and both have a bookings and "consumption" date. (Schwartz & Hiemstra, 1997).

L. R. Weatherford and Kimes (2003) considered the moving average method in a comparison with 6 other forecasting methods for the hotel industry. This comparison is discussed into more detail in Section 2.1.4. No reason has been given why the moving average is used in the comparative study.

A forecasting method for the hotel industry where the number of reservations is determined using a Monte Carlo simulation approach is proposed in Zakhary et al. (2011). Using the sub-components, which are based on the random variable distributions, the reservation process is simulated forward in time and the Monte-Carlo paths result in a forecast density. The research argues that an advantage of the method is that it results in a density forecast which is desirable to have rather than a point forecast because of the probabilistic nature of dynamic revenue management formulations. The method proposed in Zakhary et al. (2011) has not been applied or discussed in an airline revenue management perspective nor did it discuss the density distributions that result from the method in detail.

Pesch and Kovalyov (2015) studied an approach to combine existing forecasting methods with revenue optimization methods for the hotel industry. Three historical forecasting methods have been modified and considered in the model; double exponential smoothing, moving average and "same day last year" method. The study did not address the relative performance of the methods nor did it explain what were the pros and cons of the modified forecasting methods.

Stochastic models for Revenue Management

For most of the overbooking and seat allocation problems within airline revenue management, a generator for demand data is needed. In general it is assumed that the bookings process is a stochastic process. The stochastic models that have been used in research to model the bookings process are discussed in this section. Again, both the models that have been introduced for airline revenue management as for other sectors are discussed.

In Alstrup, Boas, Madsen, and Vidal (1986) an overbooking model for an airline is proposed for a flight with two types of passengers. The bookings process is modeled as a Markovian non-homogeneous sequential decision process. The Markovian property of this bookings system results in a system that changes from period to period according to transition probabilities. In the model it is assumed that the number of bookings requests follow a non-homogeneous Poisson arrival process. The study did not address the accuracy of the stochastic model but only discussed the performance optimization regarding overbooking.

The probabilistic model introduced in Lee (1990) models the airline bookings process as a stochastic process where the requests, reservations and cancellations are interspersed in the period before departure. This can be seen as a "birth-death process" of travelers. This results in a Poisson model of the bookings arrivals process and incorporates censoring of the model. The basis of this model is the non-homogeneous Markovian model developed in Rothstein (1971) that was used to determine optimal overbooking levels.

The work of Lee (1990) has been followed up by Svrcek (1991) in which an attempt has been made to characterise the stochastic properties of group bookings in the airline industry. The mathematical model that resulted from this research included a distribution of the group bookings on a flight. The study focussed on group bookings and did not discuss the performance of the model.

In Sulima (2012) a probabilistic model of overbooking is proposed for an airline. A Markov chain with discrete finite time has been used as the mathematical formulation of the air ticket sales. From this approach a distribution of the number of sold tickets at departure is achieved which makes it possible to determine the optimal overbooking limit. Again, a non-homogeneous Poisson process has been used to model the ticket requests. The number of no-show passengers is then taken into account to determine the passengers that could not take the flight due to overselling. The performance of the model has not been discussed in detail.

Application in Revenue Management of other industries In Crystal Queenan, Ferguson, Higbie, and Kapoor (2007) multiple unconstraining methods have been compared for the hotel industry. This has been done by creating sample bookings curves that have been used as an input to the unconstraining methods. These concave, convex and homogeneous curves are created by randomly generating arrival rates of a Poisson distribution for every day. Different Poisson arrival rates lead to different shapes of the booking curve.

In Weinberg, Brown, and Stroud (2007) the number of incoming calls for a call centre are forecasted. A Markov Chain Monte Carlo method has been used to forecast the call arrivals. In this study the arrival rates of calls are modeled as a non-homogeneous process. The benefit of this approach that is addressed in the paper is that is provides a measure of uncertainty of the parameter estimates and experience of the manager can be incorporated in the model.

In Haensel and Koole (2010) unconstrained demand data is estimated using an implementation of customer choice sets. This study focusses both on hotel and airline bookings and the emphasis is put on the estimation of passenger choice set parameters based on sales observations. The bookings process is modeled as a non-homogenous Poisson process where the arrival rate is assumed constant over the different intervals. This approach is similar to the approach in Lee (1990), however in Haensel and Koole (2010) the intervals are defined in such a way that at most one arrival or booking can be observed in the interval.

2.1.2. Advanced booking methods

Advanced bookings methods use the current bookings of a flight to determine the unknown number of bookings to come for a flight. It therefore considers incomplete bookings curves in order to compute the forecast. In the 80s and 90s this method has become more popular since researchers believed that the best information was in the current bookings for similar recent flights. Firstly, the methods that have been introduced for airline revenue management are discussed and then the methods introduced for the hotel industry.

A simple advanced forecasting method that applied regression to two sets of data at Alitalia was developed by Harris and Marucci (1983). The forecast aimed to predict the traffic on the routes of Alitalia in the short term. The flights have been separated on day of departure, country, aircraft type and continental/intercontinental. All of these factors, except for the day of departure, showed to have significant impact on the regression parameters of the model. Another finding was that the long-haul flights gave more reliable results than the domestic flights.

L'Heureux (1986) developed a "new twist" in forecasting short-term passenger bookings by proposing an alternative to the classical pick-up method called the advanced pick-up method. A Pick-Up (PU) method calculates the "pick-up" of bookings which is the expected number of additional bookings in an interval. The final bookings is then determined using Equation 2.2,

$$\hat{X}_0 = X_t + U_{t,0}, \tag{2.2}$$

where $\hat{X}_0$ represents the forecasted bookings at the day of departure, X_t the most recent known number of bookings for the flight and $U_{t,0}$ the bookings expected to be picked-up from the current point in time to the day of departure.

Whereas the classical pick-up only uses datasets of flights that already departed, the advanced pick-up method also uses the data of flights that still need to depart by reducing the period for which the "pick-up" is determined. According to *L'Heureux* this fulfills the maxims of forecasting which is to use all the data available and to give most weight to the most recent data. The approach in determining the pick-up value with the advanced additive pick-up method is mathematically defined in Equation 2.3.

$$U_{t,t-1} = \frac{1}{n}\sum_{f-1}^{f-n} U^f_{t,t-1}, \tag{2.3}$$

where n denotes the total number of flights in the bookings dataset for that time interval and f represents a flight in this dataset.

Sa (1987) further analyzed regression forecasting after the time series forecasts did not show promising results. The multiple linear regression model of the airline bookings process used bookings to come as the dependent variable and the bookings at hand, day-of-week index, seasonality index and the average of bookings to come in historical datasets as the explanatory variables. This model gave better results than the time series model, however, the forecasting ability and the accuracy of the models have not been tested.

Lee (1990) suggested two different advanced bookings methods: the synthetic booking curve model and the time-series of advanced bookings model. The first one aims to describe the shape of the booking curve. Lee states that the censored Poisson model can be viewed as a synthetic booking curve with a Poisson distributed dependent variable. The time-series of advanced bookings model creates a time series of total booking at earlier days before departure to determine the total bookings at time t. The model computes the missing bookings numbers until the day of departure in a sequential manner. A difficulty with both these models is that it also requires a sequential forecast of the exogenous variables.

Wickham (1995) proposed a more simple approach and used a simple linear regression model to forecast airline bookings in which the dependent variable was the number of bookings at the day of departure and the current bookings at hand the independent variable. This model can be seen as a non-causal regression model where external factors are not considered.

A similar non-causal regression model has been used in Skwarek (1996). In this work the forecasting method is combined with different detruncation and sell-up models in order to analyze the impact on the revenues of an airline.

A similar model as in Wickham (1995) has also been used in Zickus (1998). However, the focus in this work was on the impacts and revenue effects of different forecasting methods and detruncation approaches in an airline revenue management context instead of exploring the advantages, disadvantages and prediction accuracies of the forecasting methods.

Usman (2003) continued upon the work of Zickus (1998) and measures forecasting accuracy under different forecasting and detruncation combinations in an airline network simulated with a Passenger Origin-Destination Simulator (PODS). Furthermore, different seat allocation optimizations are used to analyze the performance under different conditions. The goal of this research was to investigate the relationship between revenue performance of forecasters and their forecast error. It was concluded that the combination of a regression method and projection detruncation yields a good revenue performance under various network scenarios.

Hotel industry

An innovative approach to forecast daily hotel bookings has been proposed in Schwartz and Hiemstra (1997). This method compares the shape of the bookings curve of the current day with curves of previous days and computes the forecast based on the most similar curve. A downside of the curve similarity approach that is mentioned is that this method requires a large dataset to achieve reliable forecasts.

In Tsai (2014) a modification of the curve similarity approach of Schwartz and Hiemstra (1997) was introduced that takes the influence of temporal features into account. The model is applied to railway arrivals and equipped with a learning program to deal with the fluctuations in the bookings curve. The similarity evaluation module in the model takes day-of-week variations into account.

L. R. Weatherford and Kimes (2003) compared 7 different forecasting methods for the hotel industry. A variation to the simple linear regression model of Wickham (1995) has been proposed called the *Logarithmic linear regression method*. The difference with the simple linear regression method is that it takes the logarithmic value of the independent variable bookings-at-hand to determine the final number of bookings.

In Zakhary et al. (2008) a new framework is presented for the pick-up forecasting method with 8 different variations. These 8 variations of the pick-up method are compared in the application of predicting hotel reservations. The variations that are displayed in Figure 2.2 are grouped according to the following criteria:

- Additive or Multiplicative: the additive pick-up variation calculates the average pick-up of bookings until the final bookings independently from the number of bookings already at hand. Whereas the multiplicative method assumes that the bookings to come are proportional to the current bookings. To calculate the final number of bookings the current bookings are multiplied by an average pick-up ratio in the multiplicative approach.

- Classical or Advanced: the classical pick-up method uses only data of flights that have a complete booking curve in the forecasting process. On the other hand, the advanced method, as proposed by L'Heureux (1986), uses both the complete and incomplete bookings data of flights.

- Simple or Weighted Average: the simple average sets equal weights to the observations when computing the average pick-up and the weighted average gives a higher weight to more recent observations. In this work exponential averaging has been used for the weighted average variation.

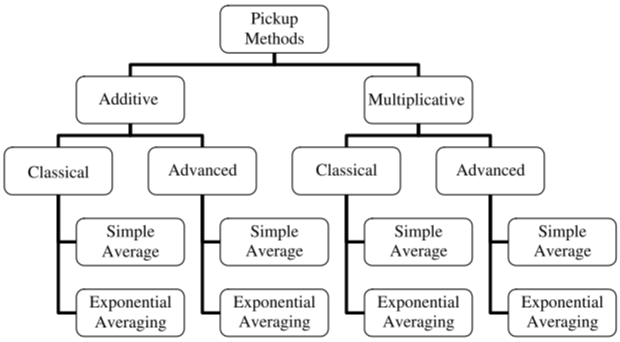


Figure 2.2: The variation of pick-up methods compared in Zakhary et al. (2008)

2.1.3. Combined booking methods

The combined booking methods use both the historical datasets and the incomplete advanced bookings data to compute the forecast. Firstly, the methods that have been introduced in studies on bookings forecasting for airline revenue management are discussed. Then, the methods that have been proposed for the hotel industry are explained.

Ben-Akiva (1987) proposed a model which was a combination of a non-causal regression model and a time-series model. The simple linear regression model used advanced bookings data and the ARIMA time-series model used historical bookings data.

The work of Ben-Akiva (1987) has been used as a starting point in Lee (1990). *Lee* proposed two different combined forecasting methods: a weighted average of an historical and an advanced bookings model and a *full-information model*. The proposed weighted average method uses a simple weighted average of forecasted values resulting from the historical bookings and from the advanced bookings model. Both the *time-series of advance bookings* model and the synthetic booking curve model have been considered as an advanced bookings model. In the *full-information model* the bookings process is viewed as a time-series of historical bookings where each element in the time-series is seen as a result of a bookings curve. The advantage is that no missing data of datapoints in between time t and the day of departure is needed to produce the forecast. However, a forecast of the exogenous factors is needed for all the intermediate points in time.

Hopperstad (1991) proposed an alternative to the *full-information combined* model called the *efficient forecaster*, which does not use the computationally intensive Maximum Likelihood Estimation (MLE) method in the model. Instead of using the MLE method *Hopperstad* incorporates the proportional relationship between the bookings at hand at two intervals in the bookings process in the model. It is argued that this proportional relationship is stable regardless of the demand level. The method has been taken into account in comparative studies by other researches which will be discussed in Section 2.1.4.

Hotel industry

A combined weighted average forecast is applied to determine unconstrained hotel reservations in Rajopadhye, Ben Ghalia, Wang, Baker, and Eister (2001). In this work a long-term forecast (one year ahead) is combined with a short-term forecast (60 days ahead) using weighting factors. The long-term forecast is computed using the Holt-Winters Method and uses only historical data. This method considers both the linear trends as well as cyclical trends in the data. The short-term forecast used a booking profile approach to arrive at an estimation of the final number of reservations. From historical data the fraction of reservations already made can be calculated and consequently the unconstrained final demand. This approach is similar to the multiplicative pick-up method. It is not addressed what the relative performance of this method is compared with traditional methods. It is suggested in this work that expert knowledge should be incorporated into the model in order to achieve more accurate results.

2.1.4. Comparative studies on bookings forecasting methods

Alongside the development of airline bookings forecasting methods during the 80s and 90s researchers started to compare the performance of the different forecasting methods. The comparative studies that have been conducted for the airline industry are discussed first in this section. Secondly, the papers that studied the relative performance of forecasting methods in other industries are explained.

Sa (1987) concluded that the linear regression model performed better than the ARIMA time series model. The forecasting ability has, however, not been tested and exact numbers of the relative performance of these methods have not been given.

Ben-Akiva (1987) compared a regression model, time-series model a and model that combined both these forecasts. The results showed that the combined model performed better than the advance booking and historical bookings methods. Furthermore, the regression model fit the data better than the ARIMA model.

Lee (1990) compared his newly developed censored Poisson model and the full-information models with regression and a moving average model. The order of most accurate forecasting models that followed from the research was as follows:

1. Full-information model
2. Censored Poisson model
3. Regression models
4. 8-week moving average

It would be better to compare the *full-information model* with a small number of intervals and the traditional methods with a higher number of intervals, to account for the higher computational time of the *full-information model* in the trade-off.

Wickham (1995) performed an evaluation of the performance of the forecasting techniques that were at that time widely used in the airline industry. These techniques were simple average, weighted average, linear regression and (classical and advanced) pick-up models. The conclusion of the study was that the pick-up model performed better than the regression and time series models. Amongst the different pick-up methods the advanced pick-up method showed the best performance.

Skwarek (1996) argued to compare forecasting methods in terms of revenue performance instead of forecasting error. Using a Passenger Origin Destination Simulator (PODS) the classical pick-up method was then compared with two forecasting methods: a regression (non-causal) method and the "efficient forecaster". From the first comparison it resulted that in the base case neither the regression nor the pickup forecast dominated. Under an increasing booking curve variability the performance of the regression method relative to the pickup method decreased and the reverse is the case for an increase in final demand variability. The latter comparison showed that pickup forecasting in general outperformed the efficient forecaster.

L. Weatherford (1998) focussed on the comparison between an additive pick-up, a multiplicative pick-up similar to the synthetic booking curve and a linear regression model for airline bookings forecasting. The set-up of the comparative study is comparable to the study of Wickham (1995). The multiplicative method performed significantly worse than the additive and regression models. There was not a large difference in the performance between the additive and regression model.

In L. R. Weatherford et al. (2003) a neural network (NN) approach has been proposed to forecast bookings for airline revenue management. The performance of a single-layer feedforward network and a functional link network are compared with three traditional forecasting methods: moving average, exponential smoothing and the regression method. The single-layer feedforward network performed best of the two neural networks. For a short-term forecast the single-layer multiple performed better than the traditional forecasting methods. For the long-term forecast the cubic regression and single-layer multiple-input neural network performed similarly and better than the other methods. The paper also discussed that the neural network method requires more data storage and computational time. Furthermore, the method results in a point forecast, it has only been tested for a forecasting horizon of maximum 3 weeks and only the final number of bookings has been of interest in this paper. Until this point the neural network approach has not been proposed in other papers that concern bookings forecasting for airline revenue management. The reason might be found in Allende, Moraga, and Salas (2002) in which statistical analysis is compared with artificial neural networks (ANN). It is concluded that the difficulties when using ANN are for example the determination of outliers, the selection of training data and establishing confidence intervals for the forecast.

Stikvoort and Van der Zwan (2010) compared five different forecasting methods in an airline revenue management perspective at *Transavia*. The first method is the weighted mean of final bookings (Littlewood, 1972). The second method is a simple linear regression model similar to the model used in Wickham (1995). The third method is the weighted pick-up similar to the method that was introduced by L'Heureux (1986) including weightings for the pick-up of different flights. Then the multiplicative pick-up method, similar to the synthetic bookings curve (Lee, 1990) is considered. The last method that is compared is the Q-forecasting method (Belobaba & Hopperstad, 2004). The weighted pick-up method showed the best performance over time in comparison with the weighted mean of final bookings, linear regression, the multiplicative pickup and the Q-forecasting method. Forecasting accuracy, calculation time, implementation and application has been used as criteria to determine the performance. It is argued that the Q-forecaster cannot be used as a stand-alone forecaster as it can only be used when an optimizing step is used. Without this step it cannot be used by the revenue controller.

Comparisons in other sectors

Schwartz and Hiemstra (1997) compared the proposed curve similarity approach with the stepwise autoregression model (AR), Higher Order Polynomial Model (HOPM) and a combined AR and HOPM model. It was found that the curve similarity approach outperformed the other three methods. It was also concluded that the combined AR and HOPM model performed better than these two models separately.

In L. R. Weatherford and Kimes (2003) state-of-the-art airline forecasting methods have been compared for the hotel industry. The research can be divided into two stages: a comparison of advanced bookings methods and a comparison of seven different forecasting methods. For the first stage the advanced pick-up and regression method performed similar and yielded a better forecast than the multiplicative methods on all datasets. In the second stage the following methods have been compared:

- Simple exponential smoothing with α ranging from 0.05 to 0.95
- Moving average method with period between 2 and 8
- Linear Regression
- Logarithmic Linear Regression
- Additive Pick-up Method
- Multiplicative Pick-up Method
- Holt's Double Exponential smoothing with α and β values ranging from 0.05 to 0.95

The main conclusions were that the simple exponential smoothing approach is overall the most robust method. This finding is in line with the conclusion of the forecasting competition held in 1982 where it was concluded that the moving average and exponential smoothing methods are the most robust (Makridakis et al., 1982) . Finally, L. R. Weatherford and Kimes (2003) advises to use either one of these most robust methods: exponential smoothing, pick-up, moving average and linear regression. Another suggestion mentioned is to generate a forecast that combines these into one model, however, it has not been addressed how this could be achieved.

Zakhary et al. (2008) compared 8 different variations of the pick-up method for the hotel industry. The forecasts were computed for four different forecasting horizons with minimum 7 and maximum 60 days. A hotel reservation data simulator was built to create three datasets with which the experiment was performed and the error was measured in MAE, MAPE, MSE and RMSE. The study showed that the multiplicative and additive pick-up variations have a comparable performance, although the additive variation was the most robust variation. The classical variations clearly outperformed the advanced pick-up variations and the exponential smoothing variations showed a comparable result as the simple average variation. It is stated that overall the multiplicative, classical, exponential smoothing variation has been identified as the best technique. However, a critical review of the results shows that the difference with the additive, classical, exponential smoothing variations is minimal.

Zakhary et al. (2011) compared his Monte Carlo simulation forecasting approach with the double exponential smoothing method and various pick-up methods when forecasting hotel reservations. The various pick-up methods included classical pick-up methods either using a moving average or exponential smoothing, advanced pick-up methods as introduced by L'Heureux (1986) and a multiplicative pick-up method which is similar to the synthetic booking curve model introduced in Lee (1990). In both these scenarios the proposed Monte Carlo approach slightly outperformed the classical pick-up method with simple moving average and the advanced pick-up method with simple moving average. The other methods performed much worse than these three. The multiplicative pick-up method clearly performed worst.

Dutta and Pachisia (2014) compared different forecasting methods for bookings in the railway industry. The forecasting methods have been categorized into two sub-groups: time-series and revenue management techniques. Single exponential smoothing, moving average and the ARIMA method have been selected as time-series methods. The additive, incremental and the advanced multiplicative pick-up method have been selected as the revenue management techniques. The incremental pick-up method is a variation on the multiplicative pick-up method where the proportion of bookings to come is determined until the day of departure instead of a relative proportion of additional bookings for every interval. The ARIMA method was the best performing time-series method and the additive pick-up method the best performing revenue management forecasting technique. These two methods were then compared with the linear regression method. It was found that the ARIMA forecasting method performed well in long-term forecasting and the pick-up and regression methods performed better in the short-term. From this result a weighted average of a revenue management and a time-series method was proposed which yielded the best forecasting performance over the entire bookings process. A downside of this comparison is that the forecast are made with a maximum forecasting horizon of 3 weeks and it does not consider group bookings or unconstraining of data.

Tsai (2014) compared the proposed self-learning curve similarity model with a linear regression and an additive pick-up method in the application of forecasting for railway arrivals. It was concluded that on average the proposed model can give results with 11% improvement compared to the benchmarks. Furthermore, an error measurement has been proposed that assigns higher weights to a errors in the beginning of the bookings period rather than closer to departure.

2.1.5. State of the art of Revenue Management forecasting methods

Following from the above discussed forecasting methods and comparisons of their relative performance which is summarized in Table 2.1, it can be concluded that the literature does not completely agree on what is the most preferred method for airline bookings forecasting. The last couple of years the research on new forecasting methods or enhancements of traditional methods in order to improve the forecasting accuracy has been limited. The pick-up method has been discussed in many papers and often compared with other forecasting methods, both for the airline and the hotel industry. In most of these comparisons an additive version of the pick-up method performs as least as one of the most accurate methods. An opportunity has been identified to develop an enhancement of the pick-up method in order to increase the information included in the forecast by giving more than only a point forecast. The forecasting methods mentioned above determine the expected number of bookings at various future time moments. However, the distribution of the bookings is not determined. This provides additional insight into the spread of the future bookings and, most importantly, enables the stochastic optimization of seat prices, seat allocation and booking limits, which, in turn, provides support for enhanced profitability. Another opportunity has been identified regarding the use of neural networks to forecast airline bookings. The performance of the method seems promising, however the method has only been analyzed in a basic neural network model for a short forecasting horizon of maximum three weeks. Furthermore, it would be an opportunity to introduce a new method for the airline industry that already showed a good performance in another industry, such as the curve similarity method.

2.2. Modeling demand uncertainty

Presenting a level of uncertainty in the forecast increases the amount of information that can be deduced from the forecast and moreover most overbooking and seat allocation models need a level of uncertainty to come as an input to their models. The level of uncertainty in demand forecasts can be expressed using demand distributions. These are firstly discussed in this chapter including a discussion of density forecasts and discrete demand distributions. Then, an alternative to include uncertainty in demand forecasts is explained which is the method of interval forecasting.

2.2.1. Demand distributions

Normally, for modeling considerations, such as the implementation of a forecast in a revenue management control system, a certain continuous distribution is fitted on the data or on the forecasted values. Firstly, the distributions used for demand forecasting in the airline industry are discussed. Secondly, density forecasts where the shape of the distribution can be flexible are discussed. Finally, discrete probability functions that can be used to express the expected demand are explained.

Empirical studies on airline demand distributions

Three empirical studies have been performed to study what distribution fits the airline demand best. Belobaba (1985) analyzed and tested demand data from an airline and tried to determine patters in the demand

Table 2.1: Overview comparative studies of forecasting methods

		Forecasting Method																					
Sector	**Authors (year)**	Linear Regression (LR))		Weighted Mean Final Bookings	ARIMA	Moving Average	Exponential Smoothing		LR-ARIMA Combined	FIC	Classic Pick-up		Advanced Pick-up		Efficient Forecaster	Q-Forecast	Curve Similarity	AR	HOPM	AR-HOPM Combined	Monte-Carlo Simulation	Single-Layer Feedford Network (NN)	Qubic regression
		Simple	Logarithmic				Single	Double			Additive	Multiplicative	Additive	Multiplicative (Synthetic Booking Curve)									
Airline	Sa (1987)	+			-																		
	Ben-Akiva (1987)	-			-				+														
	Lee (1990)	-				–				++				+									
	Wickham (1995)	-				-					+		++										
	Skwarek (1996)	o									+				-								
	L. Weatherford (1998)	+									+							-					
	L. R. Weatherford et al. (2003)					-	+															++	-
	Stikvoort and Van der Zwan (2010)	-		+							++			o		–							
Hotel	Schwartz and Hiemstra (1997)																+	-	-	o			
	L. R. Weatherford and Kimes (2003) (1)	+											+	-									
	L. R. Weatherford and Kimes (2003) (2)	o	–			+	++	+			++			–									
	Zakhary et al. (2008)										+	+	-	-									
	Zakhary et al. (2011)							-			o	–	o								+		
Railway	Dutta and Pachisia (2014)				+				++		+	-		-									
	Tsai (2014)	-									-						+						

The entries represent the relative performance of the methods in the papers. "++" indicates the best performance, "+" indicates a good performance, "o" indicates a normal performance, "-" indicates a poor performance, and "–" indicates the worst performance.

distributions. The findings were that normally distributed booking data seemed to match demand data quite accurately for moderate levels of demand. However, for low demand the data showed a significant positive skewness in the distribution and a spike in the distribution for high demand where the distribution is affected by the booking limit.

Brummer (1988) conducted a second empirical study where the objective was to determine the mean and standard deviation of an unconstrained demand distribution of airline bookings data. A lognormal distribution has been proposed in this study instead of a normal distribution because it is argued that a natural skewness is present in the demand and the distribution is already truncated at zero. Furthermore, the study derives a likelihood function of a lognormal distribution censored as a result of booking limits.

Lee (1990) continued with the work of Belobaba (1985) and Brummer (1988) and discussed two additional distributions next to the normal and the lognormal distribution: the Poisson distribution and the Gamma distribution. The Poisson distribution followed from the censored Poisson process that was introduced in Lee (1990). *Lee* stated that the Poisson distribution can easily be interpreted in a probabilistic way, that the distribution is discrete and that the Poisson distribution cannot take on negative values. The downside of the Poisson distribution is that the standard deviation equals the square root of the mean, however the empirical study done in Lee (1990) shows that this property is quite reasonable for airline demand data. The Gamma distribution has been discussed following by the introduction of the distribution in Smith and Penn (1988). The distribution is flexible because it has two variables, however it is not naturally truncated at zero which makes it less suitable for airline demand data. In the end it is stated that there is no distribution that can be used in all cases and that the best fit depends on the sample of airline bookings data. However, a normal distribution is recommended for medium demand flights that show little effect of censoring and truncation.

Distributions assumed for airline demand

Other studies have made an assumption to use a certain distribution various number of papers on airline revenue management. Littlewood (1972) approximated the final number of bookings on a flight by a normal distribution. From this distribution the expected number of off-loaded passengers could be determined based on different overbooking levels. It has not been addressed to what extend the normal distribution was a good fit to the demand data.

Maddala (1983) discussed censored and truncated regression models. It is addressed how to estimate a censored or truncated model with normally distributed airline bookings data and briefly explains an extension to the lognormal and exponential distributions.

Schneider (1986) focussed on truncated and censored data samples from normal distributions and discussed how to estimate the true distribution when a censored dataset is given. The demand data is assumed to be normally distributed when estimating the censored and truncated models.

The normal distribution has been the most popular distribution to model the demand and the normal distribution assumption of airline demand is also made in Skwarek (1996) and Zickus (1998). In both these studies unconstraining methods are compared and tested and a normal distribution has been assumed for the constrained bookings data.

Coughlan (1999) considers an overbooking problem in a multi-class case and assumes a normal distribution for both the demand and the number of bookings and assumes that the distributions are independent for different fare classes. Data of *Aer Lingus* has been used in this study.

L. R. Weatherford and Pölt (2002) considers six different methods to unconstrain airline bookings data. In the simulation analysis it is assumed that the booking observations can be generated from a normal distribution.

In Swan (2002) the normal and gamma distribution are discussed and a model combining both the normal and gamma distributions has been proposed. The model is able to indicate when the gamma shape will be dominant and when the normal shape determines the shape. It is also stated that the gamma distribution is better for revenue management (low demand numbers) and the normal shape is better for spill modeling (high demand numbers).

Taking the characteristics of the above mentioned distributions into account the best probability distribution to use depends on the demand data. A goodness of fit test, e.g. the Chi-square test, could be used to analyze what distribution fits the data best (Larsen & Marx, 1981). In most cases it is assumed that the normal distribution is the most suitable distribution for airline demand. However, most of these studies did not address why the normal distributions was used and did not discuss whether the distribution fitted the data well.

Density Forecast

A density forecast is a forecast where the range of possible future demand values is expressed as a distribution with certain probabilities attached. It describes the uncertainty of the prediction with a distribution that can take every shape. The distributions that are discussed in the previous subsection can be regarded as specific forms of a density forecast Turyna and Hrdina (2009).

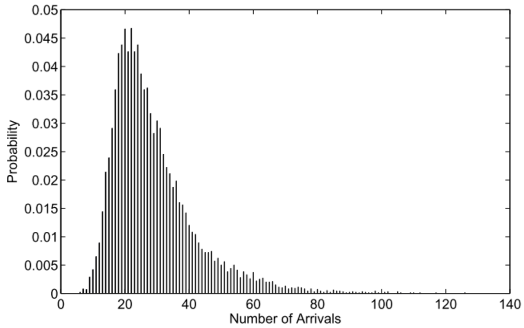


Figure 2.3: An example of a density forecast distribution for hotel arrivals (Zakhary et al., 2011)

As discussed in Section 2.1.1, a Monte Carlo simulation approach has been introduced in Zakhary et al. (2011) to forecast hotel room demand and occupation rates. These Monte Carlo paths yield forecast densities. It is stated that the density forecasts in this study show a right skewness and that information can be deduced from the density of the forecast which are useful and desired for revenue management professionals. An example that has been given in this study can be seen in Figure 2.3 that shows that there is a very low probability to have less than 7 reservations.

In Wan, Song, and Ko (2016) density forecasting was applied to predict future tourism demand. In this study density forecasts and traditional point forecasts are compared. It is concluded that the point forecasts can only give a general trend of a time-series. The density forecasts give more information and can easily be interpreted.

Discrete demand probabilities

Next to the continuous demand distributions discussed above, the demand probabilities can be expressed as a discrete function. The number of bookings can be determined separately or in groups of bookings which results in *probability buckets*. The probability function looks in that case like a histogram. Svrcek (1991) applied this approach when studying group passenger demand in the airline industry. It is argued that because requests are countable the distribution can be expressed as a probability mass function.

The advantage that the discrete and density distributions have is that no assumptions need to made of an underlying distribution. It is therefore quite straightforward to implement which makes it a good option for the airline revenue management application.

2.2.2. Interval forecasting

An interval forecast gives the probability that a future value will lie in between a certain range of values. The interval forecast consists of an upper and a lower limit which are the prediction limits and the interval is called the Prediction Interval (PI). Interval forecasts are closely related to density forecasts. In Wan et al. (2016) it is mentioned that interval forecasts can be extracted from a density forecast. The P.I.'s are computed by taking the value that corresponds to a chosen probability level. On the other hand, in Turyna and Hrdina (2009) it is discussed that an approximation of the density forecast can be created when the P.I. is constructed for several probabilities and consequently plotted with different colors or levels of shading.

The importance of interval forecasts is explained in Chatfield (1993). Several approaches to calculate P.I.'s are described. Some problems related to computing P.I's are also addressed. The most critical issue associated with P.I.'s is probably the tendency of P.I.'s to be too narrow which represents an overconfidence of the forecast. In other words, on average more than 5% of the future values fall outside the 95% Prediction Interval.

A model that uses Monte-Carlo simulations to generate interval forecasts for air passenger demand is presented in Schot (2015). In this study the effect of these intervals on the development of an airlines network is analyzed. Similar as what was mentioned in Chatfield (1993), the forecast intervals seemed to be too narrow.

In conclusion, most work performed in revenue management incorporating the forecast uncertainty assumes a certain demand distribution. Mostly, this approach is applied to final bookings only with the means to unconstrain the bookings demand. The normal distribution has been the most used distribution. A density forecast however is the most flexible demand distribution and includes interesting information the forecast. However, this method has not been found to be introduced in airline revenue management problems. The discrete distribution has the favorable characteristic for bookings forecasting that it considers only integer values. Alternatively, interval forecasts have often been discussed in papers however they yield less information about the uncertainty of bookings.

2.3. Dynamic forecasting

Updating a demand forecast over the bookings process is called dynamic forecasting. Based on new information that becomes available the forecast can be adjusted to increase the accuracy of the demand prediction. The first part of this chapter discusses some approaches to dynamic demand forecasting are addressed. Firstly, the methods that have been applied in the airline industry and secondly the methods that have been introduced in other industries.

2.3.1. Forecasting updating methods

Several updating approaches have been discussed in literature. However, little research has focussed on updating forecasts in the airline industry. The few work that has incorporated demand updating is discussed firstly. Secondly, the methods and applications of forecast updating in other industries is explained.

Methods in the airline industry

Most of the updating methods that are addressed in papers on airline revenue management update the expected number of bookings in order to update the bookings limits that result from the seat allocation models that are studied in these papers. One example is the study of Skwarek (1996) in which the forecast of the total demand on a flight is updated after every interval in order to determine the bookings limits for the next interval. It is also stated that for most U.S. airlines the number of intervals in which the forecast is updated normally lies in between 10 and 25. The forecast is updated by using an updated historical bookings database as an input to the model. Similar to the work of Skwarek (1996) the approach in Zickus (1998) and Usman (2003) used the forecast to determine the bookings limits and the accuracy improvement of forecast updating has not been discussed.

In L. R. Weatherford and Belobaba (2002) a similar approach has been used and the booking limits were updated at the beginning of each of the 15 bookings periods before departure. However, it does not use an updated historical bookings dataset but a simulation of bookings data as a Poisson process.

Markov chain method The Markov Chain method assumes that from one state all the distributions of the future states can be determined. A state transition probability matrix can be created by calculating conditional probabilities. The forecasted probabilities and the expected number of bookings can be updated based on the state at the new time step and by updating the transition probability matrix Kijima (1997).

A dynamic programming overbooking model has been used to simulate the bookings process in Alstrup et al. (1986). The airline booking process is modeled as a Markovian non-homogeneous sequential decision process. The model focussed on the overbooking level of a two class cabin. The study only discusses the performance of the overbooking model for different seat groupings that reduce the computational time. It does not address the accuracy of the replication of the booking process by the Markov process. The transition probabilities are non-homogeneous which means that they change over time.

In Hsu and Liu (2003) the Markov Chain method is combined with Grey topological forecasting to forecast the air traffic demand between two city-pairs with little available data. The method proved to predict the number of passengers accurately, however no comparison has been made with other forecasting methods.

A finite Markov Chain decision process model is developed in Goto, Lewis, and Puterman (2004) for airline meal provisioning. In this work the demand evolution is modeled as a Markov Chain and the optimal meal provisioning policy is determined using dynamic programming. The model is applied for the short-term from 36 hours before departure. It mentions that the Markov Chain method is suitable for forecast updating and it uses a combined parametric and direct estimation of passenger load changes to determine the transition probabilities.

The Markov Chain method has not been widely applied in revenue management practices of the airline or other industries to examine whether the method yield a good forecasting performance. However, Bobb and Veral (2008) identified that forecasting airline bookings using Markov Chains methods may be of value and that future research on these stochastic demand models is desired. The additional benefit of the Markov Chain method is that it incorporates the uncertainty of a prediction expressed as probabilities.

Methods applied in other industries

Most studies that considered forecasting updating have been conducted in other industries than the airline industry. These are discussed in this section. Finally the Martingale Method of Forecast Evolution is explained which is a common approach to forecast updating in the production-inventory analysis.

In Weinberg et al. (2007) a multiplicative Gaussian time series model has been proposed for the forecast of call centre arrivals. A bayesian Markov Chain Monte Carlo (MCMC) algorithm has been used to estimate and forecast the parameters. Within-day learning is used by means of a MCMC simulation in order to update the forecasts at different times of the day. A downside of the MCMC algorithm is that it is not straightforward to implement and it can take a long time before it converges.

The research of Weinberg et al. (2007) has been followed-up by Shen and Huang (2008) in which the objective is to dynamically forecast call centre arrivals by updating inter-day (day-to-day) forecasts using intraday (within-day) data. Both the inter-day and intraday dynamics have been incorporated into the model. Multiple approaches have been proposed for the intraday updating such as Historical Proportion (HP), Multiple Regression, HP in combination with Historical Average (HA) forecasting methods, Least Squares updating and Penalized Least Squares (PLS) updating which in the end performed best. The recommended Penalized Least Squares (PLS) technique has the characteristic that it gives a point forecast.

In Haensel (2012) it was argued that failing to update the forecast when information becomes available on the current Bookings Horizon (BH) could have a dramatic effect on the performance of the forecast, after it was concluded that there is a strong correlation between early and late bookings in the hotel industry. Therefore, an updating approach has been applied in this study to forecast hotel bookings for a horizon of four weeks. It compared the PLS and HP updating methods and compared the result with forecasts that are not updated. It was concluded that the PLS method gave better results in most scenarios both with low and high correlation between different parts of the bookings horizon.

Martingale Method of Forecast Evolution The Martingale Method of Forecast Evolution (MMFE) is based on the principle that a forecast of demand is maintained in all future periods of the demand forecast. It is mostly used in production-inventory analyses because it can model forecast updates from both time-series models and from information based on expert judgement. Initial forecasts are computed at the start of the first period. Then, the initial forecasts are revised with additional information that becomes available about future demand. The difference between the vector of new forecasts of demand and the forecast of the previous period is called the forecast update vector. The MMFE characterizes the sequence of update vectors and it can therefore be regarded as being a descriptive model.

Toktay and Wein (2001) models the sequence of forecast update vectors using the MMFE approach for a production inventory system. The assumptions that need to hold for a forecasting model when an MMFE approach is used are mentioned. The two most critical ones are that the forecast update vector should not be correlated and that demand should be stationary.

The MMFE used in Iida and Zipkin (2006) for a dynamic forecast-inventory model is a follow-up on Toktay and Wein (2001). Two types of forecast updates where considered: additive and multiplicative. The model eventually results in a dynamic program with a multidimensional state space. The results have shown that the forecasts are optimal with additive updates.

The MMFE is also applied to dynamically update the forecast of market demand in a news-vendor model for inventory control decisions in Wang, Atasu, and Kurtuluş (2012). A trade-off has been explored between improving the demand forecast and an increase in ordering costs. The impact of the MMFE on the dynamic forecasting performance has not been discussed in detail.

The requirement for using MMFE is that demand has to be stationary, which is most likely not the case for the airline bookings process. This could be the reason why the MMFE has not been applied to forecast airline bookings.

In conclusion little research has focused on updating demand to improve the forecasting accuracy in the airline industry. Most of the papers use forecast updating to be able to recalculate the optimal bookings limits. A common approach in the airline industry to update the prediction of airline bookings is to update the database with the most recent data and recalculate the forecast. Another method addressed in literature would be to apply the Markov Chain method that uses conditional probabilities to determine the updated forecasting distributions. This method has the advantage that it will result in a demand distribution which gives more information about the forecast.

2.4. Directions for future research

In this literature study some gaps are identified in the research field of bookings forecasting for airline revenue management. Firstly, no study on demand forecasting performed in the airline industry discussed the impact of dynamic booking forecasting on the accuracy of the prediction. Secondly, there are methods applied in other industries that could be studied for the application in airline revenue management e.g. neural network methods, curve similarity approach. Finally, the benefits of using a Markov Chain method for the purpose of airline bookings forecasting have not been studied. It has been found that the Markov Chain method enables a dynamic approach to forecasting with the additional benefit of yielding a probabilistic result.

2.5. Direction chosen for this study

This study focusses on the development of a forecasting model that is based on the Markov Chain method. From the literature review it can be concluded that the Markov Chain method possesses characteristics that make it suitable to be applied to dynamic booking forecasting. Furthermore, the result of the method includes a distribution which might be able to display the uncertainty of the future bookings situation accurately. As far as to the knowledge of the author this research is the first work that will analyze the ability of the Markov Chain method to be applied to dynamic booking forecasting with uncertainty for airline revenue management. The opportunities to apply the Markov Chain in this context will be explored to the fullest.

3
Research proposal

From the introduction and the literature study it can be concluded that is important for an airline to have a good forecasting model for revenue management and it has therefore been an important topic in research for many decades. Nevertheless, there are still some gaps in this research area, especially regarding the incorporation of the forecast uncertainty in the prediction.

In order to contribute to both the academic environment and to the revenue management practice of Kenya Airways, a sound research plan is needed. This research framework is discussed in this chapter. Firstly, the project goal and objectives for this study are mentioned. Then, the research framework and research questions are explained. Finally, the research strategy is addressed.

3.1. Research goal & objective

From the literature study it was concluded that no model has been introduced that combines forecasting uncertainty with a dynamic approach to revise and update a prior prediction. Previous models have either used methods that resulted in a point forecast or did not apply a dynamic approach to this problem. The goal is to have a model that is able to dynamically update predictions which include information of the amount of uncertainty around future values. This improves the usefulness of the forecast for the revenue management controller. This is because it is the downside of having a single point forecast that in most cases the forecast is wrong which may trigger the controller to make suboptimal decisions based on the given single value prediction. From there the research objective has been established:

"Increase the accuracy of bookings forecasts by contributing to the development of a dynamic forecasting model that includes the uncertainty in the forecast on a cabin class and flight level"

In order to measure whether the project objective has been achieved in this study two hypotheses are formulated as follows:

The forecasting model will provide forecasts with a higher accuracy than traditional methods.

The developed forecasting model can display the uncertainty in the future number of bookings.

The two hypotheses presented above are important to be answered. However, it is also important to determine to what extent both of the hypotheses are true by critically reviewing the results of multiple test cases.

3.2. Research framework and supporting research questions

In order to achieve the research goal, a research framework has been defined for this research. The research framework shows the steps that are needed in order to be able to achieve the research goal and to conclude if the hypotheses mentioned in the previous section are true. The steps for this research are based on the main research question that needs to be answered for this study which is as follows:

How does an approach combining forecasting uncertainty with a dynamic updating approach contribute to an improvement in the accuracy of airline bookings forecasts?

The main research question can be split up in sub-research questions that reflect the knowledge required to answer the main question and form a guideline towards achieving the research objective. The sub-research questions are formulated as follows:

1. What methods can be used to dynamically forecast uncertainty in bookings numbers?
   (a) What Revenue Management forecasting methods are there and what is their performance?
   (b) What updating approaches can be applied to a Revenue Management forecasting method?
   (c) What uncertainty modeling approaches can be combined with a dynamic updating forecasting method?
   (d) What methods can include the drivers of the bookings process?
2. What is the current bookings forecast situation at Kenya Airways?
   (a) What forecasting approach is currently used?
   (b) Is the forecast updated and what updating approach is used?
   (c) Is the uncertainty included in the forecast? What method is used to do this?
   (d) How are the booking limits applied during the bookings process?
   (e) How and by what divisions is the forecast used within Revenue Management?
   (f) What factors highly influence the bookings numbers?
3. What are the requirements of an airline for the bookings forecast?
   (a) How many times is an update of the forecast required?
   (b) What level of disaggregation (detail) should the forecast hold?
   (c) What level of accuracy is achieved by the currently applied methods?
4. What are the variables and drivers in the model?
   (a) What is the relationship between the variables and drivers?
   (b) What are the input variables?
5. What is the impact of the forecasting model?
   (a) What measurements can be used to determine the forecast accuracy?
   (b) What is the performance of the model in terms of accuracy?
   (c) How does the performance compare with current practice?
   (d) How valid and feasible are the results?
6. How can the model be implemented in practice?
   (a) What are the modifications needed to implement the model in operations?
   (b) What is the effect of the model on the operations of the airline?

The final sub-question is not critical to reach the research objective but is considered as an optional step to execute in order for Kenya Airways to be able to implement the model in the day-to-day operations.

The research questions that are defined above show what is needed to reach the research objective and can be summarized in a research framework. The research framework for this project is based on the methodology presented in Verschuren and Doorewaard (2010). The first column represents the *research background* that should be determined before the model is defined. The second column contains the *confrontation* which shows the input variables of the conceptual model. The third column displays the *operationalisation* which includes the testing, verification and validation of the model. The fourth column contains the conclusion and recommendations. Each main research question answers a part of the research framework as can be seen in Figure 3.1.

The *historical bookings data* has been placed in between *research background* and *confrontation* as it is both an input variable for the model as background knowledge to structure the model. The other part that is not placed in a specific column is the *consultation of experts*. This part will be used throughout the project and is therefore spread over the entire framework.

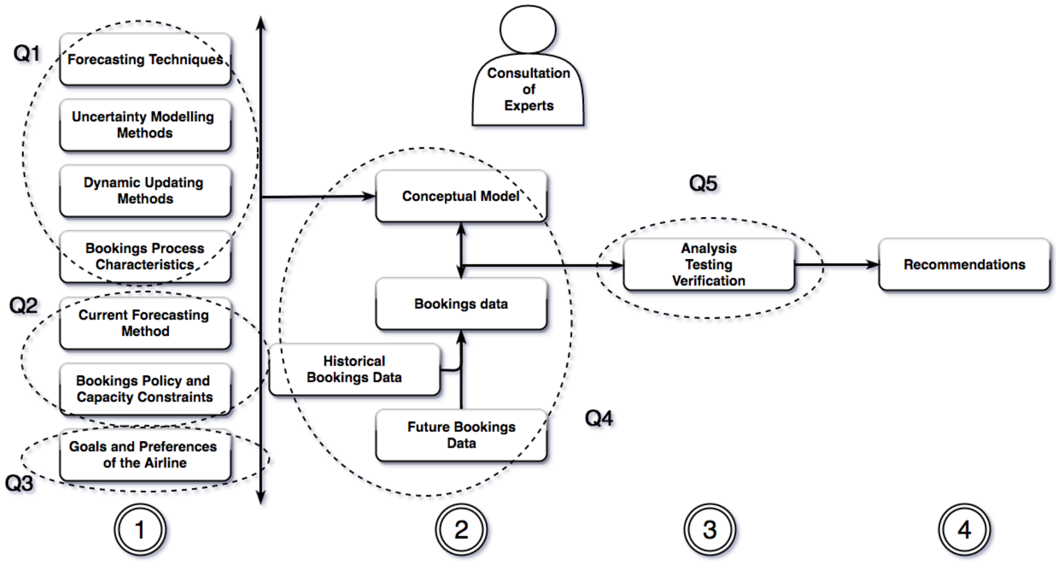


Figure 3.1: Research framework of the research project including an indication where each research question can be located

3.3. Research strategy

The research project can be defined as an empirical research because the relevant material needs to be gathered from the field at Kenya Airways. However, in the end the model should be usable for other airlines as well for which a large-scale approach is chosen.

The initial phase of the project can be considered of qualitative nature. Information is gathered from interviews with experts at the revenue management department of Kenya Airways and from the literature study. This provides insight into the characteristics of the bookings process, the current forecasting practices, constraints and the preferences of the airline.

The information from the initial phase is used in the second phase to develop the model which is of a quantitative character. The model uses numerical input data of Kenya Airways to compute results. The results of the forecasting model are in the form of an expected value and probability distributions presenting the forecast uncertainty. The input data consists mostly of bookings data, booking constraints and flight details. Firstly, the data is used to test the model and improve it in an iterative process. Then, the developed forecasting model is tested, compared to traditional methods, verified and validated.

4
Methodology

The methodology applied in this study is explained in this chapter. Firstly, the formulation of the Markov Chain model, including a discussion of its main elements, is explained. Secondly, the approach used to determine the entries in the transition matrices is discussed. Finally, the measures of performance that are used in this research to assess the accuracy of the forecasts are stated.

4.1. Model formulation

This section discusses the Markov Chain model that is proposed as the forecasting model in this research. Firstly, the main principles of a Markov Chain model are shortly addressed. Secondly, the main elements of a Markov Chain model, i.e. the probability vectors and the transition matrices, are explained into more detail. Finally, a conceptual model is presented that visualizes the separate elements of the forecasting model.

4.1.1. Markov Chains

A non-stationary Markov process concept has been identified as a method that can fulfill the research objective. The Markov process is modeled by a Markov Chain (MC) that describes the evolution of a state vector over time. Markov Chain models have the ability to include the many different possible states and the migrations from one state to another in transition matrices. It can therefore be seen as an dynamic booking forecasting tool that can predict and update the prediction of future states of the booking system. These predictions are then based on the most recent information of the current state and the evolutionary behavior of the system. It is important to note that a characteristic of a Markov Chain is that the future states of a system are only dependent on the current state. The past states of the system do not have any influence. The process in this application of an airline bookings system is considered to be non-stationary as the transitional behavior of the system changes over time. As the bookings process ends at the departure date of a flight, the Markov Chain is constructed for a finite period of time corresponding to the time until the departure date of the flight. The probability vectors can incorporate all possible states of the system at a specific moment in time. Let the Markov Chain $\{X_t\}$ denote the number of net bookings at t days before the day of departure. The state space of the Markov Chain and the time parameter indices are determined as follows:

- State Space: $\mathbb{C} = \{0, 1, ..., K\}$. Note that the state space is bound at 0 below because the system cannot hold a negative number of bookings for a flight. Furthermore, the space is bound above at K as the maximum number of bookings that can be received. This upper bound is higher than the capacity of the flight as in general an overbooking level is present and more bookings can be accepted than there are seats in the aircraft.

- Time indices: $T = \{N, N-1, ..., 0\}$, where the time index $t \in T$ represents the time remaining before the day of departure. Note that t is therefore decreasing over time and will equal 0 at the day of departure. It should be noted that the model therefore forecasts until the day of departure and not until departure time. In this study the length of a time-step equals one day. N is in this case the horizon over which the forecast is made.

4.1.2. Probability vectors

One of the goals in this study is to include the uncertainty of a forecast in the prediction. This can be achieved by defining the probability vector of the Markov Chain in a way that the uncertainty of the state of the bookings system is incorporated. The probability vector is determined to be composed of the probabilities that a certain number of seats are booked for a flight at some point in time in the bookings process. Let π_t determine the distribution of bookings at time t, as defined in Equation 4.1.

$$\pi_t = \begin{bmatrix} \mathbb{P}(X_t = 0) \\ \mathbb{P}(X_t = 1) \\ \vdots \\ \mathbb{P}(X_t = K) \end{bmatrix} \tag{4.1}$$

$\pi_t(i)$ denotes the i-th entry of the probability vector π_t as follows,

$$\pi_t(i) = \mathbb{P}(X_t = i)$$

for any $i \in \{0, ..., K\}$

According to the principles of the Markov Process the sum of the state probabilities $\pi_t(i)$ in the probability vector should be equal to one as can be represented by Equation 4.2 (Kijima, 1997).

$$\sum_{i=0}^{K} \pi_t(i) = 1 \tag{4.2}$$

Initial probability vector

The initial probability vector represents the state of the system at the point in time $t = N$ when the forecast is computed for a specific horizon N. This initial probability vector is based on the number of bookings $X_t = i$ which is known for that point in time. As it is known in what state the system is at time $t = N$, the probability state vector π_t consists of a vector of zeroes and one value equal to 1 which represents the current number of bookings in the system. For example, if a forecast is computed 100 days before departure and the net number of bookings at that time equals 20. The probability vector π_{100} is defined as follows:

$$\pi_{100}(i) = \begin{cases} 1 & i = 20, \\ 0 & i \neq 20 \end{cases}$$

4.1.3. Transition matrices

The transition matrices describe the evolution of the number of bookings in the system over time. The matrices are composed of probabilities that entail the migration of one state to another.

Let $P_{t,t-1}$ denote the transition probability matrix from time t to time $t-1$,
with $P_{t,t-1} = \left[P_{t,t-1}^{i,j}\right]$, $i, j \in \{0, \ldots, K\}$, where

$$P_{t,t-1}^{i,j} = \mathbb{P}\left(X_{t-1} = j \mid X_t = i\right)$$

and,

$$\sum_{j=0}^{K} \mathbb{P}\left(X_{t-1} = j \mid X_t = i\right) = 1 \tag{4.3}$$

for any $i \in \{0, ..., K\}$

It is important to note that one of the properties of a Markov Process that should hold is that the sum of the probabilities in the columns of a transition matrix should sum up to one which is displayed in Equation 4.3 (Kijima, 1997).

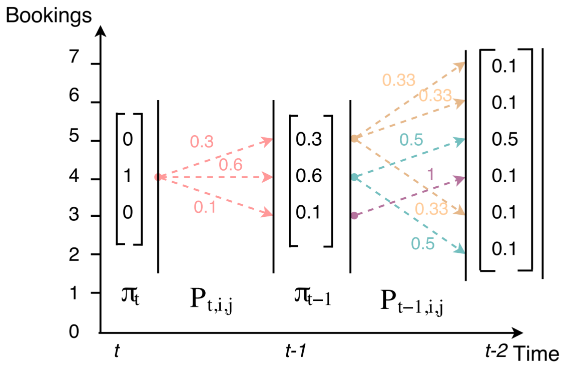


Figure 4.1: The general working principle of the Markov Chain approach

4.1.4. Determining future probability vectors

Based on the initial probability vector and the transition matrices for the remaining of the bookings process the future probability vectors are determined in a sequential order. This way the model entails a predicted state of the bookings system for all days t until the departure date of the flight. The probability vector π_{t-1} at time $t-1$ follows from the vector π_t at time t and the transition matrix $P_{t,t-1}$ at time t as shown by Equation 4.4.

$$\pi_{t-1} = P_{t,t-1} \cdot \pi_t \tag{4.4}$$

or, in general for all the probability vectors until the day of departure this transition can be expressed as follows,

$$\pi_{t-\beta} = \prod_{l=1}^{\beta} P_{t-l+1,t-l} \cdot \pi_t$$

where $\beta \in \{1, ..., N\}$

The resulting future probability vectors all hold a certain non-parametric distribution of the probability that the system will contain a certain number of bookings at that time in the bookings process. The evolution of the initial probability vector in the future is abstractly presented in Figure 4.1. In this graph the paths and probabilities that are described by the transition matrices $P_{t,t-1}$ are visualized and the probability vectors π_{t-1} that result from this sequential matrix multiplication can be seen as well for three time steps (t, t-1, t-2). In this graph it can be seen that the initial probability vector consists of one value equal to 1 and zeros for the remaining entries. This represents the situation in this case that the number of bookings at that moment in time is known and equal to 4 as can be seen in the figure.

Expected state value

For every time step in the forecasting process the values in the probability vector describe a probability mass function of the net number of bookings for the flight. From this non-parametric probability distribution, which is sequentially determined using Equation 4.4, the expected value can be calculated. Let $E[X_t]$ denote the expected net number of bookings at time t, calculated as displayed in Equation 4.5.

$$E[X_t] = \sum_{i=0}^{K} \pi_t(i) \cdot i, \tag{4.5}$$

4.1.5. Conceptual model

The proposed Markov Chain model can graphically be presented as a conceptual model which can be seen in Figure 4.2. It is seen in the diagram that the transition matrices are calculated based on a set of historical bookings data. The approach to the calculation of the transition probabilities is explained in more detail in Section 4.2 and visualized in Figure 4.7. The set of data is selected in such a way that the outliers are removed from the initial total dataset. Furthermore, the forecasts computed in this research are made for flights that

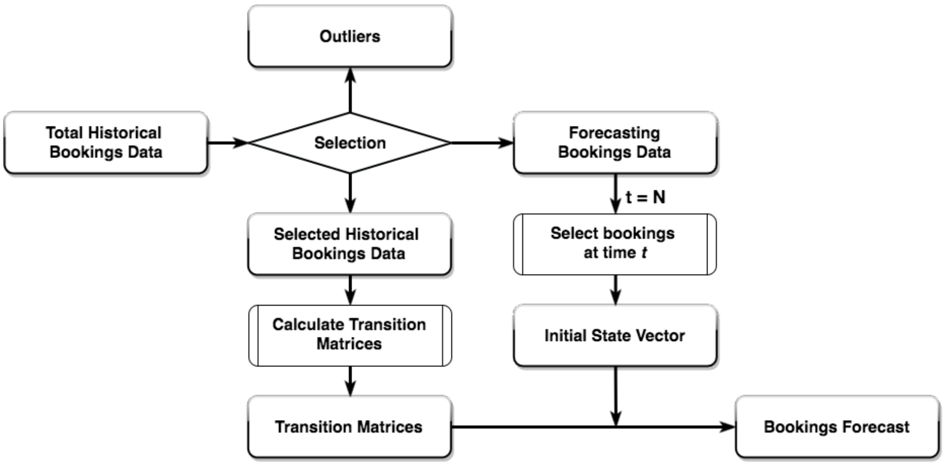


Figure 4.2: Conceptual model of the proposed Markov Chain forecasting model

already departed in order to be able to compare the outcome with the actual realized values. Therefore the bookings data of these flights has to be excluded from the set of data as this will lead to a biased forecast.

The number of bookings for the flight under consideration at a certain number of days N before departure is used to determine the initial probability vector. Based on the initial probability vector and the transition matrices the state of the system can be predicted over the length of N days until the departure date.

The model that is formulated above also assumes the following:

- The historical bookings data used as an input to the forecasting model only includes the bookings data of the same flight number. No data of other flight legs is taken into account in the forecast. Therefore, it is assumed that the booking behavior on the flight is independent from the bookings that are coming in for other flights in the network of the airline.
- The forecasts in this study are made for flights that departed in the past. This means that only complete bookings curves are available and all available data is historical bookings data. However, in reality some bookings in the historical dataset are not yet realized at the point the forecast is made. This is illustrated in Table 4.1 where the bold numbers indicate the last known bookings values for a forecast made 3 DBD of a flight that departed on the 1st of October. The shaded top rows in the table indicate the historical dataset that is used for the forecast in which it can be seen that the colored values are in fact not yet known at the point in time the forecast is computed. In the model the dataset is not updated when a forecast is made with a new horizon at another day before departure for computational considerations. Therefore only complete historical bookings curves are used in the model.

Table 4.1: Example historical bookings dataset for a forecast of a flight that departed on the 1st of October 2016

	Days before Departure							
Flight Date	0	1	2	3	4	5	6	7
26/09/16	110	109	107	104	100	100	99	96
27/09/16	103	103	100	97	92	90	89	89
28/09/16	**107**	102	101	101	99	97	92	90
29/09/16	*112*	**109**	106	102	96	92	90	91
30/09/16	*105*	*103*	**102**	96	94	95	92	87
01/10/16	*108*	*104*	*105*	**98**	92	85	85	81

4.2. Transition matrices probability estimation

The transition matrices describe the evolution of the net number of bookings for a flight in the bookings system of the airline over time. The entries in the transition matrix are based on the observed transitions that have been taking place in the historical bookings dataset fed in as an input to the forecasting model. The columns in the transition matrix can then be filled by a probability distribution of the net bookings number

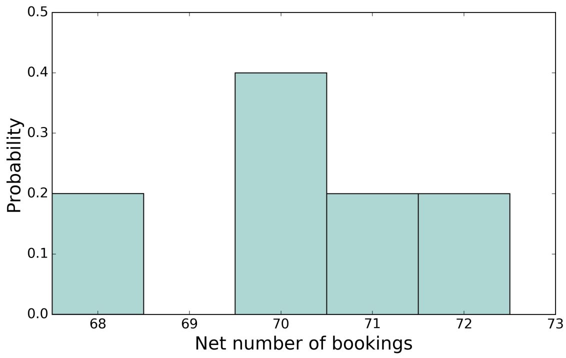


Figure 4.3: Example of empirical distribution of bookings at time *t-1*, given 70 bookings at time *t*

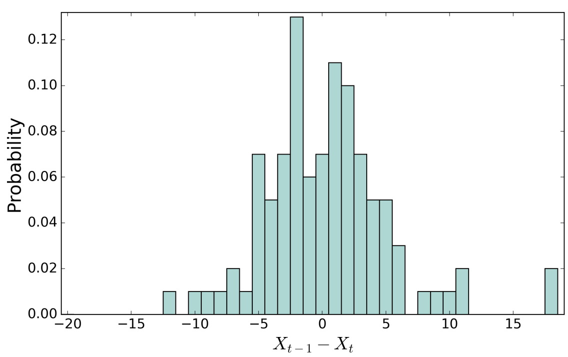


Figure 4.4: Example of differences in between time *t* and *t-1* used to compute the parametric distribution

to be in the system in the next time interval based on bookings numbers in the previous step. One of the bookings evaluations that the Markov Chain method does not cover is the transition of the state in the case a bookings number has not been observed in the dataset of the flight. In that scenario the transition matrices entail no information on the expected transition of the state to the next time interval, resulting in an empty probability vector for the next time step. This is a violation of the Markov process requirements (as mentioned in Chapter 4.1). In this study the entries in the transition matrices are determined based on the approach of Goto et al. (2004), where a combination of an empirical and parametric distribution is used in order to fill in the missing columns in the matrices and to smooth the distributions in the other columns.

4.2.1. Empirical estimation

At all the time steps until departure day, for each bookings number the probability to go to each subsequent bookings number can be determined by observing the relative frequency of the transitions. This empirical probability estimation captures the dependency of the subsequent bookings on the initial bookings in the system. Let $p^e_{t,t-1}$ denote the probability of transitioning from $X_t = i$ bookings at time t to $X_{t-1} = j$ bookings at time $t-1$, the empirical approach of calculating the matrix entries can be seen in Equation 4.6.

$$p^e_{t,t-1} = \frac{\sum_{j=0}^{K} (X_{t-1} = j \mid X_t = i)}{\sum_{j=0}^{K} (X_t = i)} \tag{4.6}$$

for any $i \in \{0, \ldots, K\}$

For example, suppose at a certain time t there are 5 observation of 70 bookings in the system which evolve to the set bookings $\{68, 70, 70, 71, 72\}$ at time *t-1*. The probability distribution of this example is seen in Figure 4.3 and the vector of column 70 ($i = 70$) of the transition matrix can be seen in Equation 4.7.

j	0	1	$\cdots$	68	69	70	71	72	$\cdots$	K
$\mathbb{P}\left(X_{t-1} = j \mid X_t = 70\right)$	0	0	$\cdots$	0.2	0	0.4	0.2	0.2	$\cdots$	0

(4.7)

Ideally, this empirical approach can be used to fill in the complete transition matrix. However, because of the large number of transition combinations and the finite amount of bookings data that is available for a flight, the empirical approach leads to zero columns in the matrix for transitions that have not been observed in the bookings data.[1] Furthermore, when only a small amount of transitions are captured from an initial bookings number, the resulting distribution may have an unusual shape or only capture extreme values.

4.2.2. Parametric estimation

In order to solve the problems that arise when determining the transition matrix entries solely based on the empirical estimation approach, a parametric estimation is proposed. The parametric approach estimates the transitions based on all the differences in bookings that are observed from time t to time *t-1*. Contrary to the

[1] To fill all the columns of a 200 x 200 transition matrix bookings data of way more than 200 flights is required, especially because because flights will repetition some transitions

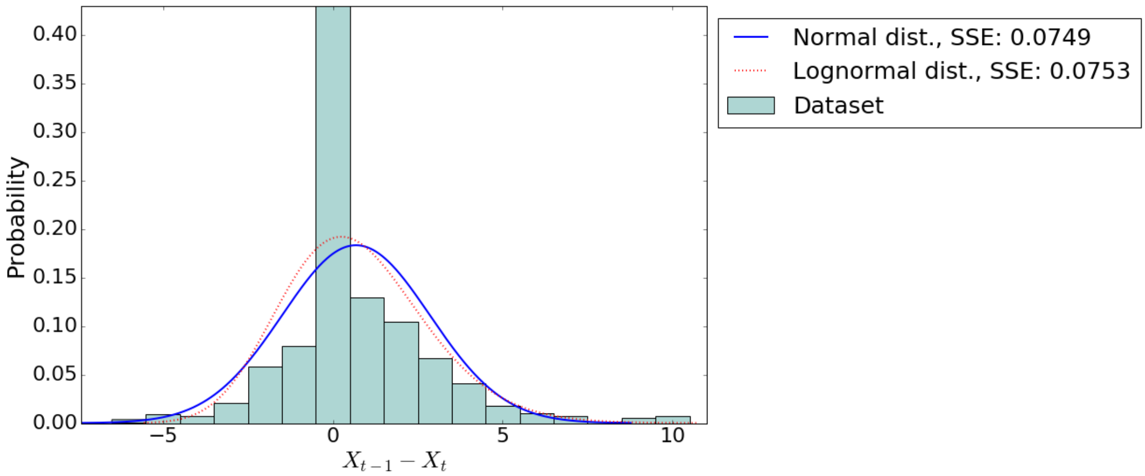


Figure 4.5: Actual distribution, fitted normal distribution and fitted lognormal distribution of difference in number of bookings for KQ101 from $t = 74$ to $t = 73$

empirical estimation, this approach assumes that the increment in number of bookings from t to time *t-1* is independent from the bookings at t.

For every time interval the frequency of the observed booking increments in the dataset is determined. For example, in Figure 4.4 the distribution of differences between two consecutive points in time are shown. To remove the effect of outliers, the tails of the distribution are removed. In this study 1% of the lower and 1% of the higher data is removed to ensure that the distributions can be fit well.

From the resulting probability distribution the transition probabilities can be defined for all the columns in the transition matrix. The decrease in net bookings resulting from the parametric estimation can, however, be larger than the initial bookings value i and a large increase in bookings added to the initial value can become larger than the limit K predefined by the state space. Therefore, the parametric transition probabilities are computed according to Equation 4.8 (Goto et al., 2004),

$$p^{p}_{t,t-1} = \begin{cases} \mathbb{P}(Z_t \leq -i) & j = 0, \\ \mathbb{P}(Z_t \leq j-i) - \mathbb{P}(Z_t \leq j-i-1) & 0 < j < K, \\ \mathbb{P}(Z_t > j-i-1 & j = K, \end{cases} \tag{4.8}$$

where $i \in \{0, ..., K\}$ and Z_t represents the random variable that describes the difference in load between time t and *t-1* ($X_{t-1} - X_t$) and results from the distribution in Figure 4.4.

Parametric distributions

In this study it is assessed if a parametric distribution can be fit to the discrete distribution that has been discussed above. Several distributions that have been discussed in literature (Section 2) to be a good fit to airline bookings data have been considered in this research such as the gamma, beta, exponential, lognormal, normal and the poisson distribution. The lognormal and the normal distribution have been identified to best fit the bookings data in an initial analysis of the data that was provided by KQ. In this analysis the Error Sum of Squares (SSE) of the fitted distribution is determined and the parametric distribution yielding the lowest error considered best. These two distributions have therefore been considered for the remaining of the studies.

The approach to fit a parametric distribution smoothens out the data and works by fitting a continuous parametric distribution on the trimmed differences in booking between two time intervals. The random variable Z_t is now assumed to follow the fitted continuous distribution after which the transition probabilities entries can be determined following Equation 4.8. This is illustrated in Figure 4.5 where a lognormal and normal distribution are fitted on the bookings data. It can be seen that the normal distribution fits the data best for this specific example based on the SSE. The distribution of the random variable Z_t will therefore follow the normal distribution and is used to determine the probability entries in the matrix.

4.2.3. Combined estimation

The empirically estimated transition matrix probability entries discussed in Section 4.2.1 and the parametric estimation of the probability entries discussed in Section 4.2.2 are combined in the final transition matrix in

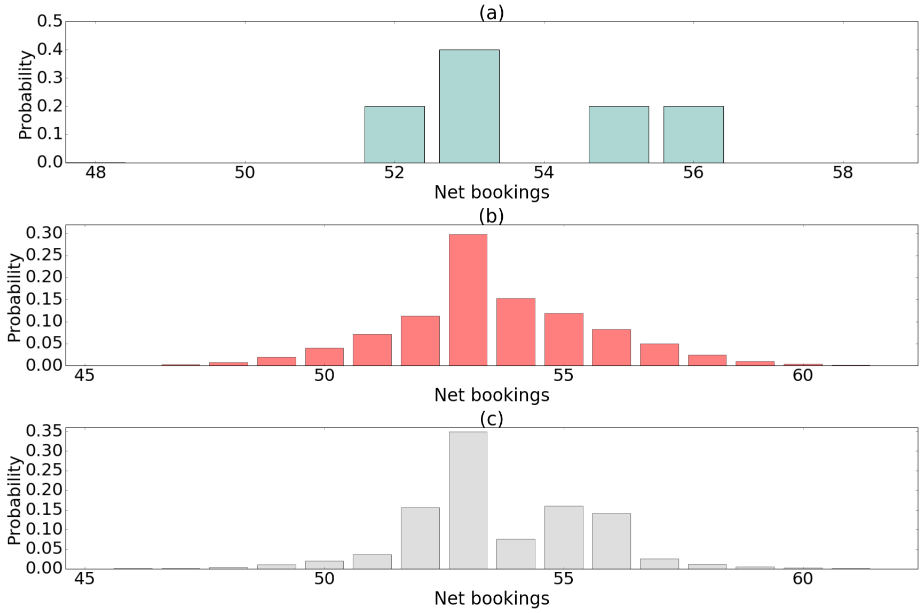


Figure 4.6: The empirical (a), normal parametric (b) and the combined distribution (c) for the bookings for KQ101 at $t = 70$ given 53 bookings at $t = 69$ days before departure

order to fill in missing cells and to smoothen the distribution. The estimated probability density functions are combined using a weighting factor α as shown in Equation 4.9.

$$p^c_{t,t-1} = \begin{cases} \alpha \cdot p^e_{t,t-1} + (1-\alpha) \cdot p^p_{t,t-1} & \sum_{j=0}^{K} p^e_{t,t-1} \neq 0, \\ p^p_{t,t-1} & \sum_{j=0}^{K} p^e_{t,t-1} = 0 \end{cases} \tag{4.9}$$

for any $i \in \{0,...,K\}$ and where, $0 \leq \alpha \leq 1$.
For example, in the bookings dataset of KQ101 (LHR-NBO) at time t there were 5 observations of 53 bookings in the system which evolved to the set of bookings $\{52, 53, 53, 55, 56\}$ at time *t-1*. The resulting empirical, normal parametric and the combined distribution with an $\alpha = 0.5$ are shown in Figure 4.6.

According to Goto et al. (2004) high values of α in general give the best results. The optimal value of α is expected to change over the bookings horizon and to be dependent on the considered flight. The influence of the value of α on the performance of the forecasting model is investigated later in this study in Section 6.1.

4.2.4. Schematic model transition matrix estimation

The above described approach to determine the probabilities in the transition matrices can be visualized in a flowchart which can be seen in Figure 4.7.

In the flowchart it can be seen that the transition matrices are sequentially computed for all remaining days until departure based on a single historical dataset. Furthermore, the several model settings that determine the way the transition matrices are constructed can be distinguished. The change in performance of the model to the following settings is further analyzed in this study:

- Seasonal data input
- Parametric fitted distribution
- Weighting factor α

4.3. Measures of performance

The forecast resulting from the new Markov Chain model is compared with the real observed bookings number on the flight to assess the performance of the method. Several statistical methods exist to measure the error of a point forecast. The error measurement used in this study is discussed firstly. Then the measurement approach used to assess the ability of the proposed model to capture the uncertainty in the demand is explained.

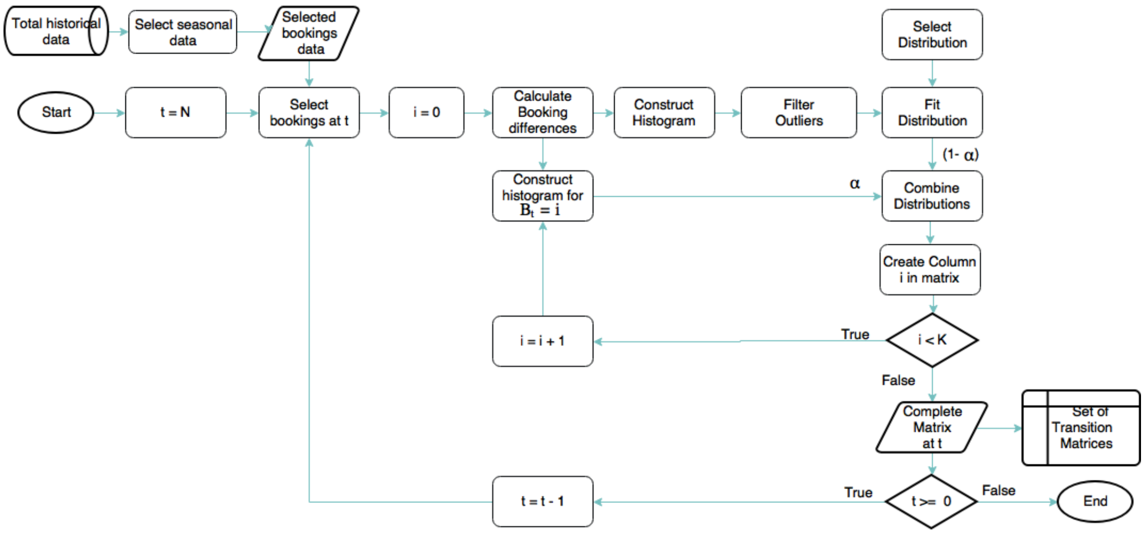


Figure 4.7: Flowchart of the calculation of the transition matrices in the proposed Markov Chain model

4.3.1. Error measurements

The forecasting error is expressed as the difference between the predicted and the actual value and is calculated using statistical measures of accuracy. In this study the accuracy of the forecast is assessed from the forecasting day until the day of departure of the flight, over the entire bookings curve. Although, the Mean Absolute Percentage Error (MAPE) is a popular error measurement technique for demand forecasting (Pesch & Kovalyov, 2015) it has not been used in this study because it is undefined for booking values equal to one. Because in this study the entire bookings horizon is considered the chance of having zero bookings is reasonable. Therefore, the MAPE is not considered in this study. The numerical method of forecasting performance that has been used in this study is the MAE.

Mean Absolute Error

The MAE is one of the most basic forms of error measuring. This statistical measure takes the average of the absolute values of the measured forecasting error. The forecasting error is defined in this error measurement as the difference between the forecasted value and the actual bookings number that is observed. Let $X_{t,f}$ and $E\left[X_{t,f}\right]$ denote the actual bookings at time t for flight f and the expected bookings at time t for flight f respectively. The mathematical definition of the MAE can then be seen in Equation 4.10 (Zakhary et al., 2008).

$$MAE(t) = \frac{1}{m}\sum_{f=1}^{m}\left|E\left[X_{t,f}\right] - X_{t,f}\right| \tag{4.10}$$

where m is the number of forecasts generated by the model.

4.3.2. Uncertainty modeling measurements

The advantage of the proposed Markov Chain method is that a level of uncertainty is included in the forecast which implies that the forecast provides more information to the revenue management controller. In this research the approach has been chosen to evaluate the ability of the model to accurately incorporate the uncertainty in the prediction with interval scores. It should be noted that the proposed method yields a probability distribution, however, it is chosen to assess the distribution for multiple probability intervals. The interval score (Gneiting & Raftery, 2007) is a method that can quantify the quality of an interval prediction. Let $S_\alpha(p_u, p_l, X_t)$ denote the interval score at time t for the $(1-\alpha)\cdot 100\%$ prediction interval, $\alpha \in (0,1)$. Then,

$$S_\alpha(p_u, p_l, X_t) = (p_u - p_l) + \frac{2}{\alpha}(p_l - X_t)\mathbb{1}_{X_t<p_l} + \frac{2}{\alpha}(X_t - p_u)\mathbb{1}_{X_t>p_u}$$

where p_u and p_l are the upper and lower limits of the interval. The score $S_\alpha(p_u, p_l, X_t)$ takes into account both the size of the prediction interval as well as the ability to cover the realizations X_t. The lower the value of the interval score, the better the forecast is defined. In the ideal case the intervals cover all realized values,

Table 4.2: Three example calculations of the interval score for a set of prediction intervals

Example nr.	**Prediction Interval**	α	p_u	p_l	X_t	$S_\alpha(p_u, p_l, X_t)$
1	80%	0.2	150	10	60	140
2	80%	0.2	190	30	20	260
3	20%	0.8	60	10	80	100

which results in the second and third elements of the formula to be equal to *0* and the interval should be as small as possible.

Consider three examples indicated in Table 4.2. In this table the upper and lower limit of the three prediction intervals is indicated as well as the observed number of bookings X_t in the system. For the first example the value falls nicely within the defined interval so no penalty is applied and the score equals the width of the prediction interval. For the second example the observed value of X_t falls below the defined interval. The penalty applied in this case is high (100) because an 80% prediction interval should in general be able to capture about all values. This can be seen more clearly when comparing the second with the third prediction interval example. For the third interval the value X_t lies further away from the limit of the interval, however, because only the 20% prediction interval is considered here, the applied penalty is lower. For a 20% prediction interval it is less bad if once in a while an observed value falls outside the interval than for the 80% prediction interval.

This approach to model the uncertainty of the forecast can not be applied to the traditional methods with which the proposed model is compared as these yield point forecasts. Therefore, the interval score measure can only be used to determine the performance of the uncertainty modeling of the proposed method for different model settings.

5
Experiment

This chapter discusses the experiment that is carried out in this study. Firstly, some background information on Kenya Airways is given. Secondly, the selected set of bookings data is discussed and the bookings data has been analyzed in order to pinpoint trends that might be present and should be considered when testing the model. Following up on that, the experimental design is discussed.

5.1. Background information on Kenya Airways

Kenya Airways, the pride of Africa, is a leading African airline and the flag carrier of Kenya. It was founded in 1977, in 1995 it entered into an alliance with KLM Royal Dutch Airlines and in 2007 Kenya Airways became a full member of the SkyTeam alliance, which opened up its network to a multiple of destinations all over the world.

Kenya Airways is operating a fleet of 30 aircraft consisting of Embraer 190, Boeing 737-800 and Boeing 787-8. It serves flights to numerous destinations across Africa, Europe, the Middle East, the Far East and Asia operating a hub and spoke network with JKIA in Nairobi as its main hub.

This hub can clearly be seen by looking at the network of KQ in Figure 5.1. The network seen in the figure excludes the destinations Kenya Airways is able to fly to using codeshares. From the map the various flights around the world for which KQ needs to manage the revenue can clearly be distinguished. The flights that are chosen as test cases for this study are visualized by a green flight path in Figure 5.1.

Figure 5.1: The network of Kenya Airways

5.2. Data selection

The data is obtained from the revenue management systems at KQ. The data on the development of the number of bookings in the system during the bookings process is extracted from *Delorean*. The revenue management system *Delorean* is a backward looking system that has originally been developed by the Koninklijke Luchtvaart Maatschappij (KLM) and currently also implemented within the revenue management operations of its *SkyTeam* partner Kenya Airways. Backward looking in this context means that the system makes no predictions on the future but only stores the newly available bookings information on a daily basis. *Delorean* can

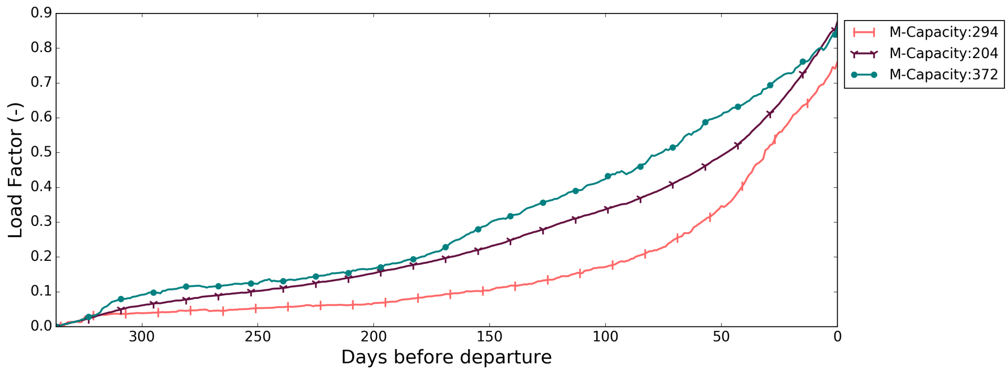


Figure 5.2: The average load factor curves for *KQ101* operating with different capacities (economy class (M))

show the bookings that had been made for a flight at a certain moment in time, called a *snapshot*. The bookings information is updated in this system overnight. More information on the data format of this system can be found in Appendix A.

The observation points that are stored in *Delorean* go up to around *336* days before departure and all these point are extracted from the system for the benefit of this research. By grouping the bookings data per flight on the snapshot date, the complete history of the net bookings in the system for the flights over the entire booking process can be determined.

Furthermore, the capacity that is deployed for the flights is extracted from *Delorean* as well. As the system is backward looking the denoted capacity numbers are most likely to be correct.

The negative aspect of *Delorean* is that it does not include information on the overbooking level (Authorization Level (AU)), the closed fare classes or the ticket prices of the different classes. This is monitored in *PROS*, a third party forward looking system that forecasts the demand for the future, adjusts the closure and opening of fare classes and sets overbooking limits for flights. As the overbooking levels and the seat allocations are not taken into account in this study, only the data from *Delorean* is used.

5.2.1. Selected flights

A subset of flights has been chosen in order to represent the varying types of flights in the network of KQ well. The set of flights consists of flights operating in different markets, with different equipment and with a different Load Factor (LF) as can be seen in Table 5.1. For these flight numbers the data of the flights departing in 2014, 2015 and 2016 has been extracted. However, only flights that were operated with the same aircraft type as the currently deployed Aircraft (A/C) indicated in Table 5.1 are considered, because the pricing scheme will be different when another capacity is deployed. Over the course of time some equipment changes have been made by KQ for their flights, which result in a different shape of the bookings curve for that flight as illustrated in Figure 5.2. The amount of departures that remain for the preselected flights can be seen in the last column of Table 5.1. Flight KQ101 denotes the flight departing from London Heathrow Airport (LHR) to Jomo Kenyatta International Airport (NBO). Flight 512 denotes the flight departing from Jomo Kenyatta International Airport (NBO) to Bamako Senou International Airport (BKO).

Table 5.1: Subset of flights used in the study to test the model

Flight	A/C Type	Market	LF	Flights in dataset
KQ101	B787-800	Europe	High	731
KQ512	B737-700	West Africa	Low	457

The variance in the datasets is significant. In Figure 5.3 the variance in the number of bookings at the day of departure in the dataset for the departures of flight KQ101 can be seen. The bookings at the day of departure for KQ101 averaged to 176 with a variance of 813. The value ranged from a minimum of 85 to a maximum of 225.

The variation in the bookings at the day of departure for flight KQ512 is displayed in Figure 5.4. For these bookings the average equals 66 with a variance of 211. The minimum observed value equals 26 and the maximum value 106. This volatility in the data makes it difficult to achieve accurate forecasts.

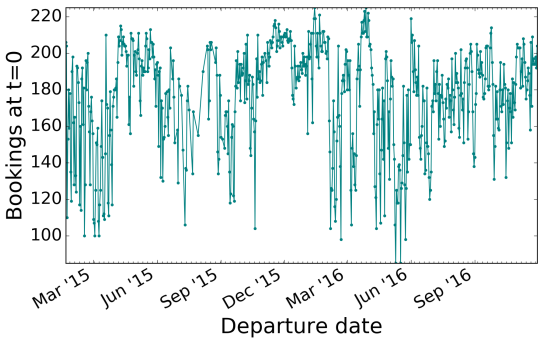


Figure 5.3: The variance in the bookings data of KQ101 at the day of departure

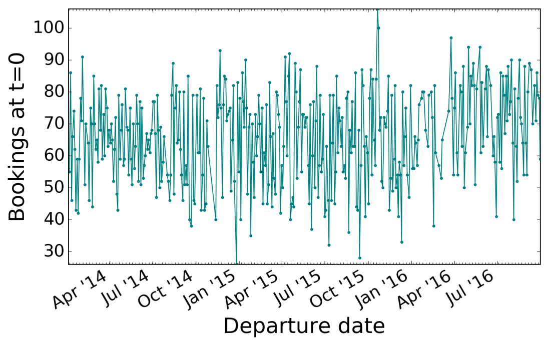


Figure 5.4: The variance in the bookings data of 512 at the day of departure

Erroneous data

After an initial processing of the data it became apparent that there exists some erroneous bookings data in the sets that are extracted from *Delorean*. The following errors have been identified:

- For some flights the number of bookings has been observed to drop suddenly to 0 closely to the day of departure. It is assumed that the state of the system has not been recorded during that day. The bookings data of the departure dates where this issue has been encountered is removed from the dataset.
- The flights for which the number of bookings in the system broadly exceeded the overbooking level that had been enforced for that flight have also been marked as erroneous data and are removed from the dataset. For these flights the capacity is most probably not stored correctly. Values that are more than 25% above the capacity of the flight are filtered.

5.3. Data analysis

The bookings data of Kenya Airways has been analyzed in order to identify trends and characteristics in the data. Taking into account these trends when computing a forecast is an important factor to the success of a forecasting model (L. Weatherford, 2016). Firstly, the assessment of seasonality in the data is discussed. Then, the exploration of other characteristics in the datasets is explained.

5.3.1. Seasonality analysis

Several seasonal trends can exist in the data. These trends are analyzed for the flights that have been mentioned in Table 5.1. All bookings data that has been provided by Kenya Airways for these flights is incorporated in the data analysis. Firstly, there might be a different booking behavior over the year. This has been analyzed by looking at monthly averages of bookings and is discussed in the first part of this chapter. Another trend that might be present in the data are day-of-week trends. Some days in the week might be more popular than others and this behavior is analyzed in the second part of this section.

Monthly seasonality

In order to analyze variations that might be present over the year in the bookings behavior of flights, the monthly booking curves have been assessed. These are an average of all the bookings curves of flights that departed in the same month. Firstly, the analysis of flight *KQ101* is discussed and thereafter the analysis of flight *KQ512*.

KQ101 The monthly bookings for flight KQ101 (LHR-NBO) are displayed in Figure 5.5. From the graph a clear difference between the high season and low season months can be identified. Half of the year (June, July, August, September, October, December) clearly shows a bookings behavior higher than average and the other half of the year (January, February, March, April, May, November) below average.

KQ512 The monthly trends for KQ512 can be seen in Figure 5.6. For this flight the difference between the high and low season period is harder to identify from the data than for KQ101. The flights in *September* and *October* clearly show an above average bookings behavior, whereas the bookings for flights in *November*

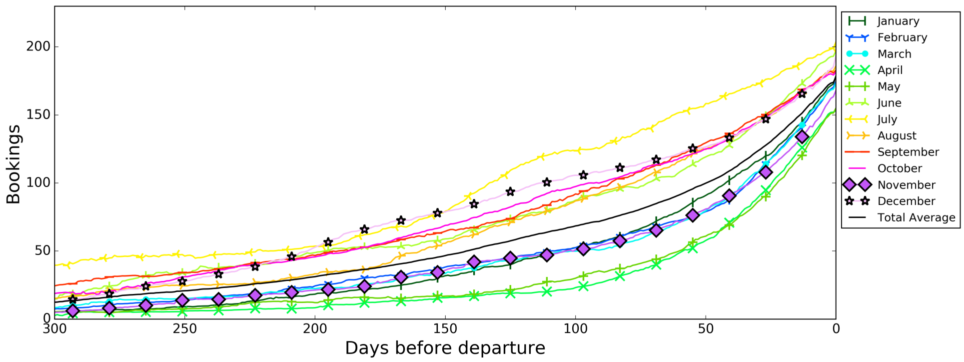


Figure 5.5: The average monthly booking curves for departures of *KQ101* per economy class (M) capacity

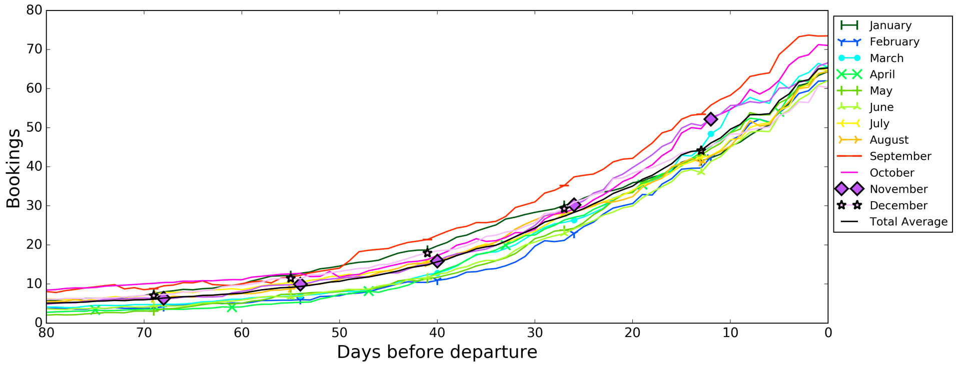


Figure 5.6: The average monthly booking curves for departures of *KQ512* per economy class (M) capacity

come in early but end up to be not that different from the average number of bookings for all flights in the year. Furthermore, the flights in *June* show a below average booking behavior and the ticket reservations for the flights in *December* come in early but end up to be clearly below average at the day of departure.

Day-of-week variations

Another variation in the number of bookings that might be present in the dataset is the development of the number of bookings in the system for flights that depart in another day of the week. Again both of the flights will be discussed in this section.

KQ101 The weekday averages of the bookings in the system for flight KQ101 can be seen in Figure 5.7. From this graph it can be concluded that the flights departing on the Fridays and Saturday are the most popular days to fly. However, the bookings for flights on Sundays come in slowly but end up at about the same level at the day of departure. The remaining days-of-the-week can be concluded to be the least popular days.

Another interesting behavior than can be observed in the data is the increase and decrease in the bookings number that repeats with an interval of around 7 days. In a more detailed investigation of this trend (which can be seen in Appendix E) it is found that in the final months of the bookings process in general the net number of bookings in the system decreases over the weekend and increases midweek.

KQ512 The difference in booking behavior for flights departing at different weekdays is displayed in Figure 5.8. This flight only departs three times a week. The flight on Sunday clearly has a different bookings curve than the other days of the week; bookings come in earlier and the final demand for the flights is also higher at the day of departure.

Furthermore, the same periodically decline in bookings is observed as for flight KQ101 where during weekends the net number of bookings in the system decreases and increases during the week.

Incorporating the above discussed trends that have been found in the bookings datasets in the forecasting model might improve the forecasting performance as the input is more representative for the situation. The

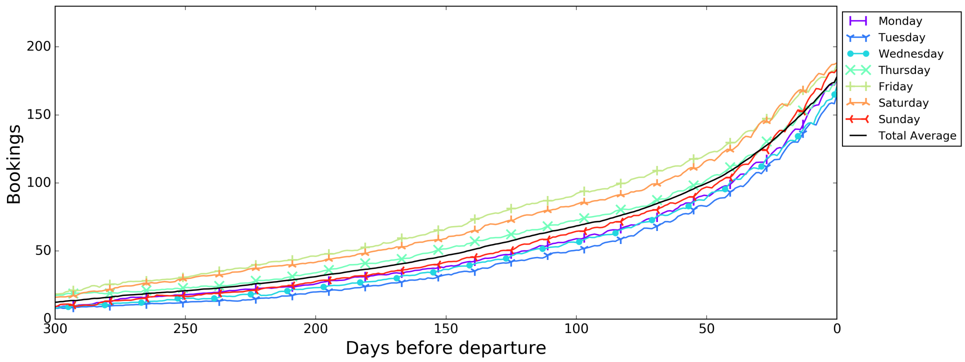


Figure 5.7: The average booking curves per weekday for departures of *KQ101* per economy class (M) capacity

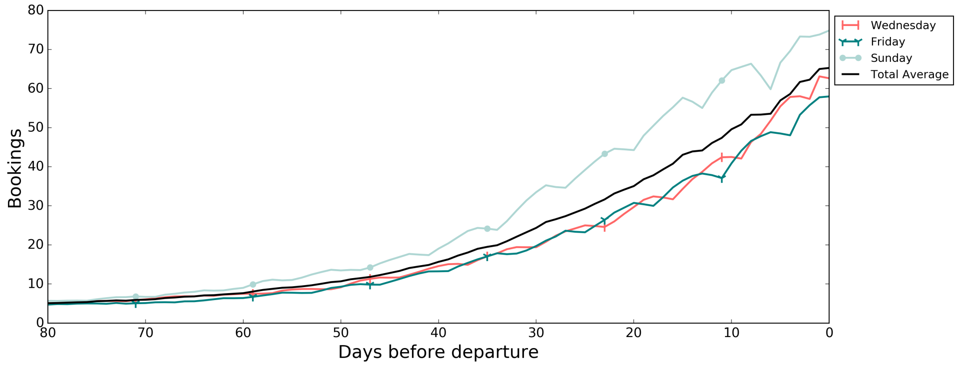


Figure 5.8: The average booking curves per weekday for departures of *KQ512* per economy class (M) capacity

Markov Chain model is tested with several selected seasonal bookings data to assess the change in the models performance as will be discussed in the next section.

5.3.2. Group bookings analysis

In the revenue management system *Delorean* a distinction is made between individual and group bookings. Group bookings are bookings made by agencies which are most often customers that are well known within Kenya Airways. From an analysis of the data is became apparent that group bookings can have a significant influence on the net bookings in the system. Firstly, it has been found that group bookings are generally made far in advance. Secondly, group bookings have a high cancelation ratio. This has to do with the first point because in most cases the early group booking is a reservation made which will be cancelled if the remaining amount has not been paid on the due date. It has been found that, especially for KQ101, the amount of group bookings that are realized at the day of departure is related to the size of the group reservation (Appendix C).

Following from this characteristic of group bookings that is present in the bookings data, the performance of the forecasting model might improve when these group bookings are handled separately from individual bookings. For this study, however, it has been chosen to assess the model first with a dataset that contains both individual and group bookings. It should be noted that the influence of large group bookings is reduced by removing the tails of the observed bookings increment distribution (Section 4.2.2).

5.4. Experimental design

This section describes the experimental design of the tests that are run in this study. The experiment consists of three parts; in the first phase the models parameters are calibrated to yield the lowest MAE. Then, the best parameter settings for the model to correctly model the uncertainty in the forecast is determined. Lastly, the performance of the proposed Markov Chain transition matrices model is compared with traditional methods.

5.4.1. Selected test cases

The total bookings data has been divided in a set of flights that are subjected to forecasting in this study and another set of flights of which the bookings data serves as an input to the model to compute forecasts. For KQ101 the flights departing in the period from the 1^{st} of October 2016 up until the 31^{st} of December 2016 are selected to be test data for the forecasting model and the remainder of the dataset is used as an input to the models. For flight KQ512 another period is selected to be test data because it has only three departures per week instead of daily. In order to make sure a sufficient set of tests are performed for this flight, forecasts are made for the period of the the 1^{st} of August 2016 up until the 31^{st} of December 2016 as can be seen in Table 5.2.

Table 5.2: Flight dates of the forecasts and the number of forecasts made for the test flights

Flight	Forecast Period	Nr. of departures
KQ101	01/10/16 - 31/12/16	92
KQ512	01/08/16 - 31/12/16	64

The performance of the forecasting model is shown and compared for a set of different forecasting horizons. Both long-term and short-term forecasts are considered. The bookings data is collected from 336 days before departure. The maximum horizon is chosen to be 300 days which is stepwise decreased to an horizon of minimum 7 days. The set of forecasting horizons for which the models are tested and compared is as follows:

$$H = \{300, 200, 100, 60, 39, 25, 14, 7\}$$

In order to give a clear overview of the results for every horizon a set of observation points V are chosen for which the performance of the model is assessed in detail, tabulated in Table 5.3. A maximum of 11 observations points are defined in a stepwise decreasing trend from the forecast horizon to the day of departure. The first observation point is the first day of the forecast horizon to observe the most short-term performance of the model. In total this comes down to 79 observation points over all considered horizons.

$$V = \{V_0, V_1, ..., V_{10}\}$$

Table 5.3: The observation points V in days before departure per forecast horizon

Horizon	V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}
300	299	270	240	210	180	150	120	90	60	30	0
200	199	180	160	140	120	100	80	60	40	20	0
100	99	90	80	70	60	50	40	30	20	10	0
60	59	54	48	42	36	30	24	18	12	6	0
39	38	35	31	27	23	19	15	11	7	3	0
25	24	22	19	16	13	10	7	4	0		
14	13	12	10	8	6	4	2	0			
7	6	5	4	3	2	1	0				

5.4.2. Model test settings

Based on the test cases that are selected, the model settings are slightly adapted. The major difference in the test cases, besides the amount of flights for which the forecasts are computed, is the determination of the state space for the forecasts. The size of the state space, and therefore the size of the transition matrices, determines the computational intensity of the model. For a low capacity flight the bookings number will be lower than for the high capacity flight. Therefore, the size of the state space is chosen to be dependent on the level of bookings that is normally experienced for a flight. The maximum state space chosen for the test cases in this study is tabulated in Table 5.4.

It can be seen that the maximum state space is chosen to be more than double the capacity for the flight. This is done because the state space actually represents the bookings that can possibly be expected for the flight. In the case of high demand the number of bookings can transcend the capacity in case no constraints are applied. Because the probability vectors should be able to capture also these high number of bookings, even in the case the probability of this booking really occurring is low, a margin is added to the economy cabin capacity. A downside of this approach is that the computational time increases significantly for an increased state space size. In Section 7 these state space sizes will be verified to determine whether the space should be increased or whether it can be decreased to reduce the computational time.

Table 5.4: The state space defined for the experiment

Flight	Capacity (M)	Maximum State Space K
KQ101	204	500
KQ512	116	250

5.4.3. Model calibration for minimum forecast error

In the first part of the experiment the new model is calibrated to yield the lowest forecast error. This calibration consists of three parts in this research. The first part will analyze the variation in forecast performance for different parametric distributions. The difference in performance of the model for changing values of the parameter α will be addressed in the second part. Finally, the variation of the model for a selective seasonal dataset is investigated.

Parametric distribution

Firstly, it will be analyzed to what extend a parametric distribution fitted on the observed values influences the performance of the proposed Markov Chain transition matrices model. As discussed in Section 4.2.2, the normal and lognormal distribution showed a best-fit on the distribution of observed bookings changes between two points in time. Therefore, the following distribution fitting approaches are considered in the calibrating part of the research:

- Normal Distribution
- Lognormal Distribution
- The best of the Normal and Lognormal Distribution
- Fitting no distribution

The third scenario takes for every day in the forecasting horizon either the Normal or the Lognormal distribution depending on which of the two best fitted the data for that time interval. For the fourth scenario it should be noted that, similar to the cases of distribution fitting, the tails of the distribution are removed as well.

It should be noted that for this analysis the model uses an α of 0.5 and the complete set of data available for the flights.

Weighting factor

The value of the parameter α determines the weight that is appointed to the empirical and parametric distribution in the construction of the combined transition matrices. The closer the value α is to *one*, the more the empirical distribution influences the expected transitions between the states. In this part of the experiment the performance of the model is assessed for the following values of α:

$$\begin{bmatrix} 0.2 & 0.3 & 0.5 & 0.7 & 0.8 \end{bmatrix}$$

Seasonal input

Some trends and seasonality have been observed in the data as discussed in Section 5.3. Therefore, the performance of the methods might improve in case only seasonal data is used by the model to compute the prediction of the number of bookings. It will be analyzed if the relative performance of the methods differs significantly when selective data is fed as an input to the model. This will be done by conducting tests where the following datasets are used:

- Total dataset
- Monthly seasonal data
- Weekly seasonal data
- Combined weekly and monthly seasonal data

5.4.4. Model calibration for optimal prediction intervals

In the second part of the experiment the forecast distributions that result from the proposed Markov Chain model are assessed. This will be done using the approach of interval scores discussed in Section 4.3.

Several fixed probability intervals are calculated based on the probability distribution of the forecast. These intervals can then be scored based on the length of the interval and the coverage of the actual observations. In order to assess the probability distribution a set of intervals has been defined that indicate the performance of the distribution both around the tails and around the center. The widest prediction interval considered in this study is the 95% prediction interval that should be able to capture almost all observed bookings. From there two other intervals are defined in a stepwise decreasing trend closer to the center of the distribution: 66% and 33%. These give an indication whether the model is able to produce an accurate prediction in case the actual bookings number lies more near the mean of the distribution.

For the calibration on optimal prediction intervals there are three times as much observation points available than for the calibration on MAE because the performance is determined for all three intervals described above. Therefore there are 228 observation points per test case when calibrating the model on the prediction intervals.

As there is no reference method and no reference interval score available, the score of the proposed method can not be compared with a score of a traditional forecasting method. Therefore, the objective of this part of the experiment is to assess how the scores evolve over the forecast horizon for the same various parameter settings that are explained in the previous section.

5.4.5. Benchmarking model

In the third part of the experiment the performance of the new method is compared with traditional methods. The first reference method that has been selected is the additive advanced pick-up method. As discussed in Section 2, the additive pick-up method has proven to perform well in comparison with other forecasting methods.

In this study the pick-up values are calculated with intervals of one day. For every interval all the available bookings data is used in this calculation. Therefore, n (in Equation 2.3) equals the size of the bookings dataset for all intervals.

Note that due to the data format, described in Section 4.1, all bookings data that is used as an input to the forecasting model are complete bookings curves. Therefore, the advanced pick-up model in this study is comparable with the classical pick-up method because the pick-up values are determined based on complete booking curves. However, because the bookings dataset includes flights that have not yet departed and the increment in bookings is determined in intervals of a day, the method is in this study defined as the advanced pick-up method.

The pick-up method is comparable with the proposed Markov Chain transition matrices model in the sense that both approaches determine the increment in bookings from one point in time to the other and add that to the current number of bookings in the system. However, the pick-up method assumes the increment in bookings to be independent from the most recent bookings in the system, whereas the proposed method assumes there is a dependency between the two. For a value of the weighting factor α of 0 the new model is actually equal to the advanced pick-up model.

Secondly, the historical average is compared with the proposed forecasting method in order to assess to what extend the proposed model is able to predict the future bookings number in case the value is significantly different from the average of the bookings dataset (a numerical example of these methods is given in Appendix D).

The relative performance of these methods is determined using the error measurements discussed in Section 4.3. The parametric distribution, weight α and the dataset used in this part of the experiment are chosen based on the results of the first two parts of the experiment where the new method parameters are calibrated.

6

Results forecast model

This chapter discusses the results of the tests performed with the proposed Markov Chain forecasting model in three steps following the experimental set-up outlined in Section 5.4. In the first part, the performance of the forecasting model in MAE is compared for different model settings. Varying values of the weighting factor α, a set of fitted parametric distributions and various seasonal datasets are considered. The next phase of the experiment assesses the performance of the uncertainty modeling by looking at the interval scores of several prediction intervals for the same varying model settings as in the first part of the experiments. In the final tests the MAE of the proposed forecasting model with the best parameter settings is compared with the traditional methods. Using a statistical test the significance of the difference in performance is tested. Finally, some forecast examples are displayed and the application to the operations within revenue management explained.

6.1. Model calibration for minimum forecast error

In this phase of the experiment the performance of the forecasting model is assessed for various model settings following the set-up as described in Section 5.4.3. Firstly, the performance of the model is assessed for a set of parametric distribution fitting approaches. Then, the variation in the forecast errors resulting from the model is discussed for a range of applied weighting factors. Finally, the MAE of the model is compared for pre-selected seasonal datasets.

6.1.1. Parametric distribution

The parametric distribution that is fitted on the observed booking increments can have an influence on the performance of the forecasting model. This section analyzes the impact of several distribution fitting approaches on the MAE. Firstly, the results for KQ101 are shown and thereafter the results for KQ512.

KQ101

For flight KQ101 the total set of results (in MAE) of the model calibration on the type of distribution approache for the set of horizons and observation points is tabulated in Table B.1. An example of the sensitivity of the model to the different applied distributions is graphically represented in Figure 6.1 for a horizon of 60 days. This graph shows that the differences in average absolute forecasting error for a forecast made with a bookings horizon of 60 days are small. Towards the end of this forecast horizon the approach not to fit a distribution yields on average a slightly higher error than fitting a parametric distribution on the observed booking increments.

It is chosen to display the results of the 60 day horizon because these represent the results found in the short-term well; small differences and the *no fit* approach yielding the highest error. In a summary of the results per horizon, displayed in Table 6.1, it can be seen that the differences between the distributions is especially small for horizons of 60 days or smaller. In these short-term forecasts the performance of the model in which a lognormal distribution is fitted performs best. For long-term forecasts of a minimum of 60 days fitting no distribution yields the best results for the largest part of the forecast horizon and the relative differences with the parametric distributions fitted to the data are larger than those present in the short-term. Overall, looking at all observation points it is found that the lognormal yielded the lowest error most often. Nevertheless, looking at the sum of all errors over the observation points and horizons fitting no distribution has the lowest sum of errors.

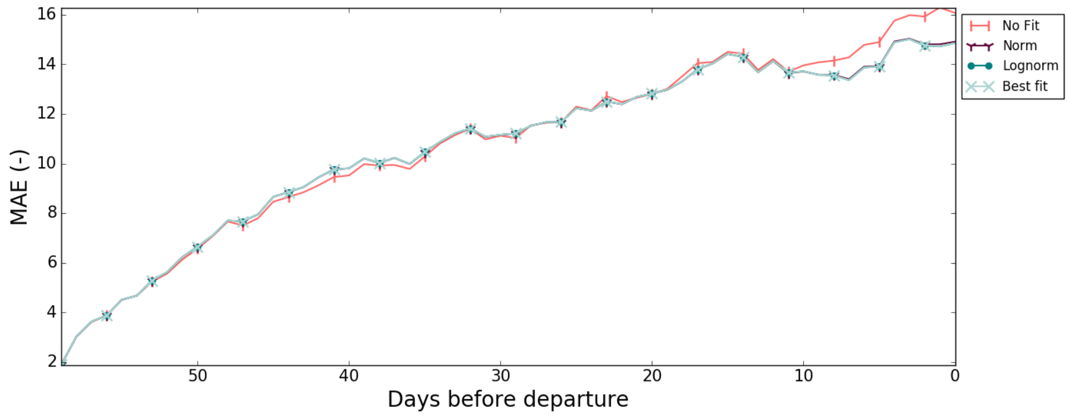


Figure 6.1: The *MAE* of a forecast for flight *KQ101* with a bookings horizon of 60 days for different parametric distributions

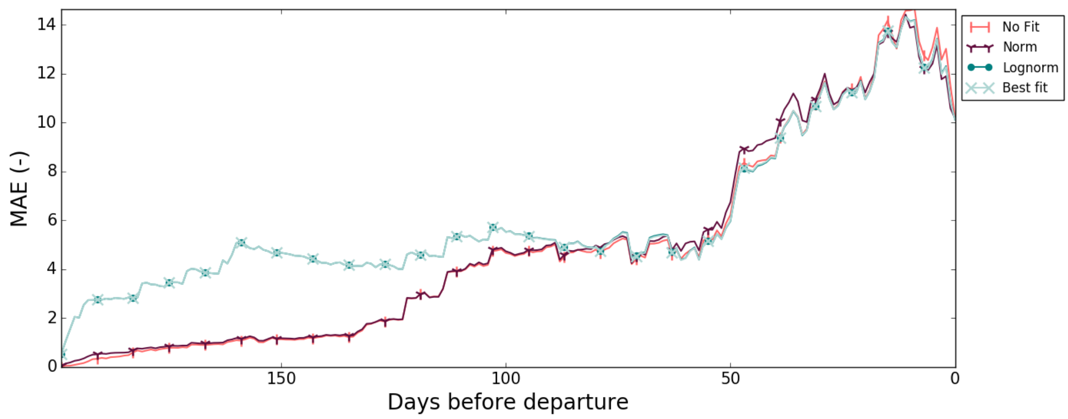


Figure 6.2: The *MAE* of a forecast for flight *KQ512* with a bookings horizon of 200 days for different parametric distributions

Furthermore, for all fitting approaches the forecasting error in the first day of the horizon increases the closer to departure the forecast is created (Table B.1). This can most likely be attributed to the increasing bookings activity closer to the day of departure and therefore a larger variation in the change in bookings in the system over these intervals.

From these results it is concluded that for flight KQ101 the type of distribution used to compute the parametric probability distribution influences the forecasting performance of the model especially in the long-term. Ideally, the optimal distribution at every time-step and for every different forecasting horizons is used. However, for computational considerations only a single fitting approach is considered in this study. Fitting no distribution performs clearly best in the long-term and for the short-term the difference with the parametric distributions is minor. Furthermore, the *no fit* approach is computationally faster as the step of fitting a parametric distribution is omitted in the forecast computation. Computing Z_t without fitting a distribution takes about 12 seconds for all the days in the forecast horizon while the computational time for a normal distribution fitting approach takes 85 seconds and determining the best distribution around 111 seconds. Therefore, for the remainder of the calibration study, no distribution is fitted on the observed data-points for flight KQ101.

KQ512

For flight KQ512 similar tests are performed as for KQ101 in order to analyze the sensitivity of the model to the application of various distribution fitting approaches. The total set of results expressed in MAE for the distribution approaches is tabulated in Table B.2. As an example, the performance of the different approaches to distribution fitting is displayed in Figure 6.2 for the entire bookings horizon of 200 days.

From the figure it can clearly be seen that the MAE of the lognormal and best fit approaches are much larger for the first 100 days of the forecasting horizon. The forecasts deviate in the first days as a result of the distribution fit on the observed bookings increments in these initial intervals that misrepresent the expected process. This lognormal distribution yields on average higher booking increments than the normal distribution. These are found to be too high which results in a deviation in the first part of the forecast horizon. Remarkably, the best fit approach follows the lognormal fitting approach for these forecasts. This implicates that the lognormal distribution fits the data better than the normal distribution in most of the intervals during the bookings process. However, both the normal distribution as just using the actual observed values without fitting yield a better forecasting performance.

Looking at the total test results of the set of distribution fitting approaches for flight KQ512, tabulated in the appendix in Table B.2, it can be seen that the above explained deviations of the lognormal and best-fit approaches are only present for the forecasts performed with an horizon of 300 days and less for an horizon of 200 days. The performance of the short-term forecasts show a much smaller difference between the four fitting approaches as can be seen in the summary of the results per horizon in Table 6.1. For the short-term horizons of 39 days and less the lognormal distribution yields good results looking at the number of times it has the lowest observation points, however, the differences for these horizons are small. This finding is in line with the observations for KQ101 where the lognormal has also been found to perform best in the short-term.

Overall, the lognormal distribution has been found to yield the lowest forecast error for most observation points. Nevertheless, this approach showed an unstable behavior for the forecasts made with a large forecasting horizon and the differences in the short-term forecasts are marginal. Furthermore, it has been found that when taking the sum of all the errors over the observations points the normal distribution and no fitting approaches have in total a lower sum of errors than the other two distributions.

For this flight the same trend can be identified as for flight KQ101, where the error of the first day in the forecast horizon increases for a decreasing forecasting horizon. Nevertheless, especially for the long-term forecasts, the size of the average error over the horizon for KQ512 is significantly smaller than the errors observed for flight KQ101 and the same horizon. This supports the suggestion made in the previous section that this error is related to the variation in booking activity because flight KQ512 is known for a late booking behavior.

From the results explained above, it could be advocated that the approach not to fit a distribution is preferred because it provides good results and is computationally faster. For the remaining of the study, the proposed Markov Chain model is tested using transition matrices that do not use a fitted normal or lognormal distribution on the observed booking increments in the system. Ideally, the model uses the optimal distribution fitting approach for every horizon and every bookings interval. Due to time constraints optimizing the distribution fitting over the complete bookings process is not considered in this study.

Subconclusion

From the tests for both KQ101 and KQ512 it can be concluded that in short-term forecasts the performance of the model is marginally dependent on the type of distribution fitted to the observed differences in bookings numbers in an interval. The lognormal distribution provides the lowest MAE for these short-term forecasts of maximum 60 days. However, the lognormal approach for flight KQ512 resulted in a large absolute forecasting error for the long-term forecasts. This unstable behavior is a negative characteristic of this fitting approach. Furthermore, the relative differences in the long-term are larger and fitting no distribution is found to perform superior for those long horizons. Given that the normal distribution fitting does not result in a major improvement in forecasting performance over the *No Fit* approach and the higher computational intensiveness, the *No Fit* is defined to be the best approach for both flight *KQ101* and *KQ512* for the remaining experiments in this study.

6.1.2. Weighting factor

The weighting factor α determines the relative weight that is given to the empirical and parametric distributions which are used to calculate the entries in the transition matrices. The higher the value of α the more weight is appointed to the empirical distribution. The tests are performed for the flights, horizons and weights of α following the experimental set-up described in 5.4. Based on the results of the previous section, the performance is analyzed without a distribution fitted to the observed booking increments.

KQ101

For flight KQ101 the total set of results expressed in MAE of the calibration on the weighting factor for the complete set of horizons and observation points is tabulated in Table B.3. An example of the relative performance of the model settings for a forecast made a 100 days in advance of departure day can be seen in Figure 6.3.

Table 6.1: Sum of MAE for the different distribution settings across the observation points of the set horizons for KQ101 and KQ512

		Flight			
		KQ101		KQ512	
Horizon	Distribution	Sum of MAE	% compared with lowest	Sum of MAE	% compared with lowest
300	No Fit	225.0	-	43.4	-
	Norm	251.2	+11.7	48.3	+11.3
	Lognorm	249.7	+11.0	644.5	+1386.3
	Best	251.7	+11.9	643.3	+1383.5
200	No Fit	164.1	-	49.8	-
	Norm	188.5	+14.9	51.7	+3.7
	Lognorm	187.8	+14.4	62.4	+25.2
	Best	188.9	+15.1	62.4	+25.2
100	No Fit	129.4	-	70.9	+0.1
	Norm	134.7	+4.1	71.0	+0.3
	Lognorm	134.7	+4.1	70.8	-
	Best	134.8	+4.2	71.0	+0.3
60	No Fit	115.0	+1.6	83.7	+0.4
	Norm	113.2	+0.1	83.4	-
	Lognorm	113.2	-	83.4	+0.0
	Best	113.2	+0.0	83.4	+0.0
39	No Fit	100.4	+5.0	94.3	+5.2
	Norm	95.9	+0.3	89.8	+0.2
	Lognorm	95.6	-	89.6	-
	Best	95.6	+0.0	89.7	+0.0
25	No Fit	70.2	+3.8	65.6	+2.6
	Norm	67.8	+0.3	64.1	+0.2
	Lognorm	67.6	-	64.0	-
	Best	67.6	+0.0	64.0	+0.0
14	No Fit	56.3	+2.4	53.4	+1.7
	Norm	55.1	+0.2	52.6	+0.2
	Lognorm	55.0	-	52.5	-
	Best	55.0	+0.0	52.5	+0.0
7	No Fit	36.3	+1.3	38.5	+0.9
	Norm	35.9	+0.1	38.2	+0.1
	Lognorm	35.9	-	38.1	-
	Best	35.9	+0.0	38.1	+0.0

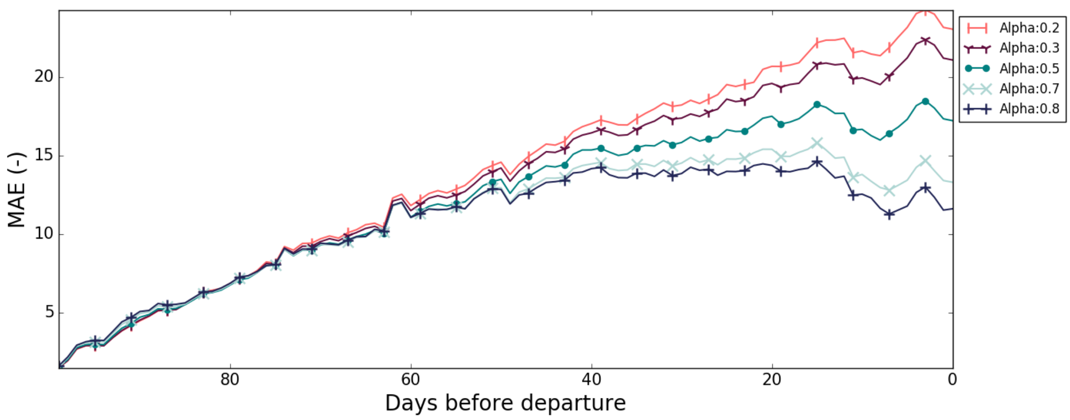


Figure 6.3: The *MAE* of a forecast for flight *KQ101* with a bookings horizon of 100 days for different values of α

From the graph it can be seen that for the booking horizon of 100 days the forecasting model with a high weighting factor results in a superior performance compared to a low weight. This difference increases closer to the day of departure. However, for the first days in the forecast horizon the model performs slightly better for low values of α.

From the complete set of results in Table B.3 it can be seen that for most forecasting horizons the same trend is visible; the first days after the forecasting day show a positive performance for the lower values of α and closer to the day of departure a higher value of α is preferred. Overall, an α of 0.8 resulted in the lowest error most of the times. Looking at the summary of the results expressed in the sum of MAE over the horizon, tabulated in Table 6.2, it can be seen that an α of 0.8 performs best for most horizons, except for the shortest horizon of 7 days for which an α of 0.7 yielded the lowest total error. Especially for the medium-term forecasts of 100 to 25 days the high weighting factor performs significantly better than the lower ones.

Based on these results it can be concluded that appointing a relatively high weight to the empirical dis-

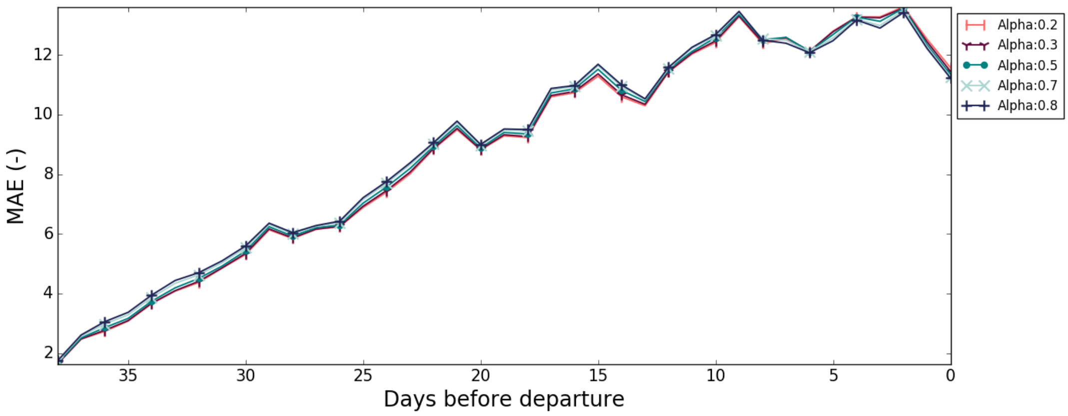


Figure 6.4: The *MAE* of a forecast for flight *KQ512* with a bookings horizon of 39 days for different values of α

tribution yields desirable results for this flight. This means that the performance of the model is positively influenced by the specific instances where the same bookings number occurred in a certain time interval. Especially in the medium-term forecasts this significantly enhances the performance. Therefore, for the remaining of the study an α of *0.8* has been used for the further assessment of the forecasting performance of the model for this flight.

KQ512

For KQ512 the performance is also expressed in the average absolute error measurement and assessed over the entire bookings horizon similar as for flight KQ101. The total set of results expressed in MAE for the different weighting factors is tabulated in Table B.4. An example of the relative performance of the model for different values of α can be seen in Figure 6.4.

For the forecasts that were made with a forecasting horizon of 39 days for flight KQ512 the differences in performance are small. The lowest weighting factor of 0.2 shows a slightly better result for the first part of the horizon and a worse forecast closer to the day of departure. For the high weighting factors the opposite performance can be observed over the horizon. Therefore, it can not be concluded, for this specific bookings horizon, if appointing a high weight to the empirical distribution gives a better representation of the expected change in bookings number.

Looking at the summary of the results in Table 6.2 it can be seen that the differences in performance for the weighting factors is small for all horizons. For the long-term forecasts the higher weighting factors yield the lowest forecast error. In the medium- and short-term the lower weighting factors are performing slightly better. Nevertheless, as can be seen in Table B.4, the higher weighting factors yield lower errors towards the end of these horizons. Therefore, the best value of the weighting factor α can not easily be determined.

Overall, looking at all observation points, the lowest weighting factor yielded the lowest MAE most often. However, looking at the relative performance in the sum over the horizons, a value of 0.5 provides good results overall, where the low weighting factors perform worse in the long-term and high weighting factors worsee in the short-term. Nevertheless, the difference in total sum of MAE is small.

Following from the observation that both low and high weighting factors yield the lowest forecasting error for some observation points and the slight difference in the total sum of the errors, it can be concluded that a moderate weighting factor is desirable. By having a closer look on the results of the tests with a weighting factor of 0.5, it can be seen that the differences with the lower value of α (0.2 and 0.3) are small and it is much closer to the positive results of the high weighting factors towards the end of a forecast horizon.

Similarly as to the results of flight KQ101 the value of the weighting factor α should ideally be optimized for every single forecast horizon and bookings interval. However, due to time constraints and computational considerations, one single α is applied for the remaining experiments in this study. For flight KQ512 a value of α equal to 0.5 has been chosen in order to have a model that performs well both in the short- and long-term.

Subconclusion

From the second part of this experiment it can be concluded that for flight KQ101 the value of the weighting factor α has a higher influence on the performance of the forecasting model than the type of distribution fitting approach used to compute the parametric distribution. For KQ512, however, the difference in performance for α is much smaller. For flight KQ101 the higher values of α clearly performed superior to the lower weighting factors, while for KQ512 no clear *winner* could be found. Looking at the sensitivity of the model to the weight α for both flights, it can be concluded that a high weighting factor is desirable towards the end of a bookings horizon. However, besides this similarity, it can be concluded that the beneficial aspect of incorporating the empirical distribution differs per flight. Therefore, the best α should be determined for all flights separately before a forecast is computed.

Table 6.2: Sum of MAE for the different values of α across the observation points of the set horizons for KQ101 and KQ512

		Flight			
		KQ101		KQ512	
Horizon	α	Sum of MAE	% compared with lowest	Sum of MAE	% compared with lowest
300	0.2	235.2	+6.4	44.4	+3.5
	0.3	230.9	+4.5	44.0	+2.5
	0.5	225.0	+1.8	43.4	+1.1
	0.7	221.7	+0.3	43.0	+0.3
	0.8	221.0	-	42.9	-
200	0.2	176.4	+11.9	50.1	+0.6
	0.3	171.5	+8.9	49.9	+0.3
	0.5	164.1	+4.1	49.8	+0.0
	0.7	159.1	+0.9	49.8	-
	0.8	157.6	-	49.8	+0.0
100	0.2	149.4	+31.9	71.0	+0.1
	0.3	142.4	+25.6	71.0	+0.1
	0.5	129.4	+14.1	70.9	-
	0.7	118.0	+4.1	71.1	+0.3
	0.8	113.3	-	71.3	+0.5
60	0.2	134.2	+34.9	81.8	-
	0.3	127.7	+28.4	82.4	+0.7
	0.5	115.0	+15.7	83.7	+2.2
	0.7	103.6	+4.2	85.1	+4.0
	0.8	99.4	-	85.8	+4.9
39	0.2	120.5	+42.1	93.7	-
	0.3	113.7	+34.0	93.9	+0.1
	0.5	100.4	+18.4	94.3	+0.6
	0.7	89.0	+5.0	94.7	+1.1
	0.8	84.8	-	94.9	+1.2
25	0.2	78.9	+21.3	65.2	-
	0.3	75.8	+16.4	65.3	+0.1
	0.5	70.2	+7.8	65.6	+0.6
	0.7	66.3	+1.9	66.3	+1.6
	0.8	65.1	-	66.7	+2.3
14	0.2	60.6	+10.6	55.3	+3.5
	0.3	58.9	+7.5	54.3	+1.6
	0.5	56.3	+2.6	53.4	-
	0.7	55.0	+0.3	53.7	+0.6
	0.8	54.8	-	54.4	+1.9
7	0.2	37.6	+3.6	38.6	+0.5
	0.3	37.0	+2.0	38.4	-
	0.5	36.3	+0.1	38.5	+0.1
	0.7	36.3	-	39.2	+2.0
	0.8	36.5	+0.6	39.8	+3.7

6.1.3. Seasonal input

In section 5.3 it is discussed that some seasonality has been observed in the bookings data. It is whether the model yields better results when a specifically chosen dataset is used as an input. This section describes the tests performed with a monthly seasonal dataset, data of the same DOW and a set where the data has to fulfill both these characteristics. The tests are performed according to the experimental set-up discussed in 5.4.

KQ101

For flight KQ101 the results of the model expressed in MAE with different datasets for the complete set of horizons and observation points is tabulated in Table B.5. As an example, the performance measured in MAE for flight KQ101 with different datasets and an horizon of 60 days is displayed in Figure 6.5. In this figure it can be seen that both the total dataset and the monthly seasonal data yield on average the lowest forecast error. Especially closer towards the day of departure the difference with the DOW and DOW-monthly combined dataset increases. This can be stated to be a remarkable result as it implies that using only a dataset of flights that departed at a similar day does not necessarily lead to the most accurate forecast.

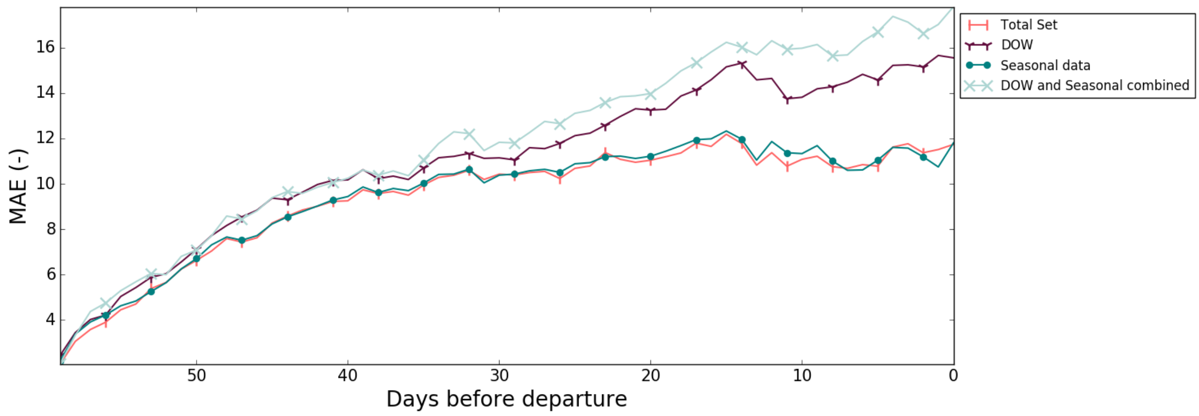


Figure 6.5: The *MAE* of a forecast for flight *KQ101* with a bookings horizon of 60 days for different datasets

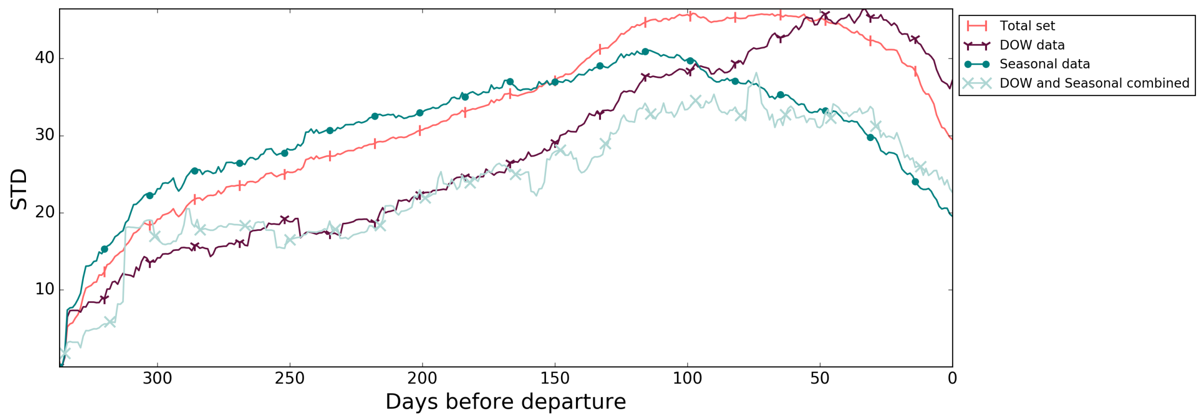


Figure 6.6: The STD of the datasets of bookings for KQ101

Looking at the summary of the results per horizon in Table 6.3 it can be seen that this does not hold for all the forecast horizons. For the long-term forecasts of 300 and 200 days the average forecasting errors of the monthly seasonal data and the DOW-monthly combined are lowest and do not differ significantly. Nevertheless, for more short-term forecasts the performance of the model with the DOW-Monthly dataset deteriorated and the performance of the total dataset improves. For the shortest horizon of 7 days the DOW dataset also yields a low error with a small difference compared to the total and monthly seasonal set.

The datasets considered all have a different variation. The STD of the four datasets for KQ101 over the entire length of the dataset are visualized in Figure 6.6. It can be seen that all these datasets show a high STD with maxima ranging from around 30 to 45 bookings.

Overall, it is observed that the total dataset yielded the lowest error most often across all observation points and horizons. By summing the errors over the observation points however, the monthly dataset shows to have the lowest total error and best relative performance. It can be stated that, looking at the MAE of all different model tests, the seasonal input provides the best results and is the desired dataset for flight KQ101.

KQ512

The total set of results expressed in MAE for the different datasets is tabulated in Table B.6. As an example the performance of the model measured in MAE for flight KQ512 at 39 DBD and different datasets is displayed in Figure 6.7. For this horizon it can be seen that when data is used of departures in the same DOW as the flight for which is forecasted, the model produces the best results. One would expect that the most specific dataset, which is both monthly and same DOW data combined, yields the best forecast performance. Nevertheless, the error when using this combined dataset is large relatively to the other methods.

The datasets used in these forecasts all have a different variation. The STD of the four datasets for KQ512 over the entire length of the dataset are visualized in Figure 6.8. It can be seen that all these datasets show a STD with maxima ranging from around 13 to 18 bookings. Furthermore, the standard deviation in the data

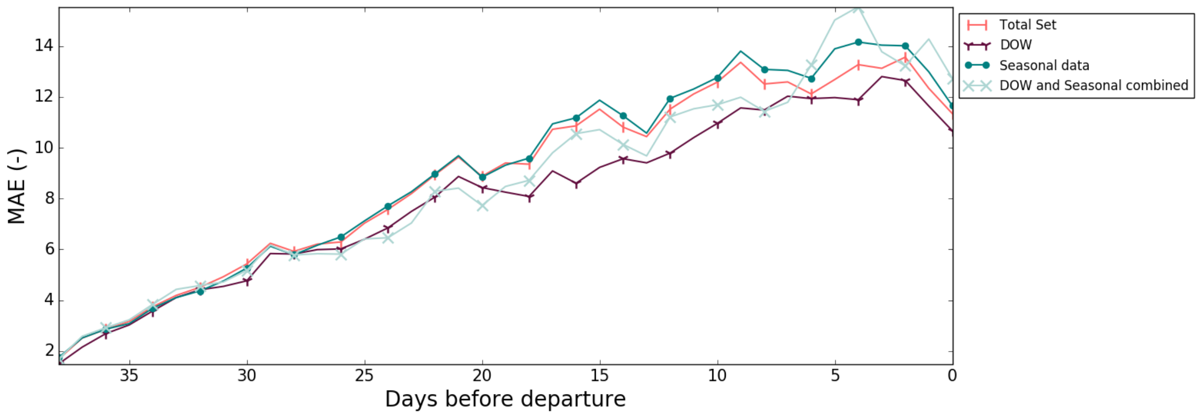


Figure 6.7: The *MAE* of a forecast for flight *KQ512* with a bookings horizon of 39 days for different datasets

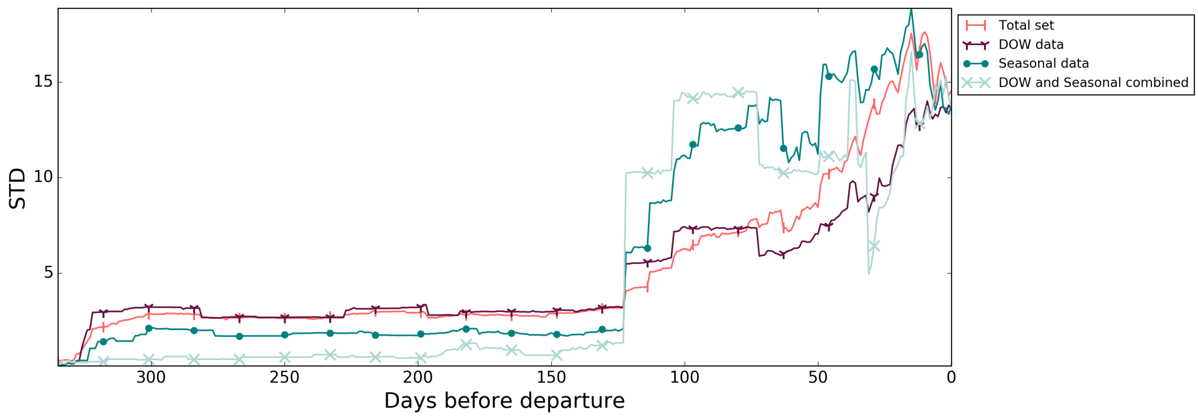


Figure 6.8: The STD of the datasets of bookings for KQ512

is relatively small and stable until around 120 days before departure. This corresponds with the low booking activity for this flight far in advance of departure.

In the summary of the results per horizon, shown in Table 6.3, it can be seen that overall the day-of-week data can be considered to be the dataset that gives the lowest mean absolute error. For all horizons, except for the two most short-term (7 & 14 days) where the total dataset yields the lowest error, the DOW data performs best. It is also observed that the forecasts resulting from this dataset have the lowest error for most observation points. Furthermore, over all horizon the relative performance of the DOW data is best. Especially for the medium-term forecasts (100-39 DBD) this dataset provides the best results relative to the other datasets.

6.1.4. Subconclusion

Altogether, it can be concluded from the calibration based on the MAE that for KQ101 the type of distribution fit to the model has less influence on the performance of the model than the weighting factor α and the dataset that is fed as an input to the model. For KQ512, however, the weighting factor has been found to influence the performance minor while larger differences in performance for the different datasets have been observed and significant differences for the distribution fitting approach in the long-term.

Both for flight KQ101 and KQ512 fitting no distribution to the observed booking increment was found to be the desired approach based on the minor differences in performance with the other approaches and the advantage of a faster computation.

The best value of α is different for the two flights. Whereas a high weighting factor attributed to the empirical distribution yielded good results for KQ101, a moderate weighting factor was beneficial for KQ512. From there it can be concluded that the best weighting factor should be separately determined for all flight legs for which a forecast is to be made. For the remaining experiments in this study a value of 0.8 is applied to the forecasts made for KQ101 and a value of 0.5 for the forecasts for KQ512.

Table 6.3: Sum of MAE for the different datasets across the observation points of the set horizons for KQ101 and KQ512

		Flight			
		KQ101		KQ512	
Horizon	Dataset	Sum of MAE	% compared with lowest	Sum of MAE	% compared with lowest
300	Total Set	221.0	+6.3	43.4	+4.9
	DOW	219.6	+5.6	41.3	-
	Seasonal	208.0	-	42.4	+2.6
	Month-DOW	211.5	+1.7	45.0	+8.9
200	Total Set	157.6	+10.6	49.8	+3.3
	DOW	164.2	+15.2	48.2	-
	Seasonal	142.5	-	50.0	+3.8
	Month-DOW	152.4	+6.9	54.3	+12.7
100	Total Set	113.3	+0.6	70.9	+9.1
	DOW	131.1	+16.4	65.0	-
	Seasonal	112.6	-	71.9	+10.7
	Month-DOW	144.0	+27.9	76.2	+17.3
60	Total Set	99.4	-	83.7	+8.8
	DOW	118.5	+19.2	76.9	-
	Seasonal	100.7	+1.3	83.9	+9.1
	Month-DOW	127.0	+27.7	84.4	+9.7
39	Total Set	84.8	-	94.3	+9.7
	DOW	101.0	+19.1	85.9	-
	Seasonal	90.7	+7.0	96.3	+12.0
	Month-DOW	109.6	+29.3	91.5	+6.5
25	Total Set	65.1	-	65.6	+4.1
	DOW	71.8	+10.3	63.1	-
	Seasonal	69.1	+6.2	69.5	+10.2
	Month-DOW	78.6	+20.8	74.9	+18.8
14	Total Set	54.8	+0.8	53.4	-
	DOW	58.2	+7.0	54.0	+1.1
	Seasonal	54.4	-	55.7	+4.3
	Month-DOW	63.7	+17.1	69.1	+29.3
7	Total Set	36.5	-	38.5	-
	DOW	36.7	+0.6	38.5	+0.1
	Seasonal	36.6	+0.3	42.1	+9.3
	Month-DOW	40.3	+10.4	50.1	+30.1

The optimal value of α varies over the forecast horizon and is different for the multiple observation points as well. Therefore, the ideal case would be to optimally determine the weighting factor for all these instances. This is not considered to be a part of this study due to time constraints.

The average forecasting error of the model is found to be dependent on the input dataset. Both for flight KQ101 and KQ512 the difference in performance is significant. For flight KQ101 the method with the lowest error varied over the forecast horizon, nevertheless the monthly seasonal data yielded the best results overall. For flight KQ512 the differences in performance between the datasets is smaller, however the DOW dataset consistently yielded the lowest error.

6.2. Model calibration for optimal prediction interval scores

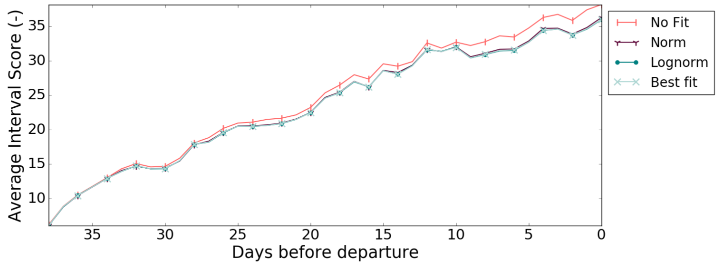
The second phase of the study assesses the ability of the model to correctly include the uncertainty of the forecast in the prediction. This is expressed in terms of interval scores which was explained in the experimental set-up in Section 5.4.4. The tests are performed for several model settings; varying values for the weighting factor α, the type of distribution fitting and the dataset used as an input to the model. This is tested for both flights KQ101 and KQ512 and the same forecasting horizons as in the previous part of this study.

6.2.1. Parametric distribution

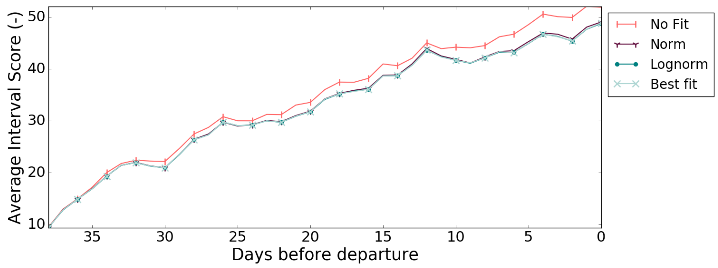
The distribution fitting approach affects the entries in the transition matrices. Therefore, the distribution that is fit on the data influences the probability distribution that results from the forecast. In this section it is discussed what parametric fitting approach leads to the best representation of the uncertainty in the forecast. This is determined for both flights KQ101 and KQ512.

KQ101

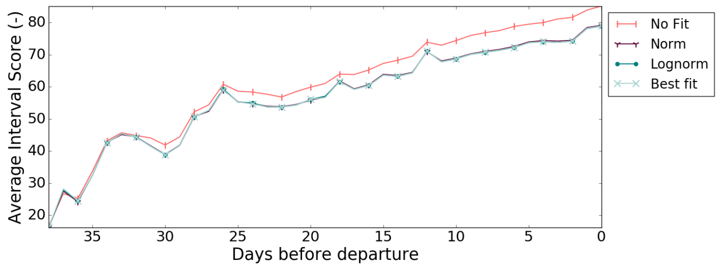
For flight KQ101 the complete set of results in Prediction Interval Score (PIS) of the tests on the distribution approaches for the set of horizons and observation points is tabulated in Table B.7. An example of the interval scores for flight KQ101 with a forecast horizon of 39 days is shown in Figure 6.9 for varying distributions. In this graph it can be seen that fitting no distribution yields the highest interval scores. This is the case for the 33%, 67% and 95% intervals. However, the first days of these forecasts the interval score of the normal



(a) 33% Prediction Interval Scores



(b) 67% Prediction Interval Scores



(c) 95% Prediction Interval Scores

Figure 6.9: The interval scores of a forecast for flight *KQ101* with a bookings horizon of 39 days for the different distribution fitting approaches

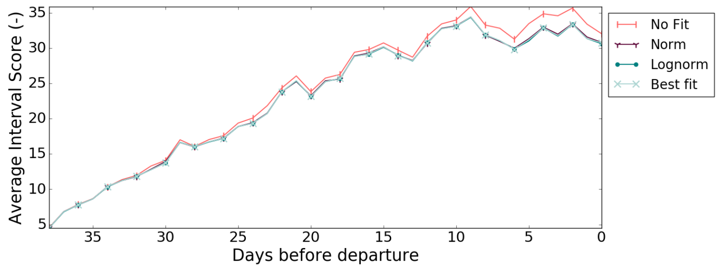
and no-fit distribution approaches yield the lowest scores. These scores indicate that especially for the 95% intervals the width of the interval is large with interval scores around 80. This implies that the model expects that a large range of bookings are possible to occur in the future. This is in line with the large variation that has been observed for flight KQ101.

From the summary of results shown in Table 6.4 it can be seen that this trend is not visible for all horizons. Especially in the long-term, fitting no distribution yields the lowest interval scores. From a horizon of 60 days or less the best-fit and lognormal distribution yield the lowest interval scores. However, it has been found that the first days of these short-term forecasts the interval score of the normal and no-fit distribution approaches yield the lowest scores, similar as what can be seen in Figure 6.9. Over all observation points and horizons it is found that fitting no distribution yielded the lowest score most often. Moreover, over all horizons the sum of the interval scores is lowest for this *No-fit* approach compared with the parametric fitting approaches, although the differences are small. From there it can be concluded that, from an uncertainty modeling perspective, it is desirable not to fit a distribution on the observed booking increments.

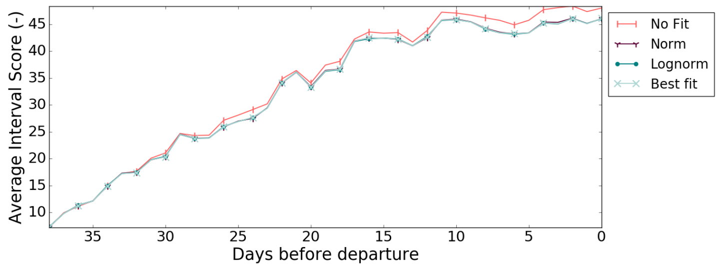
KQ512

The total setresults of the interval scores for the different distributions is tabulated in Table B.8. As an example, the interval scores for flight KQ512 with a forecasting horizon of 39 days and various fitting approaches are displayed in Figure 6.10.

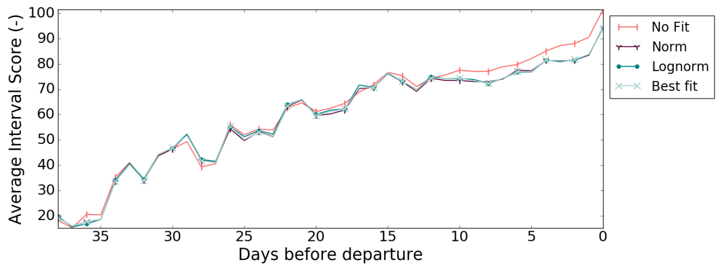
It can be seen that the differences are relatively small for the three prediction intervals. Especially in the first days of the horizon the interval scores are to a large extend similar. All three prediction intervals show the largest difference closer to the day of departure. In these last days the approach of fitting no distribution results in a slightly higher interval score than the other approaches. Solely based on this horizon one may



(a) 33% Prediction Interval Scores



(b) 67% Prediction Interval Scores



(c) 95% Prediction Interval Scores

Figure 6.10: The interval scores of a forecast for flight *KQ512* with a bookings horizon of 39 days for the different distribution fitting approaches

conclude that fitting a parametric distribution on the observed bookings increments positively influences the representation of the uncertainty in the forecast by the model.

Nevertheless, this trend is not visible for all forecasting horizons as can be seen in the summary of the results in Table 6.4. In the long-term the forecasts where no distribution was fit on the data yielded the lowest interval scores. Furthermore, the deviation in the forecasts made with the lognormal and best fit approaches, which has also been discussed in Section 6.1, leads to high interval scores for the forecasts made with an horizon of 300 days. This is the reason why over all horizons the scores for these two approaches are three times as high as the normal and $no-fit$ distribution. In the short-term the lognormal approach yields the lowest interval score, however, the differences in the relative performance are small. Over all horizons and observation points it is found that the no-fitting and fitting a lognormal distribution yielded the lowest score most often. The total sum of the interval scores over all horizons does not differ much between these two approaches, but the *no-fit* approach yielded the lowest total interval score. Because fitting no distribution yields the best result slightly more often and the approach is computationally the fastest, the approach of fitting no distribution can be concluded to be the preferred one to model the demand uncertainty.

Subconclusion

Looking at the sensitivity of the average absolute error on the fitting approach, discussed in Section 6.1, fitting no distribution yielded the lowest error. From the analysis of the ability of the model to incorporate the uncertainty of the forecast in this section it is also found that fitting no distribution is the desired approach. Both for KQ101 and KQ512 the *no-fit* approach yielded especially good results in the long-term, whereas the lognormal distribution provided the lowest interval scores in the short-term. This is similar to the findings of the MAE for the different distributions. However, because the differences are small in the short-term, fitting no distribution performs best overall. From there it can be concluded that fitting no distribution is the best setting of the model both from an absolute error and uncertainty modeling perspective.

Table 6.4: Sum of interval scores for the different distributions across the observation points of the set horizons for KQ101 and KQ512

		Flight			
		KQ101		KQ512	
Horizon	Distribution	Sum of P.I.S.	% compared with lowest	Sum of P.I.S.	% compared with lowest
300	No Fit	3111.9	-	791.8	-
	Norm	3506.6	+12.7	807.5	+2.0
	Lognorm	3488.4	+12.1	16506.9	+1984.7
	Best	3512.7	+12.9	16447.3	+1977.1
200	No Fit	2274.1	-	943.3	-
	Norm	2483.5	+9.2	972.6	+3.1
	Lognorm	2473.4	+8.8	969.5	+2.8
	Best	2484.2	+9.2	974.1	+3.3
100	No Fit	1740.9	-	1054.3	-
	Norm	1763.1	+1.3	1095.5	+3.9
	Lognorm	1762.9	+1.3	1092.6	+3.6
	Best	1763.4	+1.3	1095.4	+3.9
60	No Fit	1568.2	+3.2	1244.9	-
	Norm	1521.0	+0.1	1277.8	+2.6
	Lognorm	1519.9	-	1276.5	+2.5
	Best	1520.0	+0.0	1276.9	+2.6
39	No Fit	1297.9	+5.9	1278.5	+3.8
	Norm	1229.0	+0.3	1231.1	-
	Lognorm	1225.0	-	1233.8	+0.2
	Best	1225.2	+0.0	1231.7	+0.1
25	No Fit	910.8	+4.1	923.9	+2.7
	Norm	876.0	+0.2	899.9	+0.0
	Lognorm	874.6	-	899.9	-
	Best	874.7	+0.0	900.7	+0.1
14	No Fit	763.0	+2.2	736.8	+3.3
	Norm	750.5	+0.5	716.1	+0.4
	Lognorm	748.7	+0.2	713.5	-
	Best	747.0	-	714.2	+0.1
7	No Fit	496.9	+0.1	561.2	+3.5
	Norm	497.7	+0.3	543.3	+0.2
	Lognorm	496.4	+0.0	542.3	+0.0
	Best	496.3	-	542.3	-

6.2.2. Weighting factor

The weighting factor α balances the influence between the empirical and parametric distribution in the determination of the transition matrix entries. The value of this weighting factor influences the probability distribution that results from the forecasting model. Therefore, the accuracy of the uncertainty representation by the proposed model is assessed for varying values of α, both for KQ101 and KQ512.

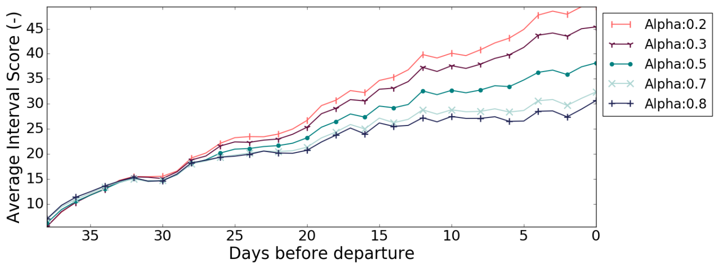
KQ101

For flight KQ101 the total set results in PIS of the calibration of the weighting factor for the set of horizons and observation points is tabulated in Table B.9. As an example, the interval scores for flight KQ101 with a forecasting horizon of 39 days are shown in Figure 6.11 for the varying value of the weighting factor α. From the graph it can be seen that higher weighting factors lead to a lower interval score, especially for the second half of the horizon. Nevertheless, for the first days of the horizon the lower values of α are performing better.

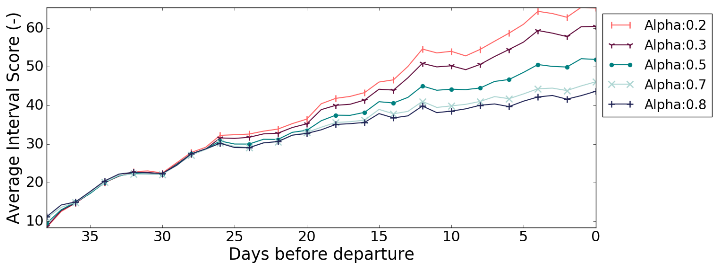
This trend is visible for all three forecasting intervals and also for all other horizons, except for the last horizon of 7 days. The first days of the horizon the lower weighting values are consistently giving lower interval scores. However, looking at the summary of the results for each forecast horizon, tabulated in Table 6.5 the higher weighting factors are performing superior. For all horizons, except for the most short-term forecast of 7 days, the weighting factor of 0.8 consistently yields the lowest error.

KQ512

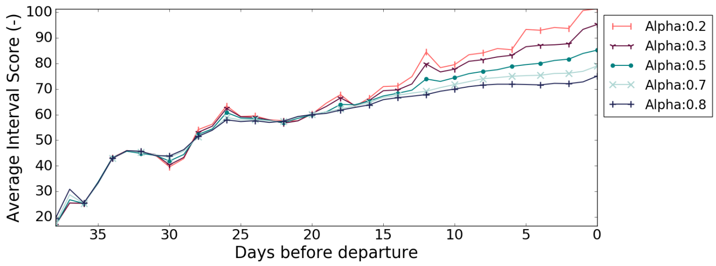
The total set of results of the interval scores for the different weighting factors α are tabulated in Table B.10. As an example, the interval scores for flight KQ512 with a forecasting horizon of 39 days and different weighting factors are displayed in Figure 6.12. From the graph it becomes apparent that for this horizon the forecasts made with high weighting factors model the uncertainty better than the lower values of α. This becomes especially apparent towards the end of the bookings horizon. Nevertheless, for the first days of the horizon the lower weighting factors yield slightly lower interval scores and therefore a better representation of the



(a) 33% Prediction Interval Scores

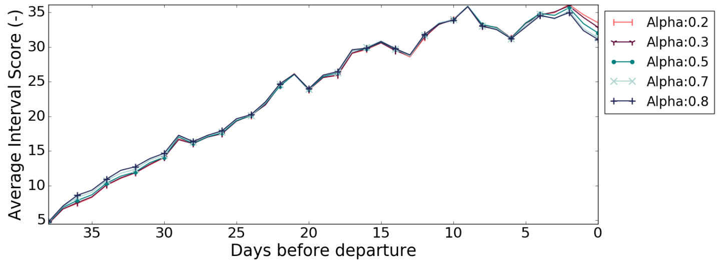


(b) 67% Prediction Interval Scores

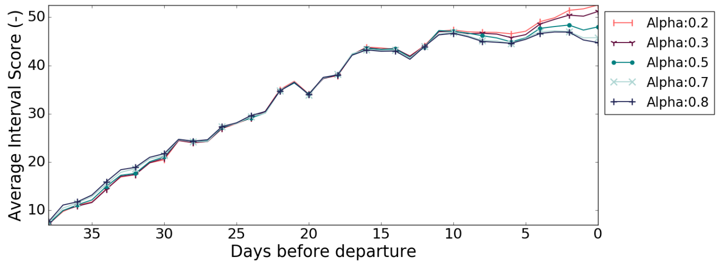


(c) 95% Prediction Interval Scores

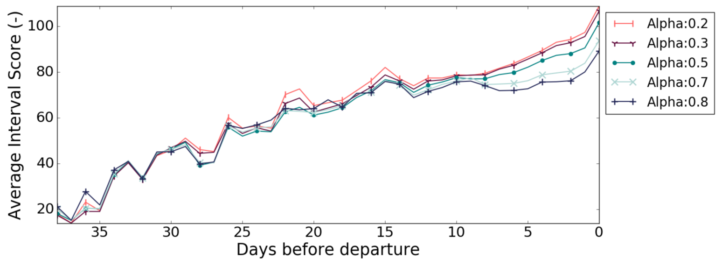
Figure 6.11: The interval scores of a forecast for flight *KQ101* with a bookings horizon of 39 days for varying values of α



(a) 33% Prediction Interval Scores



(b) 67% Prediction Interval Scores



(c) 95% Prediction Interval Scores

Figure 6.12: The interval scores of a forecast for flight *KQ512* with a bookings horizon of 39 days for varying values of α

uncertainty in the first days after the prediction is made. The differences in interval scores for the prediction intervals is smaller for the 33% and 67% intervals in comparison with the 95% prediction interval.

The summary of the results, tabulated in Table 6.5, shows that the high weighting factors consistently yield the lowest interval scores for the set of forecast horizons, except for the most short-term forecast of 7 days. The lower weighting factors have been found to yield the best scores at the beginning of the forecast horizon. These findings are similar to the performance for the different weighting factors for KQ101. Furthermore, the high weighting factors show to be superior especially for long-term forecasts where the relative performance with the other values of α is largest. In the short-term forecasts the differences between the settings is smaller. The superior performance is mostly present for the 95% and 67% intervals and less for the 33% prediction interval.

Over all observation points and horizons it is found that the interval scores for a model that applies a weighting factor of 0.8 are most often the smallest. Moreover, looking at the total sum of the interval scores over all horizons the sum of scores is also smallest for the weighting factor of 0.8. From there it could be concluded that a high weighting factor should be the preferred model setting for this flight. However, it should be noted that for the first days in the horizon the low weighting factor provides the best results. The ideal scenario would therefore be to determine the optimal α over the entire forecast horizon.

Subconclusion

For KQ101 the findings of the calibration of the model on the weighting factor α are similar to the performance in terms of MAE discussed in Section 6.1. In that sector a higher α leads to superior results closer towards the day of departure and worse results for the first days of the forecast horizon. This trend is also visible when looking at the quality of the forecasting distributions. For KQ101 the higher weighting factors show a clear superior performance overall, which also corresponds to the desired weighting factor value resulting from the first part of this study. For flight KQ512 a high weighting factor also result in the best uncertainty modeling. The lowest weighting factor yields the lowest score a significant number of times, but the high weighing factors can be stated to be superior based on an observation of the sum of interval scores over all horizons. This result is, however, inconsistent with the performance expressed in MAE where moderate weighting factors were found to be best setting for the model. This implicates that the model should be tested and calibrated separately for performance in MAE and interval scores and a trade-off should be made to obtain the best overall setting of the model.

Table 6.5: Sum of interval scores for the different values α across the observation points of the set horizons for KQ101 and KQ512

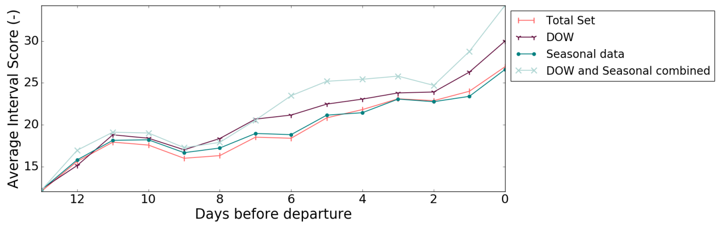
		Flight			
		KQ101		KQ512	
Horizon	α	Sum of P.I.S.	% compared with lowest	Sum of P.I.S.	% compared with lowest
300	0.2	3792.4	+30.8	853.7	+13.4
	0.3	3464.3	+19.5	829.7	+10.2
	0.5	3111.9	+7.3	791.8	+5.2
	0.7	2948.1	+1.7	767.4	+2.0
	0.8	2898.9	-	752.7	-
200	0.2	2447.5	+10.2	990.2	+9.3
	0.3	2354.5	+6.1	973.3	+7.4
	0.5	2274.1	+2.4	943.3	+4.1
	0.7	2226.6	+0.3	922.0	+1.8
	0.8	2220.2	-	905.9	-
100	0.2	1888.1	+13.6	1156.8	+12.2
	0.3	1821.3	+9.6	1117.7	+8.4
	0.5	1740.9	+4.7	1054.3	+2.3
	0.7	1681.9	+1.2	1039.7	+0.9
	0.8	1662.1	-	1030.7	-
60	0.2	1730.0	+16.4	1265.3	+3.4
	0.3	1668.3	+12.3	1257.1	+2.8
	0.5	1568.2	+5.6	1244.9	+1.8
	0.7	1504.7	+1.3	1228.4	+0.4
	0.8	1485.7	-	1223.3	-
39	0.2	1460.9	+21.3	1317.5	+4.8
	0.3	1398.1	+16.1	1301.4	+3.5
	0.5	1297.9	+7.8	1278.5	+1.7
	0.7	1232.6	+2.4	1259.8	+0.2
	0.8	1204.0	-	1257.4	-
25	0.2	980.1	+13.0	955.0	+6.4
	0.3	953.2	+9.9	943.9	+5.2
	0.5	910.8	+5.0	923.9	+2.9
	0.7	879.3	+1.4	906.9	+1.0
	0.8	867.4	-	897.5	-
14	0.2	796.7	+5.7	756.5	+4.2
	0.3	782.8	+3.9	747.4	+2.9
	0.5	763.0	+1.3	736.8	+1.5
	0.7	754.5	+0.1	727.7	+0.2
	0.8	753.4	-	726.2	-
7	0.2	507.6	+2.2	563.7	+0.7
	0.3	503.0	+1.2	564.6	+0.9
	0.5	496.9	-	561.2	+0.3
	0.7	502.4	+1.1	559.8	-
	0.8	510.7	+2.8	562.6	+0.5

6.2.3. Seasonal input

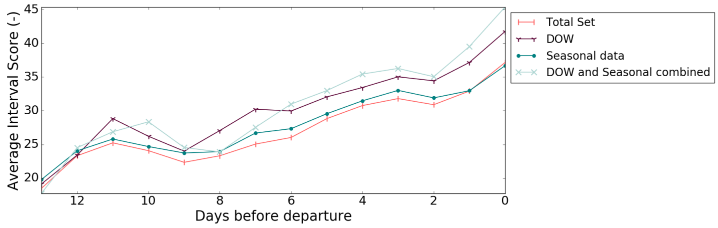
The selectively chosen seasonal dataset that is used by the forecasting model to compute the expectation of the number of bookings in the future has another variation in booking increments than, for example, the total dataset. For that reason it is tested what the influence is of the input data on the ability of the model to accurately cover the forecast uncertainty.

KQ101

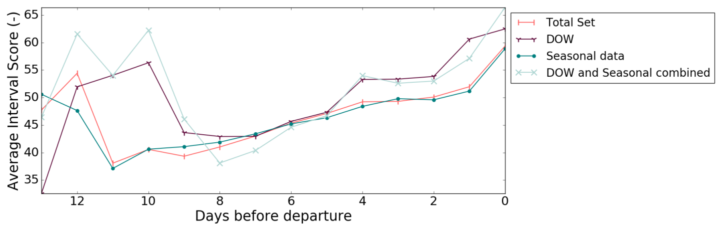
For flight KQ101 the total set of results expressed in PIS of the calibration on the different datasets used for the set of horizons and observation points is tabulated in Table B.11. As an example, the average interval scores for flight KQ512 with an horizon of 14 days are displayed in Figure 6.13 for the different sets of data that serve as an input to the proposed Markov Chain model. In this figure it can be seen that both the seasonal and total dataset yield low interval scores. This holds for all three prediction intervals that are visible in the graph. For the 95% prediction interval a large difference in interval scores is observed for the first days of the horizon. The forecasts based on the DOW-data are the only forecasts that are able to cover the uncertainty in the first



(a) 33% Prediction Interval Scores



(b) 67% Prediction Interval Scores



(c) 95% Prediction Interval Scores

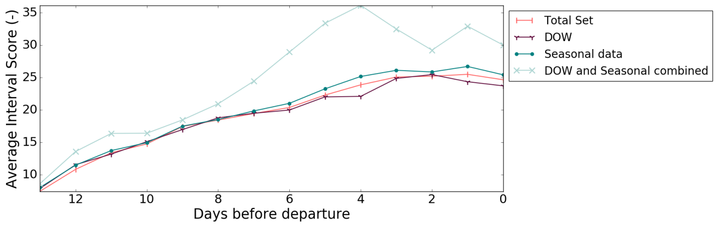
Figure 6.13: The interval scores of a forecast for flight *KQ101* with a bookings horizon of 14 days for varying datasets

day of the forecast well but thereafter the score sharply increases. This might imply that the forecasts did not cover the actual observed value with the prediction intervals which resulted in a penalty to be applied in the calculation of the interval score.

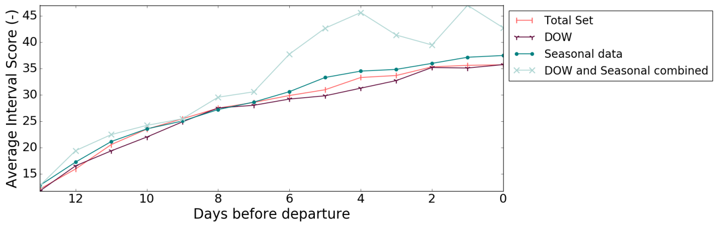
Based on the summary of the results tabulated in Table 6.6, it can be concluded that the seasonal and total datasets are superior for the other horizons as well. From a more detailed assessment of the results per observation point it is found that for the 33% and 67% prediction intervals the monthly seasonal data yields the lowest interval error for the long-term forecasts and the total dataset for the short-term forecasts. For the 95% prediction interval the same trend is visible for the long-term forecasts, however in the short-term all four datasets perform similarly. This can also be seen in the summary of the results per horizon where the seasonal data yields the best results in the long-term forecasts from 300 to 60 days and the total set performs best for the remaining horizons, except for the most short-term forecast of 7 days. Looking at all observations points the monthly seasonal data provided the lowest interval score for most instances. Furthermore, the total sum of the interval scores is also lowest for this dataset, which is mostly the result of the large difference in relative performance for the forecasts with an horizon of 300 and 200 days. From there it can be concluded that from an uncertainty modeling perspective the monthly seasonal dataset is the desired dataset to use when forecasting for KQ101.

KQ512

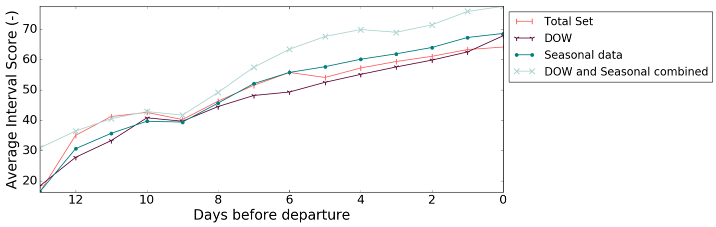
The total set of results of the interval scores for the different datasets are tabulated in Table B.12. As an example, the average interval scores for flight KQ512 with an horizon of 14 days are displayed in Figure 6.14 for the different sets of data that serves as an input to the proposed Markov Chain model. From the graph it can clearly be stated that the model using data from both the same day-of-week and similar months models the uncertainty in the forecast worst. The differences in the performance for the other datasets are smaller. Especially for the 33% probability intervals the interval scores show only a small difference between the different datasets. Nevertheless, the data of the same DOW seems to model the uncertainty slightly better. This behavior can be identified even stronger for the 95% probability interval where the day-of-week seasonal data yields the best interval score (except for the last day of the horizon).



(a) 33% Prediction Interval Scores



(b) 67% Prediction Interval Scores



(c) 95% Prediction Interval Scores

Figure 6.14: The interval scores of a forecast for flight *KQ512* with a bookings horizon of 14 days for varying datasets

This specific example for the forecast horizon of 14 days is chosen because when comparing the interval scores for all horizons the same trend can be observed. From a summary of the results, tabulated in Table 6.6 the dataset with data of flights that departed the same DOW leads to lower interval scores for most horizons. Only in the long-term the *Monthly-DOW* data (300 days) and the seasonal data (200 days) yield a better performance. For the horizon of 25 days the total set performed better, however, the difference with the DOW dataset is marginal. Over all observation points and horizons it is found that the DOW data yields the lowest score for most observation points. Furthermore, the sum of all scores over the horizons is also lowest for this dataset. From there it can be concluded that the DOW dataset is the desired set to cover the uncertainty of the prediction in the model. This finding is in line with the finding in the first part of this experiment in which it is found that the DOW dataset yields the lowest MAE.

Subconclusion

From this phase of the experiments it can be concluded that the datasets that provided the lowest mean forecast error in the first part of this study also perform best in modeling the uncertainty of the prediction. The seasonal data provided the best results in the long-term (200 & 300 days) for both KQ101 and KQ512. For KQ101 this beneficial performance continued for the other horizons as well. For KQ512 the DOW data yielded better results for the remaining horizons which is in line with the findings of the calibration on MAE. Therefore, the seasonal dataset for KQ101 and the DOW data for KQ512 are chosen to perform the remaining tests for this study.

6.2.4. Concluding remarks on the calibration

The summary of the calibration of the model for both the best performance in MAE and in interval scores is shown in Table 6.7. It can be seen that based on the findings in the calibration of the model for minimal interval scores, most found settings are correspondent with the settings that resulted from the calibration for minimal MAE. The test that yielded a different desired setting is the weighting factor for KQ512. Whereas, a moderate α was found to be desired for a minimal MAE, a high α resulted in the best probability distributions. This asks for a trade-off to be made between these two to find the optimal setting of the model. For the next

Table 6.6: Sum of interval scores for the different datasets across the observation points of the set horizons for KQ101 and KQ512

		Flight			
		KQ101		KQ512	
Horizon	Dataset	Sum of P.I.S.	% compared with lowest	Sum of P.I.S.	% compared with lowest
300	Total Set	2898.9	+13.8	791.8	+7.2
	DOW	2865.2	+12.4	780.0	+5.6
	Seasonal	2548.2	-	777.6	+5.3
	Month-DOW	2825.2	+10.9	738.6	-
200	Total Set	2220.2	+8.9	943.3	+1.8
	DOW	2294.4	+12.6	939.9	+1.5
	Seasonal	2038.4	-	926.4	-
	Month-DOW	2236.4	+9.7	969.4	+4.6
100	Total Set	1662.1	+1.8	1054.3	+0.5
	DOW	1736.8	+6.4	1048.7	-
	Seasonal	1632.2	-	1081.2	+3.1
	Month-DOW	1933.0	+18.4	1200.8	+14.5
60	Total Set	1485.7	+0.5	1244.9	+1.0
	DOW	1546.6	+4.7	1232.8	-
	Seasonal	1477.7	-	1287.6	+4.4
	Month-DOW	1724.3	+16.7	1464.9	+18.8
39	Total Set	1204.0	-	1278.5	+3.0
	DOW	1288.2	+7.0	1241.1	-
	Seasonal	1243.2	+3.3	1365.3	+10.0
	Month-DOW	1380.2	+14.6	1353.4	+9.0
25	Total Set	867.4	-	923.9	-
	DOW	971.6	+12.0	926.3	+0.3
	Seasonal	895.2	+3.2	966.5	+4.6
	Month-DOW	1026.6	+18.4	1059.7	+14.7
14	Total Set	753.4	-	736.8	+2.8
	DOW	796.8	+5.8	716.9	-
	Seasonal	755.8	+0.3	750.3	+4.7
	Month-DOW	841.7	+11.7	876.7	+22.3
7	Total Set	510.7	+0.7	561.2	+5.8
	DOW	507.4	+0.1	530.6	-
	Seasonal	507.0	-	583.8	+10.0
	Month-DOW	536.1	+5.7	647.9	+22.1

part of the experiment, the moderate α of 0.5 is used in the comparative study because the performances are compared in MAE for which an α of 0.5 yielded the best results. For the application of the model in practice it could be advocated that a high weighting factor should be used because the benefit of using such a high value for the probability intervals is much larger than the lost accuracy in MAE.

The best dataset also shows some different results for KQ512 per horizon. However, as discussed in the previous sections, the DOW data could clearly be identified to be the desired approach for this flight.

6.3. Benchmarking model

In the third part of the experiment the forecast performance of the new Markov Chain model is compared with the reference models (numerical examples are given in Appendix D) which are well known and applied in research and the industry. For this comparison only the expected value of the distributions resulting from the proposed model are considered in order to be able to compare the performance using the numerical error measurements described in Section 4.3.

Again flights KQ101 and KQ512 are considered for this part of the experiment. Following the experimental set-up discussed in Section 5.4.5 the relative performance of the methods is assessed for the dataset that followed to be the desired input from the previous parts in this study.

KQ101

The comparison of the proposed model with the traditional reference methods (the historical average and advanced pick-up method) is performed using the statistical error measurement metric MAE. Following from the conclusions in the first part of the study, the Markov Chain model uses a weighting factor α of 0.8 and no distribution is fit on the observed data-points in this comparative study. Furthermore, the monthly seasonal data is used as this dataset yielded the lowest forecasting error. Firstly, the comparative study is discussed. Then, the significancy of the observed difference in performance is assessed with a Kruskal-Wallis test (Liu & Chen, 2012).

The total set results of the comparative study of the model for KQ101 and the set of forecast horizons and observation points is tabulated in Table B.13. An example of the difference in performance is shown in Figure

Table 6.7: Summary of the calibration results

Horizon	Model setting	Approach yielding lowest MAE across all observation points V KQ101	KQ512	Setting yielding lowest P.I.S. across all observation points V KQ101	KQ512
300	Distribution	No Fit	No Fit	No Fit	No Fit
	Weighting	0.8	0.8	0.8	0.8
	Dataset	Seasonal	DOW	Seasonal	Month-DOW comb
200	Distribution	No Fit	No Fit	No Fit	No Fit
	Weighting	0.8	0.7	0.8	0.8
	Dataset	Seasonal	DOW	Seasonal	Seasonal
100	Distribution	No Fit	Lognorm	No Fit	No Fit
	Weighting	0.8	0.5	0.8	0.8
	Dataset	Seasonal	DOW	Seasonal	DOW
60	Distribution	Lognorm	Best	Lognorm	No Fit
	Weighting	0.8	0.2	0.8	0.8
	Dataset	Total set	DOW	Seasonal	DOW
39	Distribution	Lognorm	Lognorm	Lognorm	Norm
	Weighting	0.8	0.2	0.8	0.8
	Dataset	Total set	DOW	Total set	DOW
25	Distribution	Lognorm	Lognorm	Lognorm	Lognorm
	Weighting	0.8	0.2	0.8	0.8
	Dataset	Total set	DOW	Total set	Total set
14	Distribution	Lognorm	Lognorm	Lognorm	Lognorm
	Weighting	0.8	0.5	0.8	0.8
	Dataset	Seasonal	Total set	Total set	DOW
7	Distribution	Lognorm	Lognorm	Best	Best
	Weighting	0.7	0.3	0.8	0.7
	Dataset	Total set	Total set	Seasonal	DOW

6.15, where the forecast is computed 100 days in advance of the departure date. From the graph it can be seen that the proposed Markov Chain method outperforms both the other two methods. Especially closer towards the day of departure the forecasting error of the proposed method decreases significantly in comparison with the pick-up method. Nevertheless, the first days of the forecast horizon the pick-up method performs slightly better than the Markov Chain method.

The results of the historical average method are actually less relevant in the beginning of the horizon. At that point this reference method indicates how far off in general the average of the forecasted flights is from the average of the dataset that is used as an input to the models. However, especially closer towards the day of departure, it gives an indication of the ability of the forecasting models to predict to what extend the bookings for the forecasted flight will deviate from the average in the bookings dataset. For the specific horizon indicated in the figure it can be concluded that the proposed Markov Chain is able to predict this behavior well.

The summary of the results of this comparative study per forecast horizon is tabulated in Table 6.8. The difference in performance that has been explained above is also applicable to the tests for other horizons. Except for the long-term forecasts with an horizon of 300 and 200 days it is found that the pick-up model has the lowest error most often across the observation points per horizon. Looking at the sum of the errors over the horizon the pick-up method only outperforms the Markov-Chain method for an horizon of 300 days. For the horizon of 7 days the pick-up method also performs better but the difference is marginal. For the remaining forecasting lengths, the proposed Markov Chain transition matrices model outperforms the other two methods for the largest part of the horizon. Nevertheless, it is found that the first days after the forecasting day the pick-up method has a lower absolute forecasting error than the proposed model. This is the case for all forecast lengths that have been tested. It should be noted that the differences are small in those observation points with a maximum difference in average forecast error of *0.7*. On the other hand, for some other observation points the proposed Markov Chain model outperforms the pick-up method with a difference in MAE of around *13*.

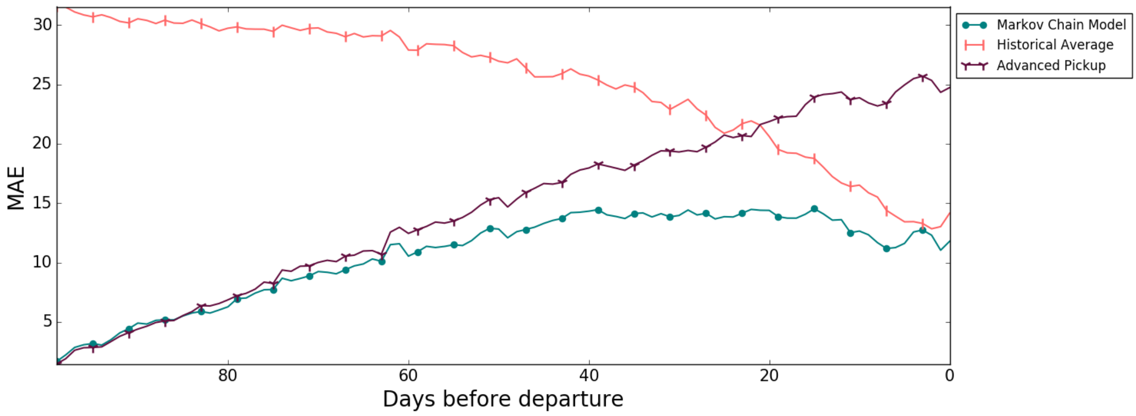


Figure 6.15: The *MAE* of a forecast for flight *KQ101* with a bookings horizon of 100 days for the set of forecasting methods

Kruskal-Wallis test

A Kruskal-Wallist test is used to test if the difference in performance is significant enough to state that one method is more accurate than the other (Liu & Chen, 2012). The Kruskal-Wallis test is used instead of the ANOVA test because the assumption of the ANOVA test that the values in each group are normally distributed is not met. This is the case because an absolute error is used so the errors are centered on the right of 0 MAE. Therefore, a non-parametric test is used in this case. The Null-hypothesis in this test is that the medians of all the groups are equal. The significance level α is set to *0.05* which means that there is a 5% risk that it is concluded that there exists a difference when there is actually no actual difference. If the *p-value* of the Kruskal-Wallis test is smaller or equal than α there exists a significant difference between the medians of the groups. If the value is larger than α the medians of the groups are not statistically significantly different.

In Table 6.8 it can be seen for each horizon how many times the difference between the errors was significant enough, following from the Kruskal-Wallis test, to reject the hypothesis that the medians are equal (the value for which this significancy is proven are indicated by the green numbers in Table B.13).

The hypothesis that the proposed Markov Chain forecasting model performs better than the other two methods can be concluded to be true for most of the forecast horizons. Especially for the medium-term forecasts (100 to 39 days) the Markov Chain method has proven to perform significantly better than the traditional methods. Altogether, the MC method has the lowest forecasting error most often across all observation points and a substantial part of these differences have been proven to be significant by the Kruskal-Wallis test. From there it can be concluded that the new Markov Chain method yields better results than the other two methods and that this difference in performance is proven to be significant enough.

KQ512

Similarly to the test for flight KQ101 discussed above, the performance of the three forecasting models has also been assessed for flight KQ512. In the first two experiments conducted in this study it is concluded that the approach of fitting no distribution to the observed values and an α of *0.5* are the best model settings for this test. Furthermore, the DOW data showed to be the best dataset to use in the forecasting model.

Again, firstly the comparative study for this dataset is discussed. Then, the significancy of the observed difference in performance is assessed with a Kruskal-Wallis test.

The total set results of the comparative study of the model for KQ512 and set of forecast horizons and observation points is tabulated in Table B.14. As an example, figure 6.16 displays the difference in performance measured in MAE for the three forecasting methods for a forecast made a 39 days before departure. From the graph it can be seen that the difference in performance between the Markov Chain and the pick-up method is much smaller than observed for flight KQ101. Only in the last two weeks before departure the proposed method starts to produce significant lower errors than the pick-up method. Furthermore, the historical average method has a lower average error at the day of departure than the other two methods.

For this specific horizon it can be concluded that the proposed Markov Chain performs slightly better than the pick-up method. Nevertheless, for the last two weeks before departure a simple average of the data in the dataset gives a more accurate prediction.

The difference in performance described above for an horizon of 39 days is not observed for the other test

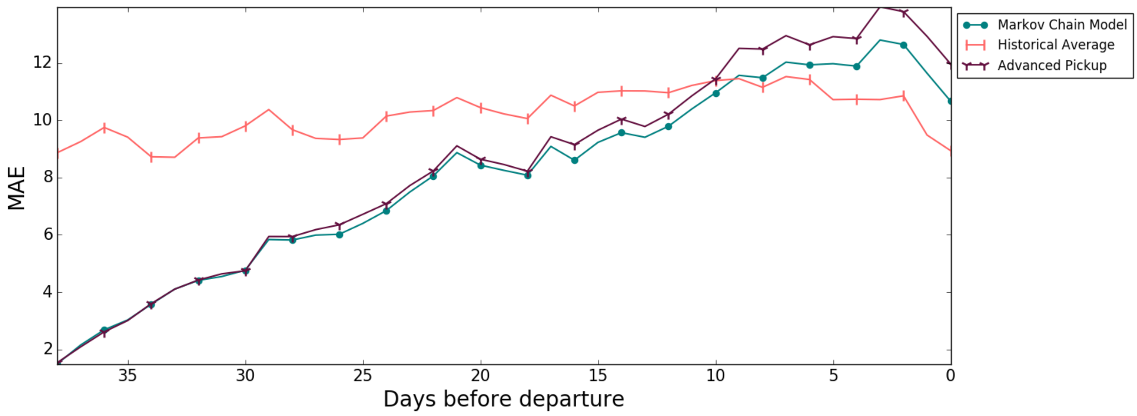


Figure 6.16: The *MAE* of a forecast for flight *KQ512* with a bookings horizon of 39 days for the set of forecasting methods

horizons as well. Especially for the long-term forecasts the best performing method changes per horizon and per test point as can also be seen in the summary of the comparative study in Table 6.8. Only in the short-term the proposed Markov Chain model seems to outperform the other two methods looking at the number of times it yields the lowest error across the observation points. Nevertheless, for the first days of the forecast horizon the pick-up method yields the lowest forecasting error. The differences are, nevertheless, small. Across all observation points and horizons the new Markov Chain method yielded the lowest error in most of the observations. Looking at the sum of the errors over the horizons it can be seen that the Markov Chain method yields a lower total than the pick-up method for most of the horizons. Where the pick-up method performs better over the horizon the difference is small. Furthermore, for the most long-term forecast the historical average performs best, however, the difference with the Markov Chain method is marginal.

In order to be able to accept the hypothesis that the proposed Markov Chain model yields more accurate results than the pick-up method, a *Kruskal-Wallist test* has been conducted in order to determine the significancy of the difference in mean forecast error.

Kruskal-Wallis test

For KQ512 the Kruskal-Wallis test is performed in a similar approach as for KQ101. For flight KQ512 the difference in forecasting performance between the methods is not found to be significant. As can be seen in Table 6.8 the number of times the Markov Chain is significantly more accurate than the other methods does not show that the method is better. Only in the long-term forecasts a few instances have been observed, however, these are divided amongst all three methods.

From these results it can be stated that for flight KQ512 the methods showed no significant difference in performance to be able to conclude that one method is more accurate than the other.

Concluding remarks on comparative study

The inconsistency in the advantageous performance of the Markov Chain method for the two test cases KQ101 and KQ512 is remarkable. One of the reasons for this might be that for KQ512 the development of the number of bookings coming in is far less impacted by the bookings that already have been observed for the flight than that is the case for KQ101. Another theory to explain the difference is the fact that the data could not be unconstrained. As the LF for KQ101 is normally higher than the LF for KQ512 (See Section 5.2.1) it can be the case that the Markov Chain method is able to model the behavior of the bookings process well in case it is constrained. Ideally, the models are compared using unconstrained datasets in order to mimic the most realistic context. Unfortunately, it was not possible to identify the constrained data points from the datasets in this study. Therefore, it is recommended to perform a study where this Markov Chain model is tested using unconstrained data.

6.4. Illustrative case

The proposed Markov Chain transition matrix forecasting model produces a forecast of the amount of bookings that is expected to be present in the system for every day in the forecast horizon until the day of departure. An example of such a forecast is visualized in Figure 6.17. This forecast is made for an historical flight

Table 6.8: Summary of the comparative study for KQ101 and KQ512

		Method with lowest MAE over observation points				MAE improvement compared with other methods (%)	
		Nr. times lowest		Nr. times p<0.05			
Horizon	Method	KQ101	KQ512	KQ101	KQ512	KQ101	KQ512
300	MC	2	*1*	0	1	-	-
	HA	0	*7*	0	1	+22.1	-0.8
	PU	9	*4*	0	1	-9.3	10.4
200	MC	4	*6*	1	1	-	-
	HA	0	*5*	0	0	48.6	16.0
	PU	7	*2*	0	1	2.6	1.4
100	MC	9	*3*	5	0	-	-
	HA	0	*1*	0	0	54.2	21.8
	PU	2	*7*	1	0	18.8	-2.4
60	MC	9	*3*	5	0	-	-
	HA	0	*1*	0	0	53.3	25.9
	PU	2	*7*	1	0	17.1	-0.1
39	MC	9	*7*	6	0	-	-
	HA	0	*3*	0	0	52.4	24.1
	PU	2	*1*	1	0	15.2	4.3
25	MC	6	*6*	2	0	-	-
	HA	0	*1*	0	0	52.9	33.3
	PU	3	*3*	0	0	10.1	-0.5
14	MC	7	*6*	0	0	-	-
	HA	0	*0*	0	0	53.4	36.4
	PU	1	*2*	0	0	9.3	3.3
7	MC	4	*5*	0	0	-	-
	HA	0	*0*	0	0	61.0	45.7
	PU	3	*3*	0	0	-0.1	1.6

similarly to the forecasts that are constructed for the three test cases above because the forecast has to be compared with the actual observed value in the system.

This graph shows the prediction intervals that can be deducted from the probabilistic forecast distribution. In this example, only the 99%, 95%, 67% and 33% intervals are displayed, nevertheless, any prediction interval can be constructed based on the probability distribution that is given as an output by the model. An example of this probability distribution is displayed in Figure 6.19a where the probability distribution of bookings at the day of departure for this flight is shown. This distribution is the main advantage of the proposed method in comparison with traditional methods. These traditional methods provide forecasts as a single point. Those are shown in Figure 6.17 as well and result in a line similar to the line in the figure representing the expected value of the forecast distribution.

The forecast in the figure is made 39 days before the departure of flight KQ101 on the 24^{th} of November 2016. It can be seen that for this specific case the predicted value quite accurately mimics the actual bookings that have been observed for this flight. Furthermore, all observed values fall within the constructed prediction intervals which means the prediction covered the uncertainty well. Nevertheless, it should be noted that the smaller the prediction intervals are the better, with the requirement that all observed values are nicely covered. A narrow interval enables the revenue management controller to trigger an action more easily. This is because the variance future number of bookings in the system presented by the distribution is smaller. As a result it can be identified more easily if the target value is within or without the range of possible future values if no action is undertaken.

A forecast made for the same flight but two weeks later than the previous forecast, at 25 days before departure, yields the output as in Figure 6.18. Still, the expected value of the forecast gives a good indication of the future bookings situation for this flight. Furthermore, it can be identified that the prediction interval is more narrow for the 25 day forecast horizon compared with the 39 day forecast. This becomes especially apparent when comparing the probability distribution at the day of departure visualized in Figure 6.19b with the distribution in Figure 6.19a. This means the model, as expected, indicates the future state of bookings with more certainty closer to departure day.

Based on this prediction the revenue management controller could conclude that there is sufficient demand for the flight and, for example, more higher fare classes could be made available. Or the controller can conclude that the demand is below the aimed load factor so that he or she has to analyze the options to increase the ticket sales for the flight.

It should be noted that the distributions computed in this study are relatively wide, as can also be seen in

Figure 6.19. This indicates a high uncertainty in the bookings to expect in the future of the bookings process. The more narrow the distribution the more likely it is that the future value will be observed within a smaller range. This increases the usefulness of the distributions to the revenue management controller. It could therefore be investigated if these distributions can be improved by for example sophisticated data pooling methods.

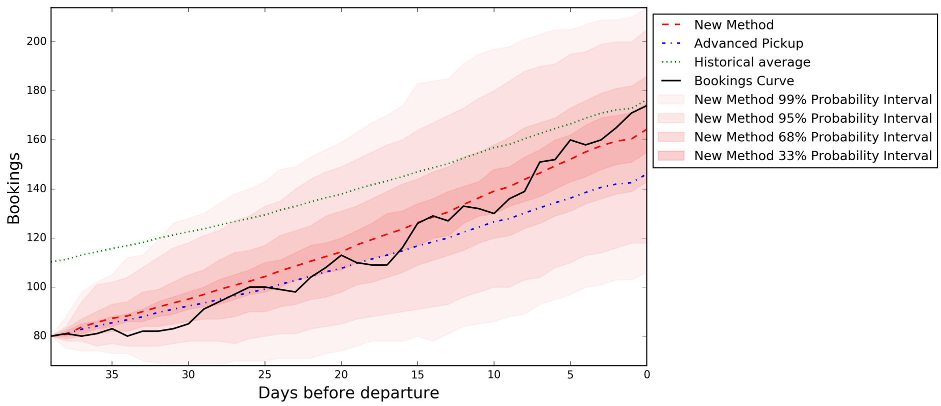


Figure 6.17: Forecast and actual bookings for flight *KQ101* on 24-11-2016 with a bookings horizon of 39 days

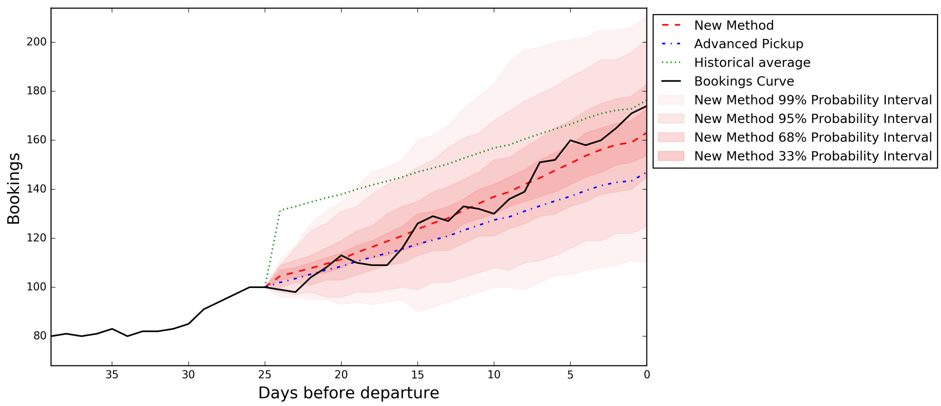
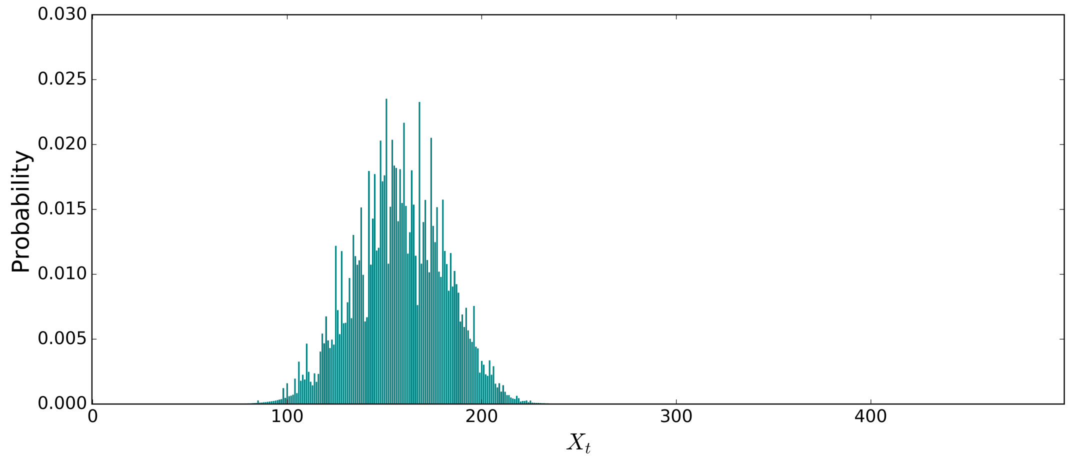
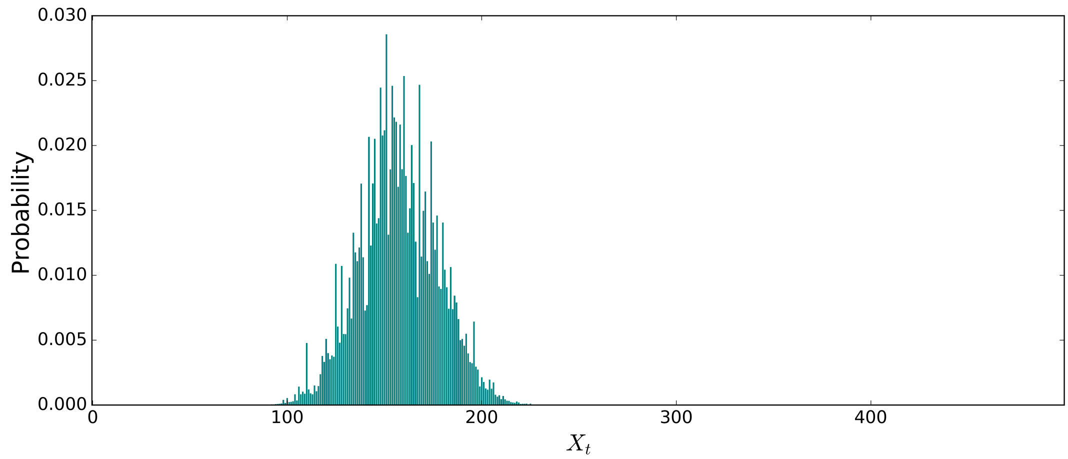


Figure 6.18: Forecast and actual bookings for flight *KQ101* on 24-11-2016 with a bookings horizon of 25 days



(a) Probability distribution at $t = 0$ for BH of 39 days



(b) Probability distribution at $t = 0$ for BH of 25 days

Figure 6.19: The probability distributions at $t = 0$ of the forecast made with an horizon of 39 (a) and 25 (b) days for flight KQ101 departing at 24-11-2016

7

Verification and validation

Now the forecasting model is developed and produces results, the next step is to verify and validate the model. This is an important part because it should be checked whether the model contains no errors and whether the results are valuable from the perspective of the industry. This chapter first discusses the verification of the model and then the validation of the model is explained.

For both the verification and validation a dataset is used that has not been applied in the test phase of this study. This is done in order to see if the model performs the way it should on data with which it has not been designed or calibrated. The model settings during the verification and validation are consistent with the conclusions of Section 6.

7.1. Model verification

The model is verified in order to check if there are errors present in the model that might influence the resulting forecast. This is done by looking at the different parts of the model. This consists of the outlier selection, transition matrix computation and the calculation of the state vectors as indicated by the shaded parts of the conceptual model in Figure 7.1. Furthermore, the behavior of the model is discussed for several scenarios.

7.1.1. Outlier selection

As discussed in Section 5 some erroneous data exists in the data extractions from *Delorean* that may not be used to compute a prediction of the future number of bookings. This consists of two parts, the erroneous zero values and the values far above the capacity number.

The first part is checked by looking at the bookings values that are present in the final *8* days of the booking dataset that is used by the model to compute the transition matrices. The minimum value present in this dataset equals *61*. Therefore, it can be concluded that this part of the outlier selection works well.

The second part of the verification assesses the upper outliers. The maximum value present in the complete dataset that is eventually fed to the forecasting model equals *225*. Since the economy class in the *Boeing*

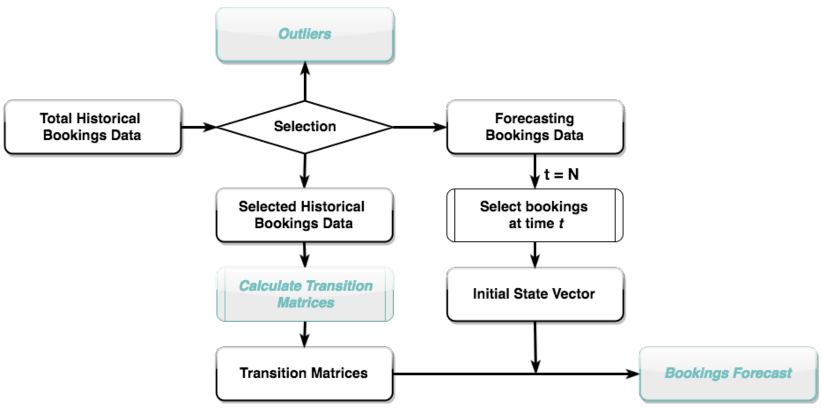


Figure 7.1: The elements of the conceptual model that are subjected to verification

787-800 aircraft of flight KQ101 has 204 seats, the maximum allowed value by the model to be observed in the dataset is *255*. Therefore, it can be concluded that the outlier selection removes the upper outliers well.

7.1.2. Transition matrix computation

One of the crucial elements of the forecasting model is the transition matrix computation. The verification of this part of the model is discussed in this section. Firstly, the parametric distribution arrays are checked. Secondly, the values in the empirical distributions are assessed. Finally, the values in the columns of the transition matrices are analyzed in more detail.

Parametric distribution entries

The parametric distributions entail the probabilities of having an incremental number of bookings in a certain time interval. This parametric distribution array should only have numbers in the range from 0 and 1 and the sum of the array has to be equal to 1 because all the possible states need to be captured. This is checked on the forecasts that are the subject in this part of the study; the flights of *KQ101* that departed in *January 2017*.

First of all the values should lie in between zero and one. Considering all distributions that are used to forecast for the flight departing in *January 2017*, the maximum observed probability in the arrays is equal to 0.943 and the minimum value $2.976 \cdot 10^{-3}$. From these numbers it can be concluded that the construction of the parametric distributions meets the requirements for probability arrays to consist of values in the range $0-1$.

Furthermore, the sum of the distribution arrays is assessed. The sum of the arrays should be equal to one. The maximum and minimum observed probability arrays are shown in Equation 7.2. Looking at the maximum and minimum sum of the distribution arrays, the values deviate slightly from 1. However, the difference is negligible and has most probably to do with the floating-point arithmetic of digital computers[1].

Max. sum	1.0000000000000002	(7.1)
Min. sum	0.99999999999999967	

Empirical distribution entries

For the empirical distribution the same requirements hold as for the parametric distribution arrays. Nevertheless, an additional requirement of the empirical distribution that should be checked is whether the state space that was defined for the model is large enough to cover all the observed booking values.

From an analysis of all the empirical distributions it was found that the maximum observed bookings number in the distributions is 225. This is far below the state space size of 500. This finding is in line with the outlier selection of the model where bookings larger than 255 are removed from the bookings data.

For the empirical distribution it is again checked if the sum of the columns equals zero. Because the empirical distribution matrix contains empty rows there exists columns of which the sum is zero. The maximum sum is slightly higher than one, however, as discussed in the previous section this is most likely due to the floating-point arithmetic.

Max. sum	1.0000000000000002	(7.2)

The minimum probability observed for the empirical distributions is 0 and the maximum value equals 1. This satisfies the requirements of the probability array, so it can be concluded that the empirical distributions are computed correctly.

Matrix column values

The transition matrices are multiplied with the probability vectors to determine the possible state of the system in the next time step. The columns of the matrix have the same requirements as the parametric distributions. The reason why the matrix columns are verified in this section is that the columns are a result of the combination of the parametric with the empirical distribution computed using the weighting factor α.

First of all the sum of the columns of the matrix are checked in order to find out if these are equal to 1 as required for probability arrays. From Equation 7.3 it can be concluded that the maximum and minimal

[1]The floating-point arithmetic problem arises when numbers need to be stored in a computer system. The floating-point numbers are stored in hardware as binary fractions. However, most decimal fractions cannot be stored as a binary fraction and as a result the floating-point numbers stored in a computer system are an approximation of the actual decimal number (Goldberg, 1991)

Table 7.1: Minimum and maximum state larger than zero observed in the probability vector

	Horizon							
	300	**200**	**100**	**60**	**39**	**25**	**14**	**7**
Minimal state value	0	0	0	0	0	13	70	110
Maximal state value	499	499	499	499	499	477	387	301

observed total column values slightly deviate from 1. Nevertheless, similar as the findings in Section 7.1.2 the difference is that small that it has most likely to do with the floating-point arithmetic problem.

$$\begin{array}{r|l} \textit{Max. sum} & 1.0000000000000002 \\ \hline \textit{Min. sum} & 0.99999999999999956 \end{array} \tag{7.3}$$

The maximum and minimum probabilities in the final transition matrices are equal to 1 and 0. Therefore, it can be concluded that the transition matrix contains no errors.

7.1.3. Probability vectors values

Similarly to the distributions discussed in the previous sections, the probability vector is a probability array and has to meet the accompanied requirements. Additionally, it should be verified whether the state space is large enough to cover all possible states.

The values in all the probability vectors have been analyzed to investigate if the probabilities are in the range $0-1$. The minimum value is found to be equal to 0 and the maximum observed probability equals 0.818.

The maximum and minimum sum observed in the probability vectors is shown in Equation 7.4. The deviations are slightly larger than those observed in the previously discussed matrices and distributions. This is most probably a result of the fact that the model repeatedly multiplies the probability vector with the matrix columns. When a slight deviation is present in the matrix this will affect the probability vector and this deviation will grow after every multiplication. Nevertheless, the deviation is small and considering the maximum horizon (and therefore number of multiplications) of 300 days, the deviation will not be much larger when the model is applied to other flights.

$$\begin{array}{r|l} \textit{Max. sum} & 1.0000000000000033 \\ \hline \textit{Min. sum} & 0.99999999999999778 \end{array} \tag{7.4}$$

The minimum and maximum state value with a probability of occurrence higher than 0 are determined to investigate if the defined state space is of sufficient size. The minimum and maximum values are determined separately for the different forecast horizons and are tabulated in Table 7.1. From the table it can be concluded that the maximum state value hits the boundary of the state space of 499 for the long term forecasts. The values of the probabilities at those boundaries are checked in order to see whether significant probabilities are missed due to the limit of the state space.

All the probability vectors containing a probability higher than 0 for the state value of 499 are analyzed and it is found that the minimum probability observed at those instances equals $9.82 \cdot 10^{-45}$ and the maximum probability observed equals $1.12 \cdot 10^{-13}$. These probabilities are that small that their influence on the expected value of the probability vector is minor. Furthermore, increasing the state space will result in a significant longer computational time because the number of additional entries in the matrix to determine grows exponentially. For example, increasing the state-space with 10 bookings leads to 10100 additional entries in the matrix[2]. Decreasing the size of the state space will logically lead to a faster computational time. It should be analyzed to what extend the size can be decreased until it negatively influences the prediction. This is not considered in this study due to time limits.

7.1.4. Expected behavior

The model is also verified by checking if the behavior of the model is as expected. Four scenarios are considered that are discussed in this section in consecutive order; a high jump in the number of bookings, a sudden drop in the number of bookings and an extraordinary high number of bookings.

[2]With the current size of 500 and an increase in size n the number of additional entries equals $(500+n)^2 - 500^2 = n^2 + 1000n$

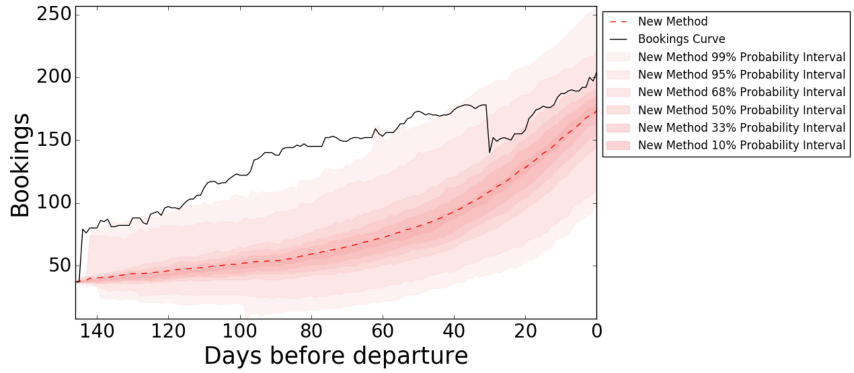


Figure 7.2: The forecast for flight KQ101 departing on 02-01-2017 that has a jump in bookings 145 days before departure

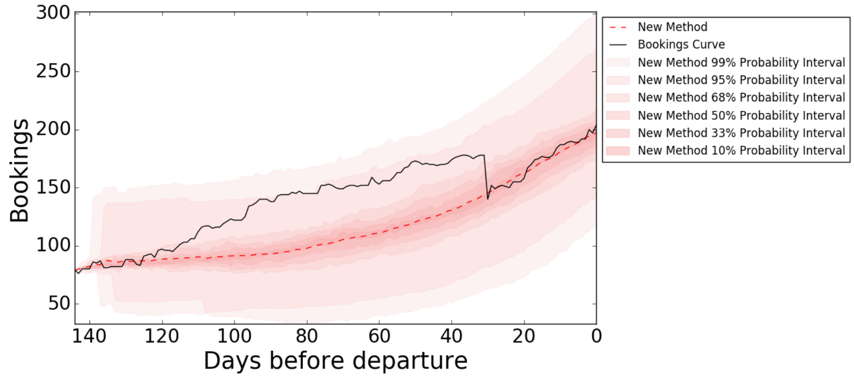


Figure 7.3: The forecast made 146 DBD for flight KQ101 departing on 02-01-2017 after the bookings suddenly increased

Jump in bookings number

A sudden jump is observed for flight KQ101 that departed on the 2nd of January 2017 at 145 DBD as can be seen in the booking curve in Figure 7.2. In this case one would expect the model to either predict the bookings curve to grow steady or to have a decline again towards the end of the bookings process. The second dynamic originates from the fact that group bookings are likely to be cancelled before the flight departs (Appendix C).

From the forecast made the day after the sudden increase in bookings number, seen in Figure 7.3, it can be stated that the model shows a steady behavior for the coming days. Furthermore, it can be observed that the model foresees a slight chance of having another jump in bookings and a drop in bookings for the coming days. This behavior corresponds to what would be expected.

Drop in bookings number

A sudden drop has been observed in the booking development for flight KQ512 on the 11th of January 2017 as can be seen in Figure 7.4. This sudden drop in number of bookings is most likely due to a group of passengers that cancelled or rebooked their tickets to another flight. After such a sudden decrease in bookings you would expect the bookings curve to develop steady, without showing a continuation of the decline of net bookings.

The forecast of the first day after the cancellation is shown in Figure 7.5. As can be seen in the forecast the prediction for the future looks steady without abnormal shocks in the expected number of bookings. As a matter of fact the forecast looks very similar to the forecast made 145 days before departure (Figure 7.4), which means that the behavior looks well.

High number of bookings

A high number of bookings has been observed for flight KQ512 on the 11th of January 2017 at around 138 days before departure. It can be seen in Figure 7.4 that the number of bookings at that day is higher than normal. In the case a high number of bookings is present in the system two things can happen dependent on the reason for this bookings number. The high bookings number can be present because of some group bookings that were made for the flight or can be the reason of high demand for individual passengers to travel on that flight. In the first case there is a significant chance that the booking will be cancelled or rebooked before the flight

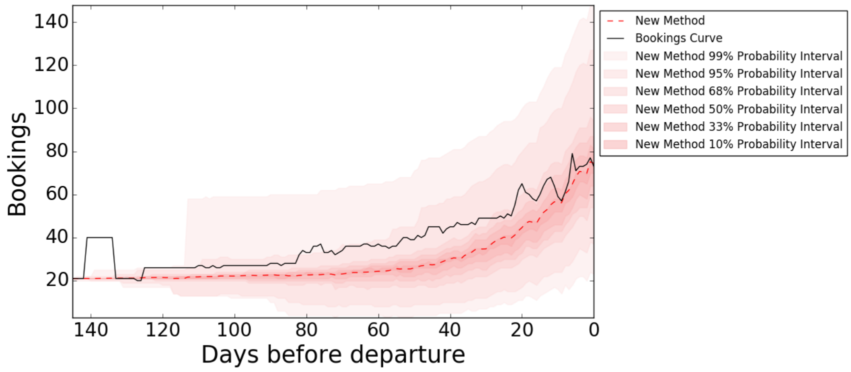


Figure 7.4: The forecast for flight KQ512 departing on 11-01-2017 that has a decrease of bookings around 134 days before departure

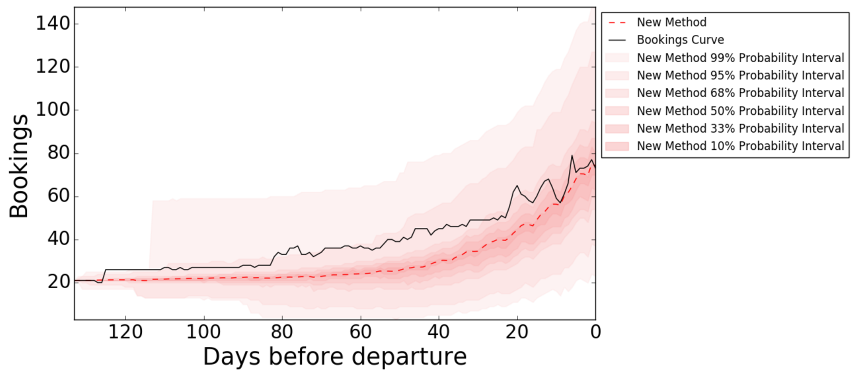


Figure 7.5: The forecast made at 133 days before departure for flight KQ512 departing on 11-01-2017 that after the bookings suddenly dropped

departs (a more detailed analysis of the group bookings can be seen in Appendix C). In the second case the bookings number is not necessarily expected to drop significantly.

These two options should be reflected in the forecast. However, because the current forecasting model does not discriminate between individual and group bookings, the model is currently not able to incorporate an expected drop due to a group booking present in the forecast. This can be seen in the forecast made 138 days before departure in Figure 7.6. However, the model 'picks-up' a sharp decreasing booking trend at around 90 days before departure which occurred for a flight in the bookings database and is as a result reflected in the forecast. It is recommended that the model discriminates between the type of bookings and accordingly adapts the forecast. This has not been considered in this study due to time constraints.

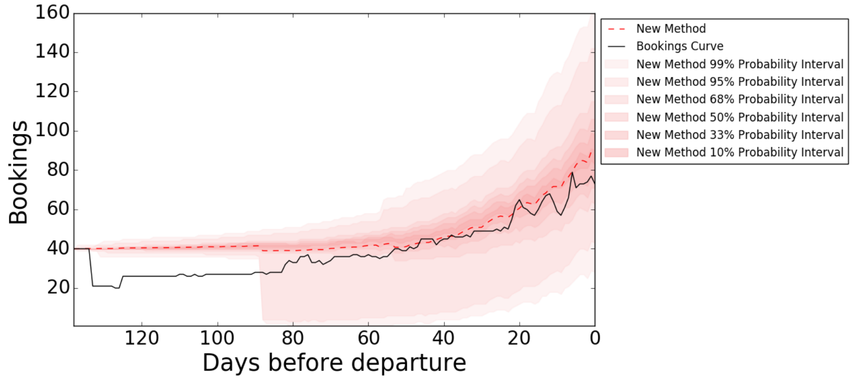


Figure 7.6: The forecast at 138 DBD for flight KQ512 departing on 11-01-2017 that had a high number of bookings

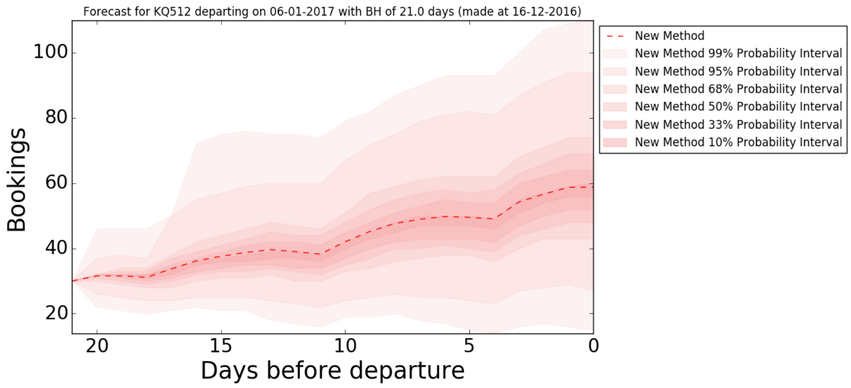


Figure 7.7: The forecast made at 21 days before departure for flight KQ512 departing on 06-01-2017

7.2. Model validation

The model is validated by computing forecasts on a new dataset. In this test the optimal model configuration is applied, which followed from the tests discussed in Section 6. The new dataset has not been used in calibrating and analyzing the models parameters and can therefore be used to determine if the proposed model provides valuable results to the airline. The value that the forecast has for the airline is determined by a discussion of the results with the flight demand analysts and commercial analysts. They have the expert knowledge and skill to judge the validity of the results. Bookings data of the month of January 2017 is extracted from the *Delorean* and used for the validation.

Firstly, the findings of the consultation of the results with the experts within the revenue management department of KQ is discussed. Then, the probability distributions resulting from the model are validated. Finally, the computational time of the model is shortly addressed.

7.2.1. Expert judgement

This study is conducted in collaboration with Kenya Airways. Besides providing data for the research, experts within the Revenue Management department of the airline have been involved in the research by providing guidance, information and insights from an applicational perspective. Two experts have validated the forecasting model using their own experience. The findings are presented in this section. Firstly, the validation of the results for KQ101 are presented before the validation of the results for KQ512.

KQ101

For the forecasts made for flight KQ101 the experts have concluded that the expected value of the forecasts seems reasonable. The prediction intervals have been found to be fine as well. Furthermore, it is validated that the model correctly captures the early booking behavior for this flight. For this flight it is necessary to have a model that is able to provide forecasts with long forecast horizons. The model developed in this study fulfills that requirement.

KQ512

For KQ512, the experts noted that the model is able to capture the pattern of the cancellations that are expected over the weekends nicely as can be seen in Figure 7.7. This is exactly as the expert would expect the behavior to look like the weeks before departure. One of the things that seemed to be different than the expectation of the expert is the pretty flat booking curve until departure date for this flight. The expert expected a more sharp increase in bookings until the day of departure. It should be noted that the degree of booking increments that the revenue management experts expect are, however, covered by the probability distributions.

7.2.2. Validation probabilistic modeling

The forecast distributions are validated in order to find out whether the probability distributions provide an accurate representation of the uncertainty in the realization of the bookings for the future. In order to validate this it is important to note that both a forecast that is too certain as well as a too uncertain forecast are undesired. This is assessed for both flight KQ101 and KQ512.

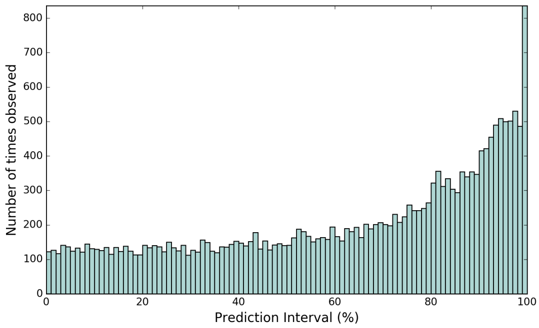


Figure 7.8: The number of times the predition interval captured the actual observation on its boundary for KQ101

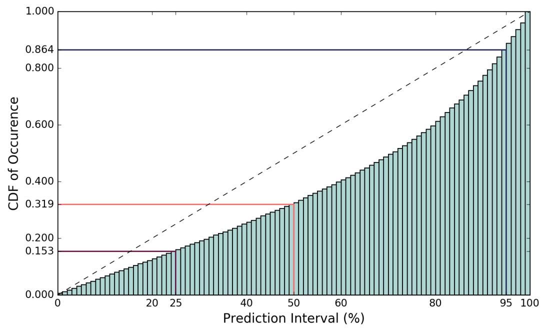


Figure 7.9: The CDF of the sample of observed prediction intervals for KQ101

KQ101

It is determined what is probability interval for which the actual observation lies just on the boundary of the interval. This is done for all the probability vectors and accompanied actual observations of the computed forecasts. The results of this analyses are illustrated in Figure 7.8. It should be noted that the probability interval of 1 indicates the instances where the observation is not captured by the prediction interval. For a probability interval of 0 the expected value exactly matched the actual number of bookings.

From the graph it can be seen that the number of times a value fell just within the probability interval increases for an increasing prediction interval. Furthermore, it can be seen that there is a spike present at an interval of 1 which means that a relatively high number of observed bookings values fall outside the probability distribution defined by the forecasting model.

By normalizing these observations and presenting them graphically in a Cumulative Density Function (CDF) in Figure 7.9, it can be seen that the lower prediction intervals are not capturing the actual observed bookings well. For example, around 32% of the values fall within the 50% prediction interval. Ideally, the CDF would follow the dotted line that indicates the case where the prediction interval probability matches the actual observed probability. From here it can be concluded that the uncertainty in the forecast is not fully accurately determined for flight KQ101. The higher prediction intervals are occurring more frequently in reality. It is investigated in more detail for which forecast horizons and for what parts of the horizon the model fails to correctly model the uncertainty. It is found that for an horizon of up to maximum 80 days the uncertainty in the forecast is modeled quite accurately. For a longer forecast horizon it is found that the probability distribution is quite accurate for the first 70 days of the horizon. For the days closer to departure, however, the model fails to capture the uncertainty that is present that many days into the future. The model should be adapted to be able to model the uncertainty in the long-term forecast better. To do this, it should be analyzed whether the high uncertainty in the future is due to external factors, group bookings, etc.. This has not been considered in this study because of time constraints.

KQ512

For flight KQ512 the same approach is used as for KQ101. The frequency of occurrence of the prediction intervals is displayed in Figure 7.10. It can be observed that there is a small spike at the prediction interval near 0 and a somewhat higher peak for the intervals close to 1. Similarly as to flight KQ101 it happens more often than desired that the observed value falls outside the probability distribution. However, it also happens that more values fall within a certain prediction interval than expected.

This is also displayed in Figure 7.11. Here it can be seen that until the prediction interval of around 80% the observed value fell within the prediction interval with a higher frequency than predicted. For example, around 57% of the values fell within the 50% prediction interval. For higher prediction intervals the observed frequency is lower than expected as is shown in the graph by the 95% prediction interval represented by around 91% of the values.

From this graph it can be concluded that the uncertainty in the forecast is modeled quite accurately by the forecasting model for flight KQ512. Nevertheless, there is still room for improvement. It can be analyzed for what horizons and parts of the horizon the prediction intervals can be made more narrow or wider.

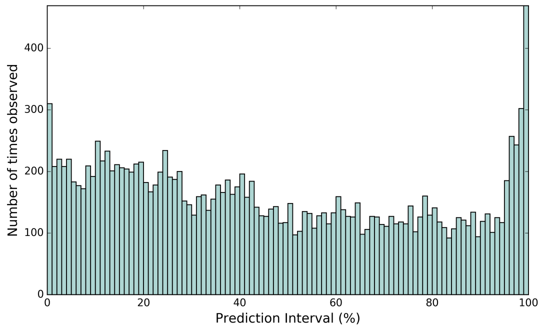


Figure 7.10: The number of times the prediction interval captured the actual observation on its boundary for KQ512

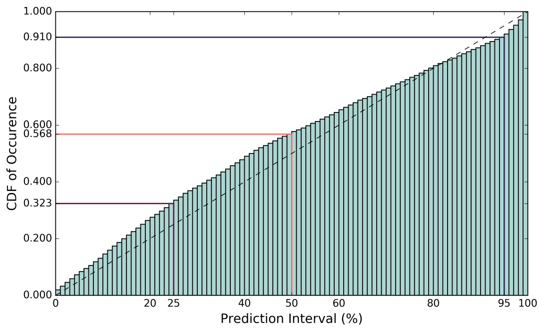


Figure 7.11: The CDF of the sample of observed prediction intervals for KQ512

7.2.3. Computational time

The computational time of the forecasts made with the Markov Chain model are analyzed. For KQ101 a total of 10 forecasts have been computed which took the model 3650 seconds to complete. This comes down to somewhat more than 6 minutes per forecast.

For KQ512 a total of 9 forecast have been computed in a timespan of 1506 seconds. Per forecast the computational time equals somewhat less than 3 minutes. The lower computational time for KQ512 corresponds with the smaller state space compared to KQ101 which requires less entries to be filled in the transition matrix.

From the computational times it can be concluded that the model is fast enough to be used in daily practice. However, since revenue management is known for its dynamic behavior, the revenue management controller should be able to extract the desired information as fast as possible. A solution could be to compute the new forecasts overnight after the bookings system *Delorean* updated the new bookings information.

8 Conclusions and Recommendations

The conclusions of this research are drawn in this chapter. Firstly, the main findings from the study are discussed. Secondly, the main limitations of the developed model are explained. Then, the future research opportunities that follow from this research project are stated. Finally, the project is finalized by evaluating the research objective that has been defined in the beginning of the project.

8.1. Concluding on the results

From the literature review it has been identified that gaps in the research area of bookings forecasting for airline revenue management exist, especially on the incorporation of uncertainty in the forecast. It is found that the Markov Chain method is suitable to perform probabilistic forecasts and it can as well be applied in a dynamic context.

Kenya Airways uses multiple forecasts within their revenue management practice. However, all these forecasts provide point forecasts with an unsatisfactory accuracy. This called for an improvement in the forecasts provided to the controllers. A forecasting model that includes a degree of uncertainty could fulfill this demand because it gives the controller more insight in the future demand for a flight.

Therefore, the following main research question was formulated:

How does an approach combining forecasting uncertainty with a dynamic updating approach contribute to an improvement in the accuracy of airline bookings forecasts?

The dynamic forecasting model in this study is modeled using the Markov Chain method. In the development of the model, it is calibrated by analyzing the performance for several model settings both in the forecast error but also in the ability of the model to incorporate the uncertainty of future predictions in the output. The conclusions on this calibration are addressed firstly in this section. After the calibration of the model, its impact has been studied by analyzing the average forecast errors the model is able to achieve and by comparing these with the errors traditional methods are able to produce. Furthermore, it is tested whether the uncertainty that is displayed by the model is valid. These analyses are discussed in the same order in the final two parts of this section and are required to be able to conclude on the hypotheses and to state if the research objective is reached.

8.1.1. Model calibration

The Markov Chain model developed in this study combines a parametric and an empirical approach to compute the transition matrix entries. An analysis on the influence of the parametric distribution used in this approach showed that for both test cases considered in this study (KQ101 and KQ512) fitting no distribution yielded the best performance compared to the model in which a normal or lognormal distribution is fit on the data. With performance in this context, both the absolute forecast error based on the expected value as well as the ability of the model to correctly model the uncertainty are meant. Furthermore, the computational time has been found to be significantly lower when no distribution is fit on the data. It should be noted, however, that when considering a short-term model the lognormal distribution yields the best results in all test cases.

Furthermore, it has been found that the relative weight attributed to either the parametric or the empirical distribution can influence the performance of the model significantly. For KQ101 a high weighting factor (0.8) is found to be the desired setting, resulting from both the lowest MAE and the lowest interval scores. For KQ512 a more moderate value of 0.5 provided a low average absolute error. Nevertheless, a high weighting factor of 0.8 yielded the best interval scores. This inconsistency in best settings for the model implicates that the model should be calibrated based on both performance indicators to determine the best setting. A trade-off should be made between having a larger absolute forecast accuracy or a probability distribution of higher quality.

To conclude on the model calibration the best dataset is determined to use as an input to the forecasts. As some significant trends have been observed in the bookings data for booking behavior over the year and during the week, the influence of taking these into account in the model are assessed using pre-selected seasonal datasets. From there it was found that monthly seasonal data provided the best results for KQ101 and data of the same DOW yielded the best performance for KQ512.

From the model calibration it can be concluded that the best setting is dependent on the flight leg for which the forecast is computed. This implicates that in the case where a forecast is created for a new flight, the model should first be calibrated for this flight before the predictions are made in order to achieve optimal results. Furthermore, it has been found that the performance differs for short-term and long-term forecasts. Altogether, depending on the application of the model the optimal model settings may be changed.

8.1.2. Model accuracy

Overall, the model produced an average forecast error in the range of maximum 30 bookings for long-term forecasts 300 days before departure and 8 bookings for short-term forecasts 7 DBD for flight KQ101. For KQ512 the maximum average forecast error ranged from 13 bookings for a forecast made at 39 DBD to 8 bookings for the forecast made 7 days in advance of the departure date.

For KQ101 these results are significantly lower than the errors the traditional methods produce. Especially for forecasts made from up to 100 DBD the method yielded clearly the lowest forecasting errors. The errors produced for KQ512 are in general also lower in the comparative study, although less dominant. However, in the first day of the horizon, the pick-up method consistently produces a lower error for both flights. This is consistent with the finding that lower values of α consistently yield better results than high weighting factors in the first days of the horizon for the Markov Chain model. The lower the weighting factor, the less influence is attributed to the empirical bookings observation. In a situation where the weighting factor equals 0 the Markov Chain method should provide similar results as the pick-up method.

By analyzing the significancy of the differences in performance using the Kruskal-Wallis test, it is found that for KQ101 the difference in performance is proven to be significant enough to conclude the hypothesis stating that the Markov Chain method is the most accurate forecasting model is correct. Especially for the medium-term forecasts of 100-25 DBD the errors in the second part of the horizon consistently supported the hypothesis. For KQ512, however, the significancy test indicated that based on those results the hypothesis could not be accepted. It could not be concluded what is the reason for the different performance of KQ512 and KQ101. It might be due to another booking behavior for KQ512 over the bookings process compared to KQ101. Or it can be that the results for KQ101 are positively biased by the data that could not be unconstrained in this study.

Looking at the difference with the traditional methods it is found that over all observation points the new Markov Chain forecasting method is on average 15% more accurate than the most competitive traditional method which is the pick-up method. Based on the analysis of Lee (1990) (Chapter 1.2) this implies that an annual revenue gain of around €4.3-25.3 Mio could be achieved [1]. Note that this is the gain that could theoretically be achieved in case the pick-up method would be the method currently used.

8.1.3. Uncertainty modeling

Besides the performance of the model in terms of average absolute forecasting error, the model has also been assessed on its ability to correctly cover the forecast uncertainty. The validity of the probability distributions is tested in order to determine whether the model is overconfident about the future of the bookings system

[1] The annual revenue of KQ for the book-year ending in March 2016 equaled 116,158 (KShs' Millions)(Kenya Airways, 2016). Based on the KSH-EUR exchange rate at 2017-04-17 17:00, this equals €1,053.46 Mio

or whether the uncertainty is undervalued. For flight KQ101 it is found that the prediction intervals are too narrow and do not cover the number of bookings that would have been expected from 70 days out in long-term forecasts larger than 80 days. It can therefore be stated that the model is overconfident for that test case. Especially the number of values that have been encountered outside of the 99% prediction interval exceeded the predicted frequency. Nevertheless, for the remaining part of the long-term forecast horizon and also for the short-term forecasts made maximum 80 days before departure, the probability distributions model the uncertainty quite accurately.

On the contrary, the probability distributions of KQ512 cover the uncertainty in the future of the bookings process quite well measured over all horizons. For the lower prediction intervals the model captures slightly more bookings than expected, which indicates that the intervals could be narrowed slightly. For the higher intervals the opposite has been observed which calls for wider intervals.

It has not been studied what is the reason for the inaccurate probability distributions of the model for the long-term of KQ101. An analysis should be carried out identifying the key factors playing a role in the high uncertainty of forecasts made for KQ101 more than 80 days in advance of the departure date.

8.2. Limitations of the model

The model developed in this study has several limitations which are discussed in this section.

Use of unconstrained data The model is developed is based on the complete set of historical data that was available. Due to inaccuracies in the datasets of KQ it was not possible to correctly identify the data points that were constrained because of enforced booking limits. As a result the forecast computed by the model might be biased when the model would be implemented in practice. An additional step is necessary in between the data selection and transition matrix calculation steps, where the constrained data is unconstrained based on a well-known statistical method. It should be tested if the method still clearly outperforms the traditional methods.

Use of partial booking curves Looking at the conceptual model, the current developed model only takes complete bookings curves into account when computing the transition matrices. When the model is implemented in practice the most recent bookings data consists for a large part of partial booking curves. The model can in that case only use the historical bookings data or the model should be slightly adapted so that it is able to use the most recent data and update the transition matrices every time a new forecast is computed.

Forecasting for special events The current developed method uses data of previous departures of the same flight to construct the transition matrices. When forecasting for special events, the remaining flights that show similar behavior to the special occurrence may therefore be insufficient. For these special events a different approach should be used to calculate the transition matrices. A solution might be to use innovative data pooling methods (Lemke, Riedel, & Gabrys, 2012) that pool similar events and feed it as an input to the model. These sophisticated data pooling methods may also result in a significant improvement of the forecasts in general, since it is found that selecting a specific dataset which includes trends relevant for the bookings forecast of a flight can improve the forecast.

Conceptual design The forecasting model is developed in-house at KQ. As a result the model is based on the bookings forecasting process at Kenya Airways. This might be a limitation when the model would be implemented at another airline.

8.3. Recommendations for further research

Some recommendations for further research are found during this study which are addressed in this section. These recommendations are an addition to the recommendations that followed from the limitations of the model addressed in the previous section.

Optimization of model parameters From the conclusion on the varying performance of the Markov Chain model over the horizon for the weights α it can be concluded that in order to obtain the best Markov Chain method, the value of α should be varied over the forecast horizon. Apparently, the benefit of incorporating the latest status of the bookings process is not consistently present over the bookings process. By using an optimization algorithm the optimal α settings for the entire forecast horizon can be determined.

Apply method on another level The Markov Chain forecasting method is applied on a cabin class (economy) and flight level in this study. First of all, it can be investigated how the method performs when it is applied to business class forecasting. Furthermore, it can be studied how the model performs when applied to fare class level or Origin-Destination (OD)-level. Another new application would be to apply the model on an aggregated level for the total bookings per week, month or year, which would make the model more useful for the commercial analysis department.

Separate bookings and cancellations It has been found that the cancellation ratio is dependent on the number of bookings that are present in the system. Therefore, a model considering bookings and cancellations separately could improve the quality of the forecast. Such a model should have a transition matrix in which the chance to observe a cancellation is dependent on the number of bookings in the previous interval. Eventually the outcome of the additional bookings and cancellations in an interval can be combined to arrive at the net number.

Consider individual and group bookings separately Group bookings have been found to result in a large variation for the number of bookings to be present in the system at a later time stage. A model that handles both booking types separately may incorporate some characteristics that are present for the group bookings, such as a high cancellation ratio and a high chance of realization after a specific day before departure. A different approach to handle group bookings is especially expected to improve the uncertainty modeling performance of the model because the current model has difficulties with incorporating the chances of a high jump or drop in bookings in the prediction.

Long-term uncertainty modeling The model showed to be unable to model the uncertainty for KQ101 for forecast horizons longer than 80 days. A study can analyze what is the reason for the bad performance in the end of long-term forecasts in terms of probability distributions.

Multiple order Markov Chain A variation of the Markov Chain model developed in this study would be to consider multiple order Markov Chain models to see if the model improves when the future bookings situation is not only dependent on the current state but the development over multiple days in the past. Consequence is that the model becomes significantly more complex and the computational time to fill in the transition matrices increases significantly.

8.4. Concluding on the research objective

Following from the conclusion on the accuracy of the forecasting model developed in this study, a conclusive answer can be formulated on the first hypothesis which was formulated as follows:

The forecasting model will provide forecasts with a higher accuracy than traditional forecasting methods.

Although the forecasts for both KQ101 and KQ512 produced overall lower errors than the traditional methods, this hypothesis can only be accepted for KQ101 following from the significancy test conducted in this study.

The second hypothesis can also be accepted or rejected based on the findings of this study. The second hypothesis was formulated as follows:

The developed forecasting model can display the uncertainty in the future number of bookings.

From the test cases it follows that the model showed to be able to display the uncertainty in the forecast quite correctly for flight KQ512, so for this flight the hypothesis can be approved. Nevertheless, for flight KQ101 quite some improvements can be gained for the long-term forecasts as the model currently underestimates the high uncertainty for this flight.

Overall, both hypotheses cannot clearly be rejected or accepted looking at both test cases. However, from these conclusions it is possible to determine if the research objective which had been established as follows has been reached:

"Increase the accuracy of bookings forecasts by contributing to the development of a dynamic forecasting model that includes the uncertainty in the forecast on a cabin class and flight level"

Since the Markov Chain model provides a level of uncertainty to the forecaster the model can be stated to be more accurate than point forecasts in case the absolute forecasting errors of the predicted value are comparable. Therefore, for flight KQ512 it can be concluded that the objective has been fulfilled. Furthermore, for KQ101 the new approach introduced in this study has proven to increase the accuracy in comparison with traditional methods. From that result it can also be stated that the objective has been reached, although there is room for improvement in the representation of the uncertainty.

Altogether, it can be stated that the study proves that incorporating the state of the system at the forecast day leads to an improved result compared to a method that uses the same mathematical method but does not take such influence into account, e.g. the pick-up method. At the same time the model has shown to be able to provide more information to the revenue management controller by including probability distributions.

A

Data Format

The bookings data is obtained from the revenue management systems at *Kenya Airways*. The data on the development of the number of bookings in the system during the bookings process is extracted from *Delorean*. This system also holds information on the capacity deployed for the flights.

Delorean

The revenue management system *Delorean* is a backward looking system that has originally been developed by KLM and now also implemented within the revenue management operations of its *SkyTeam* partner *Kenya Airways*. Backward looking in this context means that the system makes no predictions on the future but only stores the newly available bookings information on a daily basis. *Delorean* can show the bookings that had been made for a flight at a certain moment in time, called the *snapshot*. The output from *Delorean* consists of the following fields (an example output can be seen in Appendix A):

- **Snapshot Date**: the date for which the specific booking reservation is present in the system (ex. 10/03/2015).
- **DBD**: similar to the snapshot date, however now expressed in the number of days before the departure date of the flight (ex. 332).
- **Flight**: the flight number of the flight for which the booking is made (ex. KQ101_LHRNBO).
- **Flight date**: the departure date of the flight for which the booking is made (ex. 05/02/2016).
- **Day of Week**: the day of the week at which the specific flight (at the date specified above) departs (ex. Fri).
- **Cabin**: the discrimination between an economy (4) and business (2) booking (ex. 4).
- **Subclass**: the fare class in which the reservation is made (ex. N).
- **True Origin Code (TOC)**: the International Air Transport Association (IATA) airport code of the origin of the passenger (ex. LHR).
- **True Destination Code (TDC)**: the IATA airport code of the final destination of the passenger (ex. JRO).
- **PaxType:** the discrimination is made between an individual booking (ind), group booking (grp) or a duty passenger (ex.ind).
- **PNR Address**: the unique bookings code that is appointed to the reservation by the reservation system *Amadeus* (ex. 2ONXZI).
- **PNR Size**: the number of seats that are booked with the reservation (ex. 2).
- **Pax Y**: the total number of passengers booked on the flight for that entry in the data sheet (ex. 2).

Table A.1: Format of data as extracted from Delorean

Snapshot date	DBD	Flight	Flight date	DOW	Cabin	Subclass	TOC	TDC	PaxType	PNR Address	PNR Size	Pax Y
10/2/15	335	KQ0101_LHRNBO	9/1/16	Thu	4	T	LHR	NBO	ind	3WHE2S	2	2
10/3/15	334	KQ0101_LHRNBO	9/1/16	Thu	4	T	LHR	NBO	ind	3WHE2S	2	2
10/3/15	334	KQ0101_LHRNBO	9/1/16	Thu	4	T	LHR	MBA	ind	79Z7KA	1	1
10/3/15	335	KQ0101_LHRNBO	9/2/16	Fri	4	N	LHR	NBO	ind	3YR4PT	2	2
10/4/15	333	KQ0101_LHRNBO	9/1/16	Thu	4	T	LHR	NBO	ind	3WHE2S	2	2
10/4/15	333	KQ0101_LHRNBO	9/1/16	Thu	4	T	LHR	MBA	ind	79Z7KA	1	1
10/4/15	334	KQ0101_LHRNBO	9/2/16	Fri	4	N	LHR	NBO	ind	3YR4PT	2	2
10/5/15	332	KQ0101_LHRNBO	9/1/16	Thu	4	T	LHR	NBO	ind	3WHE2S	2	2
10/5/15	332	KQ0101_LHRNBO	9/1/16	Thu	4	T	LHR	MBA	ind	79Z7KA	1	1
10/5/15	333	KQ0101_LHRNBO	9/2/16	Fri	4	N	LHR	NBO	ind	3YR4PT	2	2
10/6/15	331	KQ0101_LHRNBO	9/1/16	Thu	4	T	LHR	NBO	ind	3WHE2S	2	2
10/6/15	331	KQ0101_LHRNBO	9/1/16	Thu	4	T	LHR	MBA	ind	79Z7KA	1	1
10/6/15	332	KQ0101_LHRNBO	9/2/16	Fri	4	N	LHR	DAR	ind	3YR4PT	2	2
10/6/15	333	KQ0101_LHRNBO	9/3/16	Sat	4	N	LHR	NBO	ind	347GD8	1	1
10/7/15	330	KQ0101_LHRNBO	9/1/16	Thu	4	T	LHR	NBO	ind	3WHE2S	2	2
10/7/15	330	KQ0101_LHRNBO	9/1/16	Thu	4	T	LHR	MBA	ind	79Z7KA	1	1
10/7/15	331	KQ0101_LHRNBO	9/2/16	Fri	4	N	LHR	DAR	ind	3YR4PT	2	2
10/7/15	332	KQ0101_LHRNBO	9/3/16	Sat	4	N	LHR	NBO	ind	347GD8	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

B

Complete set of forecast results

This appendix includes the complete set of test results that have been computed in this study. First of all the results of the calibration of the three model settings are shown in terms of MAE. Then the results of the calibration in terms of prediction interval scores is displayed. Finally, the complete set of results of the comparative study is shown in this appendix.

Calibration of Model Parameters to Forecast Error

This section includes the tables that present the MAE of the model for different parameter settings. Firstly, the set of results for the several distribution approaches is tabulated, then the tables with results for the different weighting factors is shown and finally the results for the four datasets considered in this study. All results are displayed for KQ101 and KQ512 separately.

Parametric Distribution

The tables related to the results of the calibration of the parametric distribution fitting approach yielding the lowest forecasting error for the new model are displayed in this section. First the results are shown for flight KQ101 and thereafter for KQ512.

KQ101

The performance of the model for different distribution fitting approaches has been analyzed for KQ101 and the results are displayed in Table B.1 below. The table shows the average MAE, for the observation points of each specific horizon, taken over the number of flights for which forecasts are computed. As can be seen in the table, for the three most short-term forecasts the number of observation points considered are less than for the long-term forecasts. The italic numbers indicate the day before departure corresponding with the observation point for that horizon.

Table B.1: Forecasting performance with different parametric distribution for flight KQ101 at various booking horizons

		Observation point										
Horizon	**Distribution**	V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}
300		*299*	*270*	*240*	*210*	*180*	*150*	*120*	*90*	*60*	*30*	*0*
	No Fit	0.57	7.36	12.30	16.11	**22.37**	**26.09**	**31.07**	**31.94**	**32.27**	**27.91**	**17.02**
	Norm	**0.54**[1]	7.09	12.06	15.74	22.67	27.22	31.94	34.96	37.81	35.77	25.45
	Lognorm	0.55	7.09	**12.05**	**15.73**	22.61	27.07	31.81	34.68	37.47	35.38	25.31
	Best fit	**0.54**	**7.07**	**12.05**	**15.73**	22.67	27.23	31.94	34.99	37.86	35.87	25.75
200		*199*	*180*	*160*	*140*	*120*	*100*	*80*	*60*	*40*	*20*	*0*
	No Fit	1.27	**7.70**	**10.68**	**13.28**	**16.42**	**19.03**	**20.42**	**21.45**	**21.57**	**18.50**	**13.80**
	Norm	**1.25**	7.74	10.99	13.88	17.49	20.27	22.49	24.55	26.44	24.22	19.21
	Lognorm	**1.25**	7.74	10.96	13.83	17.42	20.18	22.35	24.39	26.26	24.13	19.26
	Best fit	**1.25**	7.74	10.99	13.88	17.49	20.27	22.49	24.56	26.46	24.30	19.42
100		*99*	*90*	*80*	*70*	*60*	*50*	*40*	*30*	*20*	*10*	*0*
	No Fit	**1.53**	**4.70**	**6.76**	**9.30**	**11.07**	**13.48**	**15.34**	**15.82**	**17.49**	**16.66**	17.21
	Norm	**1.53**	4.72	6.87	9.63	11.74	14.19	16.37	16.79	18.27	17.88	**16.68**
	Lognorm	**1.53**	4.72	6.86	9.62	11.72	14.18	16.36	16.80	18.29	17.94	16.69
	Best fit	**1.53**	4.72	6.86	9.63	11.73	14.19	16.37	16.81	18.29	17.94	16.70
60		*59*	*54*	*48*	*42*	*36*	*30*	*24*	*18*	*12*	*6*	*0*
	No Fit	**1.89**	**4.68**	**7.66**	**9.13**	**9.78**	**11.12**	12.14	13.52	14.21	14.78	16.07
	Norm	1.90	**4.68**	7.70	9.45	9.98	11.15	**12.13**	**13.31**	**14.12**	13.90	14.91
	Lognorm	1.90	**4.68**	7.70	9.45	9.99	11.16	**12.13**	**13.31**	14.13	**13.87**	**14.85**
	Best fit	1.90	**4.68**	7.70	9.45	9.98	11.16	**12.13**	**13.31**	14.13	**13.87**	14.86
39		*38*	*35*	*31*	*27*	*23*	*19*	*15*	*11*	*7*	*3*	*0*
	No Fit	**2.31**	4.33	5.31	6.73	7.82	9.45	11.09	11.97	12.53	14.34	14.51
	Norm	**2.31**	**4.28**	**5.24**	6.49	7.54	9.09	10.66	11.62	11.60	13.43	13.64
	Lognorm	**2.31**	**4.28**	**5.24**	**6.48**	**7.52**	**9.07**	**10.64**	**11.60**	**11.53**	**13.36**	**13.58**
	Best fit	**2.31**	4.29	**5.24**	**6.48**	7.53	**9.07**	**10.64**	**11.60**	11.54	**13.36**	13.59
25		*24*	*22*	*19*	*16*	*13*	*10*	*7*	*4*	*0*		
	No Fit	2.68	5.00	6.09	7.53	8.43	9.24	9.28	10.52	11.38		
	Norm	**2.66**	4.95	5.93	7.32	8.30	8.99	8.66	10.02	10.95		
	Lognorm	**2.66**	**4.94**	**5.92**	**7.30**	**8.29**	**8.97**	**8.60**	**9.99**	**10.92**		
	Best fit	**2.66**	**4.94**	**5.92**	**7.30**	**8.29**	**8.97**	**8.60**	**9.99**	**10.92**		
14		*13*	*12*	*10*	*8*	*6*	*4*	*2*	*0*			
	No Fit	**3.67**	5.37	6.66	5.81	6.78	8.47	9.05	10.46			
	Norm	**3.67**	**5.36**	6.60	5.68	6.59	8.22	8.77	10.20			
	Lognorm	**3.67**	**5.36**	**6.59**	**5.66**	**6.57**	8.20	**8.74**	**10.18**			
	Best fit	**3.67**	**5.36**	**6.59**	5.67	**6.57**	**8.19**	**8.74**	**10.18**			
7		*6*	*5*	*4*	*3*	*2*	*1*	*0*				
	No Fit	**2.77**	4.01	4.63	5.28	5.32	6.07	8.25				
	Norm	**2.77**	**4.00**	**4.59**	5.20	**5.24**	5.92	8.18				
	Lognorm	**2.77**	**4.00**	**4.59**	**5.19**	**5.24**	**5.91**	**8.17**				
	Best fit	**2.77**	**4.00**	**4.59**	**5.19**	**5.24**	**5.91**	**8.17**				

[1] The bold numbers in the table indicate the minimum error

KQ512

The performance of the model for different distribution fitting approaches has been analyzed for KQ512 and the results are displayed in Table B.2 below. Similar as to the previous shown table this table shows the average MAE, for the observation points of each specific horizon, taken over the number of flights for which forecasts are computed. The italic numbers indicate the day before departure corresponding with the observation point for that horizon.

Table B.2: Forecasting performance in terms of MAE with different parametric distribution for flight KQ512 at various booking horizons

		Observation point										
Horizon	**Distribution**	V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}
		299	*270*	*240*	*210*	*180*	*150*	*120*	*90*	*60*	*30*	*0*
300	No Fit	**0.09**[1]	**1.52**	**1.28**	**1.49**	**1.34**	**1.68**	**3.16**	5.84	**4.62**	**12.28**	**10.07**
	Norm	0.10	1.82	2.18	2.40	2.43	2.38	3.26	**5.77**	4.95	12.59	10.39
	Lognorm	0.58	14.01	39.79	65.72	78.92	82.67	80.12	76.00	75.47	67.86	63.39
	Best fit	0.58	14.01	39.79	65.72	78.92	82.67	80.12	75.94	75.17	67.41	62.96
		199	*180*	*160*	*140*	*120*	*100*	*80*	*60*	*40*	*20*	*0*
200	No Fit	**0.01**	**0.68**	**1.09**	**1.28**	**2.82**	**4.66**	**4.80**	4.63	8.63	10.99	10.22
	Norm	0.06	0.78	1.18	1.31	**2.82**	4.73	4.97	5.05	9.36	11.22	10.17
	Lognorm	0.52	3.45	5.07	4.26	4.60	5.52	4.90	**4.45**	**8.53**	**10.94**	**10.11**
	Best fit	0.52	3.45	5.07	4.26	4.60	5.52	4.89	**4.45**	8.55	10.95	**10.11**
		99	*90*	*80*	*70*	*60*	*50*	*40*	*30*	*20*	*10*	*0*
100	No Fit	0.55	1.32	2.68	4.14	4.62	**5.61**	**7.24**	**9.84**	**10.16**	13.44	11.32
	Norm	**0.53**	**1.27**	2.66	4.10	4.63	5.88	7.68	10.12	10.37	**12.90**	**10.89**
	Lognorm	0.54	**1.27**	**2.65**	**4.08**	**4.61**	5.82	7.62	10.09	10.36	**12.90**	10.90
	Best fit	**0.53**	**1.27**	2.66	4.10	4.63	5.88	7.68	10.12	10.38	**12.90**	**10.89**
		59	*54*	*48*	*42*	*36*	*30*	*24*	*18*	*12*	*6*	*0*
60	No Fit	0.85	**2.19**	**4.76**	**5.48**	**7.60**	8.41	9.77	**10.42**	12.36	11.63	**10.21**
	Norm	**0.82**	2.21	4.95	5.69	7.83	**8.40**	**9.48**	10.50	**12.11**	11.18	**10.21**
	Lognorm	**0.82**	2.21	4.94	5.68	7.83	8.41	9.49	10.51	12.12	11.18	10.23
	Best fit	**0.82**	2.21	4.95	5.69	7.83	8.41	**9.48**	10.51	**12.11**	**11.17**	10.22
		38	*35*	*31*	*27*	*23*	*19*	*15*	*11*	*7*	*3*	*0*
39	No Fit	**1.68**	3.16	4.92	6.21	8.20	9.40	11.52	12.12	12.59	13.13	11.35
	Norm	**1.68**	3.12	**4.75**	5.98	7.72	9.09	11.05	11.64	11.72	12.17	10.94
	Lognorm	**1.68**	**3.11**	**4.75**	5.98	**7.71**	**9.08**	**11.02**	**11.62**	**11.66**	**12.10**	**10.93**
	Best fit	**1.68**	3.12	**4.75**	**5.97**	**7.71**	**9.08**	11.03	**11.62**	11.67	12.11	**10.93**
		24	*22*	*19*	*16*	*13*	*10*	*7*	*4*	*0*		
25	No Fit	2.28	4.43	4.90	7.55	7.41	9.25	9.72	10.72	9.37		
	Norm	**2.25**	**4.39**	**4.88**	**7.51**	7.30	8.99	9.25	10.24	**9.29**		
	Lognorm	**2.25**	**4.39**	4.89	**7.51**	**7.29**	**8.97**	**9.21**	**10.19**	**9.29**		
	Best fit	**2.25**	**4.39**	4.89	7.52	**7.29**	8.98	9.22	10.20	**9.29**		
		13	*12*	*10*	*8*	*6*	*4*	*2*	*0*			
14	No Fit	**2.80**	**3.87**	5.32	6.75	7.24	9.08	9.45	8.91			
	Norm	2.81	**3.87**	5.27	6.62	7.15	8.86	9.16	**8.90**			
	Lognorm	2.81	**3.87**	5.27	**6.60**	**7.14**	**8.83**	**9.12**	**8.90**			
	Best fit	2.81	**3.87**	**5.26**	**6.60**	**7.14**	**8.83**	**9.12**	**8.90**			
		6	*5*	*4*	*3*	*2*	*1*	*0*				
7	No Fit	**3.24**	4.19	5.43	5.22	6.04	6.79	**7.58**				
	Norm	**3.24**	**4.16**	**5.40**	5.09	5.94	**6.72**	7.59				
	Lognorm	**3.24**	**4.16**	**5.40**	**5.08**	**5.93**	**6.72**	7.60				
	Best fit	**3.24**	**4.16**	**5.40**	**5.08**	**5.93**	**6.72**	7.60				

[1] The bold numbers in the table indicate the minimum error

Weighting factor

The tables related to the results of the calibration of the weighting factor yielding the lowest forecasting error for the new model are displayed in this section. Firstly, the tables for KQ101 are displayed and thereafter the tables with results of KQ512.

KQ101

The performance of the model for different values of α has been analyzed for KQ101 and the results are displayed in Table B.3 below. Similar as to the tables in the previous section this table shows the average MAE, for the observation points of each specific horizon, taken over the number of flights for which forecasts are computed. The italic numbers indicate the day before departure corresponding with the observation point for that horizon.

Table B.3: Forecasting performance in terms of MAE with different values of α for flight KQ101 at various booking horizons

		Observation point										
Horizon	α	V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}
		299	*270*	*240*	*210*	*180*	*150*	*120*	*90*	*60*	*30*	*0*
	0.2	**0.45**[1]	**7.11**	12.31	16.17	22.52	26.73	31.48	32.98	34.29	30.51	20.70
	0.3	0.49	7.19	**12.28**	16.15	22.42	26.48	31.32	32.53	33.50	29.45	19.12
300	0.5	0.57	7.36	12.30	16.11	22.37	26.09	31.07	31.94	32.27	27.91	17.02
	0.7	0.65	7.51	12.30	**16.08**	**22.35**	25.90	30.90	31.53	31.47	27.07	15.97
	0.8	0.69	7.56	12.31	16.09	22.37	**25.86**	**30.85**	**31.40**	**31.27**	**26.90**	**15.72**
		199	*180*	*160*	*140*	*120*	*100*	*80*	*60*	*40*	*20*	*0*
	0.2	**1.22**	7.79	11.01	13.92	17.31	20.00	21.46	22.91	23.13	20.34	17.32
	0.3	1.23	7.75	10.90	13.71	16.97	19.62	21.03	22.35	22.53	19.56	15.87
200	0.5	1.27	7.70	10.68	13.28	16.42	19.03	20.42	21.45	21.57	18.50	13.80
	0.7	1.31	7.67	10.49	12.84	16.03	18.52	19.99	20.90	21.03	17.88	12.44
	0.8	1.33	**7.66**	**10.41**	**12.67**	**15.87**	**18.32**	**19.82**	**20.76**	**20.95**	**17.72**	**12.08**
		99	*90*	*80*	*70*	*60*	*50*	*40*	*30*	*20*	*10*	*0*
	0.2	**1.49**	**4.47**	6.86	9.69	11.81	14.57	17.02	18.20	20.66	21.65	23.03
	0.3	1.50	4.52	6.78	9.51	11.48	14.20	16.43	17.37	19.58	19.93	21.07
100	0.5	1.53	4.70	**6.76**	**9.30**	11.07	13.48	15.34	15.82	17.49	16.66	17.21
	0.7	1.57	4.93	6.79	9.33	**11.02**	12.94	14.45	14.51	15.39	13.79	13.29
	0.8	1.60	5.07	6.86	9.41	11.06	**12.83**	**14.15**	**13.84**	**14.36**	**12.53**	**11.60**
		59	*54*	*48*	*42*	*36*	*30*	*24*	*18*	*12*	*6*	*0*
	0.2	**1.81**	4.79	7.84	9.53	10.48	12.28	13.87	16.10	17.69	18.93	20.84
	0.3	1.83	4.75	7.78	9.40	10.26	11.86	13.27	15.20	16.52	17.50	19.30
60	0.5	1.89	4.68	7.66	9.13	9.78	11.12	12.14	13.52	14.21	14.78	16.07
	0.7	2.01	**4.65**	**7.56**	**8.97**	9.54	10.52	11.18	12.03	12.15	12.09	12.91
	0.8	2.07	4.71	7.59	9.01	**9.50**	**10.43**	**10.79**	**11.36**	**11.37**	**10.85**	**11.74**
		38	*35*	*31*	*27*	*23*	*19*	*15*	*11*	*7*	*3*	*0*
	0.2	**2.08**	4.29	5.55	7.52	8.91	11.11	13.10	14.49	15.95	18.50	19.00
	0.3	2.15	**4.26**	5.45	7.21	8.55	10.56	12.42	13.64	14.77	17.12	17.52
39	0.5	2.31	4.33	5.31	6.73	7.82	9.45	11.09	11.97	12.53	14.34	14.51
	0.7	2.51	4.45	**5.27**	**6.46**	7.32	8.65	9.85	10.39	10.49	11.52	12.14
	0.8	2.61	4.54	5.28	**6.46**	**7.24**	**8.37**	**9.34**	**9.69**	**9.68**	**10.28**	**11.31**
		24	*22*	*19*	*16*	*13*	*10*	*7*	*4*	*0*		
	0.2	**2.65**	**4.92**	6.35	7.94	9.39	10.30	10.71	12.76	13.92		
	0.3	**2.65**	4.93	6.25	7.79	9.06	9.88	10.21	11.99	13.00		
25	0.5	2.68	5.00	6.09	7.53	8.43	9.24	9.28	10.52	11.38		
	0.7	2.78	5.09	**6.01**	**7.44**	8.04	8.76	8.51	9.18	10.51		
	0.8	2.84	5.16	**6.01**	**7.44**	**7.95**	**8.59**	**8.24**	**8.69**	**10.17**		
		13	*12*	*10*	*8*	*6*	*4*	*2*	*0*			
	0.2	**3.53**	5.40	6.86	6.11	7.57	9.52	10.09	11.54			
	0.3	3.56	**5.37**	6.78	5.99	7.28	9.14	9.68	11.13			
14	0.5	3.67	**5.37**	6.66	5.81	6.78	8.47	9.05	10.46			
	0.7	3.85	5.45	**6.59**	**5.79**	6.47	7.97	8.70	10.17			
	0.8	3.96	5.50	6.62	5.84	**6.43**	**7.78**	**8.56**	**10.14**			
		6	*5*	*4*	*3*	*2*	*1*	*0*				
	0.2	2.70	4.11	4.74	5.56	5.56	6.41	8.52				
	0.3	**2.69**	4.04	4.68	5.45	5.46	6.28	8.40				
7	0.5	2.77	**4.01**	**4.63**	5.28	5.32	6.07	8.25				
	0.7	2.95	4.05	4.66	5.16	**5.30**	**5.96**	8.22				
	0.8	3.06	4.09	4.70	**5.13**	5.34	5.98	**8.21**				

[1] The bold numbers in the table indicate the minimum error

KQ512

The performance of the model for different values of α has been analyzed for KQ512 and the results are displayed in Table B.4 below. Similar as to the table for KQ101 this table shows the average MAE, for the observation points of each specific horizon, taken over the number of flights for which forecasts are computed. The italic numbers indicate the day before departure corresponding with the observation point for that horizon.

Table B.4: Forecasting performance in terms of MAE with different values of α for flight KQ512 at various booking horizons

		Observation point										
Horizon	α	V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}
		299	*270*	*240*	*210*	*180*	*150*	*120*	*90*	*60*	*30*	*0*
	0.2	**0.08**	1.63	1.41	1.67	1.54	1.84	3.24	5.94	4.71	**12.26**	10.08
	0.3	**0.08**	1.59	1.36	1.60	1.46	1.77	3.20	5.90	4.68	**12.26**	**10.07**
300	0.5	0.09	1.52	1.28	1.49	1.34	1.68	3.16	5.84	4.62	12.28	**10.07**
	0.7	0.09	1.46	1.23	1.41	1.26	1.63	3.14	5.81	4.59	12.29	10.08
	0.8	0.09	**1.45**	**1.22**	**1.39**	**1.23**	**1.62**	**3.13**	**5.80**	**4.58**	12.30	10.09
		199	*180*	*160*	*140*	*120*	*100*	*80*	*60*	*40*	*20*	*0*
	0.2	**0.00**	**0.67**	**1.08**	1.31	2.83	4.71	4.88	4.71	8.73	10.99	**10.21**
	0.3	0.01	**0.67**	1.09	1.30	**2.82**	4.69	4.84	4.68	8.69	**10.97**	**10.21**
200	0.5	0.01	0.68	1.09	1.28	**2.82**	4.66	4.80	4.63	8.63	10.99	10.22
	0.7	0.01	0.69	1.11	**1.27**	**2.82**	4.64	**4.78**	4.60	**8.60**	11.02	10.25
	0.8	0.02	0.69	1.11	**1.27**	**2.82**	**4.63**	**4.78**	**4.59**	**8.60**	11.03	10.26
		99	*90*	*80*	*70*	*60*	*50*	*40*	*30*	*20*	*10*	*0*
	0.2	**0.54**	**1.25**	**2.62**	**4.09**	**4.60**	**5.57**	**7.17**	**9.75**	10.25	13.48	11.68
	0.3	0.55	1.27	2.64	4.11	4.61	5.59	7.19	9.78	10.22	13.48	11.57
100	0.5	0.55	1.32	2.68	4.14	4.62	5.61	7.24	9.84	**10.16**	**13.44**	11.32
	0.7	0.56	1.37	2.72	4.18	4.66	5.60	7.31	9.92	10.24	13.52	11.03
	0.8	0.57	1.40	2.73	4.20	4.67	**5.57**	7.35	9.98	10.30	13.59	**10.94**
		59	*54*	*48*	*42*	*36*	*30*	*24*	*18*	*12*	*6*	*0*
	0.2	**0.83**	**2.17**	**4.65**	**5.37**	**7.49**	**8.21**	**9.49**	**10.15**	**11.98**	**11.20**	10.29
	0.3	**0.83**	2.18	4.68	5.40	7.52	8.27	9.58	10.23	12.11	11.35	10.25
60	0.5	0.85	2.19	4.76	5.48	7.60	8.41	9.77	10.42	12.36	11.63	**10.21**
	0.7	0.87	2.22	4.87	5.59	7.70	8.56	9.99	10.63	12.61	11.85	10.22
	0.8	0.88	2.24	4.94	5.65	7.76	8.65	10.10	10.73	12.73	11.93	10.22
		38	*35*	*31*	*27*	*23*	*19*	*15*	*11*	*7*	*3*	*0*
	0.2	**1.64**	**3.08**	**4.85**	**6.15**	**8.01**	**9.29**	**11.29**	**12.03**	12.55	13.28	11.57
	0.3	**1.64**	3.09	4.86	6.16	8.07	9.32	11.37	12.06	12.58	13.25	11.45
39	0.5	1.68	3.16	4.92	6.21	8.20	9.40	11.52	12.12	12.59	13.13	11.35
	0.7	1.73	3.28	5.03	6.26	8.32	9.48	11.64	12.23	12.50	12.99	11.27
	0.8	1.75	3.37	5.10	6.28	8.38	9.52	11.69	12.26	**12.39**	**12.90**	**11.22**
		24	*22*	*19*	*16*	*13*	*10*	*7*	*4*	*0*		
	0.2	**2.19**	**4.37**	**4.88**	**7.54**	**7.36**	**9.14**	9.64	**10.45**	9.66		
	0.3	2.22	4.39	**4.88**	**7.54**	7.37	9.18	**9.63**	10.52	9.55		
25	0.5	2.28	4.43	4.90	7.55	7.41	9.25	9.72	10.72	**9.37**		
	0.7	2.35	4.51	4.95	7.57	7.45	9.31	9.87	10.90	9.38		
	0.8	2.39	4.56	4.99	7.59	7.47	9.33	9.94	11.02	9.44		
		13	*12*	*10*	*8*	*6*	*4*	*2*	*0*			
	0.2	**2.61**	**3.82**	5.32	6.87	7.52	9.35	9.86	9.91			
	0.3	2.66	3.83	**5.25**	6.79	7.33	9.23	9.66	9.51			
14	0.5	2.80	3.87	5.32	**6.75**	**7.24**	**9.08**	9.45	8.91			
	0.7	2.97	4.03	5.48	**6.75**	7.49	9.10	**9.40**	8.54			
	0.8	3.07	4.15	5.57	6.83	7.68	9.19	9.45	**8.45**			
		6	*5*	*4*	*3*	*2*	*1*	*0*				
	0.2	**3.12**	**4.14**	**5.21**	**5.13**	**5.93**	6.95	8.13				
	0.3	3.14	4.15	5.27	5.14	5.96	6.84	7.93				
7	0.5	3.24	4.19	5.43	5.22	6.04	**6.79**	7.58				
	0.7	3.44	4.36	5.64	5.36	6.17	6.80	**7.43**				
	0.8	3.60	4.44	5.77	5.47	6.23	6.86	7.47				

[1] The bold numbers in the table indicate the minimum error

Seasonal input

The tables related to the results of the calibration of the dataset to use yielding the lowest forecasting error for the new model are displayed in this section. Again, similar as in the previous section, the tables for KQ101 are displayed first and thereafter the tables with results of KQ512.

KQ101

The performance of the model for different datasets has been analyzed for KQ101 and the results are displayed in Table B.5 below. The table shows the average MAE, for the observation points of each specific horizon, taken over the number of flights for which forecasts are computed. The italic numbers indicate the day before departure corresponding with the observation point for that horizon.

Table B.5: The comparison of MAE for different input datasets for Flight KQ101 and the set of horizons

		Observation point										
Horizon	**Data input**	V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}
300		*299*	*270*	*240*	*210*	*180*	*150*	*120*	*90*	*60*	*30*	*0*
	Total Set	0.69	**7.56**[1]	12.31	16.09	22.37	25.86	30.85	31.40	31.27	26.90	15.72
	DOW	0.95	8.09	**12.04**	**15.95**	22.44	26.19	30.08	31.14	31.20	25.36	16.15
	Seasonal data	0.85	8.44	13.39	16.79	22.07	25.03	29.50	**28.27**	**26.35**	**22.85**	**14.45**
	DOW and Seasonal combined	**0.53**	7.75	12.71	16.25	**21.69**	**24.90**	**29.31**	28.53	26.72	25.72	17.37
200		*199*	*180*	*160*	*140*	*120*	*100*	*80*	*60*	*40*	*20*	*0*
	Total Set	1.33	7.66	10.41	12.67	15.87	18.32	19.82	20.76	20.95	17.72	12.08
	DOW	1.56	8.48	10.98	13.24	16.69	19.13	20.19	21.22	21.06	18.08	13.56
	Seasonal data	1.38	**7.29**	**10.21**	**12.43**	**15.47**	17.35	16.13	16.17	18.35	**16.40**	**11.34**
	DOW and Seasonal combined	**1.17**	8.61	11.08	13.21	16.05	**17.10**	**15.96**	**16.02**	**18.11**	18.81	16.30
100		*99*	*90*	*80*	*70*	*60*	*50*	*40*	*30*	*20*	*10*	*0*
	Total Set	1.60	5.07	6.86	9.41	11.06	**12.83**	**14.15**	**13.84**	**14.36**	**12.53**	**11.60**
	DOW	**1.47**	**4.86**	6.80	9.45	11.38	13.89	16.03	16.01	17.59	16.96	16.66
	Seasonal data	1.67	4.90	**6.28**	**9.24**	**10.54**	**12.83**	14.33	13.97	14.39	12.65	11.79
	DOW and Seasonal combined	1.66	4.92	6.88	9.83	11.55	14.57	16.59	18.02	19.67	19.98	20.35
60		*59*	*54*	*48*	*42*	*36*	*30*	*24*	*18*	*12*	*6*	*0*
	Total Set	**2.07**	**4.71**	**7.59**	**9.01**	**9.50**	10.43	**10.79**	**11.36**	**11.37**	10.85	**11.74**
	DOW	2.45	5.43	8.17	9.97	10.19	11.15	12.23	13.87	14.64	14.83	15.55
	Seasonal data	2.20	4.84	7.66	9.03	9.70	**10.38**	10.94	11.69	11.87	**10.61**	11.80
	DOW and Seasonal combined	2.09	5.69	8.58	9.85	10.36	11.83	13.24	14.98	16.31	16.27	17.80
39		*38*	*35*	*31*	*27*	*23*	*19*	*15*	*11*	*7*	*3*	*0*
	Total Set	2.61	4.54	**5.28**	**6.46**	**7.24**	**8.37**	**9.34**	**9.69**	**9.68**	**10.28**	**11.31**
	DOW	**2.11**	**4.41**	5.84	7.10	7.83	10.10	12.07	11.58	12.80	13.30	13.90
	Seasonal data	2.77	4.73	5.60	6.79	7.86	9.20	10.34	10.60	10.21	10.73	11.91
	DOW and Seasonal combined	2.48	4.47	6.02	7.22	8.57	10.39	12.05	14.04	13.15	14.80	16.43
25		*24*	*22*	*19*	*16*	*13*	*10*	*7*	*4*	*0*		
	Total Set	2.84	5.16	**6.01**	**7.44**	**7.95**	**8.59**	**8.24**	**8.69**	**10.17**		
	DOW	**2.72**	**4.76**	6.59	7.65	9.15	9.57	9.57	10.37	11.39		
	Seasonal data	3.04	5.29	6.40	8.01	8.46	9.07	8.85	9.21	10.81		
	DOW and Seasonal combined	2.90	5.02	6.28	8.08	9.95	11.07	9.91	11.98	13.46		
14		*13*	*12*	*10*	*8*	*6*	*4*	*2*	*0*			
	Total Set	3.96	5.50	**6.62**	**5.84**	**6.43**	7.78	8.56	10.14			
	DOW	3.99	**5.05**	6.64	6.71	7.68	8.36	8.70	11.09			
	Seasonal data	**3.86**	5.31	6.66	6.15	6.51	**7.60**	**8.29**	**10.04**			
	DOW and Seasonal combined	4.14	6.10	6.90	6.61	8.37	9.64	9.19	12.75			
7		*6*	*5*	*4*	*3*	*2*	*1*	*0*				
	Total Set	**3.06**	4.09	4.70	5.13	5.34	5.98	8.21				
	DOW	3.27	**4.03**	**4.35**	**5.02**	**5.16**	6.27	8.64				
	Seasonal data	3.22	4.07	4.79	5.21	5.47	**5.77**	**8.10**				
	DOW and Seasonal combined	3.85	4.34	4.93	5.18	5.29	6.89	9.83				

[1] The bold numbers in the table indicate the minimum error

KQ512

The performance of the model for different datasets has been analyzed for KQ512 and the results are displayed in Table B.6 below. Similar as to the table for KQ101 this table shows the average MAE, for the observation points of each specific horizon, taken over the number of flights for which forecasts are computed. The italic numbers indicate the day before departure corresponding with the observation point for that horizon.

Table B.6: The comparison of MAE for different input datasets for Flight KQ512 and the set of horizons

		Observation point										
Horizon	**Data input**	V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}
		299	*270*	*240*	*210*	*180*	*150*	*120*	*90*	*60*	*30*	*0*
	Total Set	0.09	1.52	1.28	1.49	1.34	1.68	3.16	**5.84**	4.62	12.28	10.07
300	DOW	**0.08**[1]	**1.49**	1.27	1.44	**1.32**	1.72	3.18	5.90	**4.55**	**11.14**	**9.24**
	Seasonal data	**0.08**	1.53	**1.25**	1.38	1.34	**1.63**	3.16	5.86	4.68	12.23	9.27
	DOW and Seasonal combined	0.09	1.56	1.33	**1.29**	**1.32**	1.70	**3.07**	6.39	5.43	13.02	9.80
		199	*180*	*160*	*140*	*120*	*100*	*80*	*60*	*40*	*20*	*0*
	Total Set	0.01	**0.68**	**1.09**	**1.28**	2.82	**4.66**	4.80	4.63	**8.63**	10.99	10.22
200	DOW	**0.00**	0.71	1.12	**1.28**	2.84	4.68	**4.78**	**4.56**	8.70	**10.69**	**8.86**
	Seasonal data	0.01	0.69	1.11	1.31	**2.80**	4.70	5.01	4.89	8.68	10.94	9.89
	DOW and Seasonal combined	**0.00**	0.77	1.17	1.35	**2.80**	5.19	5.61	5.61	9.87	10.79	11.17
		99	*90*	*80*	*70*	*60*	*50*	*40*	*30*	*20*	*10*	*0*
	Total Set	**0.55**	**1.32**	2.68	4.14	4.62	5.61	7.24	9.84	10.16	13.44	11.32
100	DOW	0.56	1.33	**2.64**	**4.09**	**4.48**	**5.40**	**7.12**	**8.91**	**9.48**	**10.77**	**10.20**
	Seasonal data	0.61	1.56	2.89	4.54	4.95	5.75	7.21	9.78	9.94	13.36	11.36
	DOW and Seasonal combined	0.59	2.18	3.25	4.98	5.56	5.97	8.05	10.45	9.75	12.95	12.48
		59	*54*	*48*	*42*	*36*	*30*	*24*	*18*	*12*	*6*	*0*
	Total Set	0.85	2.19	**4.76**	5.48	7.60	8.41	9.77	10.42	12.36	11.63	10.21
60	DOW	0.74	**2.17**	4.81	5.79	**7.48**	**7.72**	**8.80**	**8.83**	**10.18**	**11.19**	**9.22**
	Seasonal data	1.02	2.45	4.93	**5.44**	7.70	8.51	9.82	10.19	12.19	11.73	9.97
	DOW and Seasonal combined	**0.72**	2.46	5.42	5.85	8.38	9.04	9.09	9.72	11.67	11.86	10.19
		38	*35*	*31*	*27*	*23*	*19*	*15*	*11*	*7*	*3*	*0*
	Total Set	1.68	3.16	4.92	6.21	8.20	9.40	11.52	12.12	12.59	13.13	11.35
39	DOW	**1.49**	**3.03**	**4.54**	5.99	7.50	**8.25**	**9.23**	**10.40**	12.03	**12.80**	**10.67**
	Seasonal data	1.75	3.08	4.75	6.16	8.27	9.32	11.87	12.31	13.05	14.04	11.66
	DOW and Seasonal combined	1.69	3.23	4.72	**5.83**	**7.04**	8.48	10.72	11.53	**11.79**	13.78	12.73
		24	*22*	*19*	*16*	*13*	*10*	*7*	*4*	*0*		
	Total Set	2.28	4.43	4.90	7.55	7.41	9.25	**9.72**	10.72	9.37		
25	DOW	2.34	**4.19**	4.80	**6.37**	**7.36**	**8.87**	9.84	**10.07**	**9.23**		
	Seasonal data	2.30	4.36	4.99	7.73	7.69	9.81	10.76	11.71	10.14		
	DOW and Seasonal combined	**2.08**	4.54	**4.65**	8.07	7.80	9.66	11.16	15.27	11.70		
		13	*12*	*10*	*8*	*6*	*4*	*2*	*0*			
	Total Set	**2.80**	**3.87**	**5.32**	6.75	**7.24**	9.08	**9.45**	8.91			
14	DOW	2.87	4.28	5.46	7.07	**7.24**	**8.57**	9.61	**8.89**			
	Seasonal data	2.86	4.16	5.36	**6.71**	7.54	9.80	9.82	9.46			
	DOW and Seasonal combined	2.94	4.96	5.91	7.23	11.19	14.12	11.50	11.20			
		6	*5*	*4*	*3*	*2*	*1*	*0*				
	Total Set	**3.24**	**4.19**	5.43	**5.22**	6.04	**6.79**	**7.58**				
7	DOW	3.76	4.38	**4.52**	5.27	**5.96**	6.94	7.68				
	Seasonal data	3.60	4.68	5.84	5.85	6.55	7.32	8.23				
	DOW and Seasonal combined	4.61	6.76	8.43	6.85	6.54	8.34	8.53				

[1] The bold numbers in the table indicate the minimum error

Calibration of Model Parameters to Prediction Intervals

This section includes the tables that present the prediction interval scores of the model for different parameter settings. The same order as in the calibration on MAE is used in this section; first the tables with results for the parametric distribution approach are shown, thereafter the results for the different weighting factors and finally for the datasets.

Parametric Distribution

The tables showing the results of the calibration of the parametric distribution fitting approach in terms of interval scores are displayed in this section. Similar as in the previous sections the results of KQ101 are shown first and finally the results of KQ512.

KQ101

The performance in terms of interval scores of the model for different distribution fitting approaches has been analyzed for KQ101 and the results are displayed in Table B.7 below. The table shows the average interval scores, for the observation points of each specific horizon, taken over the number of flights for which forecasts are computed. The italic numbers indicate the day before departure corresponding with the observation point for that horizon.

Table B.7: Interval Scores for flight KQ101 and the different distributions

Horizon	Distribution	Observation point																																
		V_0			V_1			V_2			V_3			V_4			V_5			V_6			V_7			V_8			V_9			V_{10}		
		Prediction Interval			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**		
		95%	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**
			299			*270*			*240*			*210*			*180*			*150*			*120*			*90*			*60*			*30*			*0*	
	No Fit	**9.27**[1]	1.90	**0.96**	57.72	**28.84**	**18.99**	77.77	**47.78**	**33.35**	**112.30**	**63.24**	**43.17**	**174.21**	**93.54**	**63.26**	**202.31**	**109.03**	**74.77**	**239.69**	**130.82**	**87.47**	**214.50**	**133.20**	**90.75**	**184.50**	**125.04**	**89.61**	**152.69**	**104.27**	**76.41**	148.80	**75.15**	**46.55**
300	Norm	9.30	1.88	**0.96**	57.77	29.13	19.02	77.60	49.88	33.83	114.92	65.82	44.17	195.20	98.03	65.80	234.88	120.10	80.01	283.14	143.12	94.62	263.23	147.37	103.56	230.01	145.55	110.14	168.03	125.91	101.21	140.69	87.82	63.91
	Lognorm	9.88	**1.84**	**0.96**	**56.90**	28.90	19.08	**77.35**	49.52	33.81	115.58	65.65	44.16	194.42	97.86	65.67	233.98	119.40	79.46	282.28	142.50	93.82	260.88	146.47	102.64	228.71	144.47	108.84	167.07	125.02	99.83	**140.52**	87.57	63.35
	Best fit	9.30	1.88	**0.96**	57.28	29.12	19.06	77.92	49.97	33.80	115.28	66.07	44.05	194.52	98.22	65.93	235.48	120.31	80.09	284.56	143.54	94.65	264.13	147.42	103.58	230.33	145.84	110.43	168.14	126.27	101.37	140.70	88.05	64.50
			199			*180*			*160*			*140*			*120*			*100*			*80*			*60*			*40*			*20*			*0*	
	No Fit	**18.55**	6.04	**3.21**	**71.90**	**32.91**	**21.28**	**77.65**	**43.99**	**29.84**	**100.53**	**54.12**	**37.40**	**127.30**	**70.22**	**46.67**	**120.47**	**76.75**	**52.57**	**119.16**	**77.58**	**56.81**	**126.00**	**81.18**	**59.03**	**123.22**	**80.96**	**60.36**	131.24	**73.68**	**51.47**	136.47	**67.32**	**38.22**
200	Norm	19.75	**5.77**	3.22	76.17	33.78	21.88	83.49	46.74	31.21	109.49	58.01	39.88	140.26	74.83	50.05	132.24	84.49	58.20	128.79	88.52	64.57	143.50	93.30	70.19	124.02	95.31	74.67	127.11	85.61	65.39	133.57	69.99	49.46
	Lognorm	19.73	5.80	**3.21**	75.97	33.76	21.83	82.89	46.54	31.03	109.08	57.70	39.75	140.36	74.32	49.81	131.86	83.90	57.75	128.34	88.03	64.17	142.54	92.94	69.52	123.48	95.05	74.12	**126.59**	85.32	65.17	**133.27**	70.10	49.47
	Best fit	19.75	**5.77**	3.22	76.17	33.78	21.88	83.45	46.77	31.21	109.37	58.01	39.90	140.49	74.73	49.98	132.02	84.45	58.17	129.07	88.50	64.49	143.35	93.31	70.19	123.83	95.42	74.77	126.89	85.97	65.52	133.66	70.25	49.83
			99			*90*			*80*			*70*			*60*			*50*			*40*			*30*			*20*			*10*			*0*	
	No Fit	16.55	**6.59**	**4.12**	**34.27**	**19.40**	**13.08**	47.92	**26.33**	**18.44**	**76.30**	**41.21**	**26.23**	**87.84**	**45.37**	**31.36**	**85.57**	**51.10**	**36.75**	99.63	**56.77**	**41.80**	104.34	**58.02**	**43.68**	109.35	**61.03**	**46.36**	111.65	**63.99**	**45.52**	122.91	64.22	**43.19**
100	Norm	**16.11**	6.77	4.14	34.37	19.71	13.15	45.46	26.65	18.67	81.73	42.06	27.08	88.36	46.83	32.93	87.91	54.17	39.34	98.65	61.44	45.26	**101.36**	62.87	45.97	**107.64**	64.09	48.96	107.03	64.67	47.50	116.15	**61.72**	44.39
	Lognorm	**16.11**	6.77	4.14	34.35	19.71	13.15	**45.29**	26.57	18.67	81.57	41.92	26.96	88.63	46.87	32.92	87.78	54.18	39.30	98.71	61.34	45.32	101.50	62.85	46.06	107.83	64.25	49.17	**106.65**	64.48	47.59	**115.73**	61.79	44.73
	Best fit	**16.11**	6.77	4.14	34.36	19.71	13.15	45.42	26.62	18.66	81.68	42.05	27.02	88.29	46.81	32.93	87.86	54.16	39.34	**98.50**	61.39	45.31	101.66	62.88	46.02	107.95	64.25	49.11	106.75	64.49	47.57	115.87	61.76	44.76
			59			*54*			*48*			*42*			*36*			*30*			*24*			*18*			*12*			*6*			*0*	
	No Fit	**17.91**	8.85	**5.22**	30.82	**18.40**	13.12	67.63	**31.46**	**21.44**	**77.82**	38.84	**25.54**	76.49	40.34	**27.04**	82.73	43.73	**29.87**	81.87	47.46	**32.75**	90.79	52.39	37.31	91.37	52.85	38.61	95.95	53.08	38.36	99.68	57.07	41.43
60	Norm	19.54	**8.82**	**5.22**	30.89	18.54	13.10	65.73	31.64	21.64	78.28	**38.61**	26.37	74.30	**39.99**	27.20	78.87	43.30	30.37	76.53	47.16	32.88	**86.09**	**51.56**	36.58	84.85	51.73	38.31	**87.96**	50.25	37.28	93.82	**54.35**	39.29
	Lognorm	19.54	**8.82**	**5.22**	**30.74**	18.41	**13.07**	**65.61**	31.66	21.71	78.54	38.66	26.35	**74.02**	**39.99**	27.38	**78.52**	**43.28**	30.46	**76.25**	**47.04**	32.94	86.34	51.58	**36.56**	**84.53**	51.63	38.25	88.09	50.12	**37.12**	93.82	54.44	**39.21**
	Best fit	19.54	**8.82**	**5.22**	30.80	18.44	13.08	65.67	31.65	21.68	78.59	38.67	26.40	74.14	40.00	27.33	78.61	43.34	30.46	76.36	47.06	32.94	86.37	**51.56**	**36.56**	84.62	**51.60**	**38.21**	88.13	**50.01**	**37.12**	**93.42**	54.42	**39.21**
			38			*35*			*31*			*27*			*23*			*19*			*15*			*11*			*7*			*3*			*0*	
	No Fit	16.67	9.51	6.21	33.75	17.19	11.76	44.10	22.23	14.59	54.40	28.74	18.83	57.72	31.23	21.47	61.05	36.03	25.34	67.29	40.96	29.55	72.97	43.92	31.84	77.53	46.20	33.62	81.13	50.09	36.74	85.16	51.89	38.17
39	Norm	16.30	**9.41**	6.10	32.51	16.95	**11.67**	41.67	21.32	14.32	**52.38**	27.46	18.33	53.99	30.08	20.69	56.78	34.25	24.68	63.89	38.78	28.60	68.09	42.51	**31.33**	71.67	43.38	31.67	74.25	46.70	34.73	79.17	49.06	36.27
	Lognorm	**16.28**	9.45	**6.09**	**32.47**	**16.87**	11.69	**41.52**	**21.26**	**14.29**	52.68	27.31	**18.17**	**53.75**	**30.00**	20.60	57.09	**34.13**	24.58	**63.66**	**38.66**	28.55	**67.80**	42.36	31.37	**71.40**	**43.22**	**31.42**	**73.83**	46.28	**34.61**	**78.91**	**48.75**	**35.97**
	Best fit	16.30	**9.41**	6.10	32.48	16.90	11.69	**41.52**	21.30	14.31	52.70	**27.29**	18.20	53.79	30.01	**20.59**	**56.67**	34.22	**24.56**	**63.66**	**38.66**	**28.54**	67.87	**42.30**	**31.33**	71.46	43.32	31.47	73.85	**46.21**	**34.61**	78.97	48.86	36.00
			24			*22*			*19*			*16*			*13*			*10*			*7*			*4*			*0*								
	No Fit	18.75	**10.99**	**7.19**	**32.15**	18.86	**13.26**	38.33	24.38	16.83	44.17	28.66	20.32	55.37	32.42	22.91	57.43	34.41	24.94	63.74	37.57	25.33	66.82	40.72	27.96	71.77	44.38	31.08						
25	Norm	**18.27**	11.01	7.22	33.80	**18.57**	13.28	37.49	24.10	**16.47**	41.40	27.94	**19.76**	53.51	31.54	22.48	53.80	33.67	24.64	59.42	35.62	**24.12**	62.79	38.68	**26.64**	67.36	42.44	30.01						
	Lognorm	20.20	11.01	7.23	32.84	18.71	13.32	**37.26**	24.08	16.55	**41.22**	**27.88**	19.79	**53.29**	31.56	**22.45**	**53.55**	**33.56**	24.63	**59.21**	**35.41**	24.14	**62.54**	**38.36**	26.69	**67.14**	**42.20**	**29.79**						
	Best fit	20.20	11.01	7.23	32.84	18.71	13.32	37.27	**24.07**	16.55	41.25	**27.88**	19.79	53.30	**31.53**	22.48	**53.55**	33.57	**24.60**	59.25	**35.41**	24.14	62.55	**38.36**	26.70	67.15	42.21	29.81						
			13			*12*			*10*			*8*			*6*			*4*			*2*			*0*											
	No Fit	**40.65**	**16.21**	10.42	**52.88**	**22.01**	14.90	40.74	24.05	**17.46**	40.93	23.60	16.06	46.38	27.27	19.29	51.76	32.87	22.88	54.35	33.00	23.95	64.39	38.67	28.32									
14	Norm	44.48	16.29	**10.34**	54.91	22.13	**14.85**	40.82	23.93	17.61	39.33	**22.73**	**15.85**	43.39	**26.44**	**18.68**	48.33	**32.10**	22.51	51.74	32.23	23.27	62.49	38.24	27.86									
	Lognorm	45.14	16.36	10.43	53.84	22.27	14.91	**39.88**	**23.89**	17.61	**39.10**	22.75	15.88	43.27	26.66	18.70	**48.11**	32.13	**22.39**	**51.62**	**32.00**	**23.12**	62.70	**38.19**	27.73									
	Best fit	44.48	16.29	**10.34**	53.20	22.24	14.87	39.95	23.91	17.60	39.20	22.80	15.89	**43.22**	26.61	18.71	48.12	32.13	22.41	51.64	32.05	23.13	**62.33**	38.21	**27.67**									
			6			*5*			*4*			*3*			*2*			*1*			*0*													
	No Fit	**23.25**	12.73	8.10	**31.15**	**17.13**	**11.10**	32.80	18.03	13.04	**33.92**	20.67	14.17	38.72	20.65	14.10	40.80	22.76	16.16	**52.74**	32.36	**22.55**												
7	Norm	24.26	**12.62**	**7.99**	33.67	17.25	11.13	31.87	18.01	12.98	34.34	20.55	**14.13**	36.84	20.25	**14.04**	39.64	22.14	**15.91**	55.14	32.27	22.64												
	Lognorm	24.20	12.83	8.02	33.23	17.23	11.24	**31.77**	**17.89**	**12.91**	34.25	**20.46**	**14.13**	**36.41**	**20.10**	14.05	**39.51**	22.14	15.98	55.14	**32.25**	22.64												
	Best fit	24.20	12.83	8.02	33.23	17.23	11.24	**31.77**	**17.89**	**12.91**	34.25	**20.46**	**14.13**	36.46	**20.10**	14.06	39.52	**22.03**	15.96	55.14	**32.25**	22.64												

[1] The bold numbers in the table indicate the minimum interval scores

KQ512

The performance in terms of interval scores of the model for different distribution fitting approaches has also been analyzed for KQ512 and the results are displayed in Table B.8 below. Similar as to the table for KQ101 this table shows the average prediction interval score, for the observation points of each specific horizon, taken over the number of flights for which forecasts are computed. The italic numbers indicate the day before departure corresponding with the observation point for that horizon.

Table B.8: Interval Scores for flight KQ512 and the different distributions

Horizon	Distribution	Observation point																																
		V_0			V_1			V_2			V_3			V_4			V_5			V_6			V_7			V_8			V_9			V_{10}		
		Prediction Interval			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**		
		95%	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**
			299			*270*			*240*			*210*			*180*			*150*			*120*			*90*			*60*			*30*			*0*	
	No Fit	3.23	**0.45**[1]	**0.22**	16.05	6.07	**4.18**	9.83	**4.13**	**3.17**	**5.95**	**5.38**	**3.84**	**7.35**	**3.66**	**3.17**	13.73	**6.28**	**4.19**	57.68	**14.98**	9.39	**125.15**	29.48	17.82	**50.63**	**21.22**	**14.32**	**114.80**	**49.27**	**34.30**	82.93	41.16	**27.85**
300	Norm	**1.85**	**0.45**	**0.22**	**6.45**	**5.59**	4.44	**8.38**	5.46	4.86	10.05	5.93	4.99	11.38	6.25	4.64	**13.13**	7.29	5.06	**54.08**	15.11	**8.97**	125.93	**28.77**	**17.50**	52.10	22.19	15.51	127.08	54.07	35.66	**74.98**	**40.60**	28.59
	Lognorm	7.23	1.45	**0.22**	93.95	51.00	35.04	569.80	176.03	110.69	1282.78	327.42	188.61	1755.03	404.02	227.57	1875.45	426.09	238.69	1778.45	410.64	231.08	1634.10	384.77	218.56	1505.03	381.33	218.22	777.13	333.55	197.35	236.50	246.42	182.74
	Best fit	7.23	1.45	**0.22**	93.95	51.00	35.04	569.80	176.03	110.69	1282.78	327.42	188.61	1755.03	404.02	227.57	1875.45	426.09	238.69	1778.45	410.64	231.08	1629.15	384.62	218.56	1496.08	379.01	216.90	749.55	331.23	196.31	229.85	243.39	181.47
			199			*180*			*160*			*140*			*120*			*100*			*80*			*60*			*40*			*20*			*0*	
	No Fit	**0.12**	**0.00**	**0.00**	**3.36**	2.64	**1.62**	12.32	**4.33**	2.95	10.42	4.99	**3.43**	51.22	**13.30**	8.46	100.36	23.99	**13.97**	78.27	22.31	14.26	44.42	19.51	13.58	111.78	39.53	25.52	**87.00**	**46.18**	**30.92**	82.63	41.97	**27.96**
200	Norm	1.49	**0.00**	**0.00**	4.46	**2.28**	1.84	**11.71**	4.58	**2.92**	**8.80**	**4.92**	3.55	51.39	**13.30**	**8.24**	100.78	24.26	14.70	79.98	24.65	15.63	47.61	21.63	15.21	117.85	43.58	27.79	93.93	49.14	32.25	**74.83**	40.69	28.62
	Lognorm	4.02	1.03	**0.00**	12.68	8.22	6.98	16.95	15.05	11.84	17.31	11.51	9.06	**46.17**	16.38	10.61	**84.85**	**23.51**	14.36	69.92	**20.69**	**12.96**	40.08	**18.26**	**12.67**	**110.86**	**38.59**	**25.28**	87.90	46.39	31.10	75.56	**40.62**	28.10
	Best fit	4.02	1.03	**0.00**	12.68	8.22	6.98	16.95	15.05	11.84	17.31	11.51	9.06	**46.17**	16.38	10.61	**84.85**	**23.51**	14.36	**69.81**	20.82	13.01	**39.97**	18.62	**12.67**	112.54	38.88	25.50	89.46	46.84	31.10	75.61	40.66	28.10
			99			*90*			*80*			*70*			*60*			*50*			*40*			*30*			*20*			*10*			*0*	
	No Fit	**7.05**	3.00	**1.45**	**10.94**	5.66	3.35	**32.27**	**12.63**	**7.83**	**69.38**	**19.17**	**11.75**	**66.22**	**20.84**	**13.20**	**32.59**	**24.47**	**16.16**	**57.44**	**31.43**	**20.87**	**64.30**	**39.39**	**27.36**	**68.52**	**42.32**	**28.06**	70.42	**51.12**	**36.28**	83.63	44.46	30.79
100	Norm	7.59	**2.95**	**1.45**	11.42	**5.49**	**3.19**	33.95	12.83	7.97	74.67	19.89	11.92	69.70	21.61	13.74	45.61	25.63	17.36	67.59	34.43	22.65	70.69	41.90	28.82	69.75	42.38	29.20	**65.03**	52.02	36.48	75.11	**42.59**	29.87
	Lognorm	7.59	**2.95**	**1.45**	11.44	5.54	**3.19**	33.39	12.76	7.88	74.70	19.91	11.92	69.72	21.52	13.67	45.00	25.19	17.12	67.13	34.35	22.59	71.25	41.69	28.66	69.61	42.38	29.30	65.25	51.93	36.47	**74.56**	42.61	29.90
	Best fit	7.59	**2.95**	**1.45**	11.42	**5.49**	**3.19**	33.95	12.83	7.97	74.67	19.89	11.92	69.70	21.61	13.74	45.61	25.63	17.36	67.59	34.43	22.65	70.66	41.88	28.83	69.58	42.52	29.27	65.34	51.96	36.56	74.67	42.62	**29.82**
			59			*54*			*48*			*42*			*36*			*30*			*24*			*18*			*12*			*6*			*0*	
	No Fit	**12.16**	**3.49**	2.05	20.42	**8.74**	6.14	**82.30**	**22.48**	**13.39**	**75.06**	**25.23**	**15.66**	**78.55**	**32.31**	**21.22**	**59.88**	**33.06**	**22.93**	**65.63**	**37.92**	**26.06**	**61.91**	**42.42**	**29.26**	**66.45**	**46.08**	**33.97**	73.17	**43.71**	31.18	81.84	42.33	27.93
60	Norm	12.38	3.52	2.02	19.41	9.11	**6.01**	87.34	23.28	13.79	80.00	26.79	16.31	85.92	34.12	22.17	65.20	34.51	23.73	73.20	39.19	26.41	63.75	43.07	29.80	67.14	48.28	34.24	69.36	43.81	30.60	73.80	41.54	27.97
	Lognorm	12.97	**3.49**	**2.01**	**19.25**	9.18	**6.01**	87.89	23.10	13.77	79.86	26.56	16.31	86.23	34.00	22.15	64.92	34.56	23.75	72.80	39.13	26.44	63.88	43.01	29.94	67.25	48.15	34.21	**68.58**	43.78	30.59	**73.42**	**41.48**	**27.88**
	Best fit	12.38	3.52	2.02	19.41	9.11	**6.01**	87.34	23.28	13.79	80.00	26.79	16.31	85.92	34.12	22.17	65.72	34.65	23.78	72.97	39.19	26.40	64.00	43.06	29.88	66.67	48.26	34.16	68.86	43.72	**30.53**	73.48	41.49	27.89
			38			*35*			*31*			*27*			*23*			*19*			*15*			*11*			*7*			*3*			*0*	
	No Fit	**18.23**	**7.25**	4.58	20.33	12.17	8.66	44.27	20.09	13.29	**40.58**	24.38	17.04	53.91	30.19	21.84	62.47	37.40	25.75	76.69	43.33	30.74	75.64	47.24	33.44	78.88	45.72	32.81	87.28	48.07	34.58	101.59	48.00	32.05
39	Norm	19.27	7.31	**4.53**	18.55	**12.14**	**8.63**	**43.67**	**19.77**	12.84	41.14	23.85	16.73	51.22	29.45	20.80	**60.22**	36.41	25.37	**76.23**	42.41	30.19	**73.47**	45.78	32.81	**74.03**	43.51	**30.94**	81.11	45.35	31.98	94.52	46.03	30.82
	Lognorm	19.75	7.36	**4.53**	**18.52**	**12.14**	**8.63**	44.17	19.82	**12.76**	41.48	23.84	**16.70**	52.22	29.39	**20.73**	61.73	**36.24**	**25.25**	76.33	42.41	**30.11**	74.17	**45.69**	32.76	74.23	**43.39**	31.05	81.02	**45.04**	**31.70**	**94.02**	**46.02**	**30.63**
	Best fit	19.27	7.31	**4.53**	18.55	**12.14**	**8.63**	44.22	**19.77**	**12.76**	41.03	**23.79**	16.75	**51.05**	**29.37**	20.75	61.13	36.32	25.26	76.56	**42.36**	30.19	74.25	45.77	**32.73**	74.31	43.41	31.07	**80.63**	45.08	31.80	94.11	46.08	30.77
			24			*22*			*19*			*16*			*13*			*10*			*7*			*4*			*0*							
	No Fit	**18.47**	9.42	5.71	29.84	**17.26**	12.19	38.38	21.02	**13.62**	55.27	30.21	20.61	**59.63**	32.57	21.05	58.16	39.69	25.53	61.16	39.47	26.93	70.13	41.78	28.26	78.23	42.29	27.04						
25	Norm	19.27	**9.30**	**5.70**	29.16	17.52	**12.11**	**37.53**	20.87	13.82	53.41	29.82	20.64	60.11	31.87	20.80	57.97	38.67	**25.15**	**58.14**	38.28	25.91	66.66	40.52	27.40	**72.13**	40.69	26.46						
	Lognorm	20.55	**9.30**	**5.70**	**28.88**	17.71	12.15	37.72	**20.73**	13.79	**52.81**	**29.74**	**20.59**	61.38	31.81	**20.74**	**57.48**	38.79	25.26	58.27	**38.06**	25.84	**66.23**	40.34	**27.14**	72.28	**40.34**	**26.27**						
	Best fit	20.55	**9.30**	**5.70**	28.98	17.73	12.16	37.83	20.78	13.73	52.86	**29.74**	20.61	61.42	**31.77**	20.80	57.66	**38.64**	25.21	58.36	38.09	**25.81**	66.28	**40.30**	27.23	72.52	40.37	**26.27**						
			13			*12*			*10*			*8*			*6*			*4*			*2*			*0*											
	No Fit	**16.38**	12.10	**7.43**	35.02	15.91	10.81	42.58	23.48	14.74	46.28	27.46	18.42	55.73	29.86	20.36	57.20	33.28	23.85	61.05	35.30	25.19	64.09	35.70	24.63									
14	Norm	17.05	12.02	7.55	34.69	15.77	**10.77**	42.36	23.11	14.58	**45.00**	**26.86**	18.18	54.75	29.10	19.98	54.61	**32.29**	23.38	57.05	34.01	24.61	59.41	34.45	24.52									
	Lognorm	16.92	**11.98**	7.58	34.00	**15.74**	10.87	**41.94**	**23.02**	**14.55**	45.25	26.94	**18.09**	**54.42**	**29.04**	**19.90**	**54.52**	32.35	23.32	**56.78**	33.88	**24.33**	**59.31**	**34.38**	**24.38**									
	Best fit	17.05	12.02	7.55	**33.98**	15.77	10.84	42.06	23.05	14.57	45.27	26.94	18.14	54.48	29.06	19.93	54.59	32.39	**23.29**	56.84	**33.87**	24.38	**59.31**	**34.38**	24.43									
			6			*5*			*4*			*3*			*2*			*1*			*0*														
	No Fit	24.94	13.24	9.07	33.16	17.94	11.89	39.11	22.61	15.09	42.75	23.12	15.33	46.00	24.51	16.18	52.48	29.02	18.77	54.11	31.11	20.81												
7	Norm	24.81	**13.08**	9.07	**31.36**	17.38	12.02	38.61	22.30	14.88	**40.23**	22.39	15.00	43.31	23.56	15.89	49.48	28.56	18.38	51.42	30.78	20.81												
	Lognorm	**24.80**	13.11	**9.05**	31.39	17.37	**11.86**	38.63	**22.29**	**14.79**	40.31	**22.38**	14.86	**43.19**	**23.54**	**15.81**	**49.39**	**28.48**	**18.32**	51.28	30.75	20.76												
	Best fit	24.81	**13.08**	9.07	**31.36**	**17.35**	11.91	**38.59**	**22.29**	14.84	40.27	**22.38**	**14.81**	43.22	**23.54**	**15.81**	49.41	**28.48**	18.33	**51.27**	**30.72**	**20.73**												

[1] The bold numbers in the table indicate the minimum interval scores

Weighting factor

The tables containing the results of the calibration of the weighting factor for prediction interval scores of the new model are displayed in this section. First, the results of KQ101, then the results of KQ512.

KQ101

The performance in terms of interval scores of the model for different weighting factors has been analyzed for KQ101 and the results are displayed in Table B.9 below. The table shows the average interval scores, for the observation points of each specific horizon, taken over the number of flights for which forecasts are computed. The italic numbers indicate the day before departure corresponding with the observation point for that horizon.

Table B.9: Interval Scores for flight KQ101 and the values of α

Horizon	α	Observation point																																
		V_0			V_1			V_2			V_3			V_4			V_5			V_6			V_7			V_8			V_9			V_{10}		
		Prediction Interval			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**		
		95%	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**
			299			*270*			*240*			*210*			*180*			*150*			*120*			*90*			*60*			*30*			*0*	
	0.2	**8.33**[1]	**1.80**	**0.80**	65.91	30.15	19.34	103.10	51.99	34.28	148.91	67.34	44.59	253.02	99.75	64.33	307.24	118.04	76.39	370.81	139.74	89.30	322.66	142.08	93.97	263.79	136.61	95.20	164.13	111.91	84.06	146.48	81.02	55.31
300	0.3	8.56	1.84	**0.80**	60.58	29.65	19.15	89.00	50.29	33.82	130.42	65.14	43.99	215.91	97.17	63.86	260.79	114.87	75.73	306.38	135.61	88.68	269.70	137.81	92.69	223.50	131.70	92.82	155.69	108.06	80.76	149.10	78.64	51.61
	0.5	9.27	1.90	0.96	**57.72**	28.84	**18.99**	77.77	47.78	33.35	112.30	63.24	43.17	174.21	93.54	63.26	202.31	109.03	74.77	239.69	130.82	87.47	214.50	133.20	90.75	184.50	125.04	89.61	**152.69**	104.27	76.41	148.80	75.15	46.55
	0.7	9.72	1.95	1.31	59.55	**27.90**	19.01	73.81	**46.74**	32.75	104.20	62.52	42.70	146.66	91.56	62.74	178.98	106.95	74.27	208.02	128.53	87.00	188.88	129.82	89.73	173.88	122.99	**88.29**	153.57	**104.24**	74.19	140.85	71.30	43.46
	0.8	11.37	2.00	1.33	59.79	28.44	19.10	**73.34**	46.84	**32.63**	**103.22**	**62.31**	**42.42**	**140.42**	**90.58**	**62.63**	**167.47**	**105.82**	**73.95**	**198.81**	**128.11**	**86.88**	**179.43**	**129.26**	**89.61**	**172.55**	**122.88**	88.34	155.05	104.88	**73.66**	**134.37**	**69.24**	**42.19**
			199			*180*			*160*			*140*			*120*			*100*			*80*			*60*			*40*			*20*			*0*	
	0.2	**14.35**	6.15	**3.15**	84.55	33.83	21.68	88.84	45.93	30.83	125.30	57.09	38.77	166.55	73.78	48.59	146.01	81.94	55.33	128.78	83.39	59.93	132.52	83.83	62.72	**113.41**	83.86	63.49	**123.13**	75.75	55.76	137.97	72.94	47.33
200	0.3	16.28	5.96	3.20	77.63	33.43	21.57	82.71	45.05	30.39	111.47	55.99	38.18	147.57	72.22	47.88	131.29	80.06	54.15	121.15	80.49	58.81	127.57	82.31	61.29	115.77	81.97	62.21	126.42	74.48	53.80	138.15	71.27	43.81
	0.5	18.55	6.04	3.21	71.90	32.91	21.28	**77.65**	43.99	29.84	100.53	54.12	37.40	127.30	70.22	46.67	120.47	76.75	52.57	**119.16**	77.58	56.81	**126.00**	**81.18**	59.03	123.22	**80.96**	60.36	131.24	73.68	51.47	136.47	67.32	38.22
	0.7	20.52	6.06	3.37	**71.53**	**32.52**	21.12	78.45	42.48	29.32	94.60	**53.56**	36.77	112.37	68.67	45.75	**113.67**	**74.72**	51.48	119.85	**76.56**	55.51	126.10	82.20	58.30	128.18	81.56	58.93	132.49	73.21	50.16	130.46	61.62	34.57
	0.8	23.24	**5.93**	3.38	72.39	32.56	**20.98**	80.16	**42.10**	**29.18**	**93.92**	53.61	**36.44**	**107.38**	**68.46**	**45.60**	113.77	74.76	**51.23**	121.39	76.76	**55.22**	127.68	83.68	**58.01**	129.86	82.25	**58.84**	131.13	**72.94**	**49.91**	**124.65**	**59.53**	**33.21**
			99			*90*			*80*			*70*			*60*			*50*			*40*			*30*			*20*			*10*			*0*	
	0.2	**13.52**	**6.49**	**4.02**	34.86	**19.06**	**12.75**	**42.10**	26.32	18.70	81.21	**40.52**	26.64	88.42	46.28	33.02	92.60	52.59	39.62	**98.55**	60.98	46.12	**102.26**	65.47	49.42	110.25	73.35	56.37	119.95	80.79	58.91	143.00	82.93	61.04
100	0.3	13.85	6.53	**4.02**	34.55	19.14	12.78	43.65	**26.19**	18.56	79.03	40.92	26.39	88.18	45.63	32.42	89.64	51.72	38.48	98.89	58.91	44.69	102.64	62.09	47.28	109.97	68.40	52.76	113.92	75.15	54.36	129.38	76.35	54.82
	0.5	16.55	6.59	4.12	**34.27**	19.40	13.08	47.92	26.33	18.44	76.30	41.21	**26.23**	87.84	**45.37**	31.36	**85.57**	51.10	36.75	99.63	56.77	41.80	104.34	58.02	43.68	109.35	61.03	46.36	111.65	63.99	45.52	122.91	64.22	43.19
	0.7	17.37	7.01	4.49	36.70	20.14	13.45	52.84	26.90	**18.29**	76.25	41.80	26.29	85.89	45.99	30.73	85.87	50.96	**35.64**	99.42	55.56	40.03	104.50	56.87	40.00	106.48	58.61	40.79	106.52	56.34	37.66	115.26	53.97	33.21
	0.8	22.00	7.32	4.47	38.33	20.67	13.63	54.73	27.21	18.41	**76.05**	42.24	26.44	**85.38**	46.31	**30.56**	85.70	**50.93**	35.97	101.22	**55.14**	**39.66**	103.25	**56.25**	**39.51**	**103.98**	**57.10**	**39.22**	**102.33**	**53.87**	**35.13**	**108.39**	**50.10**	**30.59**
			59			*54*			*48*			*42*			*36*			*30*			*24*			*18*			*12*			*6*			*0*	
	0.2	**15.57**	8.02	**5.07**	29.93	**18.20**	13.18	**67.11**	31.38	21.94	78.87	39.46	26.56	76.76	42.18	28.53	84.40	47.02	33.24	84.00	52.77	37.93	98.20	60.86	44.28	98.26	65.20	48.08	107.21	67.78	51.45	116.90	74.59	55.06
60	0.3	16.35	**7.90**	5.17	**29.75**	18.30	**13.07**	67.17	**31.29**	21.77	78.57	39.26	26.20	**76.12**	41.49	27.95	83.15	45.70	31.90	83.26	50.47	36.26	94.85	57.37	42.08	96.13	60.63	44.74	104.10	62.01	47.13	110.22	67.73	50.21
	0.5	17.91	8.85	5.22	30.82	18.40	13.12	67.63	31.46	**21.44**	**77.82**	38.84	25.54	76.49	40.34	27.04	**82.73**	43.73	29.87	81.87	47.46	32.75	90.79	52.39	37.31	91.37	52.85	38.61	95.95	53.08	38.36	99.68	57.07	41.43
	0.7	19.08	9.52	5.80	32.63	18.83	13.16	69.34	31.40	21.46	78.49	**38.27**	25.42	76.67	39.74	26.38	83.86	42.99	28.62	81.97	45.65	30.56	88.32	49.25	33.69	86.26	48.70	33.36	89.48	47.42	31.38	93.46	49.85	33.77
	0.8	25.47	10.00	5.86	33.32	19.28	13.48	69.87	31.76	21.57	79.57	38.87	**25.20**	77.64	**39.54**	**26.17**	83.58	**42.64**	**28.29**	**81.50**	**44.72**	**29.93**	**86.88**	**47.69**	**32.63**	**83.97**	**46.73**	**31.29**	**85.77**	**44.84**	**29.19**	**90.22**	**46.46**	**31.75**
			38			*35*			*31*			*27*			*23*			*19*			*15*			*11*			*7*			*3*			*0*	
	0.2	17.09	**8.46**	5.55	33.58	17.29	11.80	44.16	23.04	15.45	56.25	29.16	20.18	57.96	33.33	23.43	64.48	40.42	29.66	70.96	46.05	34.67	78.34	53.59	39.17	85.78	56.61	42.17	94.00	63.78	48.52	101.34	65.18	49.45
39	0.3	16.83	8.63	**5.49**	33.62	17.29	**11.76**	44.47	22.64	15.33	55.46	28.67	19.58	57.95	32.63	22.75	63.20	38.91	28.03	69.33	44.19	32.92	76.65	49.97	36.47	82.51	52.65	39.08	87.25	58.70	44.15	95.20	60.45	45.37
	0.5	**16.67**	9.51	6.21	33.75	**17.19**	**11.76**	44.10	22.23	14.59	54.40	28.74	18.83	57.72	31.23	21.47	61.05	36.03	25.34	67.29	40.96	29.55	72.97	43.92	31.84	77.53	46.20	33.62	81.13	50.09	36.74	85.16	51.89	38.17
	0.7	17.49	10.83	7.00	**33.09**	17.43	12.14	44.38	**22.17**	**14.48**	**53.88**	**28.59**	**18.57**	57.57	30.35	20.60	61.01	34.35	23.21	66.76	38.97	27.15	70.54	39.48	27.95	74.45	42.28	29.01	76.07	44.47	30.87	79.00	46.11	32.38
	0.8	19.43	11.21	7.03	33.46	17.68	12.45	**43.99**	22.55	14.55	53.90	28.73	18.74	**56.95**	**30.29**	**20.56**	**60.48**	**33.68**	**22.39**	**65.84**	**37.88**	**26.18**	**69.12**	**38.13**	**26.41**	**71.86**	**40.37**	**27.43**	**72.20**	**42.57**	**28.60**	**75.02**	**43.65**	**30.65**
			24			*22*			*19*			*16*			*13*			*10*			*7*			*4*			*0*							
	0.2	**17.62**	**10.17**	**7.01**	32.41	18.62	**13.13**	**37.66**	24.80	17.47	**43.92**	29.95	21.50	58.36	34.75	24.93	57.39	38.31	28.43	66.35	42.31	29.76	71.63	48.08	34.55	81.36	51.98	37.67						
25	0.3	19.08	10.59	7.03	33.12	**18.55**	13.32	37.77	24.35	17.19	44.11	29.38	21.05	57.05	33.65	24.17	57.23	36.70	27.24	65.50	39.90	28.21	69.76	45.29	32.00	77.17	48.72	35.12						
	0.5	18.75	10.99	7.19	32.15	18.86	13.26	38.33	24.38	16.83	44.17	28.66	20.32	55.37	32.42	22.91	57.43	34.41	24.94	63.74	37.57	25.33	66.82	40.72	27.96	71.77	44.38	31.08						
	0.7	20.29	11.72	7.65	31.32	19.49	13.60	38.54	**24.20**	**16.70**	44.21	28.21	**19.99**	53.70	31.42	21.90	57.07	33.14	23.01	61.04	35.34	23.42	63.60	37.60	25.30	67.68	41.14	28.05						
	0.8	21.08	12.58	7.72	**31.23**	20.10	13.74	38.54	24.23	16.97	44.22	**28.18**	20.13	**52.92**	**31.19**	**21.74**	**56.51**	**32.60**	**22.54**	**59.41**	**34.39**	**22.95**	**61.37**	**36.18**	**24.31**	**65.30**	**39.87**	**27.42**						
			13			*12*			*10*			*8*			*6*			*4*			*2*			*0*											
	0.2	**35.46**	**15.58**	**9.74**	**51.74**	**21.79**	14.91	43.26	24.16	18.10	**40.45**	24.17	16.86	46.10	29.57	20.84	55.20	35.97	25.47	59.21	36.44	26.83	71.30	42.83	30.71									
14	0.3	37.14	15.80	9.94	52.09	21.87	**14.82**	42.54	24.28	17.80	40.90	24.09	16.47	46.39	28.78	20.06	53.04	34.71	24.47	56.93	35.23	25.93	68.45	41.18	29.86									
	0.5	40.65	16.21	10.42	52.88	22.01	14.90	40.74	24.05	**17.46**	40.93	23.60	**16.06**	46.38	27.27	19.29	51.76	32.87	22.88	54.35	33.00	23.95	64.39	38.67	28.32									
	0.7	45.25	17.33	11.78	53.72	22.93	15.44	40.61	**23.96**	17.47	41.17	**23.12**	16.32	45.74	26.40	18.63	50.17	31.56	21.93	51.22	31.24	23.09	60.71	37.34	27.39									
	0.8	47.74	18.48	12.06	54.42	23.32	15.63	**40.58**	24.07	17.58	41.02	23.32	16.32	**45.36**	**26.03**	**18.39**	**49.23**	**30.77**	**21.81**	**50.10**	**30.90**	**22.87**	**59.37**	**37.16**	**26.92**									
			6			*5*			*4*			*3*			*2*			*1*			*0*													
	0.2	**22.37**	**12.41**	**7.67**	31.86	17.47	11.14	34.20	18.55	13.13	34.58	20.56	15.02	39.14	20.94	14.56	42.16	23.39	17.10	54.72	33.00	23.68												
	0.3	23.01	12.52	7.69	**31.12**	17.26	11.17	33.30	18.36	**13.03**	34.33	**20.41**	14.70	39.27	20.76	14.42	41.84	23.07	16.77	53.99	32.75	23.29												
7	0.5	23.25	12.73	8.10	31.15	**17.13**	**11.10**	**32.80**	**18.03**	13.04	**33.92**	20.67	**14.17**	38.72	**20.65**	**14.10**	40.80	22.76	16.16	**52.74**	32.36	22.55												
	0.7	24.23	13.58	8.45	33.97	17.19	11.33	32.85	18.29	13.08	34.48	20.76	14.29	**37.52**	20.92	14.15	40.14	**22.68**	**16.10**	53.85	**32.30**	**22.23**												
	0.8	26.28	14.06	8.72	35.68	17.39	11.60	33.63	18.67	13.27	34.85	21.16	14.43	37.63	21.51	14.11	**39.95**	22.76	16.13	53.86	32.50	22.49												

[1] The bold numbers in the table indicate the minimum interval scores

KQ512

The performance in terms of interval scores of the model for different weighting factors has been analyzed for KQ512 and the results are displayed in Table B.10 below.

Table B.10: Interval Scores for flight KQ512 and the values of α

Horizon	α	Observation point																																
		V_0			V_1			V_2			V_3			V_4			V_5			V_6			V_7			V_8			V_9			V_{10}		
		Prediction Interval			Prediction Interval			Prediction Interval			Prediction Interval			Prediction Interval			Prediction Interval			Prediction Interval			Prediction Interval			Prediction Interval			Prediction Interval			Prediction Interval		
		95%	67%	33%	95%	67%	33%	95%	67%	33%	95%	67%	33%	95%	67%	33%	95%	67%	33%	95%	67%	33%	95%	67%	33%	95%	67%	33%	95%	67%	33%	95%	67%	33%
300			*299*			*270*			*240*			*210*			*180*			*150*			*120*			*90*			*60*			*30*			*0*	
	0.2	**3.00**	**0.45**	**0.22**	21.20	8.07	5.15	12.38	5.47	4.35	8.75	6.75	4.70	**5.40**	4.70	4.08	14.15	6.51	5.05	61.83	15.32	9.51	130.75	30.47	17.75	53.63	21.47	**14.22**	132.50	49.08	**34.10**	90.40	44.07	28.20
	0.3	**3.00**	**0.45**	**0.22**	19.68	6.60	5.03	10.50	4.96	3.86	8.83	6.20	4.21	6.33	3.99	3.61	13.78	6.21	4.66	61.63	15.16	9.44	128.38	30.27	**17.70**	52.68	21.42	14.30	124.33	**48.68**	**34.10**	88.63	43.02	27.85
	0.5	3.23	**0.45**	**0.22**	16.05	6.07	4.18	9.83	**4.13**	3.17	**5.95**	5.38	3.84	7.35	3.66	3.17	13.73	6.28	4.19	57.68	14.98	9.39	**125.15**	29.48	17.82	50.63	21.22	14.32	114.80	49.27	34.30	82.93	41.16	27.85
	0.7	3.35	**0.45**	**0.22**	**13.43**	5.94	3.91	9.18	4.15	2.84	6.70	5.06	3.41	7.85	3.58	2.62	**12.93**	6.05	**4.14**	**55.53**	**14.55**	**9.26**	125.63	**29.18**	17.82	47.75	21.07	14.37	112.85	48.90	34.15	73.60	39.17	27.72
	0.8	3.35	**0.45**	**0.22**	13.50	**5.82**	**3.79**	**8.40**	4.15	**2.67**	6.85	**4.55**	**3.36**	8.23	**3.45**	**2.57**	13.08	**6.02**	4.16	55.80	14.60	9.41	125.83	29.31	18.10	**47.58**	**20.84**	14.47	**105.88**	48.93	34.20	**66.98**	**38.53**	**27.62**
200			*199*			*180*			*160*			*140*			*120*			*100*			*80*			*60*			*40*			*20*			*0*	
	0.2	**0.00**	**0.00**	**0.00**	4.90	2.99	**1.62**	12.51	4.51	2.93	11.53	5.77	3.48	54.69	14.42	8.54	107.53	24.03	13.94	82.54	23.24	14.23	46.81	20.13	13.59	117.71	39.73	25.51	91.49	46.39	31.16	90.36	45.33	28.56
	0.3	0.10	**0.00**	**0.00**	**3.12**	2.69	**1.62**	12.76	4.51	2.97	11.61	5.22	3.50	52.85	14.06	8.61	104.42	23.86	13.99	80.05	23.02	14.25	45.25	19.82	13.63	117.22	39.72	**25.43**	90.58	46.49	31.09	88.58	44.11	28.22
	0.5	0.12	**0.00**	**0.00**	3.36	2.64	**1.62**	12.32	4.33	2.95	10.42	**4.99**	3.43	51.22	**13.30**	8.46	100.36	23.99	13.97	78.27	**22.31**	14.26	44.42	19.51	13.58	111.78	39.53	25.52	87.00	**46.18**	**30.92**	82.63	41.97	27.96
	0.7	0.22	**0.00**	**0.00**	4.07	2.52	**1.62**	12.90	**4.30**	**2.92**	9.61	5.04	**3.41**	49.83	**13.30**	**8.42**	98.46	**23.41**	**13.89**	76.56	22.36	**14.18**	43.22	**19.49**	**13.54**	110.08	**39.03**	25.53	85.02	46.47	30.98	73.63	39.97	27.98
	0.8	0.22	**0.00**	**0.00**	4.07	**2.45**	**1.62**	**12.25**	4.33	2.95	**9.08**	5.04	**3.41**	**49.17**	13.33	**8.42**	**96.92**	**23.41**	13.91	**75.19**	22.50	14.21	**40.37**	**19.49**	13.60	**108.95**	39.07	25.58	**84.69**	46.40	31.11	**67.00**	**39.33**	**27.86**
100			*99*			*90*			*80*			*70*			*60*			*50*			*40*			*30*			*20*			*10*			*0*	
	0.2	7.52	**2.94**	**1.45**	11.05	**5.40**	**3.21**	34.28	12.62	**7.55**	**68.94**	19.22	**11.64**	**65.83**	**20.20**	**13.11**	68.42	24.18	**15.73**	78.06	**31.03**	20.58	79.47	39.69	**27.14**	76.78	43.04	28.24	78.67	52.10	36.44	90.59	49.36	32.29
	0.3	6.94	2.95	**1.45**	11.44	5.47	3.22	32.36	**12.38**	7.57	69.22	**19.17**	**11.64**	66.47	20.46	13.20	56.11	24.51	15.93	77.67	31.13	**20.55**	69.30	39.69	**27.14**	70.22	42.77	28.12	74.30	51.57	36.49	89.06	47.53	31.70
	0.5	7.05	3.00	**1.45**	**10.94**	5.66	3.35	32.27	12.63	7.83	69.38	**19.17**	11.75	66.22	20.84	13.20	**32.59**	24.47	16.16	**57.44**	31.43	20.87	64.30	39.39	27.36	**68.52**	42.32	**28.06**	70.42	51.12	36.28	83.63	44.46	30.79
	0.7	6.94	2.96	**1.45**	11.64	5.80	3.44	32.34	12.63	7.96	69.70	19.20	11.80	67.84	21.04	13.47	33.61	24.40	16.35	58.73	31.56	21.18	60.00	37.91	27.38	69.06	41.72	28.17	69.19	50.49	35.89	74.52	41.54	29.79
	0.8	**6.86**	2.98	**1.45**	11.88	6.00	3.55	**31.83**	12.66	8.00	70.97	19.45	11.99	68.66	20.95	13.47	34.28	**22.63**	16.43	59.25	31.64	21.43	**59.70**	**36.94**	27.51	69.06	**41.07**	28.17	**68.69**	**49.73**	**35.83**	**68.00**	**40.36**	**29.32**
60			*59*			*54*			*48*			*42*			*36*			*30*			*24*			*18*			*12*			*6*			*0*	
	0.2	**10.19**	**3.24**	**1.99**	**18.67**	8.95	**5.96**	83.75	**21.63**	**13.19**	76.19	25.08	**15.32**	82.77	**31.75**	**20.85**	62.27	32.77	**22.62**	67.97	**37.54**	**25.64**	62.34	**41.68**	**28.62**	67.86	46.27	**33.15**	77.45	44.73	**31.00**	89.02	45.74	29.07
	0.3	10.50	3.29	**1.99**	19.36	8.96	6.00	84.05	21.75	13.33	75.22	**24.79**	15.38	80.08	31.83	20.94	62.03	**32.65**	22.71	65.86	37.62	25.73	62.16	41.72	28.92	67.70	46.50	33.50	76.66	44.38	31.05	87.28	44.55	28.59
	0.5	12.16	3.49	2.05	20.42	8.74	6.14	**82.30**	22.48	13.39	**75.06**	25.23	15.66	78.55	32.31	21.22	59.88	33.06	22.93	65.63	37.92	26.06	**61.91**	42.42	29.26	66.45	**46.08**	33.97	73.17	43.71	31.18	81.84	42.33	27.93
	0.7	12.69	3.55	2.22	21.13	8.75	6.18	82.84	22.73	13.50	75.14	25.79	15.84	76.09	32.58	21.62	57.13	33.31	23.35	64.66	38.43	26.56	62.42	42.72	29.96	64.97	46.61	34.20	67.66	**43.34**	31.67	72.67	40.35	**27.74**
	0.8	13.31	3.63	2.22	20.42	**8.66**	6.21	83.75	22.68	14.09	75.75	25.92	16.24	**75.91**	32.89	21.83	**56.81**	33.40	23.65	**64.63**	38.41	26.74	63.20	43.27	30.17	**64.13**	46.92	34.62	**63.89**	43.86	31.64	**67.17**	**39.35**	27.89
39			*38*			*35*			*31*			*27*			*23*			*19*			*15*			*11*			*7*			*3*			*0*	
	0.2	17.95	7.18	**4.48**	19.45	11.74	**8.29**	**43.48**	**19.84**	**12.96**	45.27	24.26	**17.00**	55.48	30.54	21.75	66.59	37.40	**25.59**	82.02	43.65	**30.61**	77.39	47.11	33.34	81.64	46.85	32.86	92.98	49.88	34.98	108.86	52.54	33.53
	0.3	**17.50**	**7.06**	4.51	**19.09**	**11.63**	8.36	43.67	19.87	13.05	44.95	**24.25**	**17.00**	54.20	30.52	**21.64**	64.30	**37.28**	25.61	78.73	43.37	30.63	76.56	47.00	**33.26**	81.25	46.51	32.65	91.55	49.56	35.08	106.66	51.19	32.88
	0.5	18.23	7.25	4.58	20.33	12.17	8.66	44.27	20.09	13.29	40.58	24.38	17.04	**53.91**	30.19	21.84	**62.47**	37.40	25.75	76.69	43.33	30.74	75.64	47.24	33.44	78.88	45.72	32.81	87.28	48.07	34.58	101.59	48.00	32.05
	0.7	20.31	7.44	4.70	20.42	12.88	9.02	44.53	20.65	13.66	**40.33**	24.28	17.26	56.09	**30.14**	21.97	63.61	37.41	25.97	76.38	43.11	30.90	74.59	46.51	33.39	74.75	45.10	32.67	79.59	47.19	**34.05**	93.73	45.77	31.36
	0.8	21.28	7.67	4.76	21.94	13.18	9.33	45.09	21.00	13.88	40.73	24.64	17.25	59.00	30.48	22.08	67.88	37.55	25.95	**75.88**	**42.95**	30.80	**73.23**	**46.35**	33.35	**71.86**	**44.81**	**32.52**	**75.75**	**46.95**	34.16	**89.22**	**44.76**	**31.11**
25			*24*			*22*			*19*			*16*			*13*			*10*			*7*			*4*			*0*								
	0.2	**16.38**	**8.92**	**5.67**	31.58	**17.09**	**11.95**	**36.77**	21.16	**13.60**	55.39	30.26	**20.48**	63.53	32.63	21.02	60.83	39.39	25.48	65.78	40.55	26.87	76.73	43.02	**28.23**	87.70	45.59	28.44						
	0.3	18.20	9.07	5.70	30.66	**17.09**	12.10	37.36	21.35	**13.60**	55.06	29.94	20.50	62.06	**32.55**	**20.98**	58.89	39.62	**25.40**	63.78	39.96	**26.82**	74.69	42.56	28.31	85.23	44.64	27.80						
	0.5	18.47	9.42	5.71	29.84	17.26	12.19	38.38	**21.02**	13.62	55.27	30.21	20.61	59.63	32.57	21.05	58.16	39.69	25.53	61.16	39.47	26.93	70.13	41.78	28.26	78.23	42.29	27.04						
	0.7	20.67	9.83	5.92	28.98	17.77	12.29	39.20	21.06	13.93	**54.61**	30.02	20.81	59.55	32.65	21.12	57.83	39.48	25.61	58.69	38.71	26.88	65.69	41.14	28.84	68.41	40.50	26.71						
	0.8	21.78	10.11	6.06	**28.31**	18.02	12.44	40.02	21.32	14.07	55.30	**29.78**	20.75	**59.09**	32.65	21.02	**57.56**	**39.31**	25.76	**56.81**	**38.21**	26.94	**63.16**	**40.72**	29.32	**62.66**	**39.69**	**26.69**						
14			*13*			*12*			*10*			*8*			*6*			*4*			*2*			*0*											
	0.2	17.16	10.76	**7.22**	**33.97**	15.14	**10.73**	**41.45**	**22.90**	**14.52**	46.73	**26.83**	18.98	57.64	30.39	20.84	58.39	34.07	24.56	63.84	36.38	26.50	71.22	38.95	27.34									
	0.3	17.31	**10.72**	7.30	34.22	**15.08**	10.75	42.06	23.26	14.57	46.48	27.04	18.95	55.98	30.19	20.42	58.30	33.75	24.44	63.22	35.97	25.88	67.73	37.72	26.08									
	0.5	**16.38**	12.10	7.43	35.02	15.91	10.81	42.58	23.48	14.74	46.28	27.46	**18.42**	55.73	29.86	20.36	57.20	**33.28**	23.85	61.05	35.30	25.19	64.09	35.70	24.63									
	0.7	18.27	12.22	8.03	36.33	17.02	11.03	42.44	23.90	14.96	46.20	27.65	18.51	53.63	**29.11**	**20.32**	55.98	33.55	**23.70**	57.47	35.16	**24.69**	59.81	34.04	23.68									
	0.8	20.75	12.71	8.17	37.38	17.40	10.97	42.52	24.29	15.18	**45.98**	27.77	18.78	**52.34**	29.17	20.51	**55.34**	33.78	23.96	**55.33**	**35.10**	24.80	**57.13**	**33.45**	**23.45**									
7			*6*			*5*			*4*			*3*			*2*			*1*			*0*														
	0.2	**23.75**	**12.76**	**8.51**	**32.00**	17.77	**11.82**	**38.19**	**21.73**	**14.78**	**42.06**	**22.71**	15.11	46.08	**24.30**	16.09	54.64	29.98	19.21	57.20	33.06	21.91												
	0.3	24.72	**12.76**	8.65	33.05	**17.66**	11.94	39.11	22.04	14.86	42.59	22.84	**15.09**	46.22	24.48	**16.01**	53.86	29.60	18.83	56.36	32.30	21.60												
	0.5	24.94	13.24	9.07	33.16	17.94	11.89	39.11	22.61	15.09	42.75	23.12	15.33	46.00	24.51	16.18	52.48	29.02	**18.77**	54.11	31.11	20.81												
	0.7	24.97	14.41	9.71	33.25	18.34	12.00	40.39	23.00	15.59	42.70	24.00	15.31	45.23	24.73	16.54	**49.72**	**28.46**	18.92	51.72	30.34	20.47												
	0.8	26.72	15.61	9.99	32.80	18.53	12.05	40.47	23.25	15.73	42.22	24.02	15.53	**44.84**	25.19	16.91	49.92	28.48	18.88	**51.23**	**29.82**	**20.42**												

[1] The bold numbers in the table indicate the minimum interval scores

Seasonal input

The tables related to the results of the calibration for the datasets yielding the lowest interval scores for the new model are displayed in this section. Firstly, the tables for KQ101 are shown and then for KQ512.

KQ101

The performance in terms of interval scores of the model for different datasets has been analyzed for KQ101 and the results are displayed in Table B.11 below. The table shows the average interval scores, for the observation points of each specific horizon, taken over the number of flights for which forecasts are computed. The italic numbers indicate the day before departure corresponding with the observation point for that horizon.

Table B.11: Interval Scores for flight KQ101 and the different datasets

Horizon	Data input	Observation point																																
		V_0			V_1			V_2			V_3			V_4			V_5			V_6			V_7			V_8			V_9			V_{10}		
		Prediction Interval			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**		
		95%	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**
			299			*270*			*240*			*210*			*180*			*150*			*120*			*90*			*60*			*30*			*0*	
300	Total Set	11.37	**2.00**	**1.33**	59.79	**28.44**	**19.10**	**73.34**	**46.84**	**32.63**	103.22	62.31	**42.42**	140.42	90.58	62.63	167.47	105.82	73.95	198.81	128.11	86.88	179.43	129.26	89.61	172.55	122.88	88.34	155.05	104.88	73.66	134.37	69.24	42.19
	DOW	**8.43**[1]	2.45	2.78	68.92	31.60	21.44	95.34	47.22	32.77	106.44	61.66	42.98	122.97	91.74	63.24	154.44	107.27	73.78	192.27	127.14	85.78	167.38	126.79	89.78	153.92	119.81	87.82	144.74	94.80	72.64	150.72	71.90	44.21
	Seasonal data	10.34	2.60	1.92	63.64	30.47	19.49	77.97	**46.84**	33.84	**94.57**	**58.05**	42.77	**113.98**	**82.83**	**59.08**	**126.24**	**94.19**	**68.89**	**159.88**	**111.22**	**81.48**	**146.21**	**100.45**	**78.87**	**143.83**	**99.36**	**74.30**	**142.60**	**86.44**	**65.11**	**133.69**	**59.32**	**37.75**
	DOW and Seasonal combined	8.91	2.07	1.37	**58.03**	31.05	19.71	77.55	49.27	34.18	114.07	60.61	43.31	138.41	86.59	59.60	160.88	99.28	71.15	199.43	119.03	83.35	157.47	108.72	79.43	151.85	105.97	76.43	157.79	99.76	73.82	173.87	74.60	47.67
			199			*180*			*160*			*140*			*120*			*100*			*80*			*60*			*40*			*20*			*0*	
200	Total Set	**23.24**	**5.93**	3.38	72.39	32.56	20.98	80.16	42.10	29.18	93.92	53.61	36.44	107.38	68.46	45.60	113.77	74.76	51.23	121.39	76.76	55.22	127.68	83.68	58.01	129.86	82.25	58.84	131.13	72.94	49.91	124.65	59.53	33.21
	DOW	23.36	7.25	4.48	82.98	35.27	23.92	94.65	45.79	30.83	93.75	53.94	37.53	126.72	69.51	47.71	125.37	74.68	53.45	114.68	76.97	55.50	120.25	81.50	58.82	124.96	80.47	59.72	129.96	73.90	51.61	133.59	64.13	37.16
	Seasonal data	24.01	6.47	3.55	66.16	**30.80**	**20.10**	**73.21**	**39.70**	**28.21**	**80.84**	**50.74**	**34.98**	**95.68**	**64.06**	**44.03**	**103.29**	**64.15**	**47.83**	107.59	**67.52**	46.28	**119.60**	**72.64**	48.15	**122.61**	74.16	**52.73**	**125.10**	**65.26**	**45.94**	**124.57**	**55.91**	**32.57**
	DOW and Seasonal combined	30.24	6.38	**3.37**	**65.80**	35.83	23.77	74.21	43.63	30.63	98.95	54.32	36.88	118.58	67.55	45.30	106.18	66.51	47.89	**106.05**	67.59	**45.32**	126.70	72.99	**47.39**	136.12	**73.53**	53.42	142.53	78.29	53.96	156.00	74.46	46.03
			99			*90*			*80*			*70*			*60*			*50*			*40*			*30*			*20*			*10*			*0*	
100	Total Set	22.00	7.32	4.47	38.33	20.67	13.63	54.73	27.21	18.41	76.05	42.24	**26.44**	85.38	46.31	30.56	85.70	50.93	35.97	101.22	55.14	**39.66**	103.25	56.25	39.51	103.98	57.10	39.22	**102.33**	53.87	35.13	**108.39**	**50.10**	**30.59**
	DOW	16.63	**6.89**	**4.24**	**36.29**	20.96	13.93	**45.96**	**25.95**	18.96	76.93	**39.84**	26.83	87.42	46.12	32.26	92.24	52.63	39.27	105.67	58.23	44.07	100.80	56.73	43.10	**100.82**	61.82	45.58	104.07	63.37	43.62	120.76	63.04	41.70
	Seasonal data	20.49	8.18	4.78	38.77	**19.94**	**13.35**	54.21	26.59	**17.40**	**70.90**	40.52	26.67	**80.88**	**44.94**	**29.56**	**82.92**	**49.57**	**35.46**	**95.49**	**54.46**	40.12	**98.66**	**54.93**	**39.16**	102.65	**56.53**	**39.14**	103.58	**53.58**	**35.03**	111.25	51.05	31.43
	DOW and Seasonal combined	**15.97**	7.63	5.00	41.23	21.95	14.18	56.07	29.42	18.88	74.17	40.59	26.76	91.55	48.80	32.51	94.11	55.07	40.07	110.25	59.90	45.65	114.02	65.03	48.90	118.83	71.16	52.86	129.24	77.08	54.13	140.41	78.89	52.65
			59			*54*			*48*			*42*			*36*			*30*			*24*			*18*			*12*			*6*			*0*	
60	Total Set	25.47	10.00	**5.86**	33.32	19.28	**13.48**	69.87	31.76	**21.57**	79.57	38.87	**25.20**	77.64	39.54	**26.17**	83.58	42.64	28.29	81.50	44.72	**29.93**	86.88	47.69	**32.63**	83.97	46.73	**31.29**	**85.77**	44.84	**29.19**	**90.22**	46.46	31.75
	DOW	20.25	10.34	7.37	**30.18**	19.84	14.92	**65.95**	34.93	22.99	**75.87**	40.96	27.76	**74.05**	40.78	28.42	**79.73**	41.75	30.38	83.71	45.71	33.34	**84.58**	52.54	37.78	**82.76**	52.72	39.45	87.02	53.75	38.12	94.08	54.63	39.94
	Seasonal data	19.62	10.75	7.01	33.28	**18.72**	13.66	68.40	**31.37**	**21.57**	78.42	**38.36**	25.40	77.61	**39.00**	27.00	83.36	**41.61**	**28.04**	**80.53**	**43.61**	30.13	85.88	**47.17**	32.75	84.16	**46.24**	32.57	86.84	**44.68**	29.72	93.57	**45.16**	**31.48**
	DOW and Seasonal combined	**17.15**	**9.62**	6.33	43.57	22.57	15.32	77.50	36.06	23.27	87.16	41.52	27.15	87.68	41.73	27.92	92.85	47.57	33.05	87.26	51.55	35.29	95.64	56.35	39.75	95.65	59.07	43.34	101.03	57.39	42.65	109.14	65.27	46.92
			38			*35*			*31*			*27*			*23*			*19*			*15*			*11*			*7*			*3*			*0*	
39	Total Set	19.43	11.21	7.03	**33.46**	**17.68**	**12.45**	43.99	**22.55**	**14.55**	53.90	**28.73**	**18.74**	56.95	**30.29**	**20.56**	60.48	**33.68**	**22.39**	**65.84**	**37.88**	**26.18**	69.12	**38.13**	**26.41**	**71.86**	**40.37**	**27.43**	**72.20**	**42.57**	**28.60**	**75.02**	**43.65**	**30.65**
	DOW	**17.87**	**10.34**	**6.15**	33.55	17.93	12.66	44.17	23.15	16.32	57.22	29.66	19.48	57.75	31.66	21.15	**59.90**	36.43	26.62	66.17	40.80	31.84	**68.01**	43.80	32.04	72.26	47.49	34.54	74.71	49.74	35.39	79.99	52.20	37.16
	Seasonal data	18.98	11.74	7.90	35.99	19.29	13.47	**42.24**	22.96	15.88	**53.35**	29.39	19.36	**55.55**	30.62	21.47	60.18	34.84	24.20	66.54	39.32	28.55	69.65	40.82	29.82	72.52	42.03	29.32	76.35	43.56	30.18	81.51	43.66	31.93
	DOW and Seasonal combined	23.79	10.87	7.10	33.75	18.21	12.52	45.85	24.79	17.18	56.12	28.80	20.10	55.68	34.44	24.04	60.26	37.66	27.42	70.08	43.63	32.84	76.34	49.18	37.84	78.01	48.42	34.92	85.33	51.87	38.67	92.36	58.36	43.82
			24			*22*			*19*			*16*			*13*			*10*			*7*			*4*			*0*								
25	Total Set	21.08	**12.58**	**7.72**	31.23	20.10	13.74	38.54	**24.23**	**16.97**	**44.22**	**28.18**	**20.13**	52.92	**31.19**	**21.74**	**56.51**	**32.60**	**22.54**	**59.41**	**34.39**	**22.95**	61.37	**36.18**	**24.31**	**65.30**	**39.87**	**27.42**						
	DOW	**19.85**	12.68	8.05	38.57	**19.51**	**13.21**	50.76	25.50	17.91	58.89	31.47	21.11	66.35	35.02	24.66	58.55	36.71	26.10	60.61	39.69	26.12	65.10	41.08	27.70	69.95	44.76	31.70						
	Seasonal data	24.75	13.76	7.94	**30.92**	20.24	14.01	**37.25**	24.81	17.35	44.96	30.02	21.34	**52.16**	32.23	23.66	57.10	34.97	24.99	59.58	36.35	25.28	**60.34**	37.11	26.37	69.11	**39.87**	28.68						
	DOW and Seasonal combined	33.90	13.81	8.65	54.67	21.76	14.08	38.45	24.46	17.49	62.23	32.42	22.78	60.92	34.31	26.92	58.29	41.18	29.62	63.78	37.84	26.98	66.70	40.88	30.68	77.84	48.93	37.02						
			13			*12*			*10*			*8*			*6*			*4*			*2*			*0*											
14	Total Set	47.74	18.48	**12.06**	54.42	**23.32**	15.63	**40.58**	**24.07**	**17.58**	41.02	**23.32**	**16.32**	45.36	**26.03**	**18.39**	49.23	**30.77**	21.81	50.10	**30.90**	22.87	59.37	37.16	26.92									
	DOW	**32.60**	19.13	12.32	51.96	23.42	**15.14**	56.34	26.21	18.39	42.92	27.06	18.36	45.67	29.96	21.15	53.29	33.44	23.06	53.85	34.44	23.91	62.49	41.77	29.98									
	Seasonal data	50.60	19.79	12.23	**47.64**	24.05	15.81	40.62	24.68	18.20	41.90	23.93	17.23	45.22	27.33	18.82	**48.40**	31.47	**21.44**	**49.60**	31.91	**22.75**	**58.88**	**36.68**	**26.60**									
	DOW and Seasonal combined	46.47	**17.75**	12.21	61.55	24.52	16.96	62.24	28.36	19.02	**38.11**	23.89	17.90	**44.61**	30.99	23.47	53.99	35.45	25.43	53.01	35.05	24.70	66.41	45.40	34.25									
			6			*5*			*4*			*3*			*2*			*1*			*0*														
7	Total Set	26.28	**14.06**	**8.72**	35.68	**17.39**	**11.60**	33.63	18.67	13.27	34.85	21.16	14.43	37.63	21.51	**14.11**	39.95	**22.76**	16.13	53.86	**32.50**	22.49												
	DOW	**23.14**	15.51	9.72	**34.25**	18.81	11.64	33.93	**18.43**	**12.83**	**32.50**	**20.77**	**13.85**	**33.10**	**21.10**	14.51	**34.17**	25.08	17.71	57.60	35.03	23.72												
	Seasonal data	26.26	15.81	9.08	37.09	18.70	11.81	**30.93**	18.73	13.31	33.04	21.33	14.50	35.05	21.83	14.72	38.83	23.50	**15.86**	**51.91**	32.55	**22.14**												
	DOW and Seasonal combined	26.61	16.35	11.52	35.05	19.37	12.21	33.97	19.71	13.31	36.24	22.26	14.64	37.00	22.33	15.61	39.87	25.54	18.65	53.13	35.99	26.70												

[1] The bold numbers in the table indicate the minimum interval scores

KQ512

The performance in terms of interval scores of the model for different datasets has been analyzed for KQ512 and the results are displayed in Table B.12 below. Similar as to the table for KQ101 this table shows the average prediction interval score, for the observation points of each specific horizon, taken over the number of flights for which forecasts are computed. The italic numbers indicate the day before departure corresponding with the observation point for that horizon.

Table B.12: Interval Scores for flight KQ512 and the different datasets

Horizon	Data input	Observation point																																
		V_0			V_1			V_2			V_3			V_4			V_5			V_6			V_7			V_8			V_9			V_{10}		
		Prediction Interval			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**			**Prediction Interval**		
		95%	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**	**95%**	**67%**	**33%**
			299			*270*			*240*			*210*			*180*			*150*			*120*			*90*			*60*			*30*			*0*	
	Total Set	3.23	**0.45**	**0.22**	**16.05**	**6.07**	4.18	9.83	**4.13**	3.17	5.95	5.38	3.84	7.35	3.66	3.17	13.73	6.28	4.19	57.68	14.98	9.39	125.15	29.48	17.82	50.63	21.22	14.32	114.80	49.27	34.30	**82.93**	41.16	27.85
300	DOW	**2.10**[1]	**0.45**	**0.22**	22.70	6.22	**3.89**	8.85	4.48	**2.94**	4.78	5.33	3.66	**6.28**	**3.48**	3.04	12.25	**5.90**	4.29	57.88	14.95	9.34	126.10	29.28	18.15	47.85	**20.00**	14.13	113.35	48.34	**31.04**	85.70	**38.48**	**24.57**
	Seasonal data	3.18	**0.45**	**0.22**	24.33	6.38	4.58	16.55	4.61	3.12	**4.35**	4.90	3.52	6.48	3.51	2.84	12.98	6.15	**3.79**	59.18	15.13	9.11	122.03	**29.12**	17.28	44.63	20.47	**13.75**	**90.05**	**46.45**	33.68	95.48	43.15	26.21
	DOW and Seasonal combined	3.08	0.50	**0.22**	18.25	7.19	4.61	**6.43**	4.83	3.21	7.68	**3.38**	**3.04**	8.15	3.58	**2.45**	**11.80**	6.15	4.01	**56.15**	**14.16**	**8.77**	**59.78**	29.27	**17.13**	**39.25**	22.45	14.03	99.38	58.28	36.76	107.83	48.99	27.86
			199			*180*			*160*			*140*			*120*			*100*			*80*			*60*			*40*			*20*			*0*	
	Total Set	0.12	**0.00**	**0.00**	**3.36**	2.64	**1.62**	12.32	4.33	2.95	10.42	**4.99**	3.43	51.22	13.30	8.46	100.36	23.99	13.97	78.27	22.31	14.26	44.42	19.51	13.58	111.78	39.53	25.52	87.00	46.18	30.92	**82.63**	41.97	27.96
200	DOW	0.03	**0.00**	**0.00**	5.88	**2.52**	1.64	11.68	**4.24**	2.97	10.59	5.06	**3.34**	52.78	13.33	8.41	99.12	23.90	13.97	80.42	22.70	14.16	41.08	**19.06**	**13.34**	110.75	38.35	25.04	94.17	**42.19**	29.97	85.93	**38.95**	**24.31**
	Seasonal data	**0.02**	0.02	0.02	10.07	2.85	1.72	12.36	4.71	3.04	9.80	5.25	3.45	52.25	13.58	8.32	**97.54**	**23.56**	**13.84**	70.76	**21.58**	**13.96**	**39.92**	19.07	13.55	**102.42**	**37.48**	**24.62**	**79.17**	43.25	**29.32**	96.07	44.89	28.00
	DOW and Seasonal combined	0.07	**0.00**	**0.00**	9.12	2.95	1.69	**11.66**	4.67	**2.92**	**8.68**	5.18	3.65	**51.12**	**13.07**	**8.02**	109.46	24.17	14.01	**54.39**	23.18	14.25	40.05	21.13	13.58	106.39	44.03	26.24	87.97	45.84	29.71	110.58	51.32	30.29
			99			*90*			*80*			*70*			*60*			*50*			*40*			*30*			*20*			*10*			*0*	
	Total Set	**7.05**	3.00	**1.45**	10.94	5.66	**3.35**	**32.27**	12.63	7.83	69.38	19.17	11.75	66.22	20.84	13.20	**32.59**	24.47	16.16	57.44	31.43	20.87	64.30	39.39	27.36	68.52	42.32	28.06	**70.42**	51.12	36.28	**83.63**	44.46	30.79
100	DOW	9.16	**2.92**	**1.45**	**10.45**	**5.44**	3.41	33.73	**12.46**	**7.66**	70.17	**18.69**	**11.52**	**64.94**	**20.16**	**12.95**	58.56	**23.24**	**15.32**	76.47	**30.32**	**19.92**	**50.16**	**35.53**	**24.72**	**62.16**	**37.46**	**26.51**	71.11	**44.05**	**30.15**	86.42	**43.62**	**27.86**
	Seasonal data	10.41	3.03	1.48	13.03	6.03	3.60	36.52	12.59	7.76	71.22	19.92	11.85	69.11	21.36	13.69	34.81	24.48	15.85	**54.97**	30.46	20.59	62.05	38.48	26.54	68.63	40.59	27.39	77.33	48.04	34.76	94.80	48.20	31.67
	DOW and Seasonal combined	14.88	3.08	1.51	18.27	6.38	3.86	43.67	13.49	8.28	**68.86**	20.93	12.52	67.23	22.73	14.05	56.44	25.76	16.88	61.03	33.39	21.34	76.41	44.38	29.72	75.50	42.62	27.62	91.89	47.08	34.66	107.67	54.87	33.79
			59			*54*			*48*			*42*			*36*			*30*			*24*			*18*			*12*			*6*			*0*	
	Total Set	12.16	3.49	2.05	**20.42**	**8.74**	6.14	**82.30**	**22.48**	13.39	**75.06**	**25.23**	**15.66**	**78.55**	32.31	21.22	**59.88**	33.06	22.93	65.63	37.92	26.06	61.91	42.42	29.26	**66.45**	46.08	33.97	**73.17**	43.71	31.18	**81.84**	42.33	27.93
60	DOW	**9.47**	**3.11**	**1.87**	20.56	8.96	**5.78**	87.05	23.29	**13.25**	85.59	26.30	15.76	85.97	**30.99**	**20.85**	62.58	**30.77**	**21.42**	66.44	**34.93**	**23.86**	**59.73**	**35.31**	**24.45**	69.63	40.91	**27.61**	74.42	42.41	**29.39**	83.89	**40.65**	**25.63**
	Seasonal data	10.81	4.29	2.43	21.59	10.23	6.45	89.39	23.21	13.75	78.97	25.75	15.76	89.23	32.13	21.43	64.63	33.08	23.15	**65.27**	37.54	25.89	62.31	40.45	27.63	70.30	41.85	32.21	81.61	**41.36**	30.56	91.88	44.11	28.34
	DOW and Seasonal combined	18.50	4.15	2.02	29.64	10.23	6.48	112.56	24.89	14.46	105.63	28.50	16.73	126.77	38.18	24.28	66.08	39.66	26.51	73.00	38.83	26.55	66.91	40.77	27.47	78.58	**40.72**	30.56	93.08	44.36	29.58	102.52	47.85	28.86
			38			*35*			*31*			*27*			*23*			*19*			*15*			*11*			*7*			*3*			*0*	
	Total Set	18.23	7.25	4.58	20.33	12.17	8.66	44.27	20.09	13.29	**40.58**	24.38	17.04	53.91	30.19	21.84	**62.47**	37.40	25.75	**76.69**	43.33	30.74	**75.64**	47.24	33.44	**78.88**	**45.72**	32.81	**87.28**	48.07	34.58	**101.59**	**48.00**	32.05
39	DOW	**15.63**	**6.86**	**4.15**	**19.67**	**11.28**	**7.99**	**37.75**	**18.23**	**12.39**	46.48	**23.39**	**16.47**	**47.97**	**28.43**	20.08	64.58	**33.71**	**22.95**	77.19	**40.61**	**26.07**	76.16	**42.19**	**28.28**	79.72	45.82	**32.16**	89.58	**46.93**	**33.28**	105.69	48.95	**30.47**
	Seasonal data	21.45	7.14	4.51	21.59	11.79	8.54	45.44	19.95	13.10	50.14	24.83	16.70	62.89	31.28	21.80	69.13	37.71	25.78	84.28	45.11	31.21	78.91	47.05	33.13	86.55	47.97	34.57	96.36	52.69	37.42	111.31	51.71	33.20
	DOW and Seasonal combined	24.09	8.25	4.90	32.41	13.54	8.69	43.95	20.33	13.55	49.34	23.94	16.49	54.88	28.67	**19.83**	65.17	35.29	23.85	78.55	40.62	28.98	80.56	43.69	31.10	87.48	46.29	32.21	100.63	54.08	37.40	115.55	54.63	34.44
			24			*22*			*19*			*16*			*13*			*10*			*7*			*4*			*0*							
	Total Set	18.47	9.42	5.71	29.84	17.26	12.19	38.38	21.02	13.62	**55.27**	30.21	20.61	59.63	32.57	21.05	**58.16**	39.69	25.53	**61.16**	39.47	**26.93**	70.13	41.78	28.26	**78.23**	42.29	27.04						
25	DOW	**14.98**	**8.30**	6.29	33.91	**15.93**	**11.43**	39.05	21.77	13.65	55.92	**27.66**	**17.75**	61.81	32.56	**20.80**	62.56	37.66	**24.82**	61.38	**39.11**	27.08	**69.86**	**40.32**	**26.86**	86.30	**41.83**	**26.77**						
	Seasonal data	18.88	9.41	6.18	**27.52**	16.76	11.90	38.75	20.96	13.70	55.89	30.59	20.95	**56.17**	32.71	21.86	61.58	39.07	27.22	67.44	41.20	28.89	77.59	45.90	31.70	89.14	45.61	28.89						
	DOW and Seasonal combined	37.30	10.29	**5.69**	42.52	21.24	13.35	**37.94**	**20.74**	**13.10**	59.48	31.81	22.42	60.81	**32.19**	21.54	65.09	**37.50**	26.60	73.86	40.97	29.32	85.27	53.60	39.18	96.09	49.46	32.38						
			13			*12*			*10*			*8*			*6*			*4*			*2*			*0*											
	Total Set	**16.38**	12.10	**7.43**	35.02	**15.91**	**10.81**	42.58	23.48	**14.74**	46.28	27.46	**18.42**	55.73	29.86	20.36	57.20	33.28	23.85	61.05	35.30	**25.19**	**64.09**	**35.70**	24.63									
14	DOW	18.36	**11.74**	7.86	**27.77**	16.50	11.53	40.78	**21.98**	15.08	**44.50**	27.50	18.76	**49.28**	**29.19**	**19.97**	**55.09**	**31.25**	**22.07**	**59.80**	**35.18**	25.43	67.83	35.73	**23.69**									
	Seasonal data	16.45	12.79	8.02	30.61	17.22	11.46	**39.66**	23.53	14.93	45.66	**27.19**	18.50	55.73	30.59	21.00	60.09	34.49	25.14	63.97	35.97	25.84	68.58	37.47	25.40									
	DOW and Seasonal combined	30.89	12.77	8.61	36.42	19.39	13.57	42.84	24.19	16.39	49.16	29.54	20.92	63.38	37.73	28.91	69.88	45.63	36.12	71.42	39.47	29.21	77.53	42.73	29.99									
			6			*5*			*4*			*3*			*2*			*1*			*0*														
	Total Set	24.94	**13.24**	**9.07**	33.16	17.94	11.89	39.11	22.61	15.09	42.75	23.12	15.33	46.00	24.51	16.18	52.48	29.02	**18.77**	54.11	31.11	20.81												
7	DOW	27.06	14.56	9.84	**29.20**	**16.45**	**11.86**	**35.08**	**18.94**	**12.69**	**38.67**	**21.13**	**14.67**	**42.69**	**23.10**	**15.84**	**47.53**	**27.68**	18.91	**53.00**	**30.91**	**20.80**												
	Seasonal data	**23.19**	14.85	9.94	34.66	19.90	12.63	38.84	23.87	16.43	43.41	23.94	16.62	46.66	25.37	17.80	54.19	30.64	20.28	56.19	32.15	22.32												
	DOW and Seasonal combined	31.78	17.09	12.84	44.30	24.61	17.19	46.89	30.48	21.70	41.66	25.92	18.73	45.61	25.21	17.85	55.83	31.79	22.16	59.72	33.70	22.81												

[1] The bold numbers in the table indicate the minimum interval scores

Benchmarking model

The tables with results of the comparative study that has been conducted to benchmark the Markov Chain method are shown in this section. Firstly, the average errors found for KQ101 are shown. Then, the average errors for the comparative study of KQ512 are displayed.

KQ101

The performance of the three methods expressed in MAE for the forecasts performed for flight KQ101 are shown in Table B.13. The table shows the average errors, for the observation points of each specific horizon, taken over the number of flights for which forecasts are computed. The italic numbers indicate the day before departure corresponding with the observation point for that horizon. In the table the average forecast errors that are smallest and, following from the Kruskal-Wallis test, significantly different from the average forecasting errors of the other methods are shown in green.

Table B.13: The performance of the different forecasting methods measured in absolute error for Flight *KQ101* and a set of horizons with seasonal data

		Observation point											
Horizon	**Method**	V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	**Best Method (#)**
		299	*270*	*240*	*210*	*180*	*150*	*120*	*90*	*60*	*30*	*0*	
	MC	0.85	8.44	13.39	16.79	22.07	25.03	29.50	28.27	26.35	**22.85**	**14.45**	*2*
300	HA	13.25	16.37	18.58	21.32	24.77	27.79	31.95	31.17	28.75	24.13	14.49	*0*
	PU	**0.39**[1]	**7.56**	**12.54**	**16.38**	**21.76**	**24.60**	**29.18**	**27.40**	**25.97**	23.66	18.66	*9*
		199	*180*	*160*	*140*	*120*	*100*	*80*	*60*	*40*	*20*	*0*	
	MC	1.38	**7.29**	10.21	12.43	15.47	17.35	16.13	16.17	**18.35**	**16.40**	11.34[2]	*4*
200	HA	23.07	24.19	26.49	28.62	31.15	31.69	29.74	27.90	25.70	20.62	14.16	*0*
	PU	**1.23**	7.45	**10.11**	**12.40**	**15.38**	**16.80**	**15.52**	**15.61**	18.90	19.36	18.55	*7*
		99	*90*	*80*	*70*	*60*	*50*	*40*	*30*	*20*	*10*	*0*	
	MC	1.67	4.90	**6.28**	**9.24**	**10.54**	**12.83**	14.33	13.97	14.39	12.65	11.79	*9*
100	HA	31.54	30.53	29.74	29.76	27.90	26.96	25.70	23.31	20.62	16.52	14.16	*0*
	PU	**1.47**	**4.40**	6.86	10.01	12.46	15.46	17.96	19.31	21.87	23.88	24.73	*2*
		59	*54*	*48*	*42*	*36*	*30*	*24*	*18*	*12*	*6*	*0*	
	MC	2.20	4.84	**7.66**	**9.03**	**9.70**	**10.38**	10.94	11.69	11.87	10.61	11.80	*9*
60	HA	27.88	27.69	27.16	26.31	24.95	23.31	21.16	19.24	16.70	13.91	14.16	*0*
	PU	1.83	**4.83**	7.91	9.61	10.51	12.54	14.34	16.76	18.65	19.07	20.55	*2*
		38	*35*	*31*	*27*	*23*	*19*	*15*	*11*	*7*	*3*	*0*	
	MC	2.77	4.73	**5.60**	**6.79**	**7.86**	9.20	10.34	10.60	10.21	10.73	11.91	*9*
39	HA	24.95	24.79	22.90	22.41	21.70	19.52	18.77	16.41	14.39	13.29	14.16	*0*
	PU	2.05	**4.30**	5.66	7.58	9.06	11.75	13.52	15.31	15.53	17.36	18.65	*2*
		24	*22*	*19*	*16*	*13*	*10*	*7*	*4*	*0*			
	MC	3.04	5.29	6.40	**8.01**	**8.46**	**9.07**	**8.85**	9.21	10.81			*6*
25	HA	21.16	21.93	19.52	18.88	17.24	16.52	14.39	13.44	14.16			*0*
	PU	**2.66**	**4.89**	**6.39**	8.14	9.61	11.04	11.24	13.53	14.81			*3*
		13	*12*	*10*	*8*	*6*	*4*	*2*	*0*				
	MC	3.86	**5.31**	**6.66**	**6.15**	**6.51**	**7.60**	**8.29**	**10.04**				*7*
14	HA	17.24	16.70	16.52	15.51	13.91	13.44	12.84	14.16				*0*
	PU	**3.47**	5.37	6.81	6.73	7.93	9.54	10.23	12.06				*1*
		6	*5*	*4*	*3*	*2*	*1*	*0*					
	MC	3.22	**4.07**	4.79	**5.21**	5.47	**5.77**	**8.10**					*4*
7	HA	13.91	13.43	13.44	13.29	12.84	13.02	14.16					*0*
	PU	**2.71**	4.14	**4.76**	5.44	**5.45**	6.06	8.89					*3*

[1] The bold numbers in the table indicate the minimum forecast error
[2] The numbers in the table both green and bold indicate that the minimum error is significantly lower than the other errors

KQ512

The performance of the methods in MAE for the forecasts performed for KQ512 are shown in Table B.14. In the table the average forecast errors that are smallest and, following from the Kruskal-Wallis test, significantly different from the average forecasting errors of the other methods are shown in green. The table shows the average errors, for the observation points of each specific horizon, taken over the number of flights for which forecasts are computed. The italic numbers indicate the day before departure corresponding with the observation point for that horizon.

Table B.14: The performance of the different forecasting methods measured in absolute error for Flight *KQ512* and a set of horizons with seasonal data

		Observation point											
Horizon	**Method**	V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	**Best Method (#)**
		299	*270*	*240*	*210*	*180*	*150*	*120*	*90*	*60*	*30*	*0*	
	MC	**0.08**[2]	1.49	1.27	1.44	1.32	1.72	3.18	5.90	4.55	11.14	9.24	*1*
300	HA	1.16	**1.19**[1]	**1.00**	**1.22**	**1.10**	**1.61**	**3.10**	**5.79**	4.54	11.26	9.27	*7*
	PU	**0.08**	1.71	1.56	1.96	1.91	2.24	3.44	6.09	**4.49**	**10.85**	**8.82**	*4*
		199	*180*	*160*	*140*	*120*	*100*	*80*	*60*	*40*	*20*	*0*	
	MC	**0.00**	0.71	1.12	**1.28**	**2.84**	**4.68**	4.78	**4.56**	**8.70**	10.69	8.86	*6*
200	HA	1.10	1.09	1.47	1.48	2.91	**4.68**	**4.75**	4.60	**8.70**	**10.55**	**8.73**	*5*
	PU	0.01	**0.66**	**1.09**	1.32	2.88	4.77	4.88	4.61	8.87	11.05	9.13	*2*
		99	*90*	*80*	*70*	*60*	*50*	*40*	*30*	*20*	*10*	*0*	
	MC	0.56	1.33	2.64	4.09	4.48	5.40	7.12	**8.91**	**9.48**	**10.77**	10.20	*3*
100	HA	4.41	4.70	4.60	4.40	4.46	5.93	8.20	9.81	10.44	11.39	**8.94**	*1*
	PU	**0.54**	**1.24**	**2.58**	**4.06**	**4.42**	**5.19**	**6.98**	9.03	9.59	10.97	10.80	*7*
		59	*54*	*48*	*42*	*36*	*30*	*24*	*18*	*12*	*6*	*0*	
	MC	**0.74**	**2.17**	4.81	5.79	7.48	7.72	8.80	8.83	10.18	**11.19**	9.22	*3*
60	HA	4.59	5.13	7.78	8.22	9.75	9.81	10.15	10.06	10.96	11.43	**8.94**	*1*
	PU	0.75	2.20	**4.63**	**5.45**	**7.11**	**7.49**	**8.61**	**8.69**	**10.00**	11.42	10.04	*7*
		38	*35*	*31*	*27*	*23*	*19*	*15*	*11*	*7*	*3*	*0*	
	MC	**1.49**	3.03	**4.54**	**5.99**	**7.50**	**8.25**	**9.23**	**10.40**	12.03	12.80	10.67	*7*
39	HA	8.87	9.41	9.43	9.37	10.29	10.22	10.98	11.22	**11.53**	**10.72**	**8.94**	*3*
	PU	1.53	**3.01**	4.64	6.18	7.72	8.46	9.65	10.86	12.96	13.97	11.97	*1*
		24	*22*	*19*	*16*	*13*	*10*	*7*	*4*	*0*			
	MC	2.34	4.19	**4.80**	**6.37**	**7.36**	**8.87**	**9.84**	**10.07**	9.23			*6*
25	HA	10.15	10.34	10.22	10.50	11.02	11.39	11.53	10.74	**8.94**			*1*
	PU	**2.08**	**3.93**	4.83	6.38	**7.36**	9.02	10.07	10.42	9.79			*3*
		13	*12*	*10*	*8*	*6*	*4*	*2*	*0*				
	MC	2.87	4.28	**5.46**	**7.07**	**7.24**	**8.57**	**9.61**	**8.89**				*6*
14	HA	11.02	10.96	11.39	11.15	11.43	10.74	10.85	8.94				*0*
	PU	**2.64**	**3.96**	5.55	7.31	7.75	9.30	10.78	10.38				*2*
		6	*5*	*4*	*3*	*2*	*1*	*0*					
	MC	3.76	**4.38**	4.52	**5.27**	**5.96**	**6.94**	**7.68**					*5*
7	HA	11.43	10.72	10.74	10.72	10.85	9.49	8.94					*0*
	PU	**3.71**	**4.38**	**4.41**	5.34	5.99	7.38	8.48					*3*

[1] The bold numbers in the table indicate the minimum forecast error

[2] The numbers in the table both green and bold indicate that the minimum error is significantly lower than the other errors

C
Group Data Analysis

This section discusses the group data analysis that has been carried out on the bookings data for flights KQ101 and KQ512. Figure C.1 shows the percentage of group reservations that has not been cancelled before the departure day for KQ101 and Figure C.2 the same information for KQ512.

For the bookings of KQ101 a trend can be observed for the amount of group reservations that are not cancelled for the flight. For a group size from 0 to around 40, the larger the group size, the lower amount of reservations is still in the system at departure date. This trend can not be observed for KQ512. The reason for this might be that the bookings activity starts closer to the departure date which makes it unlikely that agencies had been able to make the reservations with an option to cancel later in the bookings process. Another reason is that flight KQ512 serves another market. Whereas KQ101 serves European tourists, KQ512 serves among others business passengers in the economy class. Tourists might book earlier via an agency far in advance.

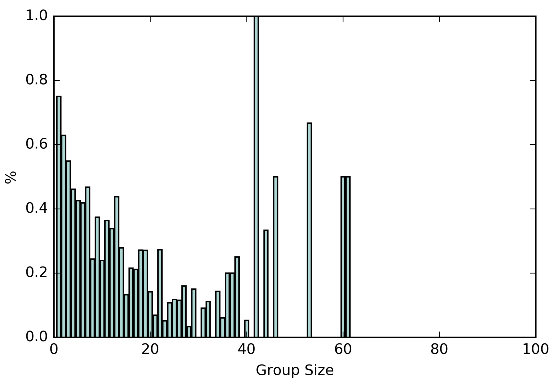


Figure C.1: The percentage of group bookings that is not cancelled or rebooked at departure date for KQ101

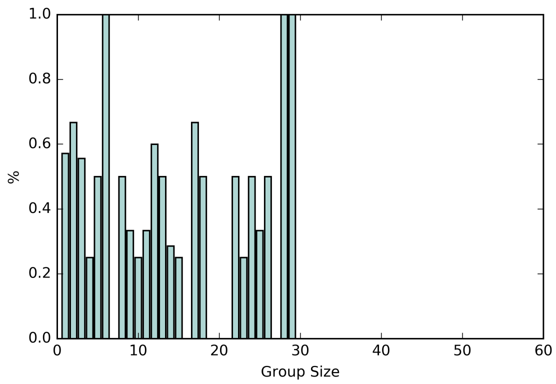


Figure C.2: The percentage of group bookings that is not cancelled or rebooked at departure date for KQ512

D
Numerical Examples Reference Methods

The working principle of the advanced pick-up method and historical average method is clarified with a numerical example in this section of the Appendix. Consider the following dataset of bookings for a flight in Table D.1. The bookings for the final four days until departure of flight f are forecasted of which the bookings data is highlighted.

Table D.1: An example set of bookings data

	Day before Departure (t)							
Flight Date	0	1	2	3	4	5	6	7
f-6	110	109	107	104	100	100	99	96
f-5	103	103	100	97	92	90	89	89
f-4	107	102	101	101	99	97	92	90
f-3	112	109	106	102	96	92	90	91
f-2	105	103	102	96	94	95	92	87
f-1	95	94	92	98	95	94	84	87
f					95	99	92	87
f+1						89	93	92
f+2							93	84
f+3								89

For $t = \{0, 1, 2, 3\}$ the historical average takes the average of the respective columns of the dataset. The resulting forecast of the number of bookings for the final four days before departure equals $\{105.3, 103.3, 101.3, 99.7\}$.

For the computation of the advanced pick-up forecast a pick-up table can be constructed from the dataset of bookings that shows the increment in bookings between the time intervals as can be seen in Table D.2. From this table the average pick-up values can easily be calculated and result in the set of values $\{2, 2, 1.7, 3.7\}$. By adding these values to the most recent available number of bookings for flight f which is 95, the forecast of the total number of bookings for the last four days before departure equals $\{104.4, 102.4, 100.4, 98.7\}$.

Table D.2: Pick-up values for the last 4 days until departure in the sample dataset

	DBD(t)			
Flight Date	*0*	*1*	*2*	*3*
f-6	1	2	3	4
f-5	0	3	3	5
f-4	5	1	0	2
f-3	3	3	4	6
f-2	2	1	6	2
f-1	1	2	-6	3

E
Weekday Analysis

Table with net increments in bookings over the last months of the bookings process. Show the trend in decline of bookings over the weekend and increase midweek. The red values in the table indicate a decline in the number of bookings for the denoted bookings DOW. It can be seen that basically all decreases in bookings are observed during the weekends. From there it can be concluded that this is important to consider in the forecasting model.

Table E.1: The average increments in bookings in the last month before departure for flight departing on the same DOW

		Bookings Week			
Departure DOW	**Bookings DOW**	*Week -1*	*Week -2*	*Week -3*	*Week -4*
Monday	*Sun*	3.75	0.31	-0.18	-0.38
	Sat	-0.17	0.74	-0.05	0.60
	Fri	0.32	1.72	1.86	1.87
	Thu	2.48	3.36	2.52	1.82
	Wed	3.39	4.41	3.04	2.67
	Tue	3.81	3.94	3.25	1.64
	Mon	4.71	3.64	3.28	2.10
Tuesday	*Sun*	-0.59	-0.49	-0.11	0.16
	Sat	0.74	-1.20	0.03	0.13
	Fri	2.55	2.34	1.24	1.50
	Thu	2.62	2.33	2.20	1.33
	Wed	2.37	3.03	2.72	2.60
	Tue	3.46	3.43	2.88	2.58
	Mon	4.45	3.23	3.61	2.28
Wednesdy	*Sun*	0.61	-1.26	-0.46	-0.78
	Sat	0.77	0.42	0.03	0.19
	Fri	3.32	1.69	1.24	1.57
	Thu	1.29	3.09	2.59	1.10
	Wed	3.97	2.54	1.78	3.05
	Tue	3.22	2.22	3.18	2.81
	Mon	5.42	3.56	2.35	1.86
Thursday	*Sun*	-0.19	-0.62	-0.52	-0.09
	Sat	0.68	0.21	0.47	-0.32
	Fri	1.79	1.88	2.30	1.53
	Thu	2.21	2.18	1.98	2.02
	Wed	2.76	2.47	2.49	2.33
	Tue	2.55	3.72	2.08	2.71
	Mon	2.86	2.70	1.83	2.64
Friday	*Sun*	-0.06	-0.74	-0.44	-0.63
	Sat	-0.07	-0.46	-0.10	-0.04
	Fri	1.24	1.53	2.11	0.63
	Thu	1.80	2.55	1.50	1.87
	Wed	2.66	1.96	2.87	1.79
	Tue	1.82	2.33	2.53	2.47
	Mon	3.45	2.21	2.37	2.18
Saturday	*Sun*	-0.59	-0.50	-0.59	-0.68
	Sat	-0.39	-0.75	-1.12	-0.33
	Fri	2.11	1.13	1.42	0.91
	Thu	2.80	1.84	2.48	1.88
	Wed	1.36	2.56	3.10	3.31
	Tue	3.07	2.85	4.20	2.43
	Mon	3.76	3.64	2.40	3.65
Sunday	*Sun*	-0.65	-0.76	-0.98	-0.54
	Sat	-0.21	-0.35	0.06	0.22
	Fri	3.21	3.38	1.98	1.24
	Thu	1.95	2.76	2.25	1.23
	Wed	2.98	2.90	3.28	2.98
	Tue	3.86	3.23	3.16	2.87
	Mon	4.82	3.75	3.80	2.34

F
Analysis of KQ101 probability distributions

An analysis has been made to identify the forecasts for which the model failed to correctly display the uncertainty. From Figures F.1 and Figure F.2 it can be seen that the probability distributions are quite accurate for forecasts that are made from maximum 80 days before departure.

From Figure F.3 and Figure F.4 it can be concluded that for the remaining forecast horizons (100 to 300 days) the model displays the uncertainty well for the first 70 days in the horizon. However, for the model fails for the part of the horizon close to departure. This is visualized in Figure F.5 and Figure F.6.

From this analysis it can be concluded that for the medium- to short-term forecasts the model is able to accurately model the forecast uncertainty. For the long-term forecast the model does not fail for the entire horizon. Only for the final part of these forecast horizons, the model needs some adaptions in order to capture the large uncertainty that many days into the future.

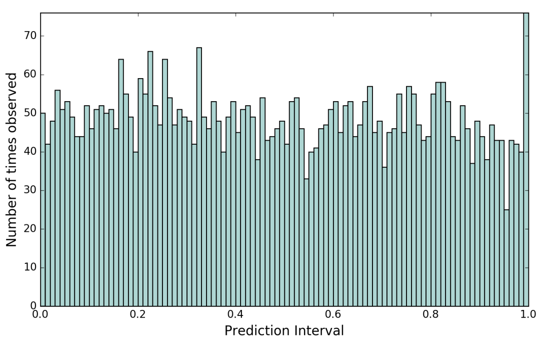


Figure F.1: The number of times the prediction interval captured the actual observation on its boundary for KQ101 for horizons shorter than 80 days

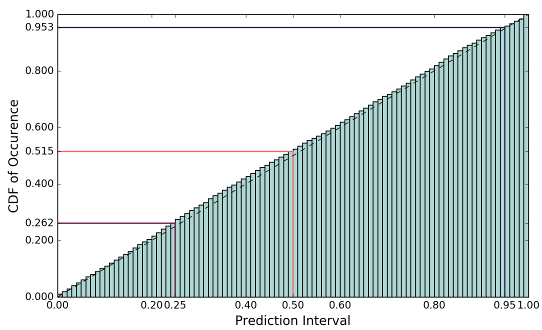


Figure F.2: The CDF of the sample of observed prediction intervals for KQ101for horizons shorter than 80 days

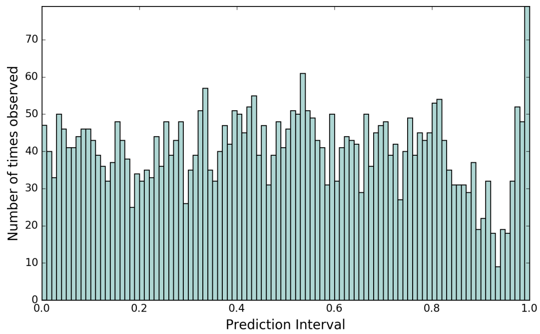


Figure F.3: The number of times the prediction interval captured the actual observation on its boundary for KQ101 for the first 70 days of horizons longer than 80 days

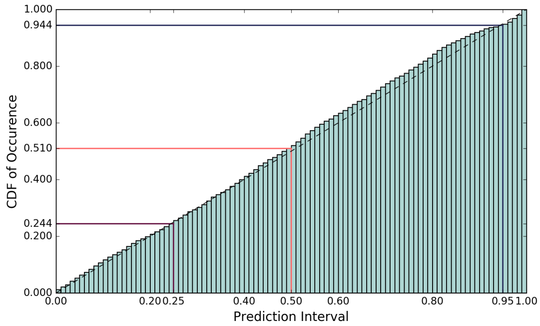


Figure F.4: The CDF of the sample of observed prediction intervals for KQ101 for the first 70 days of horizons longer than 80 days

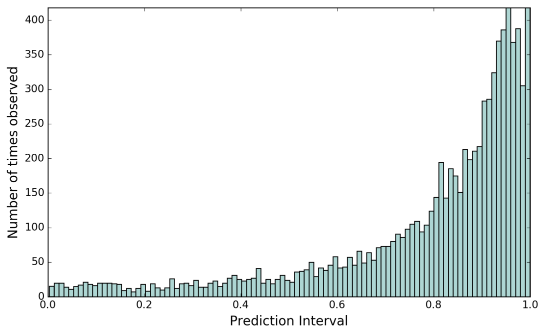


Figure F.5: The number of times the prediction interval captured the actual observation on its boundary for KQ101 for horizons longer than 80 days from 70 days out

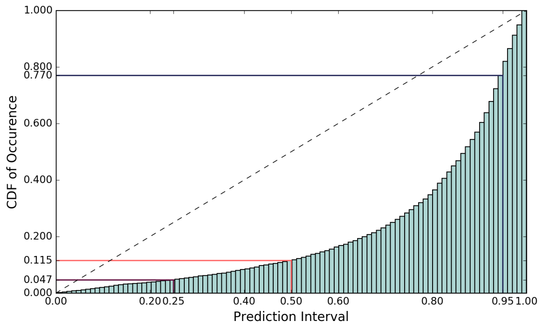


Figure F.6: The CDF of the sample of observed prediction intervals for KQ101 for horizons longer than 80 days from 70 days out

References

Adams, W., & Vodicka, M. (1987). Short-Term Forecasting Of Passenger Demand And Some Applications In Qantas. In *Agifors symposium proceedings* (pp. 240–257).

Allende, H., Moraga, C., & Salas, R. (2002). Artificial Neural Networks In Time Series Forecasting: A Comparative Analysis. *Kybernetika*, *88*(6), 685–707.

Alstrup, J., Boas, S., Madsen, O. B. G., & Vidal, R. V. V. (1986). Booking Policy For Flights With Two Types Of Passengers. *European Journal of Operational Research*, *27*, 274–288.

Belobaba, P. P. (1985). *TWA RESERVATIONS ANALYSIS: Project Update - Demand Distribution Patterns* (Tech. Rep.). Massachusetts Institute of Technology.

Belobaba, P. P., & Hopperstad, C. (2004). *Algorithms for revenue management in unrestricted fare markets.* Cambridge, MA.

Ben-Akiva, M. (1987). Improving airline passenger forecasts using reservation data. In *Fall orsa/tims conference.* St. Louis, MO.

Bobb, L. M., & Veral, E. (2008, sep). Open issues and future directions in revenue management. *Journal of Revenue and Pricing Management*, *7*(3), 291–301.

Brummer, M. (1988). *Determination of Potential Load Capacity Based on Observed Load Factor Data: A Study for Northwest Airlines* (Tech. Rep.). Northfield, Minnesota: St Olaf College Undergraduate Practicum Group.

Chatfield, C. (1993). Calculating Interval Forecasts. *Journal of Business & Economic Statistics*, *11*(2), 121–135.

Coughlan, J. (1999). Airline Overbooking in the Multi-Class Case. *Journal of the Operational Reearch Society*, *50*(50), 1098–1103.

Crystal Queenan, C., Ferguson, M., Higbie, J., & Kapoor, R. (2007). A Comparison of Unconstraining Methods to Improve Revenue Management Systems. *Production and Operations Management Society*, *16*(6), 729–746.

Duncanson, A. V. (1974). Short-Term Traffic Forecasting. In *Proceedings of the 14th annual agifors symposium.* Davos, Switzerland.

Dutta, G., & Pachisia, D. (2014). *Forecast Accuracy along Booking Profile in the National Railways of an Emerging Asian Economy: Comparison of Different Techniques.* Ahmedabab, India.

Gneiting, T., & Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

Goldberg, D. (1991). *What every computer scientist should know about floating-point arithmetic* (Vol. 23) (No. 1).

Goto, J. H., Lewis, M. E., & Puterman, M. L. (2004). Coffee, Tea, or ??: A Markov Decision Process Model for Airline Meal Provisioning. *Transportation Science*, *38*(1), 107–118.

Guo, P., Xiao, B., & Li, J. (2012). Unconstraining methods in revenue management systems: Research overview and prospects. *Advances in Operations Research*, 23.

Haensel, A. (2012). *Choice-set Demand in Revenue Management: Unconstraining, Forecasting and Optimization* (Academisch Proefschrift). Vrije Universiteit Amsterdam.

Haensel, A., & Koole, G. (2010). Estimating unconstrained demand rate functions using customer choice sets. *Journal of Revenue and Pricing Management*, 1–17.

Harris, P., & Marucci, G. (1983). A Short Term Forecasting Model. In *Agifors symposium proceedings.* Memphis.

Hopperstad, C. (1991). *Efficient Forecasting: 1/92 Equation Simplifications* (Tech. Rep.). Renton, WA: Boeing Internal Report.

Hsu, C.-i., & Liu, S.-m. (2003). Predicting City-Pair Air Passenger Traffic Using Grey Topological Forecasting Model. *Journal of the Eastern Asia Society for Transportation Studies*, *5*, 453–470.

Iida, T., & Zipkin, P. H. (2006). Approximate Solutions of a Dynamic Forecast-Inventory Model. *Manufacturing & Service Operations Management*, *8*(4), 407–425.

Kenya Airways. (2016). Annual Report & Financial Statements 2016.

Kijima, M. (1997). *Markov Processes for Stochastic Modeling* (First ed.). Chapman and Hall.

Larsen, R. J., & Marx, M. L. (1981). *An Introduction To Mathematical Statistics And Its Applications* (Fifth Edit ed.). Pearson Educatino, Inc.

Lee, A. O. (1990). *Airline reservations forecasting: probabilistic and statistical models of the booking process* (Unpublished doctoral dissertation). Massachusetts Institute Of Technology.

Lemke, C., Riedel, S., & Gabrys, B. (2012). Evolving forecast combination structures for airline revenue management. *Journal of Revenue and Pricing Management*, *12*(3), 221–234.

L'Heureux, E. (1986). A New Twist in Forecasting Short-Term Passenger Pickup. In *Agifors symposium proceedings.*

Littlewood, K. (1972). Forecasting and control of passenger bookings. *Journal of Revenue & Pricing Management*, *4*(2), 111–123.

Liu, Y., & Chen, W. (2012). A SAS Macro for Testing Differences among Three or More Independent Groups Using Kruskal-Wallis and Nemenyi Tests. *Huazhong University of Science and Technology and Springer*, *32*(1), 130–134.

Maddala, G. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., ... Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*(2), 111–153.

McGill, J. I., & Van Ryzin, G. J. (1999). Revenue Management: Research Overview and Prospects. *Transportation Science*, *33*(2), 233–256.

Pesch, E., & Kovalyov, M. Y. (2015). *Revenue management models for hotel business* (Unpublished doctoral dissertation). Universitat Siegen.

Rajopadhye, M., Ben Ghalia, M., Wang, P. P., Baker, T., & Eister, C. V. (2001). Forecasting uncertain hotel room demand. *Information Sciences*, *132*(1-4), 1–11.

Rothstein, M. (1971). An Airline Overbooking Model. *Transportation Science*, *5*(2), 180–192.

Sa, J. (1987). *Reservations Forecasting in Airline Yield Management* (MSc Thesis). Massachusetts Institute of Technology.

Scandinavian Airline System. (1978). SHOPS: A Short-term Passenger Forecasting System. *Journal of Revenue & Pricing Management*.

Schneider, H. (1986). *Truncated and Censored Samples from Normal Populations.* New York: CRC Press.

Schot, E. (2015). *Interval Forecasting for Airline Passenger Demand* (MSc Thesis). Delft University of Technology.

Schwartz, Z., & Hiemstra, S. (1997). Improving the Accuracy of Hotel Reservations Forecasting: Curves Similarity Approach. *Journal of Travel Research*, *36*(1), 3–14.

Shen, H., & Huang, J. Z. (2008). Interday Forecasting and Intraday Updating of Call Center Arrivals. *Manufacturing & Service Operations Management*, *10*(3), 391–410.

Skwarek, D. K. (1996). *Competitive Impacts Of Yield Management System Components: Forecasting And Sell-Up Models* (MSc Thesis). Massachusetts Institute of Technology.

SkyTeam. (2017). *SkyTeam Facts and Figures.* Retrieved 08-05-2017, from `https://www.skyteam.com/en/about/`

Smith, B., & Penn, C. (1988). Analysis of alternate origin-destination control strategies. In *Agifors symposium proceedings* (pp. 123–144).

Stikvoort, W., & Van der Zwan, F. (2010). Revenue optimization through demand forecasting. In *International air transport and operations symposium* (pp. 193–216). Delft, The Netherlands.

Sulima, N. (2012). Probabilistic Model of Overbooking for an Airline. *Automatic Control and Computer Sciences*, *46*(1), 49–56.

Svrcek, T. (1991). *Modeling Airline Group Passenger Demand For Revenue Optimization* (MSc Thesis). Massachusetts Institute of Technology.

Swan, W. M. (2002). Airline demand distributions: Passenger revenue management and spill. *Transportation Research Part E: Logistics and Transportation Review*, *38*(3-4), 253–263.

Toktay, L. B., & Wein, L. M. (2001). Analysis of a Forecasting-Production-Inventory System with Stationary Demand. *Management Science*, *47*(9), 1268–1281.

Tsai, T.-H. (2014). A self-learning advanced booking model for railway arrival forecasting. *TRANSPORTATION RESEARCH PART C*, *39*, 80–93.

Turyna, M., & Hrdina, T. (2009). *Calculating Interval Forecasts* (Tech. Rep.). University of Vienna.

Usman, A. S. (2003). *Demand Forecasting Accuracy in Airline Revenue Management: Analysis of Practical Issues with Forecast Error Reduction* (MSc Thesis). Massachusetts Institute of Technology.

Verschuren, P., & Doorewaard, H. (2010). *Designing a Research Project* (2nd ed.). The Hague: Eleven International Publishing.

Wan, S. K., Song, H., & Ko, D. (2016). Density forecasting for tourism demand. *Annals of Tourism Research*, *60*, 27–30.

Wang, T., Atasu, A., & Kurtuluş, M. (2012). Manufacturing & Service Operations Management A Multiordering Newsvendor Model with Dynamic Forecast Evolution. *Manufacturing & Service Operations Management*, *14*(3), 472–484.

Weatherford, L. (1998). Forecasting Issues in Revenue Management. In *Spring informs conference.* Montreal.

Weatherford, L. (2016, jul). The history of forecasting models in revenue management. *Journal of Revenue and Pricing Management*, *15*(3-4), 212–221.

Weatherford, L. R., & Belobaba, P. P. (2002). Revenue impacts of fare input and demand forecast accuracy in airline yield management. *Journal of Operational Research Society*, *53*(8), 811–821.

Weatherford, L. R., Gentry, T. W., & Wilamowski, B. (2003). Neural network forecasting for airlines: A comparative analysis. *Journal of Revenue & Pricing Management*, *1*(4), 319–331.

Weatherford, L. R., & Kimes, S. E. (2003). A comparison of forecasting methods for hotel revenue management. *International Journal of Forecasting*, *19*(3), 401–415.

Weatherford, L. R., & Pölt, S. (2002). Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues. *Journal of Revenue and Pricing Management*, *1*(3), 234–254.

Weinberg, J., Brown, L. D., & Stroud, J. R. (2007). Bayesian Forecasting of an Inhomogeneous Poisson Process with Applications to Call Center Data. *Journal of the American Statistical Association*, *102*(480), 1185–1198.

Wickham, R. R. (1995). *Evaluation of Forecasting Techniques for Short-Term Demand of Air Transportation* (MSc Thesis). Massachusetts Institute Of Technology.

Zakhary, A., Atiya, A. F., El-Shishiny, H., & Gayar, N. E. (2011). Forecasting hotel arrivals and occupancy using Monte Carlo simulation. *Journal of Revenue and Pricing Management*, *10*, 344–366.

Zakhary, A., Gayar, N. E., & Atiya, A. F. (2008). A Comparative Study of the Pickup Method and its Variations Using a Simulated Hotel Reservation Data. *ICGST-AOML*, *8*, 15–21.

Zickus, J. S. (1998). *Forecasting For Airline Network Revenue Management: Revenue And Competitive Impacts* (MSc Thesis). Massachusetts Institute Of Technology.