FACULTY OF SCIENCE AND TECHNOLOGY

MASTER THESIS

Study programme / specialisation:
Industrial Economics/
Project management and
Investment and finance

The spring semester, 2022

Open

Authors:
Hatem Hamdani
Denis Lajqi

Denis Lajqi
Hatem Hamdani
(signature authors)

Supervisor:
Sigbjørn Landazuri Tveteraas

Thesis title:
A comparative study of forecast combinations in tourism

Credits (ECTS): 30

Keywords:
- Tourism demand forecasting
- Forecast combination
- Empirical analysis
- Neural Networks

Pages: 59

+ appendix: 0

Stavanger, 14.06.2022

Preface

This thesis marks the final chapter of our master's degree in industrial economics at the Department of Safety, Economics, and Planning, at the University of Stavanger. First, we would like to direct our gratitude towards our supervisor, professor Sigbjørn Landazuri Tveteraas at the University of Stavanger, for being a source of inspiration. Sigbjørn introduced us to the domain of research. He was always available for inquiries and advice, and eager to assist us with anything.

Our work would not be completed without data provided by d2o. Therefore, we also owe our gratitude to d2o and Henrik Gerdin for their kind collaboration. Finally, we want to thank our friends and family for their helpful advice, warm discussions, encouragement, and support throughout our studies.

Abstract

With the rapid expansion of the tourist sector in many nations, tourism forecasting has piqued the interest of marketers and academic researchers. Despite that interest, necessary knowledge about tourism forecasting is still lacking. This study addresses crucial questions to better understand the mechanisms underlying forecasting in tourism. In the present study, we investigate whether forecasting performance could be improved by merging tourism forecasts given by two different models: LF and PMI. The investigation measures forecasting of two different metrics over four distinct lead periods. Two error measures are employed to assess forecast accuracy: the mean absolute percentage error and the mean squared error.

To combine the forecasts, we used four established methods, i.e. The simple average method (SA), the geometric mean method (GEOM), the inverse of the mean squared forecast error method (INVM), and ultimately the variance-covariance method (VACO). Our results show remarkable consistency. For the first metric, the combination of forecasts ranks between the two single model forecasts for both error measures. The findings of the other metric reveal that the forecast combination gives the most accurate forecast.

In addition to the four different weighting methods, this study proposes a method of combining forecasts using neural networking. This latter approach shows results that differs from the other four methodologies. The neural networks reveal inconsistent and erroneous results when the mean absolute percentage error is used to rank forecast accuracy. However, when using the mean squared error, the approach rates first out of all the other methods for all lead times.

Altogether, we demonstrate that combining two reasonably accurate forecasts decreases forecast error. Across all forecasting horizons, the combined forecast is much more accurate than the worst single model forecast. Furthermore, the results reveal that when two relatively accurate forecasts are merged, as with metric 2 in the current study, the combined forecast has more minor errors than both single forecasts. The findings indicate that a forecast combination in tourism might yield positive outcomes.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviations	Meanings
LF	Live forecast
PMI	Performance management intelligence
MAPE	Mean absolute percentage error
MSFE	Mean squared forecast error
RN	Room nights
ARR	Average room revenue
AVG	Simple average
GEOM	Geometric mean
VACO	Variance-covariance
INVM	Inverse of the mean squared error
PERC	Perceptron

1 Introduction

1.1 Background

Scientist Francis Galton saw people at a cattle show in 1906 participating in a contest to predict an ox's weight in pounds (Galton, 1907). Although no one, including cattle specialists, could determine the ox's precise weight, the mean value of all the visitors' estimates was just a pound higher than the animal's actual weight (Galton, 1907). In other words, the group estimate was significantly more accurate than any individual estimate. Galton had discovered how a collection of individuals might be wiser than any of its members (Surowiecki, 2005). The scientist's discovery is the essence of the wisdom of crowd's theory. The theory might be considered a forerunner for a particularly significant topic in today's society: forecast combination.

Forecasting is a method of predicting future behavior by evaluating historical trends and data. (Kun, 2021). The method is a prerequisite for properties to accurately analyze their demand, price fluctuations, trends, and seasonality. Many industries have benefitted from forecasting (Arsmtrong, 2001), and the tourism sector is no exception (Song *et al.*, 2009). With the rapid expansion of the tourist sector in many established and emerging nations, tourism forecasting has sparked the interest of marketers and scholars alike. Accurate tourist demand estimates are critical for the private sector regarding business planning and investment or destination governments regarding tourism policy creation and execution (Song *et al.*, 2009). Due to the nature of the sector and its operational characteristics and problems, demand forecasting in the hotel industry has become relatively significant (Song *et al.*, 2009). Forecasting demand is vital because of the enormous changes that may occur and because it is difficult to assess the success of attempts to increase occupancy rates (Yüksel, 2007).

Various quantitative strategies have been used to anticipate tourism demand throughout the years. Time series modeling, live forecasting, and regression are examples of these practical approaches (Song *et al.*, 2009). This study investigates whether forecasting performance is enhanced by merging different forecasting models' tourism forecasts. Chen and Chuang (2000) observed that the days between the forecast and the projected day significantly affected the amount of demand uncertainty. Given the emphasis that tourist planners and commercial decision-makers place on forecasting accuracy, academics must investigate the best methodologies for tourism demand forecasting.

Combination forecasting strategies combine separate forecasts given by several models using suitable weighting algorithms described in the general forecasting literature. Previous research on forecast combination indicates that combining distinct forecasts might increase forecasting accuracy (Bates & Granger, 1969). Although forecast combination has received much attention in the general forecasting area, its role in tourism forecasting remains unsettled.

In this thesis, we investigate whether forecasting performance is enhanced by merging tourism forecasts from different forecasting models. Furthermore, we examine how forecast combinations may enhance the overall forecasting accuracy of tourism demand models. The main goal of the present study is to go deeper into the possibilities for improving forecasting outcomes in the hotel industry by merging forecasts from two distinct types of individual forecasts. The present work is based on the demand for various hotels in Norway, Sweden, and Finland within a highly volatile timeframe; the Covid-19 pandemic. The research issue addressed in this thesis is: Does combining two different demand predictions result in more accurate demand projections?

1.2 Outline

The analysis of this thesis follows four steps. First, the data supplied for the thesis is evaluated, and the forecast inaccuracy for each of the four lead periods is investigated. Second, five alternative combination algorithms are employed to provide combination forecasts across four periods: 7-days, 14-days, 28-days, and 42-days. Third, the statistical significance of the variations in accuracy between combination predictions and single forecasts is determined. Fourth, differences in predicting accuracy are investigated in terms of the combination technique utilized and the length of the forecasting horizon.

The thesis is broken into seven sections. Section one (Introduction) introduces the topics discussed within the present work and establishes the goal of the thesis. Section two (Theory) presents the relevant theory for the thesis. The third section (Data) presents the dataset and describes some of its essential aspects. The fourth section (Methods) presents the methods utilized to solve the thesis question. The findings of the approaches mentioned in section four are presented in section five of the thesis (Results). Following that, section six (Discussion) addresses the outcome concerning the thesis's question. Finally, section seven (Conclusions) presents our main conclusions.

2 Theory

2.1 Forecasting

Forecasting has always been crucial to decision-making and planning. The future's unpredictability challenges individuals and organizations to minimize risks while maximizing benefits. To address real-world difficulties, the enormous variety of forecasting applications necessitates a vast collection of forecasting approaches (Petropoulos *et al.,* 2022). In the 21st century, forecasting has grown tremendously in both theory and practice. Rapid computer developments have made it possible to analyze more extensive and complicated data sets, sparking interest in analytics and data science.

Consequently, the forecaster's toolbox has expanded in quantity and sophistication (Petropoulos *et al.,* 2022). The current expansion in machine learning has permitted the development of dense and complex prediction algorithms, which are gaining popularity among forecasters. Other approaches, such as statistical methods like Bayesian forecasting and complicated regression models, have also profited from computer breakthroughs (Petropoulos *et al.,* 2022). Moreover, advancements have not been restricted to those based on technological developments. For instance, the literature on judgmental forecasting has grown significantly, owing primarily to the "wisdom of crowds" concept (Petropoulos *et al.,* 2022).

Forecasting theory is founded on the assumption that current and historical information may be used to create predictions (Petropoulos *et al.,* 2022). A forecast is a prediction regarding the future value of a variable. Forecasts, or predictions, rely on data about variables that change through time, introducing new difficulties and possibilities. Multiple regression analysis enables us to quantify past relationships, determine if those relationships have remained stable over time, create quantitative projections about the future, and evaluate the correctness of those forecasts (Stock, 2020).

An essential stage in forecasting is often determining when something can be reliably forecasted and when projections will be of little value. A practical and good forecast can capture authentic patterns and linkages in historical data while not replicating previous occurrences that will not occur again (J.Hydman & Athanasopoulos, 2018). A competent forecasting model can distinguish between a random fluctuation in primary data that can be ignored and a genuine trend that should be modeled. Some believe forecasting is impossible in a changing environment; however, this is incorrect. Every environment changes, and a good forecasting model reflects the way things change. Forecasts seldom presume that the environment remains constant (J.Hydman & Athanasopoulos, 2018). A typical forecasting assumption is that the environment will continue to evolve in the same manner it has in the past. A highly volatile environment will continue to be extremely volatile in the future. Conversely, a nonvolatile environment will continue in the same way in the future (J.Hydman & Athanasopoulos, 2018).

Several forecasting models may be used to anticipate the future. The availability of historical data determines the optimal model, the strength of the correlations between the forecast variable and any explanatory factors, and the intended application of the predictions (J.Hydman & Athanasopoulos, 2018). The essential forecasting models in this thesis are classified as quantitative forecasting. According to Hydman and Athanasopoulos (2018), quantitative forecasting may be used when two requirements are met. Firstly, numerical information about the past must be available. Secondly, it must be realistic to expect that certain features of historical patterns will persist in the future.

2.1.1 Demand forecasting

Demand forecasting in the production of goods may be seen as a function connected to projecting the consumption of items so that they can be created correctly to fulfill demand. Demand forecasting is a typical statistical activity in business. It influences production, transportation, and staff scheduling choices and provides a direction for long-term strategic planning (Archer, 1987). Proper demand forecasting provides organizations with helpful information about their potential in their present and other markets, allowing managers to make educated pricing, company growth plans, and market potential decisions (Archer, 1987).

The prediction for the future may differ depending on what type of demand forecasting the company chooses. There are several different ways to conduct the forecasting process within demand forecasting. Utilizing several demand forecasts may enhance forecasting performance (Archer, 1987). Using many forecasting models can expose discrepancies in forecasts. These disparities may indicate a need for more study or improved data inputs (Archer, 1987).

The data provided in this thesis comes from short-term demand forecasting. Short-term demand forecasting is limited to the next three to twelve months (Rheude, 2020). This can help manage the just-in-time supply chain. Looking at short-term demand may change estimates depending on real-time sales data. It enables the ability to react swiftly to changes in client demand (Rheude, 2020).

Forecasting is divided into two main parts: quantitative and qualitative forecasting. Within the different types of demand forecasting, there are several ways to create the forecasts quantitively and qualitatively. This thesis will go deeper into the econometric demand forecasting approach. The econometric technique necessitates data computation (Rheude, 2020). The method combines sales data with information about external influences that influence demand. After the combination, a mathematical formula is developed to forecast future client demand. The econometric demand forecasting approach considers economic factor interactions (Rheude, 2020).

2.1.2 RMS and Hotel Forecasting

Revenue management has expanded dramatically in the lodging sector over the last three decades and is now regarded as an essential component of hotels' marketing and operational strategy (Cross R. C., 2011). Revenue management uses disciplined analytics to forecast customer behavior at the micro-market level and optimize product availability by utilizing price elasticity to maximize revenue growth and profit. Revenue management's primary goal is to sell the right product to the right consumer at the right time, at the right price, and with the appropriate bundle (Cross R. G., 1997). Hotel revenue management strategies were created to economically match or control variable demand with the hotel's limited and perishable capacity. This is accomplished by utilizing a variety of room pricing and allocation mechanisms and addressing key revenue management issues. These notions include reserving a portion of the capacity for higher-paying customers later, efficient pricing discrimination tactics to collect as much of the consumer surplus as feasible, and rigorous overbooking regulations to prevent consumers from canceling at the last minute (Koupriouchina *et al.*, 2014).

A vital part of the revenue management tasks for hotel businesses is forecasting demand. Forecasting in hotels varies for individual and group guests. Because of group consumers, management better understands demand's quantitative and qualitative components. Meanwhile, individual demand is more difficult to foresee (Yüksel, 2007). As the day of stay approaches, hotel room occupancy must be anticipated and re-predicted several times. Forecasts calculated many weeks in advance enable revenue managers to create marketing tactics that may be deployed in reaction to potential low-demand times, generally collaborating with the sales team. Short-term projections decide between a few tactical options, such as staff management or inventory modifications. Consequently, it is critical for an effective demand management system to evaluate the accuracy of a forecasting method at multiple distinct prediction horizons and track how accuracy changes as the date of stay approaches (Koupriouchina *et al.*, 2014).

Ultimately, a revenue management system requires forecasts of several quantities. These quantities are cancellation probabilities, price elasticity, demand, and revenue. The performance of the revenue management system for a hotel business depends very largely on the accuracy of these forecasts (Koupriouchina *et al.*, 2014). As defined previously, a forecast is a prediction regarding the future value of a variable. Consequently, a hotel business cannot rely on a forecast's prognosis as uncertainty is always connected to the future. Improving and sustaining forecast quality is a critical challenge for revenue managers and automated tools. Ultimately, forecasting accuracy has emerged as one of the most pressing issues in the lodging industry. Subpar execution of the forecasts will impede the hotels' efforts to optimize income (Koupriouchina *et al.*, 2014).

The hotel sector is subject to swings in demand. Due to the nature of the sector and its operational characteristics and obstacles, demand forecasting has become very significant in the hotel industry (Yüksel, 2007). Hotel forecasting varies depending on whether the consumer is a solitary or a group. Due to group clients, management better understands demand's quantitative and qualitative components (Yüksel, 2007).

Conversely, individuals make forecasting demand more challenging. Forecasting challenges in hotels can be approximated using either downward or upward forecasting. Downward forecasting employs aggregation, which means that the difficulties are forecasted together. Upward forecasting assesses the concerns independently. According to Yüksel (2008), downward forecasting is less expensive and more accurate during stagnant demand periods. In nonstationary demand periods, however, upward forecasting is recommended (Yüksel, 2007). The Covid-19 epidemic has significantly impacted the data used for the present study. Consequently, this thesis will concentrate on upward forecasting.

Forecasting demand at a high level is critical for a successful hotel operation. A robust prediction model may assist purchasing choices, action plans, marketing strategy, hotel upkeep, personnel decisions, and inventories. However, a faulty forecasting model may cost a hotel business much revenue, making preparing for the future challenging. Consequently, the value of improved forecasts is relatively high in the hotel industry (Yüksel, 2007).

2.1.3 Forecasting accuracy

Because the future is unpredictable, errors in forecasting are inevitable. The error made by the forecast is realized after time has passed, and the actual value is measured. The discrepancy between the actual value Y_T and the forecasted value $\check{Y}_T$ is known as the forecast error (Stock, 2020).

$$\text{Forecast error} = Y_T - \check{Y}_T \qquad [1]$$

Since forecast errors are unavoidable, the forecaster's goal is to minimize them as much as possible to make the forecasts achieve maximum accuracy. A quantitative measure is needed to reach this goal and understand what it means to have a small forecast error.

Various forecasting accuracy measures have been suggested and widely explored in the generic forecasting literature. Although there are numerous forecast accuracy measures, there is no universally accepted measure for all situations because they all have their own set of issues (Koupriouchina *et al.*, 2014). In this study, we decided to use the mean absolute percentage error and the mean squared forecast error to evaluate and determine the performance of the different forecasting methods.

The mean absolute percentage error (MAPE) is one of the most frequently used measures for forecast accuracy. This measure was selected because it is easy to communicate and interpret by the hospitality industry and helpful in comparing forecasts (Koupriouchina *et al.*, 2014). It measures the accuracy as a percentage and is found by taking the sum of the individual absolute errors $|e|_t$, divided by the actual value Y_t, and finding the average (Vandeput, 2019).

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|e_t|}{Y_t} \qquad [2]$$

Another of the most commonly used measures is the mean squared forecast error (MSFE), measured by squaring the individual forecast errors and then finding the average sum of the squared errors.

$$\text{MSFE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \qquad [3]$$

The mean squared error is often used because squaring the error numbers yields all positive values. Because mean squared deviations are easier to work with mathematically, they are frequently utilized in statistical optimization. In practice, significant forecast errors can be far more costly than small ones. A sequence of minor forecast errors usually generates only minor issues for the user, but a single significant forecast inaccuracy might confuse the entire forecasting process. The MSFE captures this principle by calculating the square of the forecast error, which means that huge errors are penalized significantly more than small ones (Stock, 2020).

Significant prediction errors can cause severe issues due to misallocating resources in tourism demand and the hospitality industry. Therefore, having multiple minor deviations rather than a few major ones is preferable. Therefore, knowing which forecasting methods to avoid can be interesting for those in the industry. This is also why the MSFE was chosen as an error measure.

2.2 Wisdom of crowds

The wisdom of crowds' theory holds that big groups of individuals are more innovative than individual specialists in problem-solving, decision-making, inventing, and forecasting (Surowiecki, 2005). The notion is that an individual's viewpoint is intrinsically skewed (Halton, 2021). However, using the average knowledge of a population can reduce the bias or noise to provide a more precise and coherent outcome. James Surowiecki (2005) popularized the wisdom of crowd's notion by investigating how huge groups have produced superior judgments in various industries. Despite its current popularity, the notion of wisdom of crowds may be traced back to ancient Greece and, more particularly, Aristotle's theory of collective judgment (Maskivker, 2013).

There is always some noise in any demand forecasting situation, making forecasting difficult. The wisdom of crowds' idea holds that because idiosyncratic noise is associated with any individual answer, taking the average of numerous replies tends to balance out the noise. The wisdom of crowds' notion does not only hold to human estimate. The idea may also be applied to computer estimations and algorithms in current times. The wisdom of crowds' concept influences the notion of combining forecasts (Petropoulos *et al.*, 2022).

2.2.1 Collective intelligence

Collective intelligence, according to (Kurvers *et al.*, 2016) is one of the most promising techniques for improving decision-making. In relation to solving complicated cognitive challenges, collective intelligence refers to a group's ability to outperform individual decision-makers. In an investigation performed by Kurvers and co-workers (2016), they looked at two significant areas of medical diagnostics: breast and skin cancer detection. They investigated whether integrating the independent judgments of numerous doctors outperformed the best doctor in a group. The study utilized simulation research based on massive real-world datasets containing more than 140 doctors making more than 20,000 diagnoses (Kurvers *et al.*, 2016). The study discovered that diagnostic accuracy similarity is necessary for collective intelligence. When doctors' diagnostic accuracy was relatively similar, aggregating their independent judgments outperformed the best doctor in the group. However, when doctors' diagnostic accuracy differed too much, the study found that aggregating the judgments did not result in more accurate detection (Kurvers *et al.*, 2016).

2.3 Combining forecasts

Initially, the forecasting process was based on a single approach from the available alternatives. However, because of the market's complexity, single outcomes may not be adequate for decision-making in many cases. Aiming towards increasingly accurate projections with the lowest possible error, the theory of combining forecasts erupted. Combining forecasts is a process of composing predictions using some objective or subjective procedure to generate a final combined forecast (Mancuso & Werner, 2013). Over the years, a substantial body of literature on forecast combinations has accumulated. This line of research aimed to decide whether prediction accuracy may be significantly improved by combining multiple individual forecasts. The following section of the thesis gives a detailed evaluation of some discoveries made on the subject of forecast combining.

The work by Bates and Granger (1969) is regarded as a critical piece on combining forecasts. Slightly earlier, Crane and Crotty (1967) published an article where they suggested that combining forecasts through regression might yield positive findings. In their study, Crane and Crotty (1967) combined two forecasting techniques; time series analysis and multiple regression. The two-stage model was successfully applied to an asset management problem within banks, forecasting demand deposits. The authors concluded that the two-stage model effectively integrates knowledge from the dependent variable's previous patterns with information from casually connected variables (Crane & Crotty, 1967). Despite the two authors' findings, it was the work of Bates and Granger (1969) and Reid (1968, 1969) that supplied the first push for the creation of prediction in combining forecasts theory (Clemen, 1989).

J.M. Bates and C.W.J Granger (1969) analyzed whether combining two forecasts would better predict future behavior. The authors observed that merging two forecasts of the same event might result in fruitful findings if the goal was to get the best forecast possible. Their work combined two different forecast sets of airline passengers to create a composite set of forecasts. The issue for Bates and Granger was determining how much weight to give each prediction. The goal was to select an approach likely to provide minimal errors for the combined forecasts. They proposed combining forecasts using a linear combination of two non-biased objective projections. The first projection was given a weight k, and the other was given a weight $(1-k)$. Ultimately, the following equation was created:

$$C = kf_1 + (1-k)f_2 \qquad [4]$$

where f_1 is the value of the first forecast, f_2 is the value of the second forecast, C is the value of the equation, and k is a factor that minimizes the error variance (Bates & Granger, 1969).

Essentially, Bates and Granger used the historical mistakes of each of the original forecasts to calculate the weights to assign to these two original forecasts when constructing the combined forecasts. They eventually determined that the composite forecasts produce a lower mean-square error than the original ones. Most forecasters regard Bates and Granger as foundational authors in combining predictions (Clemen, 1989). In some ways, this is accurate. These researchers were the first to construct a broad mathematical model for optimally integrating forecasts and use their methodologies in real-world circumstances (Clemen, 1989).

Reid (1968) discovered three actual gross domestic product growth rate estimations. The publication released one of the earliest articles where the study aimed to combine forecasts for an optimal error variance measure. In the upcoming year, Reid published another article on the combining forecast subject. Reid (1969) developed the mathematics for combining more than two forecasts. The development was then used to combine multiple forecasts for several economic variables. He discovered that it was possible to gain some improvements in forecasting by combinations of forecasts.

After the discoveries by Bates and Granger (1969) and Reid (1968,1969), the field of forecast combination got a large amount of attraction. The number of studies and discoveries made in the field spiked substantially. A cumulative number of research articles published on combined forecasts from 1960 to 1990 are represented in Figure 1.

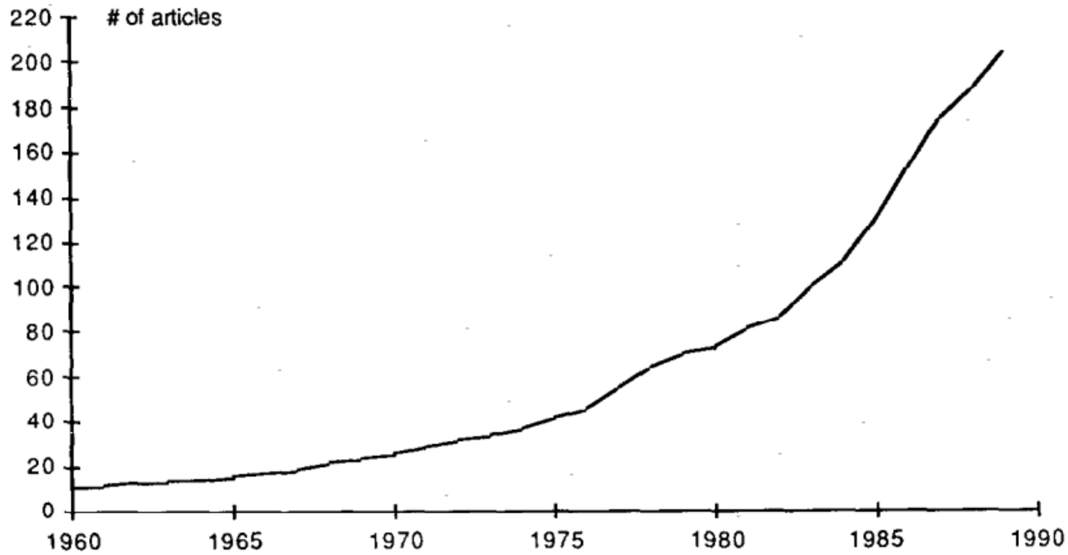

Figure 1: Graph showing the number of articles published on combined forecasts from 1960 to 1990.

(Clemen, 1989)

The discoveries made by Bates and Granger and Reid caused a stream of published articles. This stream of publishment included important work by Dickinson (1973, 1975), Bunn (1975, 1977), and Öller (1978). The articles are theoretical, emphasizing building distinct sections of statistical models for forecast combinations. These works are considered essential additions to the forecast combination field (Clemen, 1989).

Madriakis and Hibon (1979) and Newbold and Granger (1974) conducted seminal works comparing forecasts' performance for various forecasting methods. Madriakis and Hibon (1979) analyzed why some forecasting methods outperformed others under various conditions. The studies showed that basic forecasting models outperformed advanced ones when considerable randomness was included in the data (Makridakis & Hibon, 1979). In the following years, Spyros Makridakis conducted a forecasting competition called the "M-competition", which is discussed in detail in Makridakis (1982). In this competition, 1001 different economic time series were forecasted using various time series forecasting methods. Forecast performance was assessed using several error metrics. While the competition's primary goal was to assess the predicting performance of various time series approaches, two distinct combination strategies were investigated. Both combinations fared well compared to the different strategies, with the simple average outperforming the two (Makridakis *et al.*, 1982; Clemen, 1989).

Applications have grown in number as the concept of combining forecasts, and the usefulness of doing so has expanded. According to Clemen (1989), using a combination of forecasts increases forecast accuracy. Combining forecasts has proven to be a valuable tool for any industry that needs demand predictions. Many theoretical and empirical concerns in forecasting have been addressed, and distinct fundamental issues have been answered to a large extent. However, several difficulties must be addressed throughout various businesses. When combining forecasts, the most pressing challenge is determining how to assign weights to each forecast. This study will go more into this issue for the hotel industry.

3 Data

This thesis examines the accuracy of hotel demand forecasts. Actual – and predicted – demand must be evaluated to develop conclusions relating to the thesis aim. d2o is the innovator and market leader in real-time performance management software. d2o's objective is to provide hotels with analytics and optimization to quickly deploy resources where and when they are most required. d2o creates algorithms and tools for precise forecasting using real-time data (d2o, 2020). d2o provided the data used in this study. In addition, live forecasting data is provided by each hotel.

Each lead time comprises two projections: LF and PMI. The LF is a regularly updated prognosis by individual hotel revenue management teams. In certain circumstances, the LF and PMI are similar in the data. In this scenario, the hotel companies elected to rely on the PMI projection rather than their own forecast.

3.1 General information and labeling

This section explains the organization of the data. The dataset contains 1346 rows of data with information about 32 hotels. The data contains various information and labeling, which is explained in this part. The data provided in this thesis is labeled as shown in the list below.

- **Chain**: The three different types of hotel chains that is represented in the data:
    - ***Chain A:*** A large hotel chain with at least three different brands
    - ***Chain B:*** A small hotel chain with unbranded individual properties
    - ***Chain C:*** A small hotel chain with a franchise and independent properties
- **Brand**: Chain's brand, or a brand operating as a franchise. The same brand always has the same label in the data.
- **Fake name:** Chain + brand + name = fake property name
- **Type**: The three different types of hotels represented in the data:
    - ***Type B&B:*** The properties whose primary source of revenue is by far their rooms

- *Type M&E:* The properties that get much revenue from rooms and rely on meetings and other events.
- *Type XX:* The properties that benefit mainly from their location (for example, airport hotels) or other reasons.

- **City:** The city in which the hotel is located
- **Capacity:** The property's current maximum capacity of rooms.
- **LF:** Live forecast; this is the saved forecast used in other parts of PMI. The live forecast can be created manually, imported by an external source, or automatically updated by PMI's algorithms.
- **PMI:** PMI's forecast using ML and statistical algorithms, without human intervention. D2o provided it.
- **XX 7, XX 14, XX 28, and XX 42:** Forecast type and the number of days before it was created (lead time).
- **Actual:** The actual value of RN and ARR.

3.2 Room Nights and Average Room Revenue

In the present study, forecasts of two separate metrics are considered: room nights and average room revenue. Each hotel's number of booked rooms is abbreviated as room nights (RN). The average room revenue (ARR) forecasts represent the average room price for which a hotel may sell its rooms. Interestingly, the overall forecasts for room nights perform much worse than the overall forecasts for average room revenue. Table 1 presents the accuracy of the two forecasts, each with the four different lead times. The accuracy is measured by using mean absolute percentage error (MAPE).

Room nights MAPE							
LF7	LF14	LF 28	LF42	PMI7	PMI14	PMI28	PMI42
52,9 %	60,8 %	70,0 %	77,7 %	38,2 %	39,7 %	50,8 %	64,0 %

ARR MAPE							
LF7	LF14	LF 28	LF42	PMI7	PMI14	PMI28	PMI42
12,2 %	13,4 %	15,5 %	15,5 %	11,5 %	12,9 %	16,0 %	16,1 %

Table 1: MAPE measurements of different forecasts presented in the present study

Looking at the percentages in Table 1, the mistake in the ARR estimates is significantly lower than the error in the room nights projected. The difference in accuracy between the room nights prediction and the average room revenue projection is attributable to the fact that the room nights forecast is substantially more sensitive during highly volatile forecasting periods, such as the period from which the data in this thesis was projected. Even in a very unpredictable timeline, a hotel must always have some reference to the price they charge for a hotel room to remain in business. The room nights variable, in contrast, is significantly more delicate, as several variables might influence the number of people that check into a hotel. Covid-19 made it nearly impossible for hotel owners to forecast customers in this situation. Thus, the room nights projections perform significantly worse than the ARR estimates.

3.3 Forecasting horizon

Forecasts for one, two, four, and six weeks are included in the data. This section of our study demonstrates the accuracy of the data within the different time horizons and shortly discusses the impact of time horizons in forecasting. Research shows that a short lead time decreases the risk of demand forecasting (De Treville, 2014). Longer lead times increase risk since the further distant the projection, the more uncertainty around the prognosis. Demand forecasting risk requires optimizing the production schedule, which may be controlled via lead time compression. According to the influence of demand forecasting accuracy, the shorter the lead time, the more minor the inaccuracy of demand uncertainty (Chen & Chuang, 2000).

Table 1 clearly illustrates that the forecast is less accurate the further away the projected day is. This poses a significant problem as it is much more beneficial for a revenue management team to have good predictions for larger lead times. The greater the time interval between the forecast date and the actual date (*i.e.*, the forecast horizon), the more successful the revenue management response to the projection (Benavides-Velasco *et al.*, 2014). For example, if it is forecasted many months in advance that a hotel will have poor occupancy during a specific week, revenue management and sales teams might utilize this time to develop and implement initiatives to rectify the issue. However, if the projection is only known a few days in advance, only a few revenue management tactical choices remain possible, such as altering inventory constraints (Benavides-Velasco *et al.*, 2014).

3.4 Data representation

Each hotel has two sorts of forecasts: LF and PMI. LF is each hotel's prediction. Each hotel has unique conditions, and the prediction accuracy of one hotel may differ significantly from that of another. Hence, the accuracy of the LF is highly dependent on the skill of the various hotel management revenue teams. A plot of the seven-day projections for hotel AC1 is created to understand the forecasts and their accuracy better. Forecast values are plotted versus actual values; the plot is shown in Figure 2. It is noteworthy that this is not a general representation as the plot for another hotel may be substantially different.

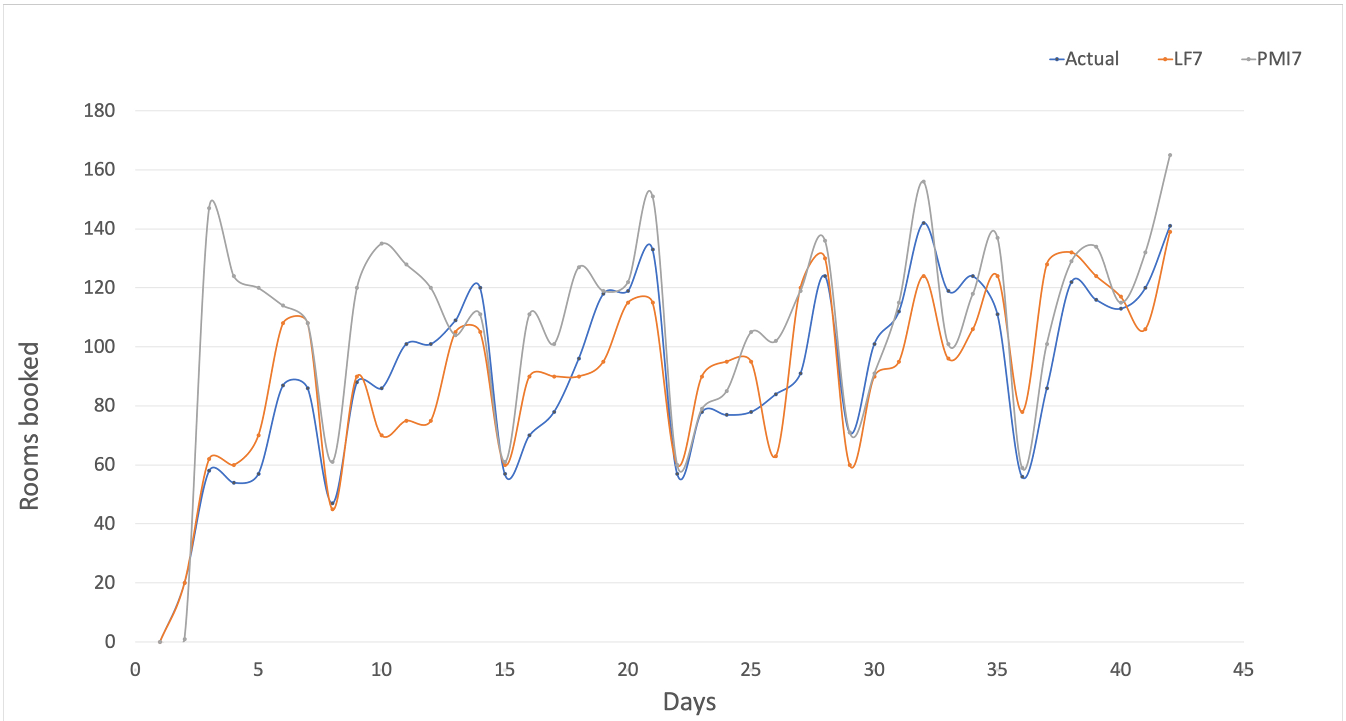

Figure 2: Plot of the 7-day forecasts for hotel AC1 (LF and PMI). The forecasts are plotted against the actual value. The days are along the x-axis, and the number of rooms that are booked at the hotel is along the y-axis

A box plot of the error values is constructed to understand the data better. Box plots are valuable because they give a visual overview of the data, allowing for easy determination of mean values, data set dispersion and skewness. Box plots split data into portions that include around 25% of the data in that set. The plot is presented in Figure 3.

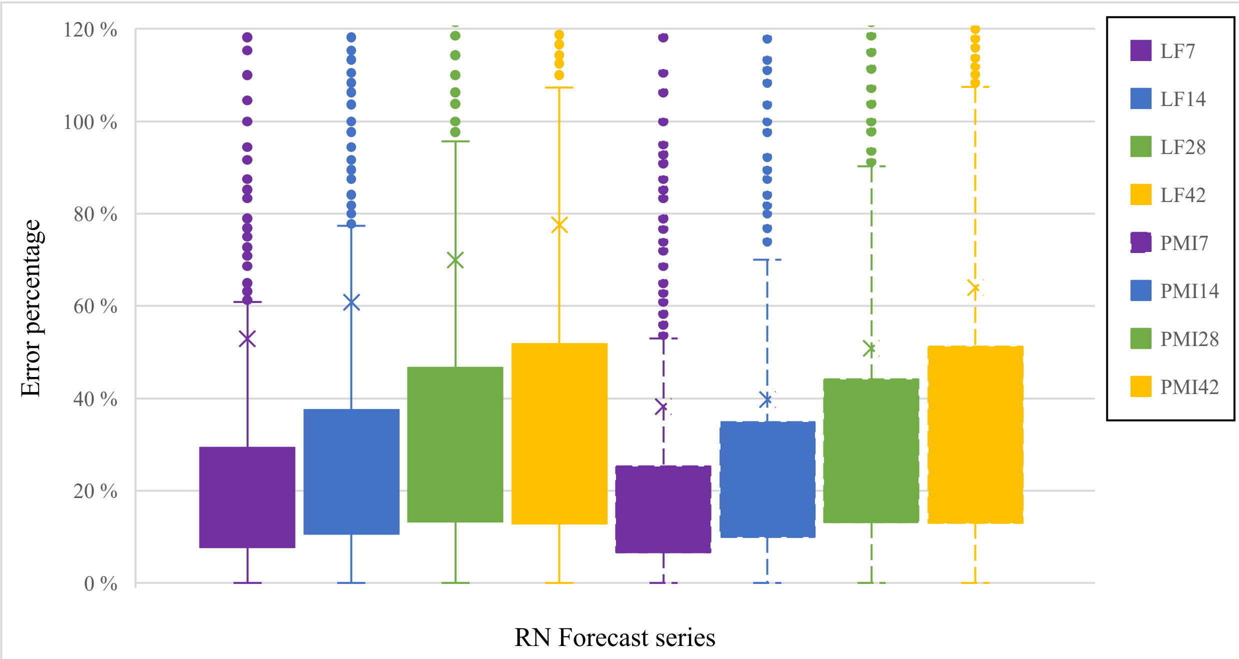

Figure 3: A boxplot of the MAPEs for each RNs forecast series. The gap between the bottom line and the box reflects the data's 0-25th percentile. The gap between the box and the top line reflects the 75th-100th percentile, while the box represents the 25th-75th percentile.

The supplied dataset has unpredictable data as the covid-19 epidemic made forecasting demand challenging. Consequently, there are some considerable outliers in the data. The significant outliers are factored into the calculation of the combinined forecasts. Thus, it is critical to visualize the significant outliers to have a solid grasp of the influence such outliers might have on the computation, particularly with a smaller dataset. The y-axis size is modified to visualize these outliers; the results are shown in Figure 4. The numerous dots show the data outliers over the 100th percentile in the box plot in Figure 4.

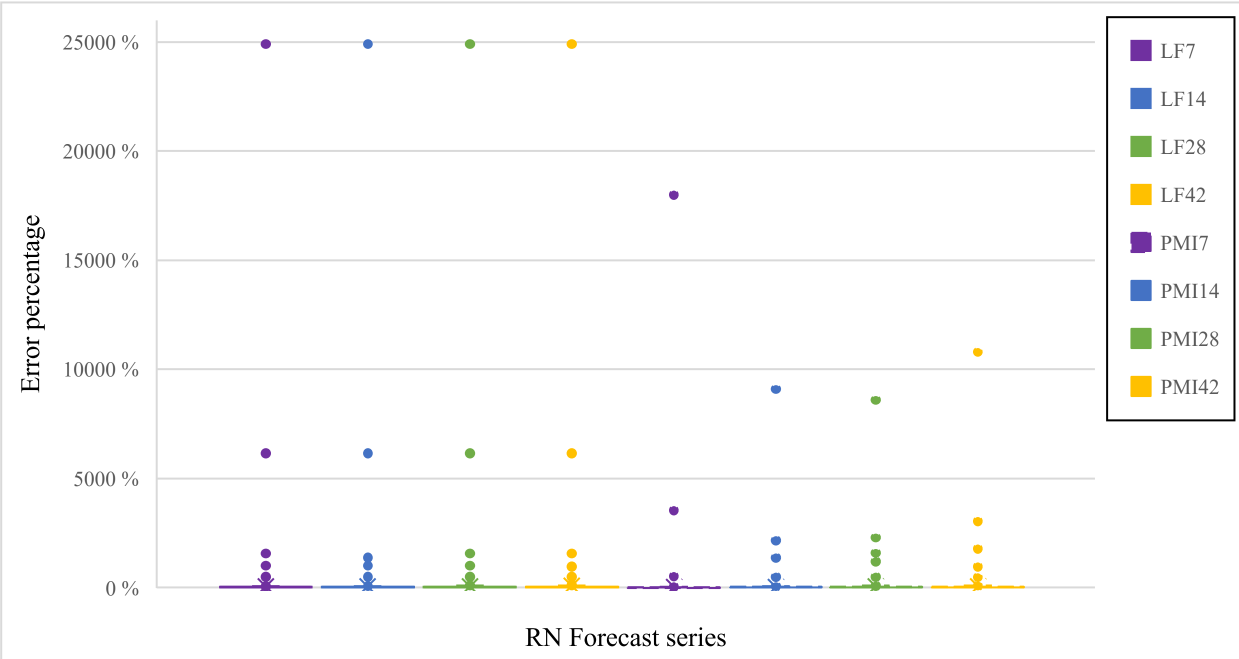

Figure 4: Representation of significant MAPE outliers on RN.

As shown in Figure 4, the most significant outliers vary from 4000 percent to 25000 percent, which is exceptionally high. The high outlier values will influence the data's subsequent computation. The plot in Figure 3 and Figure 4 are of the MAPEs in the RN forecast. A similar plot is made of the ARR forecasts errors. The plot is shown in Figure 5.

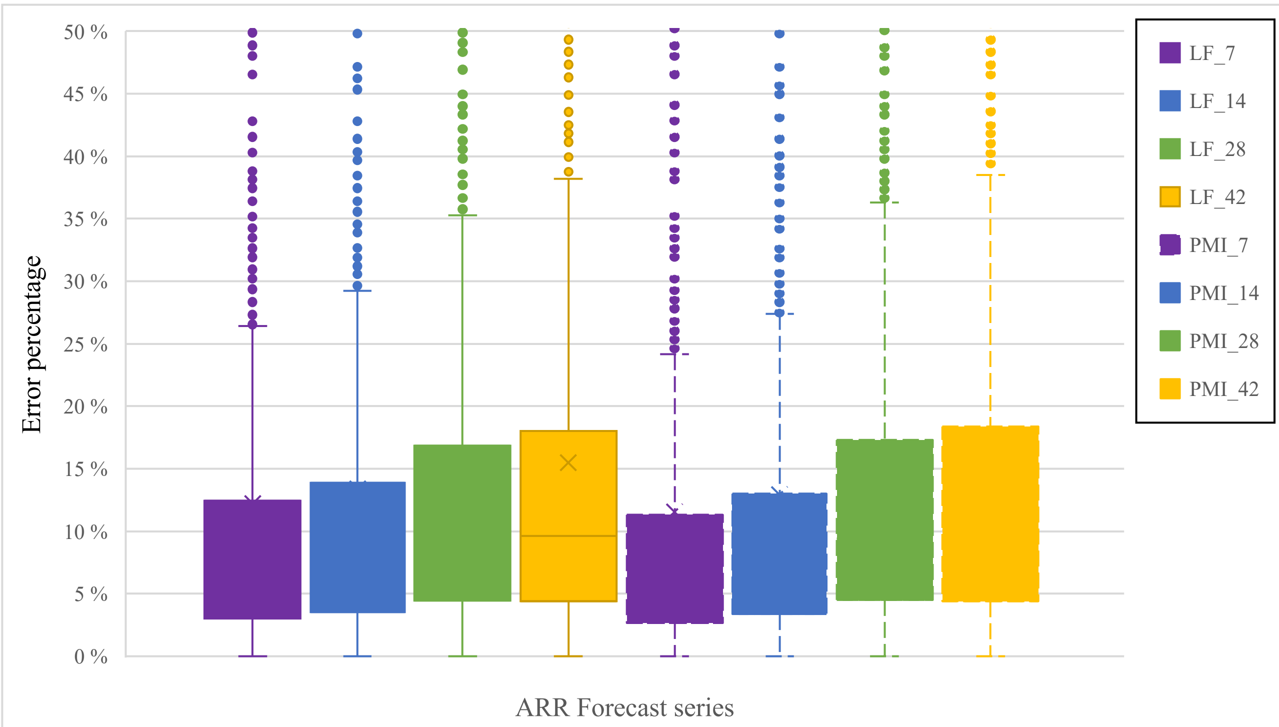

Figure 5: A boxplot of the MAPEs for each of the ARR forecast series.

When the y-axis in Figure 3 and Figure 5 are compared, the ARR projection is more accurate than the RN forecast. Figure 5's error levels are significantly lower than Figure 3 and Figure 4. Outlier values in the ARR forecast are not nearly as significant as those in the room nights forecast.

3.5 Considerations based on pre-analysis data characteristics

Due to the nature of the data, specific changes to the design of this illustrative research are judged essential. The changes are discussed in this section as they mirror the reality of revenue management techniques. The changes also emphasize the significance of error messages and the implications of addressing such failures. Furthermore, the revisions emphasize the need for accurate prediction numbers when merging forecasts.

Segment-level data should be used ideally since it best depicts the reality of daily occupancy forecasting for hotel revenue management, where choices are segment-based and require segment-level prediction and accuracy evaluations (Koupriouchina *et al.*, 2014). This method would result in significantly more accurate forecasts with fewer significant anomalies. However, as the following discussion demonstrates, the reality of operating property level data shows that this may not always be possible due to the nature of some error measurements (Koupriouchina *et al.*, 2014). Another essential factor to consider is that there will always be some ambiguity in the data because no one can foresee the future with absolute precision.

The MAPE measure is employed as one of the vital error measures throughout this study. The MAPE measure's formula is presented in equation [5]. According to the formula, the MAPE measure divides by the actual value at period t. When a segment's actual occupancy is zero, the MAPE is undefined; that is, it cannot be computed due to the zero value of its denominator, resulting in the incomputable value message "#DIV/0!" within MS Excel. Clearly, the problem of missing observations (of computed error measures) owing to zero daily actuals defines the MAPE measure and all forecasting accuracy measures with a percentage error component (Koupriouchina *et al.*, 2014).

There are various solutions to the problem of zero values. If the number of incomputable days is considerable, the hotel might not employ any percentage-based error metrics. This strategy can influence the reliable forecast since different types of forecast error metrics provide different outcomes. Another appropriate method is to disregard the incomputable periods. The main disadvantage of the latter strategy is that it ignores a potentially large number of forecast errors. Many of the incomputable periods could have had a non-zero forecast, implying a non-zero forecast error for that incomputable period that should have been included in the accuracy assessment (Koupriouchina *et al.*, 2014). Despite the disadvantage, this technique may be the best when just a tiny fraction of the dataset contains zero values. There were 24 of 1346 rows in the dataset used for this study that included such zero values; hence it was decided to discard the incomputable periods.

4 Methods

Five methods of combining forecasts have been used in this study. In this section, we first examined the single model forecasts and provided various prediction error measurements. Furtherly, we present the five methods we used to integrate the forecasts.

4.1 Performance of PMI and LF

4.1.1 Forecast accuracy

Performance was evaluated using two of the most popular accuracy measures in the hospitality forecasting literature; the MAPE and the MSFE (Koupriouchina *et al.*, 2014). The error measures for each data row were calculated in MS Excel. Furtherly, the mean of all the measures was calculated. **Feil! Fant ikke referansekilden.**

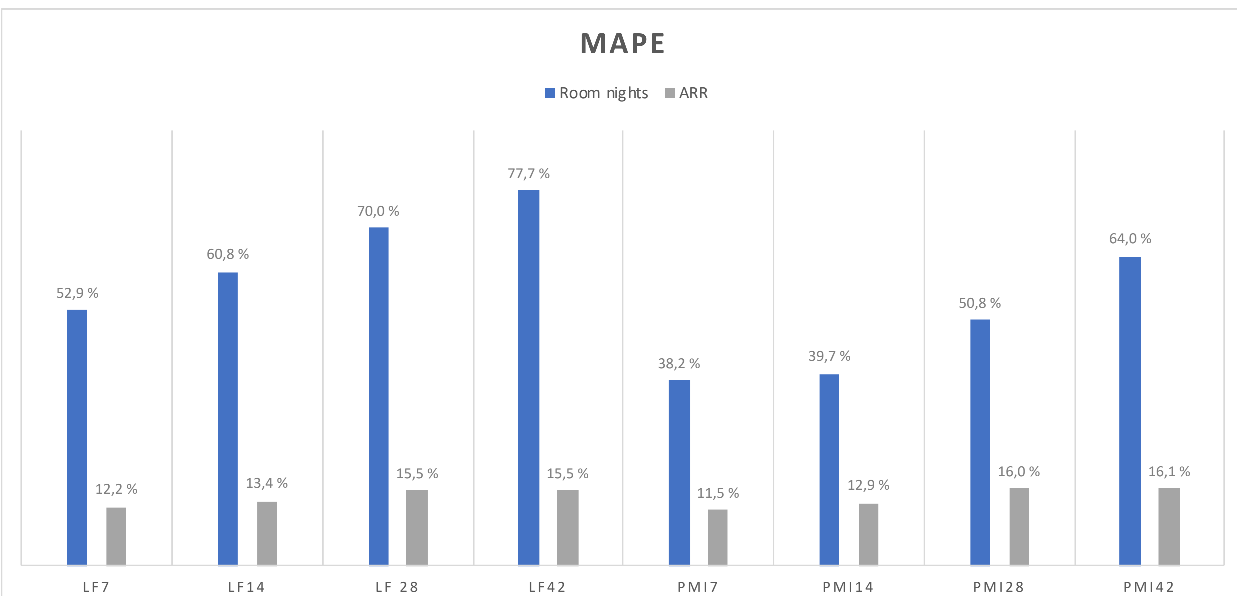

Figure 6: LF and PMI MAPEs for RN and ARR

Figure 6 represents the MAPEs for each forecast on RN and ARR. The figure shows that PMI7 performed the best on both RN and ARR, having the lowest MAPEs of 38,2% and 11,5%. The

PMI performed better than LF for all time horizons on RN. However, the PMI only performed better on the seven- and fourteen-day time horizons on ARR. Figure 7 represents the MSFE measures. The MSFEs provided the same results as the MAPEs, with the PMI performing the best on the seven- and fourteen-day time horizon. The PMI7 performed the best for both room nights and ARR, according to the MSFEs.

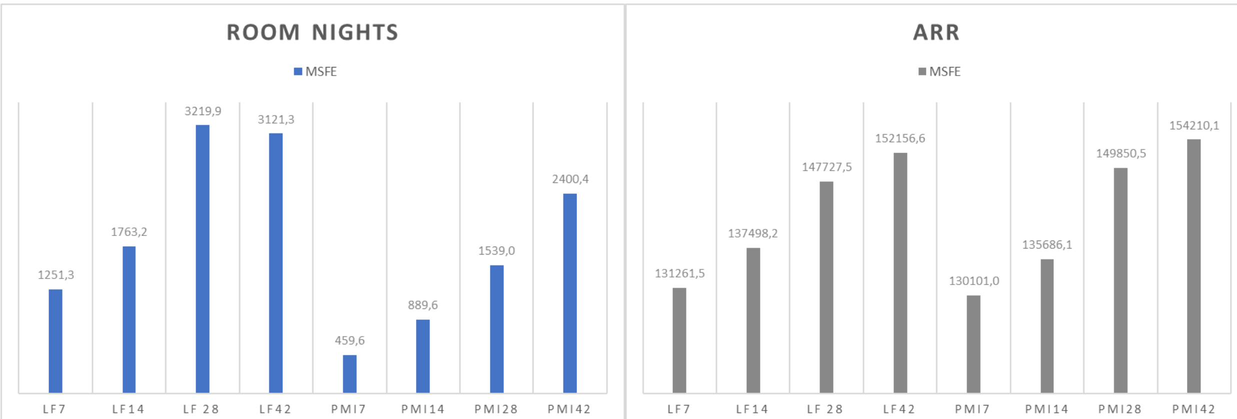

Figure 7: LF and PMI MSFEs on RN and ARR

4.1.2 Forecast statistics

In order to understand the characteristics of the data, statistics of the error measurements were analyzed. The PMI7 forecast was thoroughly analyzed as it was the most accurate forecast. The standard deviation for each of the different prognoses was calculated, and the result of PMI7 was further used to check the distribution of errors around the mean value of 38%. The standard deviation measurements are shown in **Feil! Fant ikke referansekilden.**.

Room nights MAPE St.dev							
LF7	**LF14**	**LF 28**	**LF42**	**PMI7**	**PMI14**	**PMI28**	**PMI42**
707,9 %	709,0 %	710,3 %	712,1 %	505,5 %	263,4 %	256,8 %	322,9 %

ARR MAPE St.dev							
LF7	**LF14**	**LF 28**	**LF42**	**PMI7**	**PMI14**	**PMI28**	**PMI42**
63,1 %	63,5 %	73,4 %	54,2 %	62,3 %	61,1 %	73,4 %	60,6 %

Table 2: MAPE Standard deviations of LF and PMI

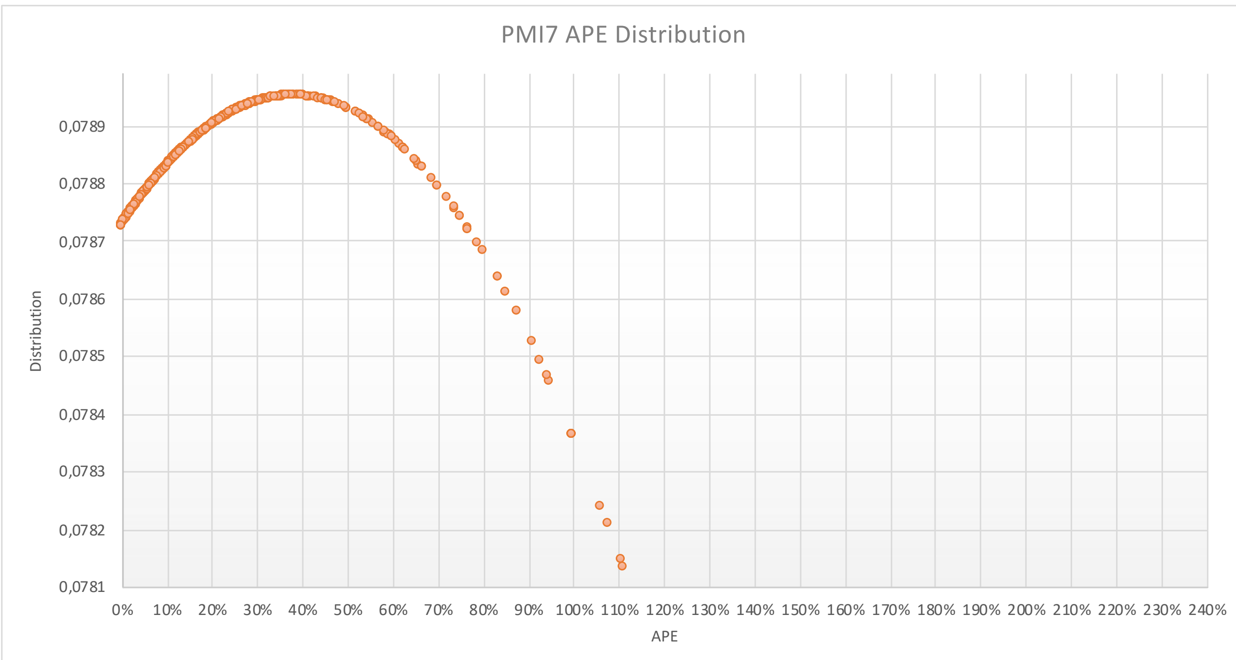

Figure 8: PMI7 Absolute Percentage Error distribution

The calculated standard deviations are shown in **Feil! Fant ikke referansekilden.**2, pointing to a massive spread in the absolute percentage errors. Although the PMI7 had the lowest MAPE and was most accurate, the errors are more spread out than for other prognoses. The reason for these vast deviations were three extremely high dataset errors. Figure 8 shows that the distribution of the absolute percentage errors positively skewed as there were more errors below the mean value than above. Consequently, the graph skewed to the left.

4.1.3 Observations and considerations

The MAPE and the MSFE values indicated that the PMI and LF forecasts performed poorly. The best performing prognosis was the PMI7. Although the PMI7 performed the best, a MAPE of 38% on room nights is still a significant percentage error. Overall, the MAPE and MSFE measurements were unnaturally high, indicating a significant improvement potential.

The forecasts had significant errors due to some extreme values in the dataset. This resulted in skewed data with high error measures and massive standard deviations. The MAPEs were measured without the extreme value dates to comprehend the magnitude of which those values offset the data. The calculations showed a 17% decrease in MAPE for PMI7. In addition, the standard deviation decreased by 650%. It was considered to ignore the extreme value dates when evaluating the forecasts. However, it was ultimately decided to keep the dates in the dataset as the errors were fundamental and represented how inaccurate the forecasts may be in today's unpredictable environment.

4.2 Combining forecasts

The performance evaluation showed that the single model forecasts were poor. Hence, a way of producing more accurate forecasts was necessary. This study conducts a set of different combination approaches to study the possibility of achieving more accurate forecasts. Specifically, we investigated four different combination methods from the forecasting literature. In addition, we created a way of combining forecasts using neural networking.

4.2.1 Simple average

The most straightforward procedure for combining forecasts was to take an arithmetic average. This process offered a reasonable starting point and was found to outperform more complex combination methods (Clemen, 1989). According to Clemen (1989), the simple average (SA) offers the advantages of impartiality, robustness, and an excellent "track record" in economic and business forecasting. Therefore, the SA is considered a popular choice in many forecast combination studies and a valuable benchmark. Thus, the SA method was the first method used to combine the forecasts in this study.

Because the combined weight was assigned equally to each of the individual forecasts, the simple average combination approach calculated the composite forecasts without considering the previous performance of the individual forecasts. The simple average combination method is expressed as,

$$f_{ct} = \sum_{i=1}^{n} \frac{f_{it}}{n} \quad [5]$$

where f_{ct} was the combined forecast, f_{it} was the ith forecast in time t, and n was the number of individual forecasts that are combined.

In this study, the PMI and the LF are combined by the SA procedure, and the SA was found for each time horizon. The SA was found in MS Excel by first finding the arithmetic average of the predictions for all the dates. The combined forecasting predictions were denoted with AVG followed by the time horizon. The simple average for a seven-day time horizon would then be AVG7. Table 3 shows a section of the calculated output using the SA combination method.

Room nights					ARR					Room nights Absolute Percentage Error				ARR Absolute Percentage Error			
Actual	AVG-7	AVG-14	AVG-28	AVG-42	Actual ARR	AVG:7	AVG:14	AVG:28	AVG:42	AVG7	AVG14	AVG28	AVG42	AVG_7	AVG_14	AVG_28	LF_42
8	7,0	11,0	46,0	55,0	824,2	362,7	463,6	1 016,9	979,6	12,50 %	37,50 %	475,00 %	587,50 %	55,99 %	43,76 %	23,38 %	18,85 %
14	28,0	29,5	46,0	59,5	1 190,7	875,3	1 174,7	1 142,0	1 135,4	100,00 %	110,71 %	228,57 %	325,00 %	26,49 %	1,34 %	4,09 %	4,65 %
10	25,0	25,0	44,0	48,0	858,2	779,0	1 011,3	1 025,4	1 030,7	150,00 %	150,00 %	340,00 %	380,00 %	9,22 %	17,84 %	19,49 %	20,11 %
15	26,0	38,0	71,0	77,0	964,1	946,6	949,4	960,0	966,8	73,33 %	153,33 %	373,33 %	413,33 %	1,82 %	1,53 %	0,43 %	0,28 %
22	34,5	48,5	47,5	51,5	1 241,6	782,6	784,1	867,8	887,8	56,82 %	120,45 %	115,91 %	134,09 %	36,96 %	36,84 %	30,11 %	28,49 %
30	38,5	51,0	47,0	50,5	676,3	743,7	783,0	878,0	877,1	28,33 %	70,00 %	56,67 %	68,33 %	9,97 %	15,79 %	29,83 %	29,70 %
16	18,5	15,0	18,5	20,0	676,9	868,5	703,2	862,1	885,5	15,63 %	6,25 %	15,63 %	25,00 %	28,30 %	3,87 %	27,35 %	30,80 %
34	24,5	40,0	46,5	41,0	1 270,6	912,8	904,8	888,8	940,3	27,94 %	17,65 %	36,76 %	20,59 %	28,16 %	28,79 %	30,05 %	25,99 %
30	37,0	40,0	47,5	44,5	865,4	849,1	1 217,1	1 128,0	1 113,2	23,33 %	33,33 %	58,33 %	48,33 %	1,89 %	40,63 %	30,34 %	28,62 %
30	35,0	37,0	40,0	43,5	1 090,8	923,1	1 169,2	1 102,2	1 107,9	16,67 %	23,33 %	33,33 %	45,00 %	15,37 %	7,18 %	1,04 %	1,56 %
14	30,0	41,0	34,0	37,5	1 012,3	862,5	999,1	973,2	1 039,8	114,29 %	192,86 %	142,86 %	167,86 %	14,80 %	1,31 %	3,86 %	2,71 %

Table 3: MAPE calculation with SA combination method

Further, the procedures of finding the error measures were conducted. The MAPE and the MSFE of the AVG forecast were measured to analyze the combined forecasts' performance. The orange-colored cells in Table 3 show a section of the calculated MAPE for the SA combined forecasts.

4.2.2 Geometric mean

Using the arithmetic average is only one of the conventional ways to combine forecasts. Other sorts of averages could also be valuable in forecast combinations. The geometric mean was calculated and used to combine the forecasts of LF and PMI. The geometric mean has the advantage of always returning a lower value than the arithmetic mean. As a result, it provides some shrinking, which is a desired feature (Andrawis *et al.*, 2010). Consequently, this method of combining was included as a combination method.

The geometric mean z of the two forecasts was found by taking the square root of the product of the two forecasting predictions x and y on time t.

$$z_t = \sqrt{x_t y_t} \qquad [6]$$

The method was abbreviated as GEOM and was conducted for all the time horizons on MS Excel, similarly to the simple average method.

4.2.3 Inverse of the Mean Squared Forecast Error

Stock and Watson (1999) devised a system in which the weights of the individual forecasts were proportional to the inverse of the MSFE. This method was abbreviated as the INVM method. The weighting of one forecast in the combination mix was the proportional size of the sum of its means squared forecast errors. The weights for each of the forecasts were computed as follows:

$$w_1 = \frac{\sum_{t=0}^{t} MSFE_t^{(2)}}{\sum_{t=0}^{t} MSFE_t^{(1)} + \sum_{t=0}^{t} MSFE_t^{(2)}} \quad [7]$$

$$w_2 = \frac{\sum_{t=0}^{t} MSFE_t^{(1)}}{\sum_{t=0}^{t} MSFE_t^{(1)} + \sum_{t=0}^{t} MSFE_t^{(2)}} \quad [8]$$

Where $MSFE_t^{(i)}$ was the mean squared forecast error for forecast i on time t. Therefore, the MSFEs of the LF and PMI forecasts were used to determine the combination weights. The weights for the two individual forecasts were computed in MS Excel and are presented in Table 4.

	Room nights				ARR			
Horizon	7	14	28	42	7	14	28	42
LF	26,9 %	33,5 %	32,3 %	43,5 %	49,8 %	49,7 %	50,4 %	50,3 %
PMI	73,1 %	66,5 %	67,7 %	56,5 %	50,2 %	50,3 %	49,6 %	49,7 %

Table 4: Weights computed with the INVM method

4.2.4 Variance-covariance method

Bates and Granger (1969) were the first to present the variance-covariance approach. The variance-covariance approach of forecast combining has a simple underlying logic. In essence and in its most basic form, it addresses the potential that one may reveal information that the other does not when comparing two rival projections. Consequently, a combined forecast based on them can outperform each of the individual forecast estimates. Bates and Granger (1969) assumed that the individual projections would perform consistently over time to derive the variance-covariance formula. A linear combination of the two sets of forecasts would provide the combined forecast, with the first set receiving a weight of k and the second receiving a weight of (1-k). Section 2.3 examined the equation in-depth (equation [4]). By assuming consistency in the forecasts, the variance of the two forecasts could be denoted by σ_1^2 and σ_2^2 for all values of time t. The variance of errors in the combined forecast could then be written as follows:

$$\sigma_c^2 = k^2\sigma_1^2 + (1-k)^2\sigma_2^2 + 2\rho k\sigma_1(1-k)\sigma_2 \quad [9]$$

Where k was the proportional weight assigned to the first set of forecasts and ρ was the correlation coefficient between the mistakes in the first and second sets of forecasts. The value of k should be chosen so that the total forecast errors are minor. We selected to reduce the total variance σ_c^2 in particular. By differentiating with respect to k and equating to zero, we obtain the minimum σ_c^2 occurs when:

$$k = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad [10]$$

The preceding calculation was computed in MS Excel to try and find optimal weights for the forecasts. The variance and covariance of the various forecasts were determined using the built-in features in MS Excel. The variance and covariance were then utilized in equation [10] to compute the weights between the LF and PMI forecasts. The values used are shown in Table 5.

	Room nights							
Forecast	LF7	LF14	LF28	LF42	PMI7	PMI14	PMI28	PMI42
Variance	50,12	50,27	50,46	50,71	25,55	6,94	6,59	10,43

	ARR							
Forecast	LF7	LF14	LF28	LF42	PMI7	PMI14	PMI28	PMI42
Variance	0,40	0,40	0,54	0,29	0,39	0,37	0,54	0,37

	Room nights				ARR			
Horizon	7	14	28	42	7	14	28	42
Covariance	35,69	18,48	17,80	22,57	0,39	0,38	0,54	0,33

Table 5: Variance and covariance for LF and PMI. Green tables show variances, and the blue table shows the covariances.

	Room nights				ARR			
Horizon	7	14	28	42	7	14	28	42
LF	19,95 %	1,17 %	1,01 %	3,27 %	50,82 %	52,41 %	49,92 %	44,31 %
PMI	80,05 %	98,83 %	98,99 %	96,73 %	49,18 %	47,59 %	50,08 %	55,69 %

Table 6: Weights calculated from the VACO combination method

The weights assigned to the LF and PMI are shown in Table 6. After determining the weights, the results were applied to the forecasts, and the variance-covariance method values were calculated. The MAPE and the MSFE of the combination were calculated to map the accuracy of the forecast combination. The error calculations are discussed in the results chapter.

4.2.5 Using a neural network to find optimal weights

Neural Networks, also known as Artificial Neural Networks, are mathematical formulations inspired by the work and operation of biological neurons. They distinguish themselves by their capacity to simulate nonstationary, non-linear, and very complicated datasets. With improved computer capacity, this feature propelled neural networks to the forefront of study in almost every science sector, including demand forecasting (Petropoulos *et al.*, 2022).

A standard neural network structure consists of three layers: input, hidden, and output. Each layer is made up of nodes. The input layer is the initial layer in any neural network, and the number of nodes corresponds to the number of explanatory variables, also known as inputs. The last layer is the output layer, with the same number of nodes as the number of response variables (forecasts). There are one or more hidden layers between the input and output layers where the nodes specify the degree of complexity the model can fit (Petropoulos, et al., 2022). The structure of a neural network is shown in Figure 9.

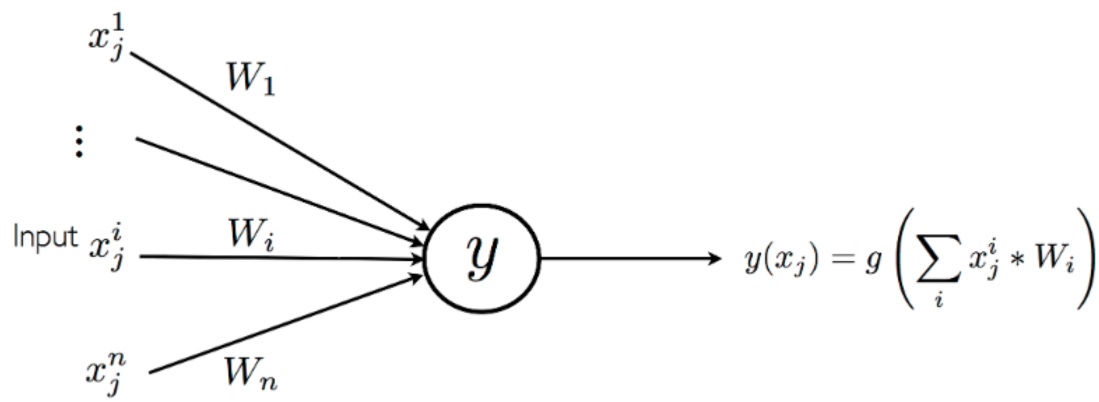

Figure 9: Figure of a standard neural network structure that consists of the input layer, hidden layer (y), and output layer (Bringsjord, 2018)

Each node in one layer is connected (weighted) to all or a subset of the nodes in the following layer. Neural networks process the information: the explanatory variables are contained in the input nodes. The connections between the input and the first hidden node weight these variables, and the information reaches the hidden nodes as a weighted sum of the inputs. A non-linear or linear algorithm in the hidden nodes frequently transforms the information received. This procedure continues until the information reaches the forecast/output layer. Within the neural network, the weights that connect the nodes are adjusted such that the network maps the input value of the training data to the matching output value. This mapping is based on a loss function determined by the forecasting problem (Petropoulos *et al.*, 2022).

The model built in this study utilized the two types of forecasts presented in the data and found optimal weights for each by comparing the forecast values with the actual value. The network would continually try to understand the pattern of the forecasts concerning the actual value by utilizing some of the data as training data. Following training, the neural network would be provided solely with the forecast values as input. Next, the neural network would try to predict the actual value based on the pattern learned in the training process by weighting the two inputs. The weighting and mathematical calculations were all made inside the neural network. As this model was made to present the weights as outputs, the neural network could not have any hidden layers. By adding hidden layers to the network, the complexity would rise, and it would not be possible to extract the weights used by the network. A perceptron is a neural network without hidden layers, so the model was a perceptron.

Like a neural network, a perceptron assigns weights to each input. The perceptron then adds a bias, which is a hidden value, to the weighted inputs. The perceptron would generate a weighted net sum by summing the inputs and adding a bias. Because there are no hidden layers, the weighted sum was the perceptron's output, and the sum was measured with the actual value. After comparing the total and the actual amount, the perceptron would use an error function to determine how close the sum was to the actual value. The perceptron then used the information to modify the weights. Ultimately, the perceptron would have found the weights that gave the least amount of error. The construction of a perceptron is displayed in Figure 10.

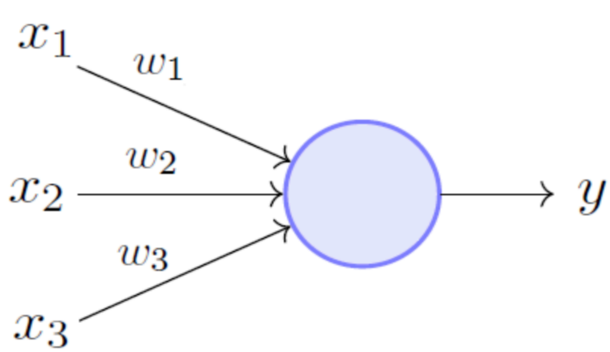

Figure 10: The construction of a perceptron

Importing the data from the excel sheet into the Python workbook was the first step in creating the perceptron. Following that, it was determined which forecasts would be used. As mentioned in section 3.5, some of the data provided have an 'Actual' value of zero in the dataset. The dataset provided for the thesis is relatively small for a neural network. Therefore, such an anomaly in the dataset might have created significant disruptions in the weighting of inputs, ultimately leading to non-representative outcomes. Consequently, it was decided to eliminate any row in which the 'Actual' value was equal to zero.

The next stage in establishing the perceptron was to indicate which rows would be used as training data and which would be utilized as targets. The algorithm predicts the results of the testing samples using the parameters computed with the training samples and then compares the forecast to the target information to determine how well the prediction went. It would utilize this data to determine how to recalculate based on the training samples. The test size needed to be set after specifying that the perceptron would use the forecast values to predict the target values ('Actual' values). The test size represented the proportion of the dataset to include in the test split. The test size decided how much data was allocated for strictly training the perceptron. Usually, the test size is set to 25% of the datasets, but because the dataset presented in this thesis is relatively small for a neural network, the test size was reduced to 20%.

The neural network must be layered once the training and target data have been provided to the model. As previously stated, the model is a perceptron with no hidden layers. Therefore, the only layers that must be defined are the input and output layers. Some hyperparameters need to be set when building a neural network: the learning rate and the loss function. The learning rate may be the most crucial hyperparameter when constructing a neural network. The learning rate is a hyperparameter that specifies how much the model should change in response to the predicted error each time the model weights are updated. Choosing the learning rate was problematic since a value too low may have resulted in a complex training process that may become stuck. In contrast, a high value may have resulted in learning a suboptimal set of weights too quickly, creating an unstable training process (Brownlee, 2019). The learning rate used in this study was found by trial and error. Several dataset areas have significant error measurements, as described in the data chapter. Setting the learning rate to a "high" number may make these differences in the data significantly impact the weighting, resulting in a chosen learning rate of 0.01.

We employ two forms of loss functions in the present work: MAPE and MSE. As a result, the perceptron's natural loss function would be either MAPE or MSE. Because MSE is commonly employed in neural networks, this perceptron employed MSE as the error function. The stacking of the perceptron was completed once the hyperparameters are specified.

The perceptron struggled to identify effective weightings due to a lack of sufficient and consistent data. In order to improve consistency in weight computation, the entire code was placed in a while loop, and the perceptron was executed one hundred times for each forecast. The outcomes would then be stored in an array. Consequently, the array would have one hundred weights for each of the forecasts. Finally, the array's average value was determined. The possibility of insufficient training data influencing weighting would be significantly reduced in this manner. The code used to combine the forecasts is shown in Figure 11.

```python
1  import tensorflow as tf
2  import pandas as pd
3  import numpy as np
4  from sklearn.model_selection import train_test_split
5
6  # Retrieve the data from the excel doc
7  data = pd.read_excel("/Users/bruker/Downloads/University-Data.xlsx", sheet_name=1, header=1)
8  data = data[["LF-7", "PMI-7", "Actual"]]
9  # Remove rows where actual == 0
10 data = data[data['Actual'] != 0]
11 data = data.to_numpy()
12 #Create variables that specify which part of the data that are training data and which that are targets
13 X = data[:, [0, 1]]
14 Y = data[:, [2]]
15 # Create the array that holds the weights
16 averages = [[],[]]
17 p = 0
18 #while loop start and will run until p is equal to 100
19 while p <= 100:
20     # Use the variables in line 13 and 14 to split into train and targets
21     x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.2)
22
23     # Create input layer for network based on size of training data
24     input_layer = tf.keras.layers.Input(shape=len(x_train[0]))
25     x = input_layer
26     # Uncomment this line to add hidden layer
27     #x = tf.keras.layers.Dense(32)(x)
28     # Output layer
29     out = tf.keras.layers.Dense(1)(x)
30     model = tf.keras.models.Model(
31         inputs=[input_layer], outputs=out)
32
33
34     # Build model with specified hyperparameters
35     model.compile(optimizer=tf.keras.optimizers.Adam(lr=0.01), loss="mse", metrics=["mse"])
36
37     # Run training process
38     model.fit(x_train, y_train, validation_data=(x_test, y_test), batch_size=4)
39
40     #Add the weights to the array
41     averages[0].append(model.layers[1].weights[0][0])
42     averages[1].append(model.layers[1].weights[1][0])
43     #Increment p-value
44     p += 1
45
46 #Take the averages of the weight-arrays and print the values to the terminal
47 weight1 = np.array(averages[0])
48 print(weight1.sum()/len(weight1))
49 weight2 = np.array(averages[1])
50 print(weight2.sum()/len(weight2))
```

Figure 11: The perceptron code with explanatory comments

5 Results

This section presents the results of employing the various forecast combination methods. The MAPEs and MSFEs were calculated for each approach. We chose to only display the findings for the MAPE error measure in situations when the two error measurements had the same outcome.

5.1 Combination methods

5.1.1 Simple Average (SA)

We achieved interesting results by combining the LF and PMI forecasts with the SA approach. The arithmetic average between the two forecasts for all time horizons was calculated. Further, the error measures were conducted to compare the performance of the combined forecast to the individual ones. This comparison is visualized by the two figures below, where Figure 12 displays the MAPE for the combined and individual forecasts on room nights, and Figure 13 presents the same on ARR.

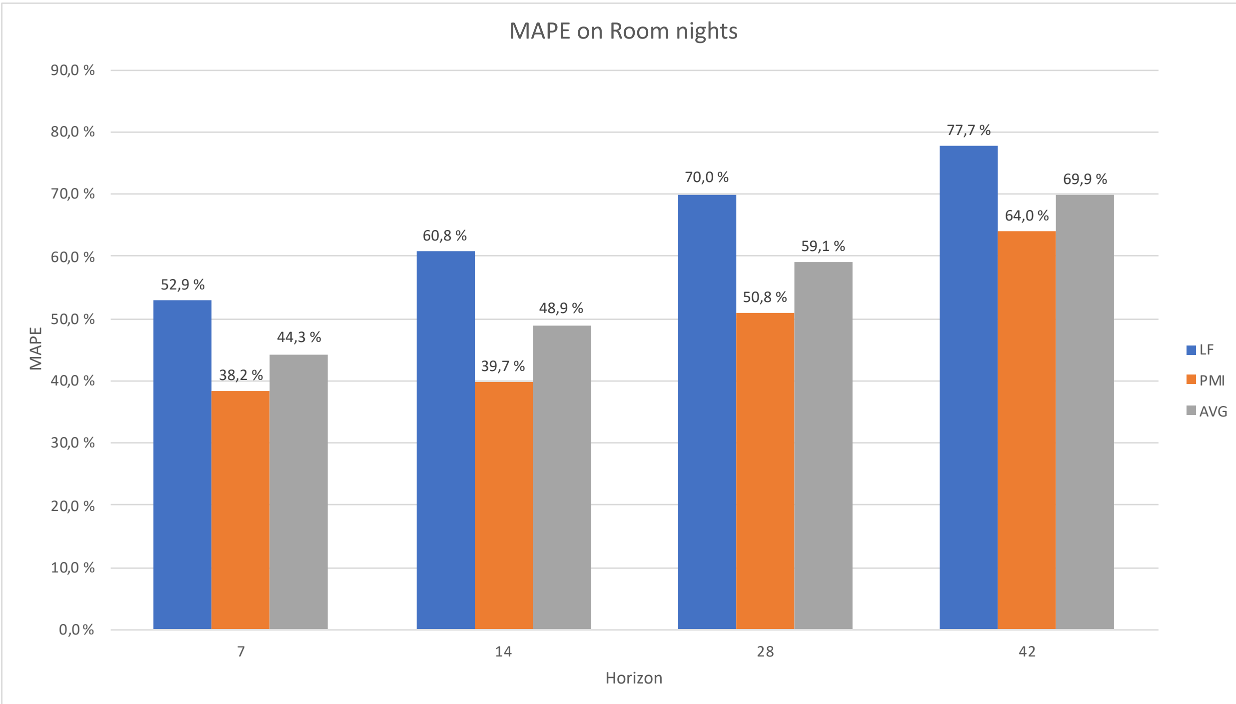

Figure 12: MAPE on RN for LF, PMI, and SA

Comparing the accuracy results on room nights, we observed that the PMI forecast still was the most accurate among the three forecasts on all time horizons. The AVG forecast was also observed to perform better than the LF forecast across all time horizons, which ultimately makes the LF the least accurate forecast. The AVG forecast fell between the two individual ones on performance on RN, as shown in Figure 13.

While the accuracy results put the AVG forecast between the PMI and the LF on room nights, the AVG forecast performed the best on ARR according to the MAPE and the MSFE measures. The MAPE results are displayed in Figure 13 and show that the AVG outperforms both single model forecasts on nearly all the time horizons except for the 42-day lead where the LF performed as well as the AVG.

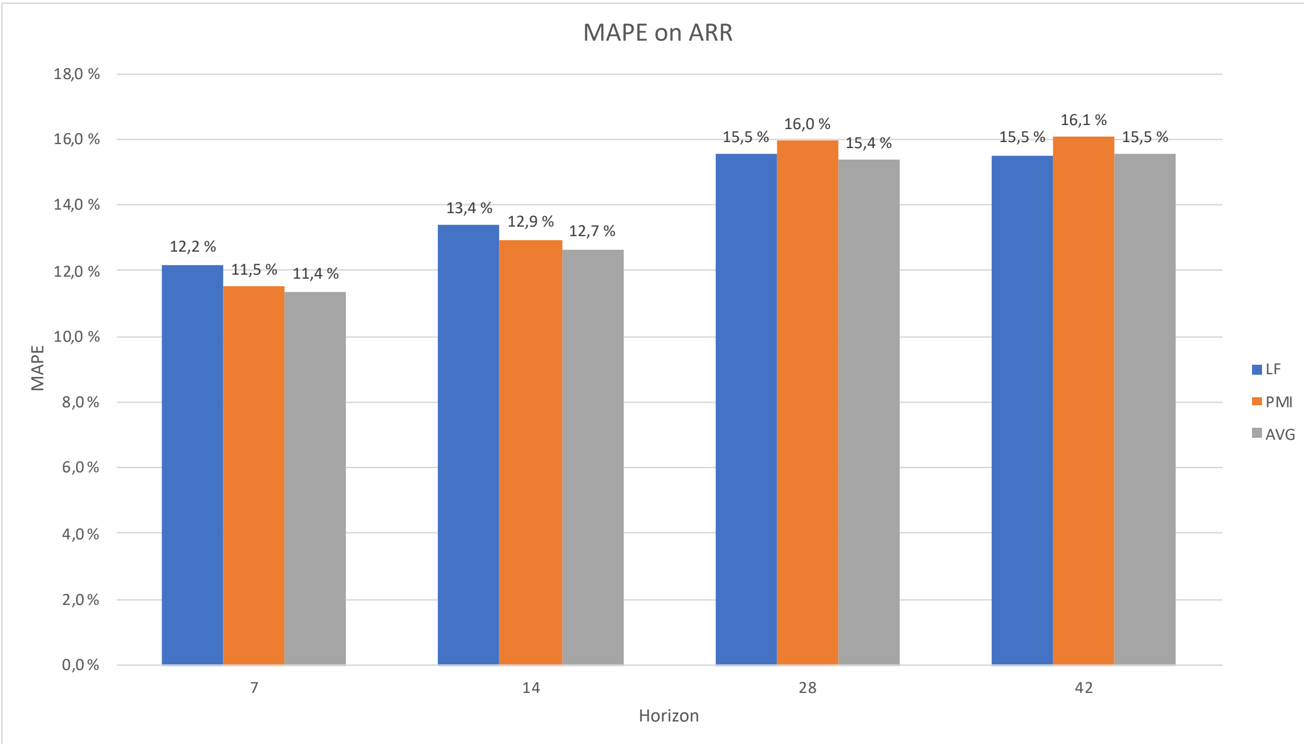

Figure 13: MAPE on ARR for LF, PMI, and SA

5.1.2 Geometric Mean (GEOM)

As well as with the SA method, combining the forecasts with the geometric mean also proved to outperform the LF forecast across all time horizons for RN. As depicted in Figure 14, The GEOM fell short compared to the PMI for room nights.

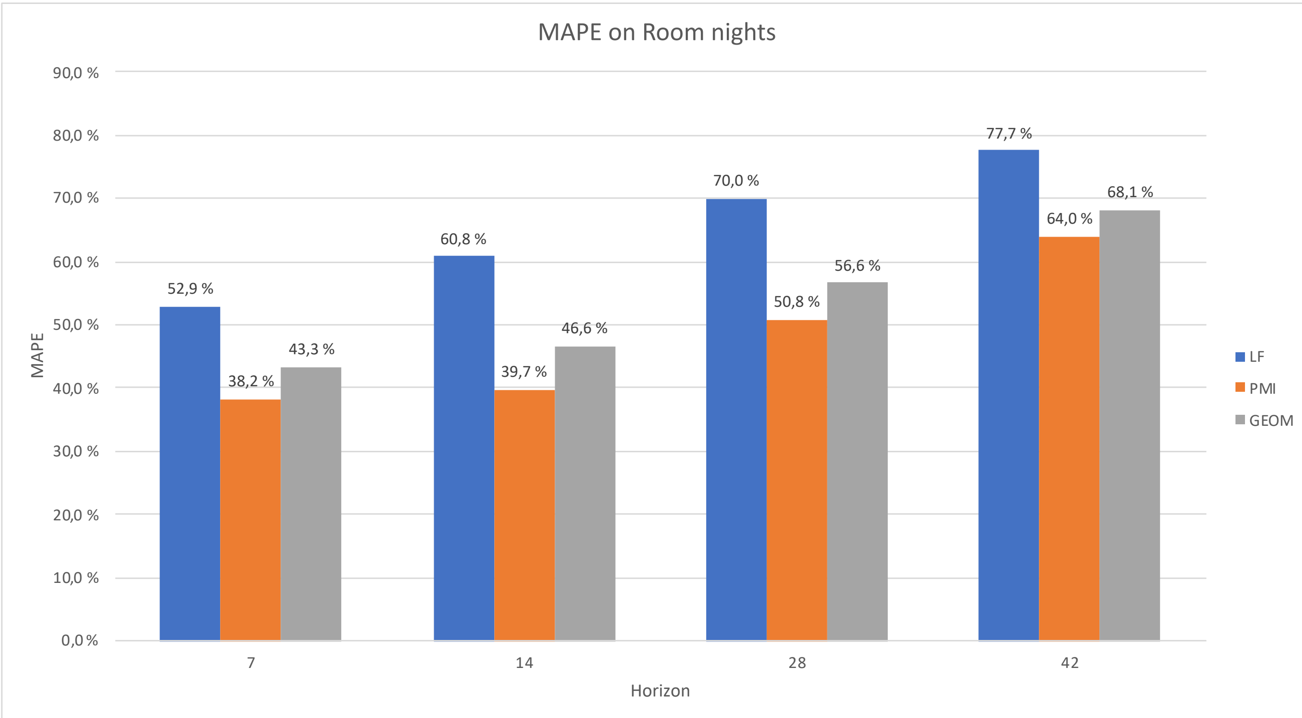

Figure 14: MAPE on RN for LF, PMI, and GEOM

On ARR, however, GEOM outperformed both individual forecasts almost identically to the simple average combination approach according to the MAPE and the MSFE.

5.1.3 Inverse of the MSFE (INVM)

The INVM method outperformed LF on all RN forecasts, with the best MAPE being 41% on a 7-day lead versus LFs 52,9%. The PMI forecasts were not outperformed on RN, which ultimately placed the INVM forecast in-between the individual forecasts according to the MAPE and MSFE. Figure 15 displays how the INVM method outperforms the LF forecast on RN and how close the INVM method is to the PMI on a 7-day lead.

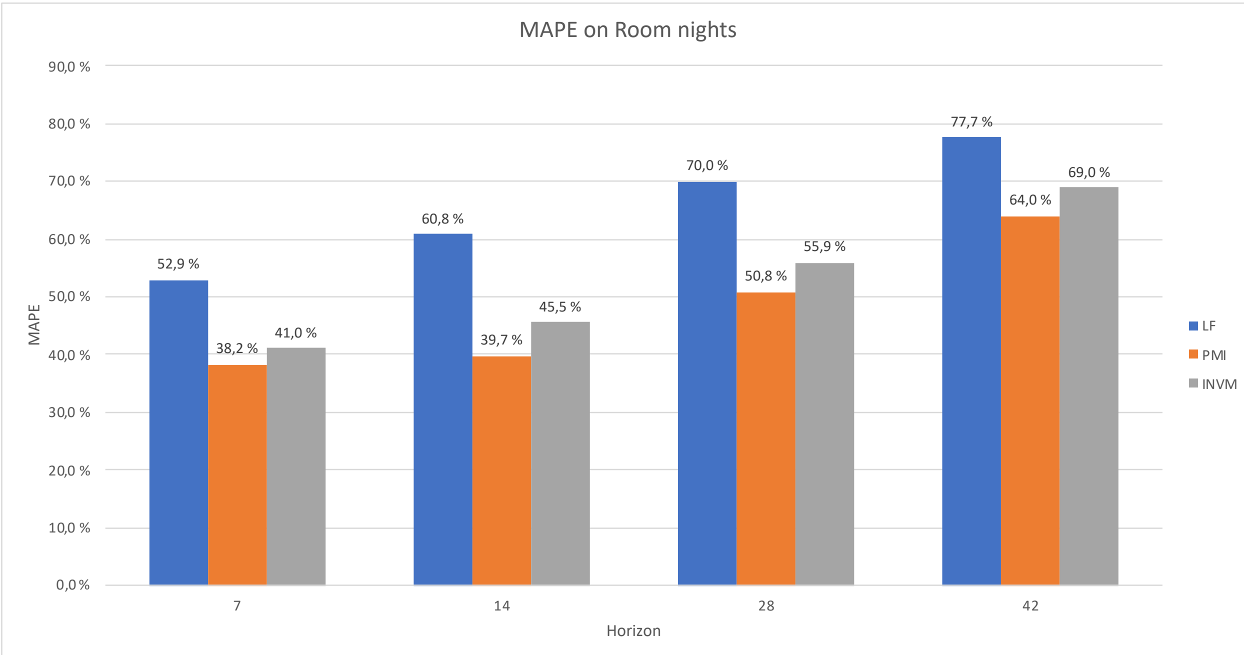

Figure 15: MAPE on RN for LF, PMI, and INVM

The MAPEs of the INVM method on ARR are visualized in Figure 16 and show how much the INVM outperformed the individual forecasts. Equivalently with the average-based combination methods, the INVM method outperformed the LF and the PMI on ARR across all time horizons. This further supports the theory that the ARR are more accurately predicted by combining the PMI and LF instead of using them individually.

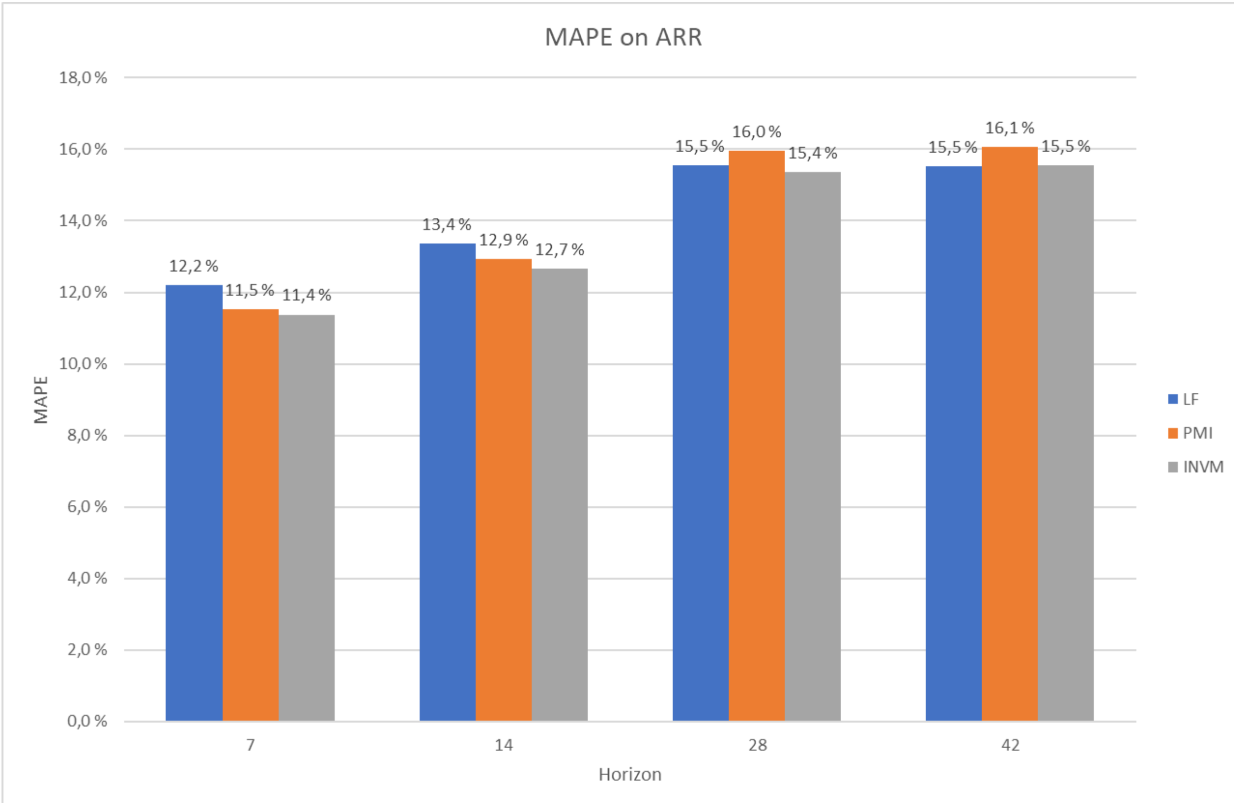

Figure 16: MAPE on ARR for LF, PMI, and INVM

5.1.4 Variance-Covariance method (VACO)

The weighted computation of the LF and PMI projections for the various time horizons was calculated, and the error measurements were performed to demonstrate the accuracy of the combination approach. Positive findings were obtained by combining the different forecasts using the variance-covariance approach. The outcomes of the combinations were compared to the individual forecasts to get a good picture of the total performance. The measurements are depicted in the graphs shown in Figure 17.

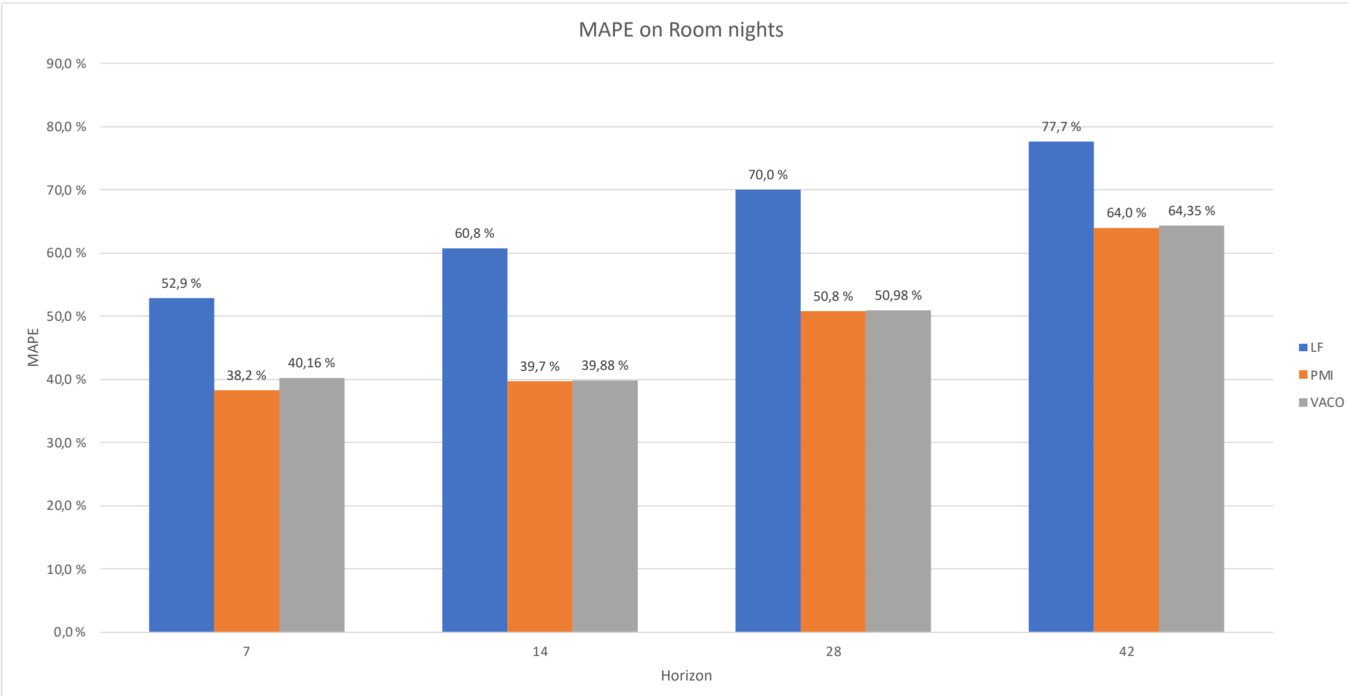

Figure 17: MAPE on RN for LF, PMI, and VACO

Figure 17 shows that the PMI was marginally better than the VACO combination for each lead time. However, LF was significantly poorer than the other two. Table 6 demonstrates that the weighting attributed to the LF room nights projection was relatively low for each lead time. Thus, the VACO approach would be more like the PMI forecast than LF. Because of the modest weight provided to the LF for the combination, the forecast combination performed marginally worse than the PMI forecast. Combining forecasts with the VACO approach for RN projections was ineffective since the PMI outperformed the forecast combination. However, using the VACO method on the ARR metric yielded interesting results. The findings are displayed in Figure 18.

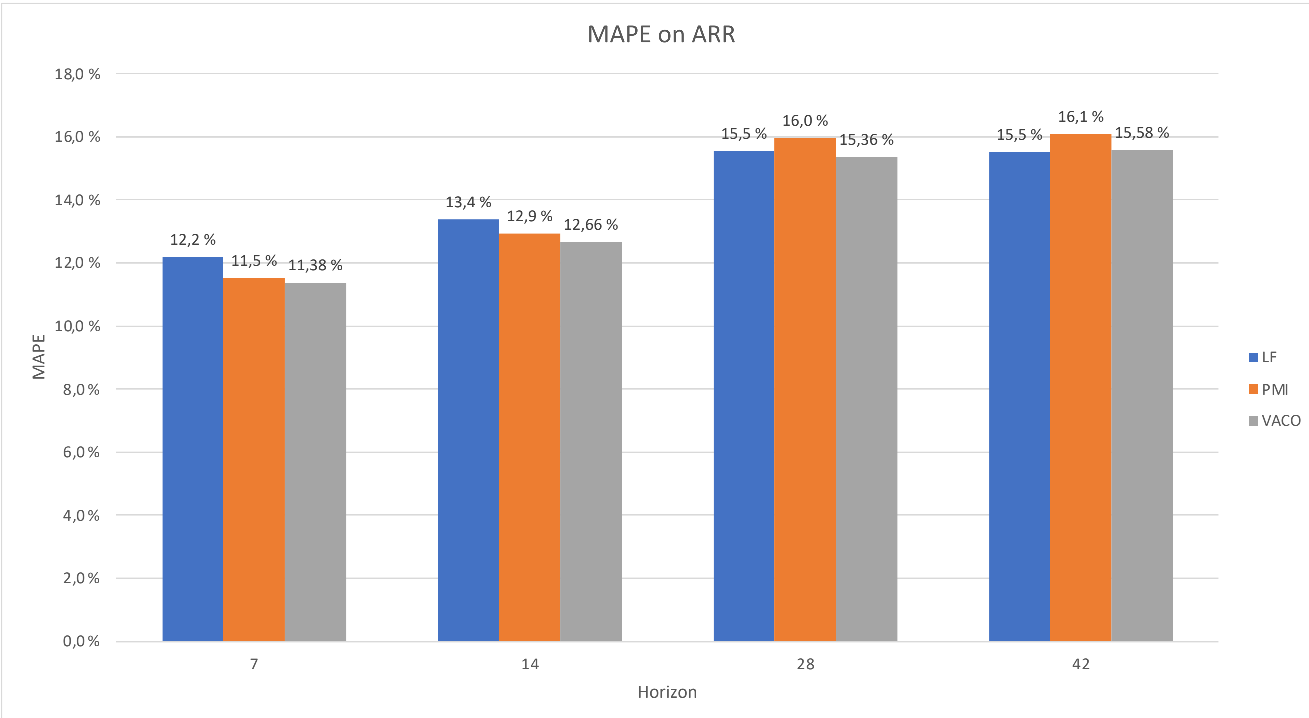

Figure 18: MAPE on ARR for LF, PMI, and VACO.

Figure 18 shows that, in contrast to the RN predictions, the VACO forecast combination for ARR was the most accurate projection. The MAPE reveals that the VACO combination of forecasts surpassed LF and PMI for predictions with lead times of 7, 14, and 28 days. However, for the 42-day prediction, LF surpassed the others. As a result, the VACO method's forecast combination is advantageous for ARR estimates.

5.1.5 Perceptron (PERC)

The perceptron was created to calculate the optimal weights for each forecast combination. By introducing two forecasts into the perceptron and comparing them to the actual value, the algorithm aimed to learn the pattern and eventually present the user with the optimal weights for the forecast combination. The dataset supplied for the study is too small for a neural network. Consequently, the neural network did not get the appropriate quantity of training data. The size of the dataset resulted in lackluster training sizes, ultimately leading to inconsistent conclusions. Generally, the more data utilized as input in a neural network, the more accurate the network is. In contrast, scarce data may lead the neural network's conclusions to deviate significantly. Such anomalies were seen in the perceptron. The perceptron randomized the training data every time the code was ran, as mentioned in section 4.2.5. Since the training data is limited, the network would learn a different pattern each time the code was executed.

Consequently, the perceptron would output different weights each time. An average of one hundred weights for each forecast was used to improve consistency, but the results were still insufficient. A plot of the MAPE for the combined forecasts compared to the MAPE of LF and PMI is provided in Figure 19.

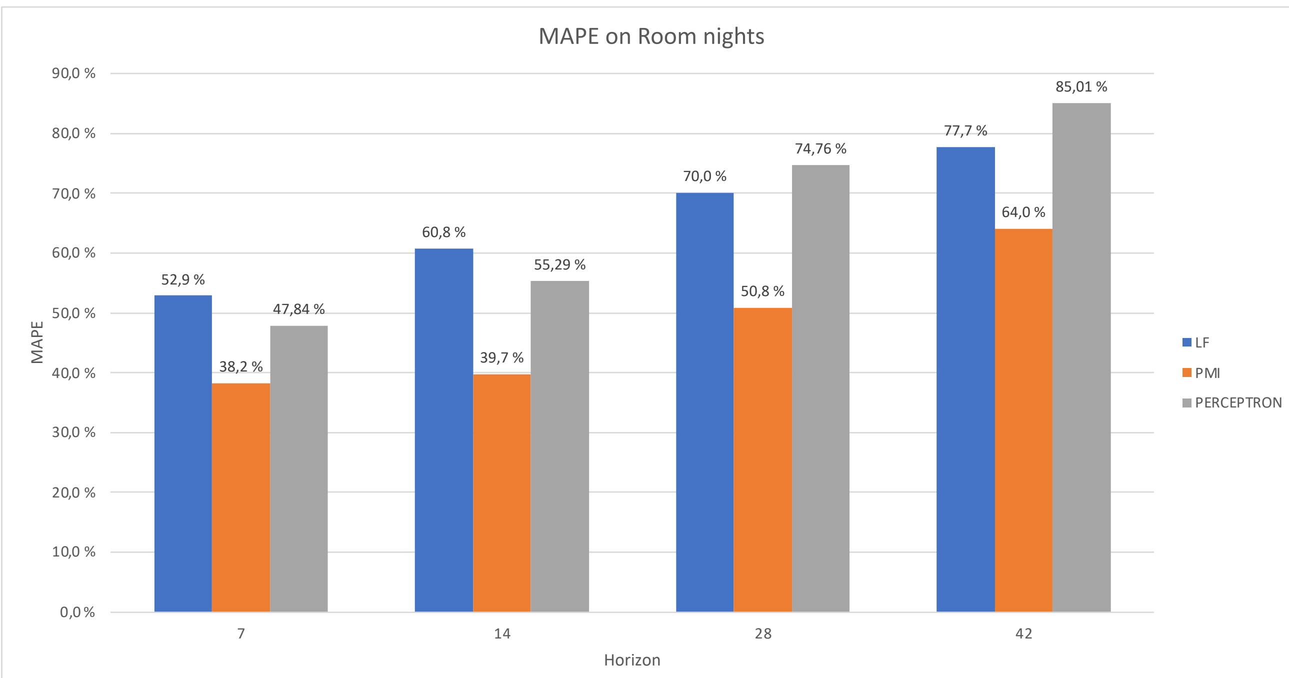

Figure 19: MAPE on RN for LF, PMI, and PERC.

The perceptron showed higher accuracy for the ARR metric. The MAPEs of the forecast combination compared to the MAPEs of the LF and PMI for ARR is displayed in Figure 20.

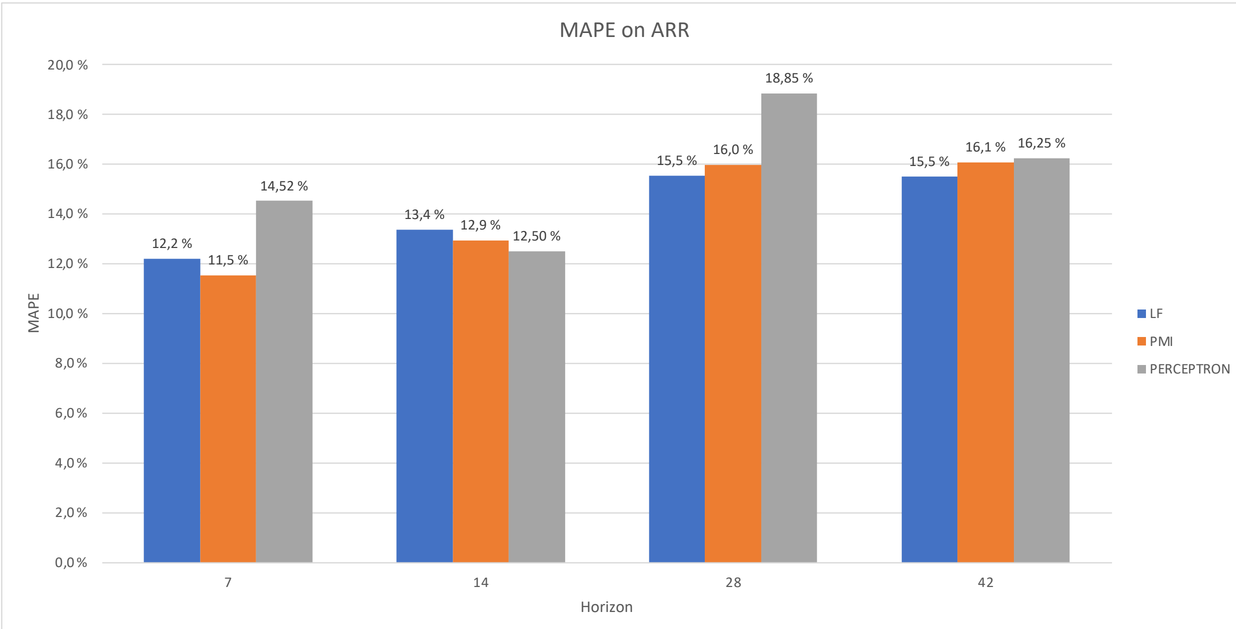

Figure 20: MAPE on ARR for LF, PMI, and PERC.

The fundamental purpose of this approach was whether we could improve prediction accuracy by combining forecasts. It is difficult to respond to this issue based on the graphs shown in Figure 19 and Figure 20. In Figure 19, the PMI beat the combination approach. However, the combination technique outperformed LF during the first two lead periods. The results of the second graph are more ameliorating than those of the first. The PERC combination for the 14-day lead time outperformed the individual forecasts. However, the combination technique performed the worst for other lead times. The discrepancy in the forecast can be attributed to the model's lack of consistent training data. The data is inconsistent in its projection and, more crucially, is too limited. Thus, depending on which part of the dataset is utilized as training data, specific predictions will perform better while others will perform worse than the individual forecast.

Despite not being able to get concrete answers to the research question, the perceptron produced a fascinating result when using the MSFE to measure the accuracy of the forecasts. The results are shown in Figure 21.

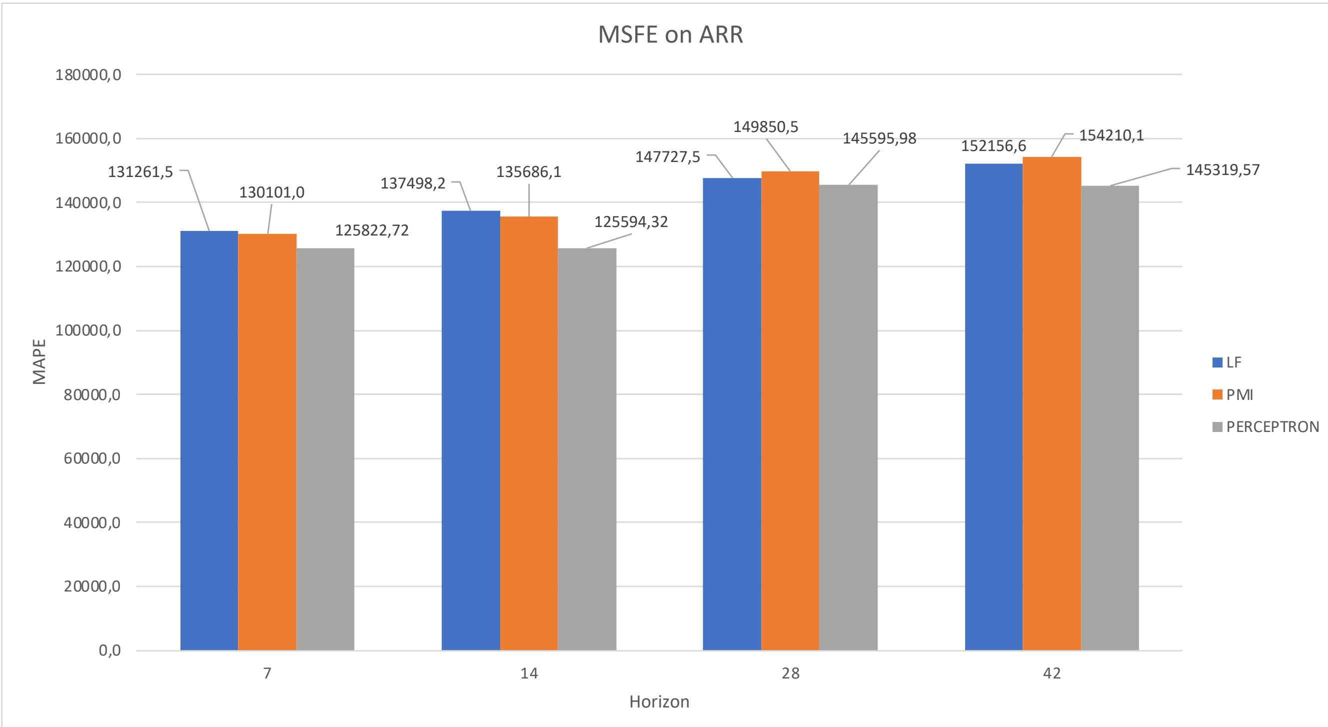

Figure 21: MSFE on ARR for LF, PMI, and PERC.

When utilizing the MSFE to gauge forecast accuracy, the forecast combination beat both the PMI and the LF, as shown in Figure 21. The MSFE squared the predicted error compared to the actual value. Consequently, big mistakes were severely punished by the approach. This indicated that the PERC errors are more consistent and have smaller values than the two single model forecasts. The perceptron employed MSE for determining weights, which explains the outcome.

5.2 Summaries and rankings of forecast performances

Two summary tables and a ranking table were established to overview the forecast performances of all the different forecasts. Table 7 summarizes forecast performances on room nights, and Table 8 does the same for average room revenue. Table 9 is a performance ranking of each forecast for all time horizons according to their MAPE and MSFE results. Furtherly, the results are discussed in section 6.

ROOM NIGHTS

Accuracy measure	Forecasting horizon	Forecasting method						
		LF	PMI	AVG	GEOM	INVM	VACO	PERC
MAPE	7 days	52,9 %	38,2 %	44,3 %	43,3 %	41,0 %	40,2 %	47,8 %
	14 days	60,8 %	39,7 %	48,9 %	46,6 %	45,5 %	39,9 %	55,3 %
	28 days	70,0 %	50,8 %	59,1 %	56,6 %	55,9 %	51,0 %	74,8 %
	42 days	77,7 %	64,0 %	69,9 %	68,1 %	69,0 %	64,4 %	85,0 %
MSFE	7 days	1251,3	459,6	704,4	619,8	553,5	521,0	545,0
	14 days	1763,2	889,6	1163,7	1161,2	1025,2	1034,3	973,2
	28 days	3219,9	1539,0	1938,8	1837,0	1690,1	1714,3	2498,9
	42 days	3121,3	2400,4	2593,0	2543,7	2537,3	2560,9	3602,0

Table 7: Summary of forecast performances on RN

AVERAGE ROOM REVENUE

Accuracy measure	Forecasting horizon	Forecasting method						
		LF	PMI	AVG	GEOM	INVM	VACO	PERC
MAPE	7 days	12,2 %	11,5 %	11,4 %	11,5 %	11,4 %	11,4 %	14,5 %
	14 days	13,4 %	12,9 %	12,7 %	12,8 %	12,7 %	12,7 %	12,5 %
	28 days	15,5 %	16,0 %	15,4 %	15,4 %	15,4 %	15,4 %	18,8 %
	42 days	15,5 %	16,1 %	15,5 %	15,6 %	15,5 %	15,6 %	16,2 %
MSFE	7 days	131261,5	130101,0	128693,3	128825,7	128690,7	128703,4	125822,7
	14 days	137498,2	135686,1	133924,4	135466,4	133909,9	133966,8	125594,3
	28 days	147727,5	149850,5	147222,0	147538,3	146946,6	147223,9	145596,0
	42 days	152156,6	154210,1	151907,2	152266,0	151671,8	152037,7	145319,6

Table 8: Summary of forecast performances on ARR

RANKING

	Room Nights				ARR			
Rank	Forecast	MAPE	Forecast	MSFE	Forecast	MAPE	Forecast	MSFE
1	PMI7	38,2 %	PMI7	459,6	INVM7	11,379 %	PERC14	125594,3
2	PMI14	39,7 %	VACO7	521,0	AVG7	11,380 %	PERC7	125822,7
3	VACO14	39,9 %	PERC7	545,0	VACO7	11,385 %	INVM7	128690,7
4	VACO7	40,2 %	INVM7	553,5	GEOM7	11,480 %	AVG7	128693,3
5	INVM7	41,0 %	GEOM7	619,8	PMI7	11,524 %	VACO7	128703,4
6	GEOM7	43,3 %	AVG7	704,4	LF7	12,193 %	GEOM7	128825,7
7	AVG7	44,3 %	PMI14	889,6	PERC14	12,496 %	PMI7	130101,0
8	INVM14	45,5 %	PERC14	973,2	INVM14	12,651 %	LF7	131261,5
9	GEOM14	46,6 %	INVM14	1025,2	AVG14	12,652 %	INVM14	133909,9
10	PERC7	47,8 %	VACO14	1034,3	VACO14	12,665 %	AVG14	133924,4
11	AVG14	48,9 %	GEOM14	1161,2	GEOM14	12,790 %	VACO14	133966,8
12	PMI28	50,8 %	AVG14	1163,7	PMI14	12,937 %	GEOM14	135466,4
13	VACO28	51,0 %	LF7	1251,3	LF14	13,370 %	PMI14	135686,1
14	LF7	52,9 %	PMI28	1539,0	PERC7	14,519 %	LF14	137498,2
15	PERC14	55,3 %	INVM28	1690,1	INVM28	15,353 %	PERC42	145319,6
16	INVM28	55,9 %	VACO28	1714,3	AVG28	15,355 %	PERC28	145596,0
17	GEOM28	56,6 %	LF14	1763,2	VACO28	15,356 %	INVM28	146946,6
18	AVG28	59,1 %	GEOM28	1837,0	GEOM28	15,400 %	AVG28	147222,0
19	LF14	60,8 %	AVG28	1938,8	LF42	15,513 %	VACO28	147223,9
20	PMI42	64,0 %	PMI42	2400,4	INVM42	15,540 %	GEOM28	147538,3
21	VACO42	64,4 %	PERC28	2498,9	AVG42	15,542 %	LF 28	147727,5
22	GEOM42	68,1 %	INVM42	2537,3	LF 28	15,547 %	PMI28	149850,5
23	INVM42	69,0 %	GEOM42	2543,7	VACO42	15,577 %	INVM42	151671,8
24	AVG42	69,9 %	VACO42	2560,9	GEOM42	15,590 %	AVG42	151907,2
25	LF 28	70,0 %	AVG42	2593,0	PMI28	15,960 %	VACO42	152037,7
26	PERC28	74,8 %	LF42	3121,3	PMI42	16,071 %	LF42	152156,6
27	LF42	77,7 %	LF 28	3219,9	PERC42	16,247 %	GEOM42	152266,0
28	PERC42	85,0 %	PERC42	3602,0	PERC28	18,845 %	PMI42	154210,1

Table 9: Forecast rankings based on MAPE and MSFE measures

6 Discussion

Results from the present study indicate that the combination of forecasts can outperform single individual forecasts on forecast accuracy, especially for the ARR metric. Table 9 shows that according to the MAPE and MSFE measures, every forecast combination method performed outperforms the individual forecasts of LF and PMI on ARR prediction. The performance evaluation of the PMI and LF forecasts shows that PMI is the most accurate one, with a MAPE of 11,52%. Four of five combined forecasts beat this accuracy, with INVM being the most accurate one, yielding an 11,37% MAPE. All combined forecasts outperform the individual forecasts on MSFE, with two PERC forecasts being the most accurate.

Despite the ARR results strongly favoring forecast combinations over individual forecasts, the RN results are not as convincing. Between the two individual forecasts, it was found that the PMI has the lowest MAPE of 38,2% and the lowest MSFE. Table 9 shows that neither was beaten by any of the five combination forecasts. The closest one is the one of VACO, with a 39,9% MAPE. However, all five combination forecasts beat LF on all time horizons. Consequently, the combined forecasts fall between the two individual forecasts on performance as they only manage to beat one of the individual forecasts. The ultimate goal was to achieve better accuracy than both individual forecasts, which was not achieved on RN.

A question that arises is why forecast combinations are favorable on ARR prediction and not on RN. According to Bates and Granger (1969), a combined forecast will offer a lower forecast error when one of the forecasts has information that will help the other forecast produce a better prediction. Regarding ARR prediction, the performance of the PMI and LF forecasts was relatively good and quite similar. However, for RN, the overall performance is poor, and the accuracy of LF and PMI are somewhat dissimilar. Consequently, combining LF and PMI for RN prediction did not show good results, providing robustness to the findings of Kurvers *et al.* (2016). They found that collective intelligence only applies when there is similarity in prediction accuracy. Our findings show no benefit for the PMI forecast to utilize any of the information LF provides. These results support the study of Bates and Granger (1969), as the LF forecast is too inadequate to have a beneficial impact on RN predictions. Conversely, for ARR, the forecast combination is suitable. Ultimately, our results show that forecast combination has limitations when producing and achieving better forecasts.

Five combination methods were conducted and tested against the individual LF and PMI. The rankings show that overall, the VACO and the INVM methods performed the best, closely followed by the simple average methods. Some methods perform consistently at predicting both metrics, while others only show strong performance on one target.

Makridakis and Hibon (1979) demonstrated that simple forecasting models beat sophisticated ones when there was significant unpredictability in the data. In the present study, the more sophisticated combination methods, such as the VACO and INVM, have the best performance. These methods take the historical performance of the individual forecasts into account and outperform the simple methods, contradicting the findings in the studies of Makridakis and Hibon (1979). However, the results also side with the mentioned study when it comes to the method of the perceptron. The use of the perceptron is by far the most advanced combination method in this study and is heavily outperformed by the simpler models on four out of five rankings.

Even though the perceptron did not provide the desired consistency, the data, not the perceptron, constrained the model. The model is functional in principle, and the desired consistency should be achieved given the appropriate quantity of data. However, there is significant ambiguity about the model's capacity to achieve consistency. In this study, the model has only been tested on a small data set, making it hard to anticipate the outcome when large data sets are used.

If more extensive datasets are unavailable, one solution may be to combine all the predictions into a single list before running the perceptron. In this case, the input data will be two lists, one including all the LF forecasts (LF7, LF14, LF28, and LF42) and the other containing all the PMI forecasts. The training data will be quadrupled in this manner. The main issue with this strategy is that the perceptron will weigh each prediction equally while training. By evenly weighting the forecasts, the perceptron will alter the same amount for a 7-day prediction and a 42-day prediction. Thus, the findings may not give helpful information.

According to the MSFE results on ARR, the perceptron has the least amount of error across all five combination methods. This is because the perceptron calculates optimal weights using the mean squared error. It is interesting to note that it is possible to develop the desired error measurement in neural networking. Consequently, creating a neural network model that utilizes MAPE as the error measurement may lower the mean absolute percentage error.

Only two error measurements are used to get the results of this study. There are, however, a plethora of other error metrics to pick from, each of which may provide different findings. Yüksel (2008) emphasizes the need for accurate forecasts in good hotel business. Consequently, one of the essential tasks for a revenue management team is determining which error metric to use when assessing prediction accuracy.

In the present study, the different error measurements result in different findings. Stock (2020) states that a series of modest forecast errors typically cause only minor problems for the user, but a single significant forecast error might doubt the entire forecasting process. Stock (2020) furtherly discusses that MSFE encapsulates the principle by computing the square of the forecast error, meaning significant errors are punished more than minor ones. Consequently, for a revenue management team in the hotel sector, choosing the mean squared error as the error measurement may result in more beneficial forecasting.

7 Conclusions

This study investigated hotel forecasts and their accuracy in relation to the actual value. We used data provided by d20, a company that provides real-time performance management tools. The data predicted two different metrics: room nights and average room revenue. Each of the metrics had two different forecasts over four different lead times. We used five methods to aggregate the projections; these were: The simple average (SA), the geometric mean method (GEOM), the inverse of the mean squared forecast error method (INVM), the variance-covariance method (VACO), and ultimately neural networking. The main goal of this study was to unravel the following question: *Does combining two different demand predictions result in more accurate demand projections?*

7.1 Main findings

The statistical comparison of the forecasting accuracy of the combination and single-model predictions was a fundamental contribution of this thesis. By comparing forecasts, we demonstrated that forecast combination did not constantly improve forecasting accuracy. Only the combined forecasts for one of the metrics, average room revenue, outperformed the individual forecasts. Our findings were consistent with earlier forecast combination research, which state that the individual predictions included in a forecast combination must contain information that benefits the other forecast (Bates & Granger, 1969). This was not applied for room nights because the LF for all lead periods was relatively poor. In addition, our findings showed that all combination predictions beat the worst single model forecast, implying that forecast combination reduces the chance of complete forecasting failure.

We demonstrated combined projections to be substantially more accurate than the worst single model forecasts across all forecasting horizons and combination strategies. However, our forecast combination produced the best average room revenue forecasting results. Our findings imply that tourism practitioners should use forecast combinations to improve forecasting accuracy. This is especially true when the two single model forecasts are accurate.

7.2 Future research

Because our findings are limited to the 32 hotels from which data were collected, our conclusions are far from being definitive. Consequently, additional research with larger datasets is required before concluding that forecast combination produces consistent outcomes. According to our findings, sophisticated methods outperform more simple combination methods. Therefore, it is critical to look at this topic's issues. Accordingly, several concerns need to be investigated further: (1) why do the sophisticated methods work so well, and (2) under what situations do more simple approaches perform better?

References

Andrawis, R., Atiya, A., & El-Shishiny, H. (2010). Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting 27*, pp. 870-886.

Archer, B. (1987). Demand forecasting and estimation. In B. Archer, *Travel, tourism, and hospitality research. A handbook for managers and researchers* (pp. 77-85). Wiley.

Arsmtrong, J. (2001). *PRINCIPLES OF FORECASTING: A Handbook for Researchers and Practitioners.* Kluwer Academic Publishers.

Bates, J., & Granger, C. (1969). The combination of forecasts. *Operational research society*, pp. 451-468.

Benavides-Velasco, C., Quintana-Garcia, C., & Marchanta-Lara, M. (2014). Total quality management, corporate social responsibility and performance in the hotel industry. *International Journal of Hospitality Management*, pp. 77-87.

Bringsjord, S. (2018). *Neural Nets*. Retrieved from Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/entries/artificial-intelligence/neural-nets.html

Brownlee, J. (2019). *Understand the Impact of Learning Rate on Neural Network Performance*. Retrieved from Machine learning mastery: https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/

Chen, M., & Chuang, C. (2000). An extended newsboy problem with shortage-level constraints. *International Journal of Production Economics*, pp. 269-277.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of forecasting*, pp. 559-583.

Crane, D., & Crotty, J. (1967). A Two-Stage Forecasting Model: Exponential Smoothing and Multiple Regression. *Management Science*, pp. 5017-507.

Cross, R. C. (2011). Milestones in the application of analytical pricing and revenue management. *Journal of revenue and pricing management*, pp. 8-18.

Cross, R. G. (1997). How revenue management can work for your business. *The Cornell hotel and restaurant administration quarterly*, pp. 32-43.

d2o. (2020). *About d2o*. Retrieved from d2o: https://www.d2o.com/company/about-d2o/

De Treville, S. (2014). Valuing lead time. *Journal of operations management*, pp. 337-346.

Galton, F. (1907). Vox Populi (The wisdom of crowds). *Nature*, pp. 450-451.

Halton, C. (2021). *Wisdom of crowds*. Retrieved from Investopedia: https://www.investopedia.com/terms/w/wisdom-crowds.asp

J.Hydman, R., & Athanasopoulos, G. (2018). *Forecasting: Principle and practice 2nd edition.* OTexts.

Koupriouchina, L., van der Rest, J.-P., & Schwartz, Z. (2014). On revenue management and the use of occupancy forecasting error measures. *International journal of hospitality management* , pp. 104-114.

Kun, M. (2021). *What is Hotel Forecasting - Models and Best Practices*. Retrieved from Hotelminder: https://www.hotelminder.com/what-is-hotel-forecasting-models-and-best-practices

Kurvers, R. H., Herzog, S. M., Hertwig, R., Krause, J., Carney, P. A., Bogart, A., . . . Wolf, M. (2016). Boosting medical diagnostics by pooling independent judgments. *Proceedings of the National Academy of Sciences*, pp. 8777-8782.

Makridakis, S., & Hibon, M. (1979). The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition. *Journal of the Royal Statistical Society*, pp. 97-145.

Makridakis, S., Andersen, A., Fildes, R., Hibon, M., Lewandowski, R., Parzen, E., . . . Winkler, R. (1982). The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition. *Journal of Forecasting*, pp. 111-153.

Mancuso, A., & Werner, L. (2013). Review of combining forecasts approaches. *Independent journal of management & production*, pp. 248-277.

Maskivker, J. (2013). Review: The Wisdom of Crowds. *The Review of Politics, Vol. 75, No.4*, pp. 702-705.

Petropoulos, F., Daniele, A., Vassiolos, A., Mohammed, B., Devon, B., & Souhaib, T. (2022). Forecasting: Theory and practice. *International Journal of forecasting*, pp. 705-871.

Reid, D. J. (1968). Combining three estimates of gross domestic product. *Economica*, pp. 431-444.

Reid, D. J. (1969). *A comparative study of time series prediction techniques on economic data.*

Rheude, J. (2020). *Demand forecasting: Types, methods, and examples*. Retrieved from Red stag fulfillment: redstagfulfillment.com/what-is-demand-forecasting/

Song, H., Witt, S. F., Wong, K. F., & Wu, D. C. (2009). An empirical study of forecast combination in tourism. *Journal of hospitality & tourism research*, pp. 3-29.

Stock J., W. M. (1999). A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. *Oxford University Press*, pp. 1-44. Retrieved from https://scholar.harvard.edu/stock/publications/comparison-linear-and-nonlinear-univariate-models-forecasting-macroeconomic-time

Stock, &. W. (2020). *Introduction to Econometrics (Fourth edition.; Global edition.).* Pearson.

Surowiecki, J. (2005). *The Wisdom of Crowds: Why the many are smarter than the few.* Doubleday; Anchor.

Vandeput, N. (2019). *Forecast KPIs: RMSE, MAE, MAPE & Bias*. Retrieved from https://towardsdatascience.com/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d

Yüksel, S. (2007). An integrated forecasting approach to hotel demand. *Mathematical and computer modelling*, pp. 1063-1070.