LAPPEENRANTA-LAHTI UNIVERSITY OF TECHNOLOGY LUT
School of Business and Management
Business Administration

Nisse Nurmi

FORECASTING AIRPORT PASSENGER TRAFFIC IN THE ERA OF COVID-19 PANDEMIC

Examiners: Professor Satu Pätäri
Professor Kaisu Puumalainen

ABSTRACT

Lappeenranta-Lahti University of Technology LUT
School of Business and Management
Degree programme in Accounting

Nisse Nurmi

Forecasting airport passenger traffic in the era of COVID-19 pandemic

Master's thesis
2021
89 pages, 28 figures, 9 tables, 3 appendices
Examiners: Professor Satu Pätäri and Professor Kaisu Puumalainen

Keywords: passenger forecasting, airport, COVID-19, time series

Air passenger forecasting is a critical activity in determining future financial performance, optimizing operational activities, and assessing future infrastructure needs of an airport. The vital part of business management is endangered by the coronavirus pandemic, which has caused an unprecedented fall in global air travel demand and created a shadow of uncertainty over the aviation industry for years to come. The level of uncertainty caused by the pandemic and varying government policy responses to fight against it, such as international travel controls, have significantly weakened the ability to forecast future passenger volumes at airports. Although plenty of research has been conducted on air passenger traffic forecasting, also in the context of airports, the predicting power of forecasting methods during such an exogenous shock as coronavirus pandemic has yet remained unexplored.

The thesis aims to fill this gap by approaching the problem by comparing five different forecasting methods (ARIMA, TBATS, Prophet, multiplayer perceptron, extreme learning machine) before and during the pandemic and assess the relevance of refining them by pandemic-related exogenous variables. In addition to evaluating the performance of forecasting methods during the global financial and health crisis, the thesis also sheds light on the current status of research by conducting a systematic literature review on airport passenger traffic forecasting, which has received increased attention from academia lately.

While the research findings showed promising results regarding forecasting accuracy before the crisis, the results expectedly provided less promising results during the pandemic. Forecasting accuracies were slightly improved when a pandemic-related variable reflecting European international travel controls was included in the models. However, in general, the quantitative methods included in this research showed weak performance even when the pandemic-related variable, which highly correlated with Helsinki Airport passenger development, was introduced to the data. Despite the results, the topic is worth to be explored further. Since this thesis focused on comparing the forecasting tools using their default settings, a more profound analysis could be conducted focusing on the extended possibilities of new automated forecasting tools included in free statistical programs.

TIIVISTELMÄ

Lappeenrannan-Lahden teknillinen yliopisto LUT
School of Business and Management
Laskentatoimi

Nisse Nurmi

Lentomatkustajamäärien ennustaminen COVID-19 -pandemian aikana

Pro gradu -tutkielma
2021
89 sivua, 28 kuviota, 9 taulukkoa, 3 liitettä
Tarkastajat: Professori Satu Pätäri ja Professori Kaisu Puumalainen

Hakusanat: lentomatkustajat, ennustaminen, lentoasema, COVID-19, aikasarjat

Lentomatkustajamäärien ennustaminen on kriittinen tekijä lentoasemille, joiden tavoitteena on ennustaa niiden taloudellista suorituskykyä, optimoida operatiivista toimintaa, ja arvioida tulevaisuuden kasvutarpeita. Tämä tärkeä aktiviteetti on vaarantunut vuonna 2020 julistetun koronaviruspandemian seurauksena, mikä on ennen näkemättömällä tavalla vähentänyt lentoliikenteen maailmanlaajuista kysyntää ja tuonut mukanaan pitkäkestoisen epävarmuuden ilmailualan ylle. Epävarmuus pandemian kehittymisestä sekä valtioiden vaihtelevat toimet pandemian vastaisessa taistelussa, kuten kansainväliset matkustusrajoitukset, ovat merkittävästi heikentäneet lentoasemien mahdollisuuksia ennustaa tulevaisuuden matkustajamääriä. Vaikka lentomatkustajamäärien ennustamisesta on tehty lentoasemakontekstissakin paljon tutkimusta, ei ennustemenetelmien soveltuvuutta koronakriisin kaltaisessa tilanteessa ole aikaisemmin riittävästi tutkittu.

Tämä pro gradu -tutkielma lähestyy tätä tutkimusaukkoa vertailemalla viittä eri ennustemenetelmää (ARIMA-mallit, TBATS-malli, Prophet-algoritmi, monikerroksinen perseptroniverkko MLP, äärimmäinen oppimiskone ELM) toisiinsa ennen pandemiaa ja sen aikana. Vertailun lisäksi tutkielmassa arvioidaan pandemia-aiheisten muuttujien vaikutusta valittujen menetelmien ennustetarkkuuteen. Empiirisen menetelmävertailun ohella tutkielmassa valotetaan lentomatkustajamäärien ennustamiseen liittyvän tieteellisen tutkimuksen nykytilaa lentoasemakontekstissa toteuttamalla systemaattinen kirjallisuuskatsaus aiheesta, joka on viime vuosina lisännyt suosiotaan akateemisessa maailmassa.

Tutkimustulosten näyttäessä lupaavia tuloksia ennustetarkkuudella mitattuna ennen kriisiä, tulokset odotetusti osoittivat heikompia tuloksia pandemia-ajan ennustamisessa. Ennustetarkkuus parani hieman, kun Euroopan kansainvälisiä matkustusrajoituksia kuvaava pandemia-aiheinen muuttuja sisällytettiin malleihin. Yleisesti voidaan kuitenkin todeta, että tutkielmaan valitut kvantitatiiviset menetelmät osoittavat heikkoa ennustetarkkuutta pandemian aikana, vaikka vahvasti Helsinki-Vantaan lentoaseman matkustajamäärien kanssa korreloiva pandemia-aiheinen muuttuja sisällytettiinkin malleihin. Tuloksista huolimatta tutkielma antaa aihetta jatkotutkimukselle, jossa ennustemenetelmiä ja ilmaisten tilasto-ohjelmien mahdollistamia automaattisia ennustamistyökaluja voitaisiin yksittäin analysoida tätä vertailevaa tutkielmaa syvällisemmin.

Forewords

The pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is currently emerging and rapidly evolving. The thesis reflects the situation of the pandemic at the end of 2020. We already know how steep the fall in passenger volumes was, but no one can predict the length of the recovery or future travel patterns. The unprecedented level of uncertainty, on the other hand, has provided this fascinating opportunity to examine air passenger demand forecasting during a crisis the industry has never experienced before. It is a significant opportunity to take since the pandemic will most certainly not be the last one. The saying of a 6th century BC Chinese philosopher Lao Tzu provides a solid justification for the relevance of choosing the topic from the field of predictive analytics during these uncertain times:

“Those who have knowledge, don't predict. Those who predict, don't have knowledge.”

-Lao Tzu

We do not have knowledge.

Acknowledgments

In 2019, when I left the booming aviation business to focus on my master's studies in Lappeenranta, I could not even imagine a pandemic and the condition in which it has left the aviation industry. Now when I am about the step back onboard, I am thankful for all the insights and tools LUT has provided me during my last nearly two years here. First and foremost, I want to thank you, Satu and Kaisu, for your invaluable guidance that improved the quality of this thesis.

I never thought the student experience, although an untypical one, could have had such a significant impact on my life. I am thankful for all the life-long friendships I made. Thank you, new and old friends, for providing balance to an otherwise busy life with studies. I hope there will be more time together now as my studies are over.

Gaining knowledge is a path-dependent process. Therefore, I wish to thank my former educational institute, JAMK University of Applied Sciences, for providing me with a solid platform to succeed in my master's studies at LUT. It is also evident how my professional background in the aviation industry has supported my academic journey and writing this thesis. Thus, I want to take this moment to express my deepest gratitude to my employer Finavia and all those colleagues who have supported my academic and professional development.

Eljas, thank you for always being there for me. Your unparalleled wisdom in life and science has had a profound impact on my life and this thesis. You are inspirational. Veera, thank you for inspiring me to become better at writing. Although I did not become a novelist, which I dreamed about as a kid, I have always been inspired by how you write. The rest of the Salmi family, Sami, Severi, and Frida, thank you for the fun and sometimes spontaneous trips to Lapland and elsewhere during my studies. They were much-needed breaks during which I could not even think about my studies.

Finally, Äiti, Isä – this thesis is dedicated to you. Thank you for your endless support and love.

Lappeenranta, 17.4.2021

Nisse Nurmi

TABLE OF CONTENTS

FIGURES

TABLES

1. Introduction

1.1 Background

Traffic forecasting is a crucial activity in airport business management. Whether the forecasts are produced for strategic planning, monthly revenue or cost projections, project appraisals, or daily service level optimizations, the common aim is to estimate future outcomes as accurately as possible for sound decision-making. Passenger volumes generate a significant share of airport revenues through air traffic fees and retail activities (Twinn, Qureshi, Rojas & Conde 2020, 2). Thus, air passenger demand is the key driver for airports' financial performance, making passenger volume forecasting an essential activity for practitioners.

The coronavirus pandemic, announced in 2020, has caused unprecedented market shock in the aviation industry, significantly reducing demand for global air travel (see figure 1) and creating one more dimension to consider in forecasts. For comparison, in 2002, global air passenger traffic was down by twelve percent (year-on-year, YoY) six months after the 9/11 terrorist attacks (Gerrish & Baggaley 2020). In 2003, following the outbreak of SARS epidemic, Asia-Pacific airlines lost approximately 35 percent of monthly passengers, but the whole year ended up only eight percent down[1], which illustrates a rapid recovery of just nine months (IATA 2020c). Globally, passenger traffic grew by 2,3 percent compared to 2002 (The World Bank 2020). A few years later, in the wake of the financial crisis, global air passenger traffic growth ground to a halt for two years, 2008-2009.

This time the magnitude of the shock is different: nine months after the outbreak in September 2020, analysts expected 2020 global air passenger traffic to remain 60-70 percent down from the 2019 levels (Gerrish & Baggaley 2020). In February 2021, IATA (2021) released an unprecedented number: international passenger demand fell by 75,6 percent[2] compared to 2019 levels. The link between the number of air passengers traveling through an airport and revenues of the airport operator is evident: The crisis was, already in May 2020, expected to cut airport operators' 2020 revenues by more than half compared to the levels prior to the pandemic (Twinn 2020, 3). This expectation became realized as Finavia Corporation, the Finnish airport operator, reported a 61,3 percent fall in its revenues compared to

[1] measured in revenue passenger kilometers (RPK), number of revenue passengers x total distance traveled
[2] RPK

2019, amounting to 238,6 million euros[3] loss (Finavia 2021). The fall in passenger numbers was 75,4 percent compared to 2019, when 26 million passengers traveled through 21 Finavia-operated airports.

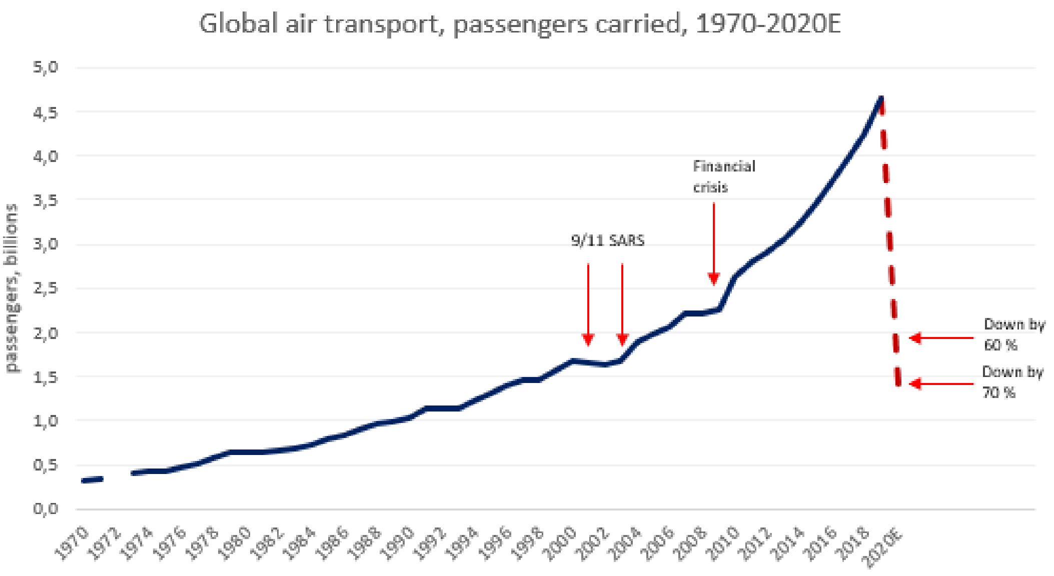

Figure 1. Development of global air passenger traffic (data obtained from ICAO 2020; The World Bank 2020)

Furthermore, industry advocates (ACI 2020; IATA 2020b) look forward to an abnormally slow recovery: the passenger traffic is not expected to surpass 2019 levels earlier than at least 2024. Thus, for the first time in aviation history, the industry faces such a steep fall in demand, together with a recovery line that does not seem to follow the typical V-shape. Instead of a typical rebound, the industry expects a long "swoosh-shaped" gradual recovery, referring to the shape of the world-famous Nike logo (Yokota et al. 2020). Even more gradual, an L-shaped flatline recovery has been suggested by, for example, Gallego & Font (2020), who estimated future passenger demand by using searches and picks on a flight ticket search engine. In economic terms, L-shaped recovery indicates depression, and in the aviation context, it indicates a possibility of aviation not returning to trend line growth (ICAO 2020). The present uncertainties have made traffic forecasting a complex challenge without off-the-shelf solutions being available.

---

[3] 2019 revenue: 150,6 millions, 2020 revenue 150,6 millions

1.2 Purpose and significance of the research

Managerial significance

Air passenger traffic forecasting is vital for airports since the most significant share of revenues is generated through passenger volumes (Twinn et al. 2020). Following the outbreak, traffic forecasting has been demanding, which is partly claimed to be a result of continuously evolving policy responses, such as international travel restrictions set by governments (Gerrish & Baggaley 2020). In April’s survey, IATA (2020a) discovered that 86 percent of travelers were concerned about quarantine requirements, and 81 percent would not consider traveling if it involved a 14-day quarantine upon arrival. Figure 2 illustrates the development and stringency of global travel controls from January to August 2020.

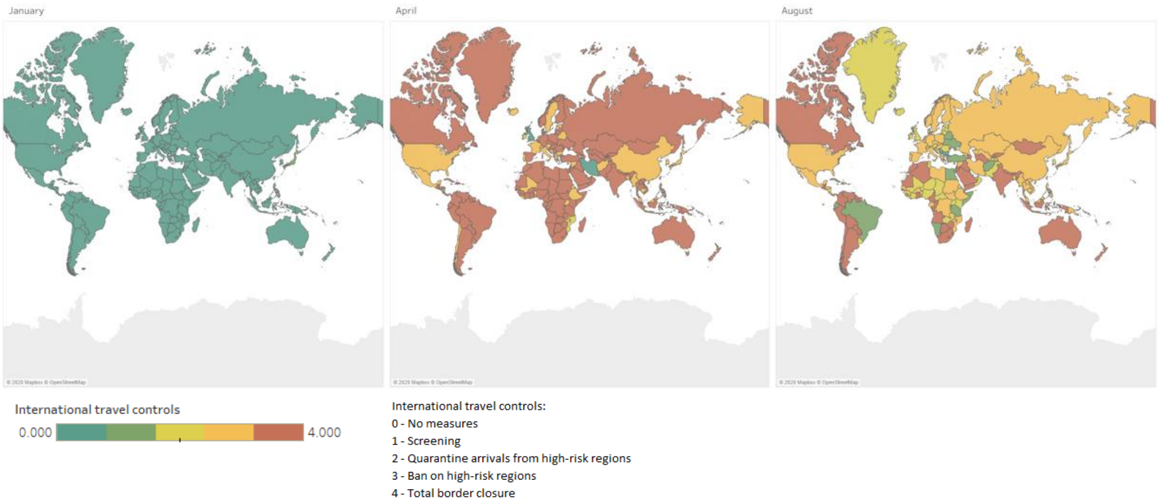

Figure 2. Global travel restrictions from January to August 2020, monthly averages for stringency are used (Data obtained from OurWorldInData.org)

In Finland, quantitative measures guiding decisions on government travel restrictions have been enforced since June 2020, first with the 14-day COVID-19 incidence limit being eight and then, as of September 19th, 25 cases per 100 000 inhabitants (Ministry of Foreign affairs 2020). According to Finnish institute of health and welfare THL (THL 2020), no single country in Europe met the Finnish government's strict COVID-19 incidence requirements on 21st October 2020. According to the same source (THL 2021), only Iceland, out of those European countries with an airport, met the limit on 1st March 2021. Helsinki Airport, the case airport of this research, is one of the worst-hit by the coronavirus measured by the air traffic recovery (see table 1).

Table 1. YoY %-change in air passenger traffic at selected airports in 2020 (data obtained from Eurostat and websites of individual airports)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Helsinki Airport	2,7	-2,2	-57,8	-98,8	-98,1	-96,0	-89,8	-86,7	-91,6	-92,3	-92,3
Copenhagen Kastrup	-0,3	-0,4	-63,7	-99,0	-98,5	-94,9	-81,8	-78,8	-83,1	-85,5	-90,9
Tallinn Lennart Meri	4,2	8,7	-55,4	-98,9	-96,5	-92,1	-76,2	-74,1	-85,4	-88,9	-89,3
Paris Charles de Gaulle	2,8	-0,3	-58,5	-98,0	-96,8	-91,0	-76,7	-71,3	-80,2	-81,7	-88,8
London Heathrow	2,9	-0,7	-52,4	-97,0	-96,6	-95,2	-88,8	-81,5	-81,5	-82,4	-88,0
Stockholm Arlanda	-5,9	-5,2	-59,5	-97,7	-97,4	-95,3	-87,3	-82,8	-83,0	-79,0	-87,0
Frankfurt am Main	-0,7	-3,9	-62,0	-96,9	-95,6	-90,9	-80,9	-78,1	-82,9	-87,0	-83,4
Riga	12,3	13,6	-55,9	-99,5	-98,6	-92,9	-77,5	-77,0	-84,5	-87,1	n/a
Amsterdam Schiphol	1,5	-2,4	-56,0	-97,9	-96,8	-92,8	-80,1	-72,8	-79,4	-82,2	n/a
Oslo Gardermoen	-0,4	0,9	-55,8	-94,5	-91,0	-85,8	-72,7	-73,0	-78,3	n/a	n/a

The future uncertainties recognized by International Civil Aviation Organization ICAO (2020) are related to the length and severity of the pandemic, depth and length of the global economic recession, stringency and duration of lockdowns and movement restrictions, restoration of consumer confidence in air travel, and airlines' ability to survive from liquidity crises and the possible structural paradigm shift in travel behavior. Air travel, which is often perceived as non-essential activity compared to other mass transport options, is sensitive to long-lasting demand reductions (Sung & Monschauer 2020). Lamb, Winter, Rice, Ruskin & Vaughn (2020, 4–5) found a significant negative relationship between willingness to travel by air, for both business and pleasure purposes, and perceived COVID-19[4] threat and fear. Their findings in the US aviation industry (ibid., 5–6) suggest that air travel demand will not recover until the pandemic gets under control and passengers are convinced of air travel safety.

The significance of such factors as health and safety cannot be dismissed with a shrug: Forsyth, Guiomard, and Niemeier (2020, 1) discuss how health and safety factors can hold even a more significant effect on demand than GDP. Traditionally five percent change in GDP has translated to a 5-10 percent change in air transport demand, they argue. Sung and Monschauer (2020) discuss how history holds many examples of people shifting to other transport modes post-crisis. They also discuss how business travel, which previously

[4] COVID-19 is the acronym of Corona Virus Disease 2019, which is a disease caused by the novel coronavirus SARS-CoV-2. The year 2019 refers to the year of outbreak. (WHO 2020)

has been considered an essential activity, holds a risk of demand reduction in the aftermath of extended travel restrictions and technological improvements.

In their recent article, Brown and Kline (2020) critically analyze and discuss the preparedness of the top management teams of U.S. commercial airlines for the coronavirus pandemic. According to them, the pandemic is one of the exogenous shocks among, for example, certain macroeconomic events and terrorism, which pose a significant adverse threat to airlines' operational and financial performance. Since airports' financial performance is reliant on air traffic, the shocks are equally critical to airports, too. The authors argue that the past severe viral epidemics (such as SARS, MERS[5]) should have been considered wake-up calls by the managers responsible for environment scanning. Although the magnitude of the currently emerging pandemic is something that the industry has never experienced, a pandemic was still a somewhat predictable event, Brown and Kline (ibid.) argue.

Considering the expected long duration of the crisis and uncertain future demand for air travel, this "new normal" requires airports to consider new highly dynamic variables in their forecasting models to gain insights from the fuzzy future (Khurshid & Chandrasekhar 2020). The insights are essential for financing, planning and meeting the new competitive norms, which are not expected to return to normal anytime soon, maybe never again. Moreover, since epidemics and pandemics are deemed regularly occurring non-black swan[6] events (Browne & Kline 2020, 8), airport operators should improve their preparedness and ability to make predictions during such exogenous shocks in the future. Thus, this thesis aims to provide valuable insights for airport management teams by examining the possibilities to make predictable events a bit more predictable.

Societal significance

Aviation has a significant impact on the economy and global economic growth (ATAG 2020, 14). In a report commissioned by ATAG (2018, 13), Oxford Economics estimated that the global GDP impact of the aviation industry was 2,7 trillion United States dollars (USD) in 2016 when the direct (704,4 billion), indirect (637,8 billion), induced (454,0 billion) and tourism catalytic (896,9 billion) effects were accounted. The direct GDP impact accounted for 0,9 percent of the global GDP. The direct impact on the economy of the European Union

[5] Severe Acute Respiratory Syndrome, Middle East Respiratory Syndrome
[6] black swan event refers to an unpredictable rare event (Brown & Kline 2020, 2)

(EU) was 144 billion euros. Including all direct, indirect, induced, and tourism catalytic effects, the aviation industry contributed 624 billion euros, an equivalent of 4,2 percent share, in the EU GDP in 2016. (ATAG 2018, 52–53). In their revised analysis, ATAG (2020, 5) expects both direct and indirect annual impacts on global GDP to reduce by approximately 50 percent due to the pandemic.

The industry employs directly 10,2 million people, out of which 2,0 million jobs are supported in the EU. It represents 4,1 percent of all employment (ATAG 2018, 52–53). In their report published on 30th September 2020, ATAG (2020, 5) estimated a potential loss of 4,8 million global aviation jobs due to COVID-19, out of which 220 000 jobs at airports (a reduction of 34 percent from pre-COVID levels). In light of the current industry knowledge, the magnitude of loss seems to be even higher than the numbers suggested in April 2020 by Iacus, Natale, Santamaria, Spyratos & Vespe (2020), who estimated a loss of 25-30 million jobs due to air travel restrictions. By taking into account both direct and indirect jobs supported by the global aviation industry, ATAG (2020, 5) estimates that 46 million jobs are at risk of disappearing, a 52,5 percent reduction to pre-COVID levels. Government policy responses regarding international travel represent a significant role in the future demand for air travel. Considering the aviation industry's economic impact, the effects of the pandemic and policy responses on air passenger volumes should be of great interest to government bodies and policymakers, too.

Scientific significance

Traditionally causal-explanatory statistical methods, which aim to test hypotheses and evaluate the explanatory power (measured by, e.g., R^2), have held a dominant role in empirical research (Schmueli & Koppius, 554). This thesis, on the contrary, approaches the phenomenon primarily from the perspective of predictive analytics. Although the approach has its primary goal of reaching high predictive power (i.e., maximizing the accuracy) instead of testing hypotheses, Schmueli and Koppius (2011, 554) argue how predictive analytics may equally be used to generate new theory, comparing and improving the existing ones, and assessing the relevance or predictability of empirical phenomena. This thesis aims to improve the current body of knowledge by evaluating air passenger demand predictability during the pandemic by adopting the predictive analytics approach.

Passenger demand forecasting in the transportation industry has gained increased attention in academia lately. Banerjee, Morton, and Akartunali (2020) reviewed 120 scientific publications from the airline, railway, bus transport, and maritime sectors. The findings demonstrate how demand forecasting of passenger transport has received increased attention, especially in the aviation industry, which is the most popular industry in academic research related to passenger demand forecasting. The findings of the upward trend are supported by Ghalehkhondabi, Ardjmand, Young, and Weckman (2018, 77), who noted an increased number of publications that combine demand forecasting with such terms as tourism, transport, travel, and passenger.

To the best of my knowledge, only one scientific article has yet been published discussing the effects of government restrictions on air passenger traffic (see Iacus et al. 2020). In addition to that, only one (see Gallego & Font 2020) was initially found focusing on the effects of the COVID-19 on air passenger demand in general. Forsyth et al. (2020) has contributed to the novel topic by analyzing airport pricing responses in the wake of demand reduction. There are only a limited number of papers focusing on air passenger demand in past crises, and those had mainly focused on estimating the duration of impact on demand. Gudmundsson, Cattaneo & Redondi (2020, 12), for example, have concentrated on the recovery length of past crises and noted that the typical recovery length in the aviation sector had been a maximum of four years. Lee, Oh, and O'Leary (2005) analyzed the impact of the 9/11 terrorist attacks on passenger demand and estimated that the terrorist attacks rather have a short- than long-term impact on demand. Within the same 9/11 context, Blalock, Kadiyali, and Simon (2007) suggested that new airport security screening measures introduced after the attacks had an adverse effect on air travel demand.

Based on comprehensive literature searches, no papers focusing on the accuracy of forecasting methods amid crisis were found, which, however, may be explained by the comparably short crisis recovery times in the past. This assumption is supported by Njegovan (2006), who in his article discussed how the shocks, transitory in nature, do not typically affect long-term air travel demand. Thus, there is a prominent research gap in the research of air passenger forecasting during exogenous shocks. To fill this gap, the thesis aims to answer such questions as to whether traditional forecasting models can provide accurate forecasts during the ongoing crisis and whether pandemic-related factors could improve forecasts' predictive power.

This thesis is one of the few contributing to the theory and practice by discussing air passenger traffic forecasting during an exogenous shock and one of the first ones, if not the first, empirically testing the predictive power of the models during a pandemic. Moreover, based on an exhaustive literature search, the thesis seems to be the first one conducting a systematic literature review on passenger volumes forecasting in the context of airports. Only one (see Wang & Song 2010) systematic review focusing on air travel demand forecasting was discovered, which reviewed the methods in the whole aviation sector. This approach has a risk of generalizing findings by giving too much weight to airlines' and governments' perspectives. Only six studies included in the review focused on an airport (ibid. 38).

1.3 Research problem and questions

The aim is to compare methods used to forecast commercial air passenger volumes at airports during the pandemic and, besides comparing the overall accuracy of the models, evaluate the effect of including COVID-19 related variables in the models. The objectives of this research are 1) to map methods for air passenger traffic forecasting at airports, 2) to test the accuracy of the selected models during the ongoing coronavirus pandemic, and 3) to estimate the relevance of refining the models with COVID-19 related factors.

Plenty of research has been conducted on air passenger transport forecasting. However, *the performance of quantitative passenger traffic forecasting methods in the era of COVID-19 remains unknown*. Based on this research problem and the objectives described above, the main research question has been formulated: **How quantitative forecasting methods are able to predict airport passenger volumes in the era of COVID-19 pandemic?** The main research question is supported by three sub-questions:

> **Sub RQ1**: What quantitative methods have been used in passenger volumes forecasting at airports, and how past industry shocks have been handled in the models?
>
> **Sub RQ2**: What is the accuracy of selected forecasting methods during the COVID-19 pandemic?
>
> **Sub RQ3**: How does including COVID-19 related variables in the models affect forecasting accuracy?

The research questions have been sorted in chronological order. To start estimating the accuracy of models, a review of potential forecasting methods must be conducted. To estimate the relevance of adding exogenous pandemic-related variables in the models, baseline models must first be constructed, and their predicting power estimated. Figure 3 on page 14 illustrates the steps of the research project and the relationships between the research problem, questions, aim, and objectives.

1.4 Data and methodology

Daily data of aggregated passenger traffic at Helsinki Airport were collected from January 2016 until December 2020. In addition to daily data, monthly data was collected from the same source for 2010-2020.

The research applies two methodological approaches: a systematic literature review (SLR) and quantitative empirical research. The former aims to answer the first sub-question of this thesis: *"What quantitative methods have been used in passenger volumes forecasting at airports, and how past industry shocks have been handled in the models?".* An exhaustive literature review on all available scientific publications related to air passenger forecasting methods at airports was conducted.

The empirical section aims at answering the remaining two research questions. To answer the second sub-question, which aims to evaluate the accuracy of selected forecasting methods during the pandemic, five time-series methods are implemented, and their predictive power examined. ARIMA models are widely covered in the scholarly literature and are well-recognized as providing a solid base for benchmarking with more sophisticated models. TBATS, unlike ARIMA, is able to handle multiple seasonalities. Facebook's Prophet represents a modern approach for automated forecasting. Forecasting with artificial intelligence based models, such as neural networks, has become commonplace in the aviation industry (see pp. 40). Thus, two neural network methods are utilized.

The third research question aims to evaluate the relevance of including COVID-19 related variables to the constructed models. The relevance is evaluated by comparing the predictive power of the univariate models with SARIMAX, Prophet, and neural networks, all suitable for forecasting with multiple variables. The selection of the models stems from the theory. However, the exploratory nature of this phase stems from the novelty of the pandemic and choosing the COVID-19 related variables. To the best of my knowledge, no scholarly assessment has yet been conducted on COVID-19 related variables in air passenger demand forecasting models.

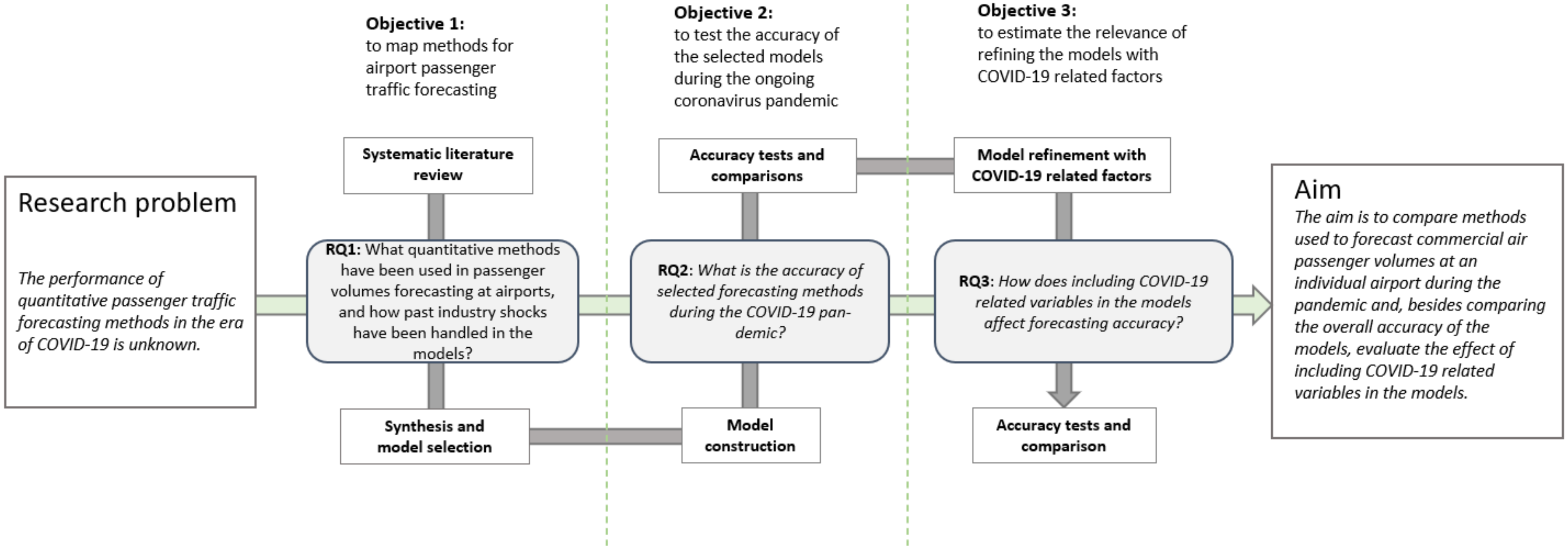

Figure 3. The relationships between research problem, questions, objectives, and the aim of the research

1.5 Delimitations and theoretical framework

The thesis relies on both peer-reviewed primary and non-peer-reviewed secondary literature. The latter, also referred to as “grey” literature, consists of institutional reports, company documents, and other documents that may not have been subject to editorial control or scientific peer review (Hemmingway 2009, 4; 7). Due to the novelty of COVID-19 pandemic and the scarcity of academic publications related to this research topic, secondary literature plays a significant role in discussing the relationship between COVID-19, air passenger traffic, and forecasting. The practice of including both primary and secondary sources in this thesis follows Nielsen (2018, 172), who argued it is an appropriate approach when novel phenomena are being explored.

The thesis explores air passenger traffic forecasting from the perspectives of airports and quantitative methods during an exogenous shock, namely COVID-19 (see figure 4). The methodological delimitation is based on the research objectives that support selecting quantitative forecasting methods: to construct forecasting models based on available time-series data, and to estimate the predictive power of the models. Moreover, according to Armstrong (2001a, 7), quantitative methods are expected to provide more precise predictions when a sufficient amount of objective data is available on the dependent and independent variables.

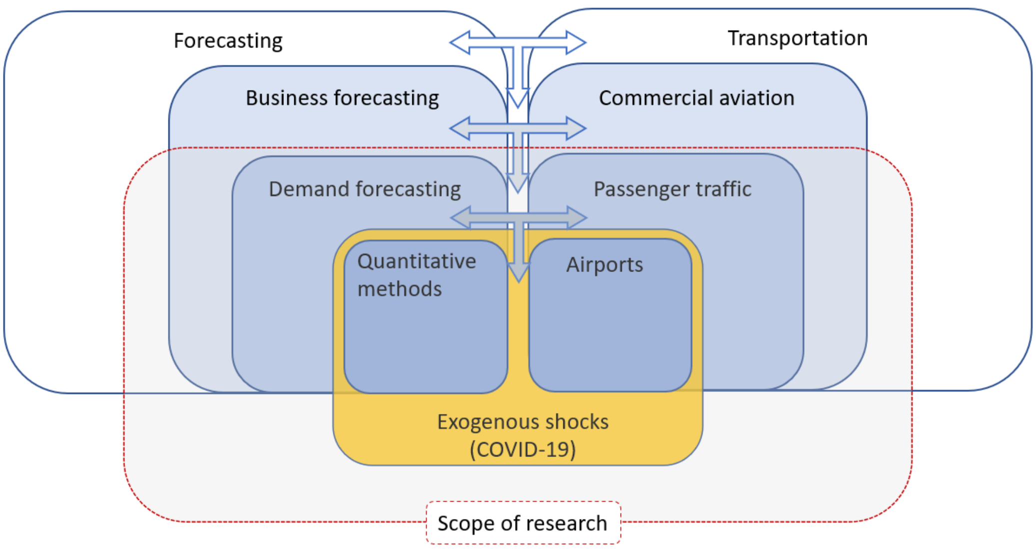

Figure 4. Theoretical framework

Aviation, more specifically civil aviation, was chosen as an industry sector because it is claimed to hold a significant role in facilitating global economic growth (ATAG 2020, 14) and has been one of the pandemic's worst-hit industries. Civil aviation can be broadly divided into commercial and general aviation (see figure 5). Commercial air transport has been defined by The European Parliament in the regulation (EU) 2018/119 as "an aircraft operation to transport passengers, cargo or mail for remuneration or other valuable consideration". In this thesis, only commercial air transport is considered. The delimitation was done because passenger volumes, which are of interest in this thesis, are among the most significant revenue streams in commercial aviation. Based on my professional industry knowledge, passenger numbers do not play such a significant role in estimating future financial performance in non-commercial general aviation. Also, sample data supports limiting the study to commercial aviation as the data is available of commercial air transport passengers only.

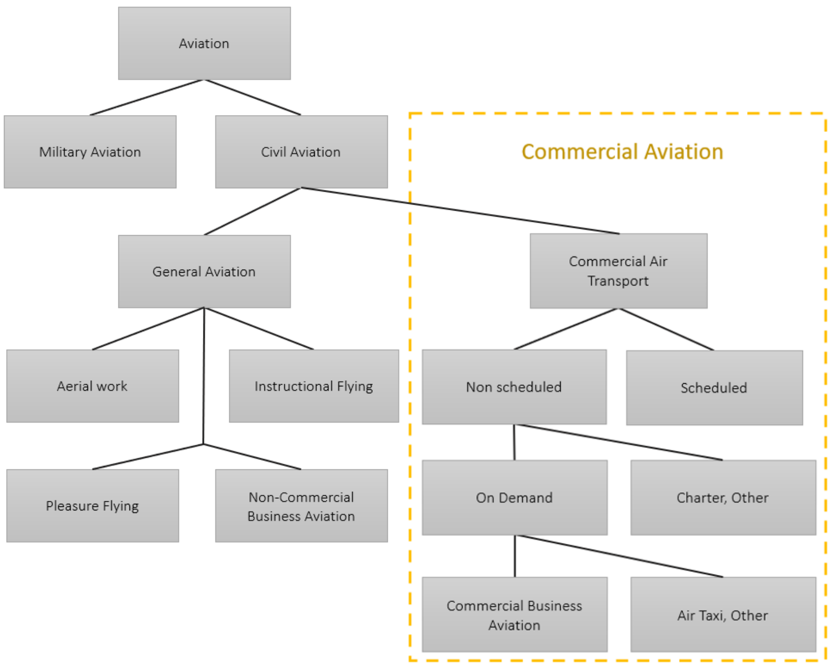

Figure 5. Classification of flight types (adapted from ICAO 2009)

The airport perspective was chosen since data availability is different from that of airlines who have access to forward-looking internal sales and booking data, for example. Therefore, airports must rely on historical data and open-source or paid information in their forecasts. Also, approaching the phenomenon from an airport perspective is not tied to a particular airline and its business model (low-cost vs. traditional). Instead, the selected perspective allows estimating the overall effect of the pandemic through multiple different airlines. Although the aim is to maintain objectivity by analyzing passenger traffic with aggregated passenger numbers of different airlines, it should be noted that passenger numbers may, at times, be influenced mainly by a single airline, typically a flag carrier of a country.

The focus on air passengers was selected because a large share of airport revenue is generated through passenger volumes (Twinn et al. 2020). In the context of demand forecasting, it is meaningful to concentrate on air passengers whose future demand, in the end, defines the future revenues and financial performance of airport operators. Also, it is the passengers whose willingness and ability to travel are subject to government responses and the progress of the pandemic. While the long-term effects of the pandemic on travel demand remain unknown, this thesis will focus on short-term forecasting during the pandemic. The short-term horizon in this thesis translates to 3-12 months projections, which can be deemed suitable for budgeting and forecasting monthly revenues (ACI 2016, 2; Banerjee et al. 2020, 798).

1.6 Research outline

The thesis is divided into six distinct chapters, including this introductory part where the background, motivation, and justifications of this research were discussed. Following this, the theoretical background of passenger demand forecasting is scrutinized. In this second chapter, the purpose of forecasting is defined, typical passenger demand forecasting methods reviewed, and method selection and performance evaluation scrutinized.

In the third chapter, theoretical foundations are narrowed down to the context of air passenger demand forecasting at airports. In this chapter, the SLR is conducted in which the scholarly research is reviewed, and potential forecasting methods for the empirical study mapped. The fourth chapter is dedicated to the methodological part of the empirical section. In this chapter, data and pre-processing are discussed, and methodological aspects of the selected forecasting methods are presented. The fifth chapter presents the forecasting results. Finally, the results and their limitations are discussed in the sixth chapter. The findings are intertwined with the theoretical and practical aspects, and future research avenues are proposed.

2. Passenger demand forecasting

This section discusses the theoretical foundations of forecasting by reviewing different forecasting methodologies, method selection, and performance evaluation. In addition to the general section, the theoretical discussion is expanded to the following chapter, where the findings of the SLR on airport-specific passenger traffic forecasting methods are presented.

2.1 Definition and purpose of forecasting

Hyndman and Athanasopoulos (2019, 1.2) define forecasting as a practice of “predicting the future as accurately as possible, given all of the information available, including historical data and knowledge of any future events that might impact the forecasts”. The purpose of forecasting stems from the needs, in this case of business needs, of understanding future events as accurately as possible.

There are different time-spans in which the forecasts are needed. One way is to classify forecasts into long-, medium-, and short-term forecasts (Hyndman and Athanasopoulos (2019, 1.2). However, there is no clear definition of what the forecasting horizon for each is. It varies even within the industry. Airport Council International (ACI 2016, 3) illustrates this from the perspectives of airlines, airports, aircraft manufacturers, and civil aviation authorities (CAA): While the long-term forecasts for airlines are prepared for 3-5 years, it is up to 20-25 years for airports and aircraft manufacturers, and up to 30-40 years for CAAs. Medium-term forecasts for airlines are considered for the next 12 months, while for the other industry players, medium-term horizon means the next five years. Airports and airlines need forecasts even for a very short-term horizon, which in the airline business means next flight and in the airport context can mean anything from the next day to current IATA[7] season (summer/winter) (ibid., 3).

Forecasts can also be classified into operational, tactical, and strategic planning, roughly reflecting short, medium, and long-term forecasting, respectively (Næss & Strand 2015, 41). Banerjee et al. (2020, 797-798) divide forecasting into micro and macro forecasting. They

[7] The International Air Transport Association

consider strategic planning as an application of long-term macro forecasting, which spans out predictions over a year ahead (ibid., 797). In addition to strategic planning, long-term forecasts in the aviation industry are typically influenced by long-term investment plans such as infrastructure developments (ACI 2016, 3). Another activity for macro forecasting is budget planning that, according to Banerjee et al. (2020, 798), occurs annually or even as often as quarterly, with often monthly income and expense projections.

Archer (1994, 105) defines demand in economic terms “as the quantity of a product or service that people are willing and able to buy during a given period of time”. He continues describing demand forecasting as “the art of predicting the level of demand that might occur at some future point or period of time” (ibid. 105). Ashford (1985, 101) describes air transport forecasting as a mechanism through which the future demand may be analyzed locally and globally. Demand forecasting is vital for companies involved in the transportation and travel industry. In their articles, Banerjee et al. (2020, 798) and Ghalehkhondabi (2018, 77) explain how travel-related products are considered highly perishable, highlighting the importance and need for accurate demand predictions.

Ashford (1985) recognizes four types of users for air traffic forecasts: aircraft manufacturers, airlines, governments, and airports. In scientific literature, air passenger forecasts in the airport context are typically described to be used for capacity planning (e.g., Rodriguez, Pineda & Diaz Olariaga 2020, 10; Li, Han, Liu & Li 2018, 442), infrastructure project feasibility studies (Wadud 2011, 59), and service level optimization (Felkel, Steinmann & Follert 2017, 444; Wu et al. 2020). Felkel et al. (2017, 451) also demonstrated how passenger demand forecasting techniques could be used to predict airport retail revenues with just two parameters: number of passengers and time. Frankfurt Airport has built a simulation model to calculate the opportunity costs of suboptimal aircraft parking positions and to propose “retail-optimized” positions for connecting flights (ibid.). Thus, there are several applications for which air passenger forecasts are needed.

Demand forecasts contain plenty of uncertainty, which poses a significant risk to the financial viability of, for example, infrastructure projects (Flyvbjerg, Skamris Holm & Buhl 2005, 131). Flyvbjerg et al. (2005, 140) examined inaccuracies in demand forecasts in 210 rail and road infrastructure projects and found out that rail passenger forecasts have typically been overestimated in 9 out of 10 cases. The average overestimation was more than 100 percent in rail projects, and half of the road traffic forecasts were off by 20 percent (ibid., 140). According to the findings of Suh & Ryerson (2019), who in their study examined the accuracy of 704 long-term (10-years) U.S. airport forecasts, found out that 85 percent of the forecast errors were overestimations, with the mean forecast error being 39,5 and median 27,6 percents. Thus, optimism-bias seem to be commonplace in demand forecasting in the transportation industry, which may have significant financial consequences. On the other hand, underestimating demand may lead to an over-cautious approach in capacity expansion projects, which may restrict future earnings. Therefore, companies should be careful in selecting a suitable forecasting method and variables to predict future demand as accurately as possible.

2.2 Forecasting methods

Different methodologies exist for projecting future passenger demand and estimating the effects of uncertainties: Market research, scenario planning, time-series analysis, and econometric models, to mention few (ICAO 2006). According to Li, Han, Liu, and Li (2018), there are about 300 methods, out of which 30 are commonly and ten widely used. There is no precise classification of forecasting methods, but a general way is a broad classification to quantitative and qualitative methods. Chambers, Mullick and Smith (1971) classified forecasting models into three main classes: qualitative, time series analysis and projection, and causal models. The former two are typically considered sub-categories of quantitative methods that both have one thing in common: statistical analysis and predictions based on historical data (Banerjee et al. 2020, 799). ICAO (2006, I-2) has proposed three classes for air passenger demand forecasting: quantitative, qualitative and decision analysis methods. Perhaps the popularity of using quantitative methods for air passenger demand forecasting has led Kim and Shin (2016, 98) to propose a general classification into time series and causal analyses. Banerjee et al. (2020) classify passenger demand forecasting methods into four main categories: quantitative, qualitative, mixed models, and ancillary tools. Figure 6 represents this taxonomy of passenger demand forecasting methods proposed by Banerjee et al. (2020), which serves as a methodological framework for classifying forecasting models in this thesis. The taxonomy of quantitative methods is in line with Dantas, Oliveira, Luiz, Repolho and Miguel (2017, 117), who claim causal econometric, time series and artificial intelligence based methods the most popular ones in air transport demand forecasting.

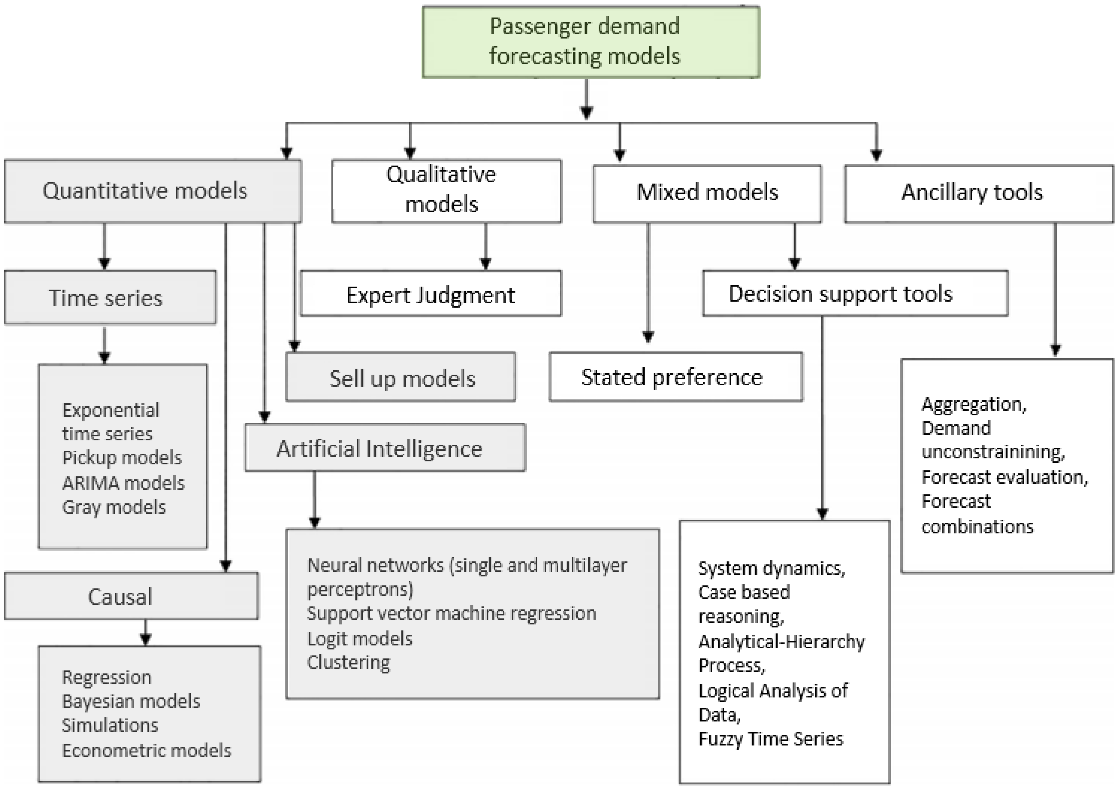

Figure 6. A taxonomy of passenger demand forecasting methods (adapted from Banerjee et al. 2020, 799)

Causal models are the most used demand forecasting methods in scheduled passenger transportation research (Banerjee et al. 2020, 802). Causal models, such as regressions, simulations, and econometric models (ibid., 804), are built from factors that aim to explain future demand with historic causations (ibid., 803). Suryani, Chou, and Chen (2009) have illustrated causal effects of internal (industry factors) and external factors (economic conditions and demographic factors) on airport passenger demand with a causal loop diagram (see adapted figure 7). Although their research presented a system dynamics model, the key factors are well documented to illustrate variables used in the causal models. Econometric models are suitable for airport, city, region, or national level forecasting. Based on the selected literature of Wadud (2011, 62), gravity models, a type of econometric model, typically use aggregate data of cities or countries to forecast demand for city pairs or at the national level.

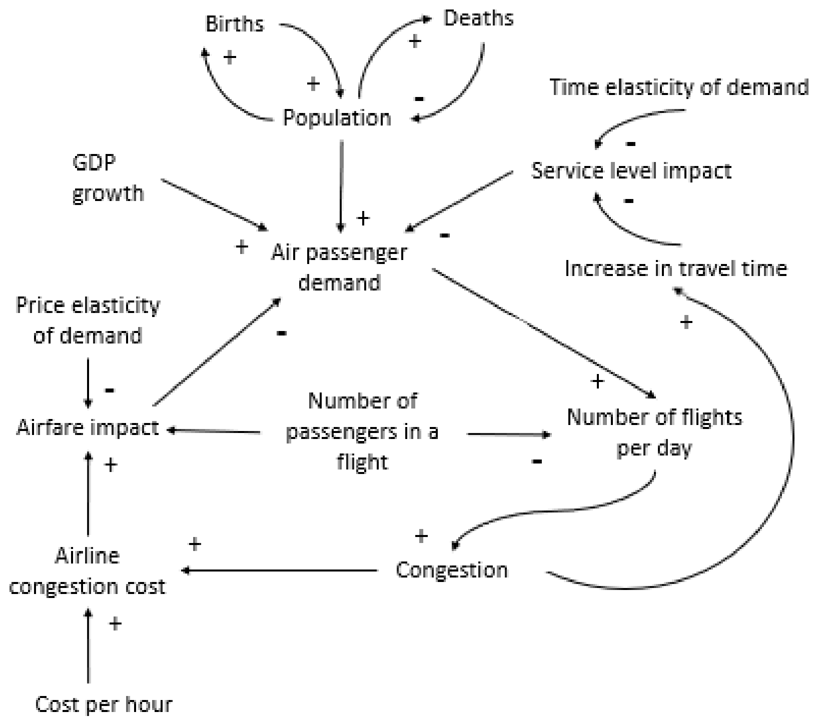

Figure 7. Simple causal loop diagram of air passenger demand (adapted from Miller & Clarke 2007, 21; Suryani et al. 2009, 2327)

The diagram can be interpreted by following the arrows indicating causality and plus or minus signs, which indicate positive and negative effects. For example, air passenger demand increases the supply, the number of daily flights. Increased traffic causes congestions. Congestions increase airlines' costs and, consequently, airfare, which has a negative impact on air passenger demand due to price elasticity. On the other hand, congestions may also increase travel time, making air travel non-competitive compared to other transport forms (air travel substitutes). This has a demand decreasing impact through decreased service level (Miller & Clarke 2007, 23). Miller and Clarke (ibid., 23) have applied estimations for price elasticity of -1,6 for leisure travel and -0,8 for business travel and time elasticities of -0,8 and 1,6, respectively. Thus, airfare impact can be estimated by multiplying air travel costs by price elasticity and service level impact by multiplying the change in travel time by time elasticity.

A time series model uses its previously observed values to predict future outcomes (Banerjee et al. 2020, 803). Time series are widely adopted in passenger demand forecasting and, according to the findings of Banerjee et al. (2020, 802), are also one of the oldest set of methods in passenger demand forecasting in scheduled transportation. Perhaps the popularity stems from the conceptually simple but powerful methodology (Spitz & Golaszewski 2007, 20). The methods range from simple YoY trend projections and exponential smoothing methods, where the most recent observations are given more weight, to more sophisticated ARIMA (Autoregressive Integrated Moving Average) models (ibid., 21). This section focuses on ARIMA models since they are popular methods within passenger demand forecasting and often used as benchmarks in comparison to others (Banerjee et al. 2020, 803).

ARIMA, first introduced by Box and Jenkins (1976), consists of three individual components: autoregressive (AR), integrated (I), and moving average (MA). The AR model is similar to the multiple regression model, but the predicted value is correlated with its own past values. That is, the AR model uses lagged values of the predicted values as regressors (see equation 1). MA component analyses previous prediction errors and uses a similar kind of regression model as AR, but instead of using lagged observed values as regressors, it uses lagged errors as regressors (see equation 2). Differencing, I, is used to convert non-stationary time series (i.e., time series with seasonality or trend) into stationary, which is the requirement for analyzing time series (see equation 3). Non-stationarity may not disappear after first-order differencing and, thus, second-order differencing may be required, which represents a change in the change (see equation 4) (Hyndman & Athanasopoulos 2019, 9.1). Finally, the ARIMA model (see equation 5) can be presented as ARIMA(p, d, q), where p is the lag order of observation (AR), d represents the order of differencing (I), and q denotes the lag order of observation errors (MA). See, for example, Hyndman & Athanasopoulos (2019) for more comprehensive details on ARIMA models.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t , \qquad (1)$$

where,	y_t	is the value at time t,
	c	the average of the changes between consecutive observations,
	ε_t	white noise,
	ϕ	the coefficient of lagged observation, and
	p	the order of lagged observation.

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} , \qquad (2)$$

where,	θ	is the coefficient of past forecast errors, and
	q	the order of lagged error.

$$y'_t = y_t - y_{t-1} \tag{3}$$

$$y''_t = y'_t - y'_{t-1} \tag{4}$$

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \text{ ,} \tag{5}$$

where y'_t is the differenced series

Regardless of the model, no single time series model can be generalized as the best, as the mixed results of a forecasting competition in 1980 demonstrate (see Makridakis et al.1982, 123). Zhang (2004, 2) proposes that the reason for the mixed results may be with the linear forecasting methods attempting to predict non-linear real-world problems.

Artificial intelligence (AI) based methods have been recognized to handle better these complex non-linearities in the data than traditional causal and time series methods (Jin, Li, Sun & Li 2020, 2). The predictive power of AI models lies behind computationally intensive algorithms that are able to fine-tune their predictions iteratively by self-learning (Banerjee et al. 2020, 803). The literature recognizes multiple uses for AI-based forecasts, such as passenger demand forecasting, wind speed forecasting, and electricity price forecasting, as listed by Jin et al. (2020, 2). This relatively new approach to demand forecasting has also been applied in scheduled passenger demand forecasting too, but no significant growth in popularity can be detected from the results of Banerjee et al. (2020, 802). Although AI-based tools are often considered superior in contrast to traditional methods, they are also criticized for overfitting, slow learning speed, and not always being able to produce the global minimum in prediction accuracy (Jin et al. 2020, 2).

Artificial neural networks (ANN) are typical examples of AI-based complicated systems that aim to resemble the human brain in producing outputs (Liu, Huang, Chen, Qui, Chen 2017). Ghalehkhondabi et al. (2019, 86) claim ANNs were first used in tourism demand forecasting studies in the late 1990s and have become commonplace during the last decade. In business forecasting in general, ANNs have become useful due to their versatility in modeling both linear and non-linear problems (Zhang 2004, 2-3). In simple terms, the most widely used feedforward type ANN consists of an input layer, one or more hidden layers, and an output layer, each containing a certain number of nodes, neurons (see figure 8). The nodes in the input layers represent the independent variables that are used to predict the dependent variable, output Y. Information from the input layers is weighted and sent to the hidden layer, where the data is being processed and sent again to the output layer for prediction by using individual weights. (ibid., 3–4)

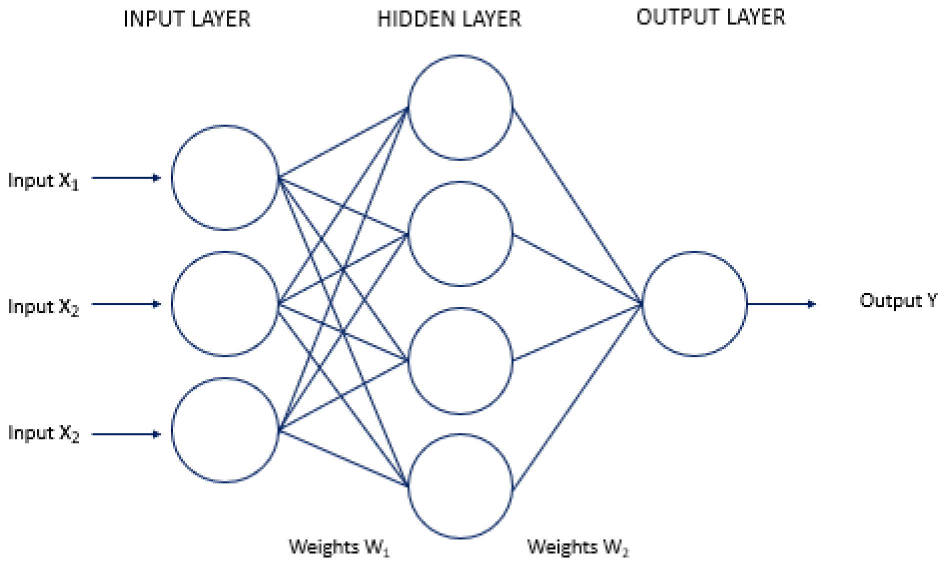

Figure 8. An architecture of a simple (feedforward) ANN (adapted from Zhang 2004, 4)

Spitz and Golaszewski (2007, 7) explain how demand forecasts are not objectives themselves. Instead, they are supposed to reflect the demand for, in this case, aviation services. However, they highlight that historical aviation activity at an airport is not only influenced by demand but also the supply of aviation services. Thus, historical observations may not fully reflect the historical demand, which may have been constrained by, for example, capacity. That is an important aspect to understand. The differences between descriptive and predictive models are often overlooked (Ashford 1985, 103). That is, a typical confusion is to consider a good descriptive (causative) model to be a good predictive model, too (ibid., 103). Thus, it is not indifferent which method and variables to use in predictions.

2.3 Method selection and evaluation

There can be significant differences between model performances, as was indicated in the study of Jin et al. (2020, 9). According to Spitz and Golaszewski (2007, 20), time-series analyses are expected to perform well in short-term predictions for airport operational planning and budgeting purposes when the environment is stable, and long time-series data is available. However, as Makridakis et al. (1981, 123) have demonstrated, there are significant differences in performance even between different time series methods. Underlying data and forecasting horizon has much to do with performance differences (ibid., 127). While one model may perform well with monthly data, it may not be the case with annual data. Also, models that do not take into account a trend tend not to perform that well, Makridakis et al. (ibid., 127) argue.

It is also typical that multiple accuracy measures are used since they all tend to produce different results depending on the situation (Makridakis et al. 1981, 115). Mean absolute error (MAE), root mean square error (RMSE), and Mean absolute percentage error (MAPE) are among the most common methods in airport forecasting (Spitz & Golaszewski 2007, 26). MAE (see equation 6), which measures the mean of absolute errors, is not suitable for comparing data with different scales to each other. RMSE (see equation 7) gives more weight to large errors (Spitz & Golaszewski 2007, 26). Thus, it is more sensitive to outliers than MAE (Hyndman & Koehler 2006, 682). Armstrong (2001b, 18) advises against using RMSE. Since both of the measures, MAE and RMSE, use absolute values in measuring error, they are not suitable for comparing time series of different scales (Armstrong 2001b, 11). MAPE (see equation 8), on the other hand, considers relative errors, which makes it more suitable for comparing different forecasts. However, MAPE weights positive errors more heavily than negative ones (Hyndman & Koehler 2006, 683), and it is not suitable for forecast problems containing zero or close to zero values as it produces infinite or indefinite error values in these cases (Kim & Kim 2016, 669)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \quad , \tag{6}$$

where y_i is the observed value at point i,
$\hat{y}_i$ the predicted value at point i
n number of observations

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^{\ 2}} \tag{7}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{8}$$

There are many variations of error measures that aim to tackle the issues with the most traditional ones. Some further developments include median absolute percentage error (MdAPE), symmetric mean absolute percentage error (sMAPE), symmetric median absolute percentage error (sMdAPE), median relative absolute error (MdRAE), geometric mean relative absolute error (GMRAE), mean absolute scaled error (MASE) (Hyndman & Koehler 2007, 680–681), and mean arctangent absolute percentage error (MAAPE), which aims to address the issues with zero and near-zero values in MAPE (Kim & Kim 2016, 669). Since each method has its ups and downsides, Armstrong (2001b, 15) suggests using various error measures. Nevertheless, Armstrong (ibid., 16) urges against using R-squared (R^2) in assessing forecasting accuracy in time series due to overlooking bias. R^2, which is often used to reflect the strength of a causal relationship in explanatory models, does not necessarily reflect a models' predictive power (Schmueli 2011, 4; Armstrong 2001b, 16). That is, high explanatory power (measured by R^2) may lead to inaccurate forecasts, and vice versa. (Armstrong 2001b, 16).

Albeit statistical error measures are valuable in defining a forecast model, there are other perspectives for choosing the model, too. Unnecessarily high accuracy may often involve unnecessarily high costs, drifted away from the optimal solution (Chambers et al. 1971). Therefore, Armstrong (2001b, 17) suggests conducting a cost-benefit analysis for the candidates. In addition to the statistical perspective, Armstrong (2001a, 1) proposes five more perspectives: convenience, market popularity, structured judgment, track record, and guidelines from prior research. Convenience refers to selecting an understandable model that is not unnecessarily sophisticated (ibid., 2). However, he advises not to use this criterion alone in other than stable environments (ibid., 14). Market popularity refers to the widely adopted methods by individuals and comparable institutions, which, however, may not be the best indicator for determining which should be used (ibid., 2–3). The forecast method may also be selected by using predetermined criteria for comparing the methods. Such an approach is defined as a structured judgment by Armstrong (ibid., 4). Comparing past performance, track record, of different methods may be helpful, but one needs to bear in mind that the superior past performance of a model may not hold in the future (ibid., 5). Finally, Armstrong (ibid., 6–8) proposes principles of previously published research: use structured, quantitative, causal, and simple methods, provided that enough data exists (quantitative), when large changes are expected (causal), and when there is no urge to use complex model (simple).

3. A systematic literature review on airport passenger traffic forecasting

SLR was chosen as a method to answer the first sub-RQ related to airport passenger forecasting: *What quantitative methods have been used in passenger volumes forecasting at airports, and how past industry shocks have been handled in the models?* SLR was initially designed for evidence-based research in the medical field (Tranfield, Denyer & Smart 2003, 208-–209) but has increasingly gained interest in business research (Snyder 2019, 334). SLR is a process used to "map and to assess the existing intellectual territory, and to specify a research question to develop the existing body of knowledge further" (Tranfield et al. 2003, 208). The benefits stem from its ability to identify relevant literature in a replicable, scientific and transparent manner (ibid., 209). Based on the general objectives of the SLR, this review aims to map the current status of research on airport passenger volumes forecasting and determine the development of forecasting methods, their applications, and forecasting performance.

3.1 Scoping review

Preceding the full SLR, a scoping review was conducted on Scopus and Web of Science (WoS) databases. The scoping review aimed to help determine the scope of a full systematic literature review and identify keywords. Scoping reviews can be used, for example, to map existing literature from a broader perspective to identify research gaps and evaluate the potential value of a full SLR (Peters, Godfrey, Khalil, McInerney, Parker & Soares 2015, 141–142). The objective of this preliminary search was to find out "*what is the current status of air passenger demand forecasting*". Based on the search strings on Scopus and WoS, *TITLE-ABS-KEY ("air passenger" OR ("airport" AND "passenger") AND ("demand forecasting" OR forecast*))* and *TOPIC: ((("air passenger" OR (airport AND passenger)) AND ("demand forecasting" OR forecast*))),* respectively, a total of 642 documents were found (Scopus: 449, WoS: 193) after the search results were limited to English language. The articles were then exported to reference management software Mendeley and JabRef, where duplicates (n=165) were removed. In the next step, documents (n=44) without author information were examined and removed. Removing these records was necessary to ensure the quality of the dataset: the articles without author information resulted in problems with the exported CSV file, which was required in the next steps of the analysis. However, before removing them, it was confirmed that the excluded articles did not meet the criteria for discussing air passenger forecasting. Finally, 433 documents were included in the next phase of title-abstract level analysis.

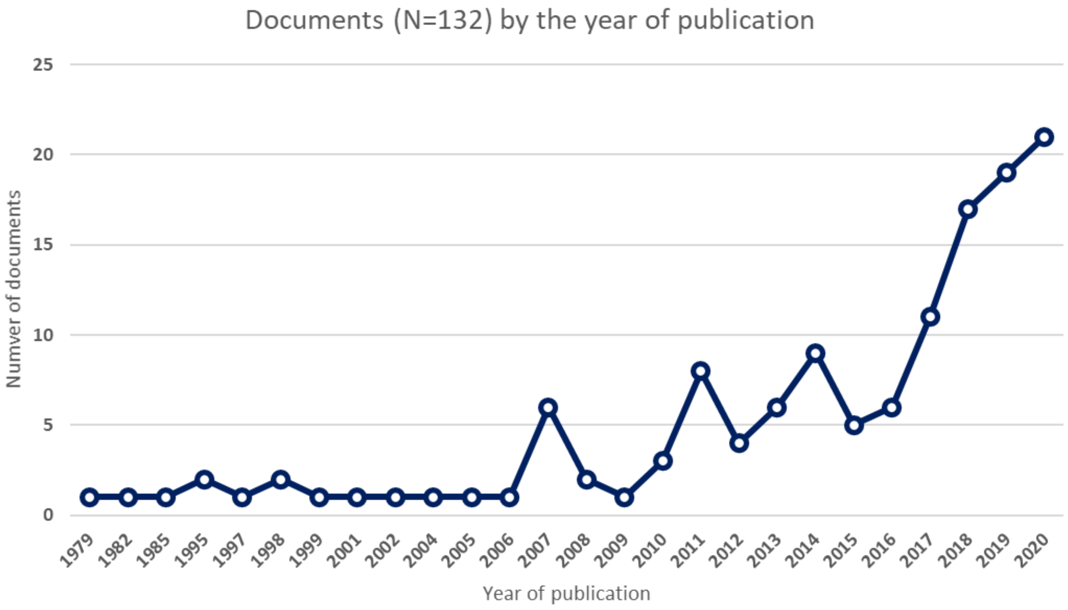

Figure 9. Documents (N=132) related to quantitative air passenger traffic forecasting (missing years not included in the graph)

The inclusion criteria in this phase were, under loose scrutiny, to exclude any document that seemed not to focus on air passenger demand forecasting. Those articles with a quantitative or mixed-method approach were included in the final sample of articles (N=132). The final number of publications, illustrated in figure 9, show increased attention to the topic, which is in line with the findings of Banerjee et al. (2020, 802) and Ghalehkhondabi et al. (2018, 77).

The scoping review was targeted at all documents related to air passenger traffic forecasting. Thus, in this search, the results were not limited to documents discussing airport forecasting only. The sample consisted of two book chapters, 44 conference papers and 86 journal articles. Based on this preliminary literature search, it was understood that no comprehensive literature reviews on passenger forecasting research in the context of airports had been conducted. Cheng and Mengting (2017) have attempted a narrative literature review for domestic (China) and international forecasting research, but it does not meet the criteria for systematic reviews.

3.2 SLR process

The scoping review represented the first of three stages proposed by Tranfield et al. (2003): **planning the review**, which provided guidelines for the review protocol. According to Okoli (2015, 889), a review protocol is a plan "which details the specific steps and procedures to follow in the particular review being conducted". This step was followed by **conducting the review**, which comprised search and exclusion phases in which relevant documents were identified. Finally, the third step was dedicated to **reporting** the results by synthesizing findings.

The SLR adopted the steps proposed by Tranfield, Denyer & Smart (2003) and quality considerations by Snyder (2019, 338), and followed the example of Banerjee et al. (2020) in applying a snowballing method for improving literature coverage. The complete process has been documented in figure 11 on the next page. The scoping review helped assess the scope of the research field and identify relevant keywords (illustrated in figure 10). In addition to the initially determined keywords *"air passenger", "airport", "passenger", "demand forecasting",* and *"forecast*"*, one additional relevant keyword was identified during the scoping review: "predict*".

Keyword	Frequency
Forecasting	18
Forecast	9
Air passenger demand	6
Prediction	5
Airport	5
Air traffic	4
Passenger flow	4
Air passenger	3
Seasonality	3
SARIMA	3
Demand forecasting	3
Big Data	3
Time series	3
Scenarios	2
Air transportation	2
SVR	2
Domestic Air Transportation	2
COVID-19	2
Airports	2
Short-term forecasting	2

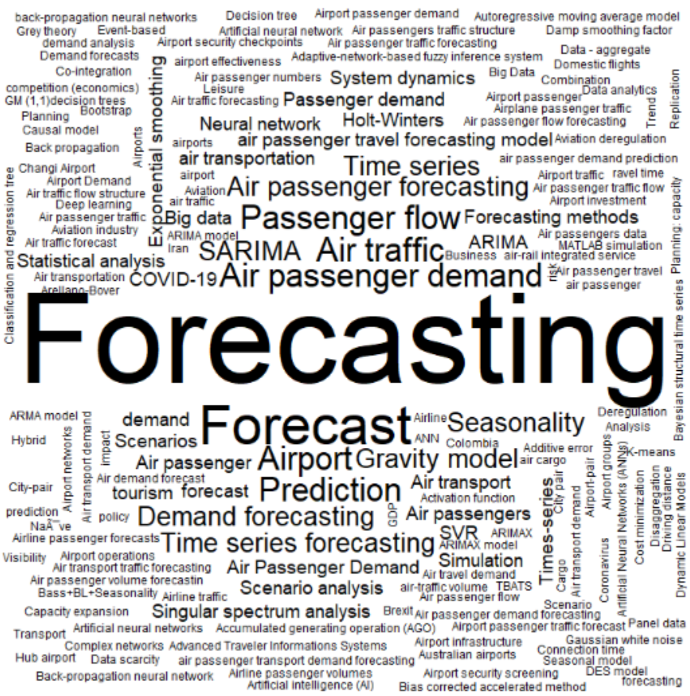

Figure 10. Keywords (N=339) used in the literature

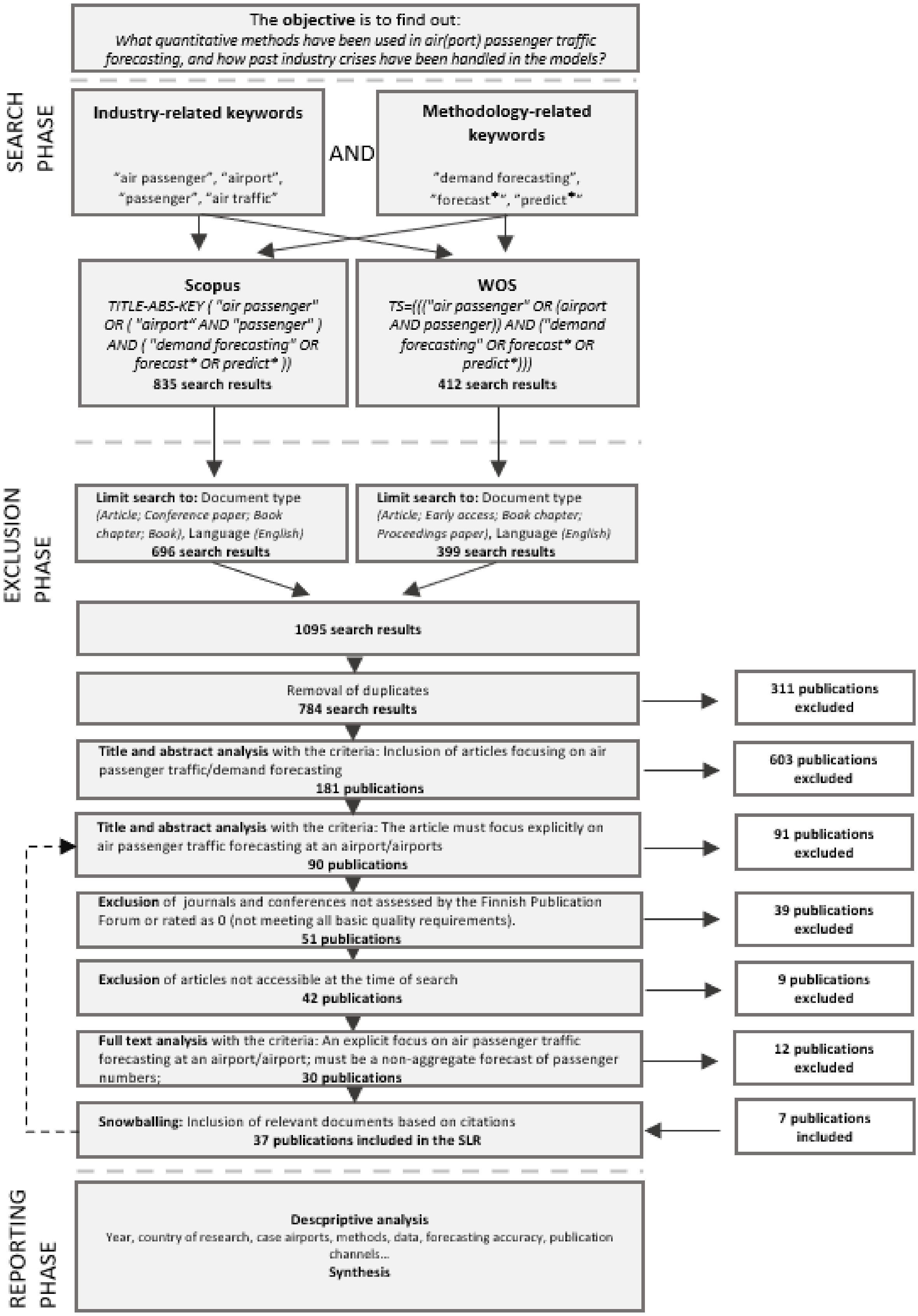

Figure 11. The process of the SLR

The search was initially started by forming search strings, in which the keywords were nested by using Boolean operators OR and AND (see figure 11). The searches were run on Scopus and WoS on 6th November 2020. WoS was selected following the example of Banerjee et al. (2020), and Scopus was added to improve literature coverage. Search strings were chosen to identify selected keywords from the title, abstract and list of key-words. Search results were limited to the English language, and after removing duplicates (n=311), a total of 784 search results were shortlisted for title and abstract level analysis.

In the first round of title and abstract level analysis, publications (n=603) not focusing on air passenger forecasting were excluded. Excluded documents included topics such as move-ments forecasting (number of flights), the spread of diseases, delay forecasting, emissions forecasting, analysis and forecasting of passenger process times (check-in, boarding, se-curity screening), baggage amount forecasting, capacity planning (when the focus was not in air passenger forecasting), airport access studies, and weather forecasting, to mention some.

In the second round of title and abstract level analysis, the articles were screened for airport context. Only those articles explicitly focusing on airport passenger forecasting were in-cluded in the sample for further review. It was determined in the review protocol that the airport context should become apparent when the reading was conducted on the publica-tions' title, abstract, and keywords. Some of the publications excluded at this stage included research papers whose focus was on airline passenger forecasting, forecasting demand at the national level, or forecasting demand for city pairs (airline demand or economic per-spective). Although forecasting methodology and data would have been suitable for airport forecasting as such, these kinds of contributions were removed from the review if they did not comply with the strict and precise inclusion criteria of discussing air passenger forecast-ing in the airport context. A total of 90 publications were identified discussing the topic in the desired context, out of which 52 were journal articles, 35 conference proceedings, and three book chapters.

The quality of publication channels for the remaining documents was screened against the Finnish Publication Forum ratings (JUFO). Assessing quality through a rating of a particular publication channel instead of scrutinizing each document's quality individually is a widely adopted practice in management research (Tranfield 2003, 216). All publications were ex-pected to meet at least the basic requirements for quality, which translates to JUFO levels

1-3. JUFO 1 level represents the basic level, and levels 2-3 are dedicated to a limited number of high standard or leading publication channels (Publication forum, n.d.). Based on the quality assessment, 41 journal articles, seven conference papers, and three book chapters, corresponding to 52 documents in total, were selected for the next round for full-text analysis. However, eight articles and one book chapter were inaccessible[8] at the time of search on 15th November 2020. Therefore, only 42 publications proceeded to full-text level analysis.

The inclusion criteria for a publication to be included in the SLR were: passenger forecasting must be discussed in the airport context, and the forecasting model must be applied to an individual airport or airports. Those studies predicting air passenger demand for a region or nation, by aggregating passenger traffic figures from multiple airports, were excluded. The SLR aimed to review publications that would assist in selecting forecasting methods for an individual airport. After the full-text analysis, 30 publications were included in the SLR based on database searches.

Those 12 publications that were excluded during the full-text analysis typically had either too vague approach to airport passenger traffic forecasting or the studies used aggregated numbers to forecast passenger demand at a city, region, or national level. Such examples include studies by Wu et al. (2020), who used aggregated monthly passenger figures of all U.S. international routes. Kagan Albrayerak, Özkan, Can & Dobruszkes (2020) studied determinants of air passenger traffic in Turkey at the provincial level using aggregated statistics. Aggregate numbers were used at national level forecasting in a German study as well (see Gelhausen, Berster & Wilken 2019). An excluded article by Liu, Liu, Liu, Chen, Feng, Xiong & Huang 2017) used data (passenger name record, PNR) from airline reservation systems that is not typically available to airport operators. The study by Liu, Li, Liu, Zhang, Rong, Yang, and Xiong (2018) was excluded since it did not entirely focus on air passengers; instead, the authors aimed to predict total occupants of an airport, including, for example, greeters in the arrival's hall. The excluded articles were deemed not to provide added value in meeting the objective of the SLR and this research.

Snowballing was performed once the final articles were chosen based on database searches. Both forward and backward snowballing methods were applied to ensure the

[8] No open access, green open access, or access through university

inclusion of as many relevant scientific works as possible. Snowballing aimed to identify all other relevant articles that were not found from the databases with the keywords. Backward snowballing means finding new references by scrutinizing the references cited by the given article, while forward snowballing refers to the practice where all documents citing the given document are examined (Wohlin 2014).

Backward snowballing, which was conducted by screening literature reviews and reference lists of the selected publications, highlighted 12 potential articles that met the quality standards measured by the JUFO rating. Out of those, five articles were considered meeting the requirements at title-abstract-keyword level analysis and, thus, were included in the full-text analysis. They were all found relevant to the SLR. Forward snowballing was conducted on Google Scholar and Microsoft Academics, which improved literature coverage and provided the details about the number of citations. Although both Google Scholar and Microsoft Academics also return non-scholarly citations, which may lead to over-estimation in numbers of scholarly citations, Harzing (2020, 5) considers the slight over-estimation more realistic than significant under-estimation of search results in Scopus and WoS. JUFO ratings controlled the quality of potential publications. After the title-abstract-keyword analysis, two potential articles were selected for full-text analysis and finally in the SLR. An article by Xie, Wang & Lai (2014) was initially excluded in the title-abstract-analysis in database searches as it did not indicate airport context. However, due to the significant number of citations to relevant articles included in this SLR, the article was selected for full-text analysis, where it was found relevant to this study. According to Tranfield et al. (2003, 215), the review protocol should allow flexibility and creativity in the process. However, any changes and rationale for doing so need to be reported (ibid., 215).

Finally, based on comprehensive backward and forward snowballing, seven more articles focusing on air passenger traffic forecasting at airports were included in the SLR. Thus, the total number of articles included in the review was 37, which is significantly more than Wang & Song (2010) managed to recognize in their SLR in 2010: back then, only six studies were categorized as focusing on a particular airport (ibid., 38). This emphasizes the importance and relevance of a fresh snapshot of the status quo.

3.3 The current state of research

Figure 12 illustrates the growing trend of research on air passenger demand forecasting also in the context of airports. The majority (84 %) of publications have been published during the past decade (2010-2020), one (Profillidis 2000) during the previous decade 2000-2009, and only a handful (n=6) during 1982-1999. The finding is in line with the SLR conducted by Wang & Song (2010).

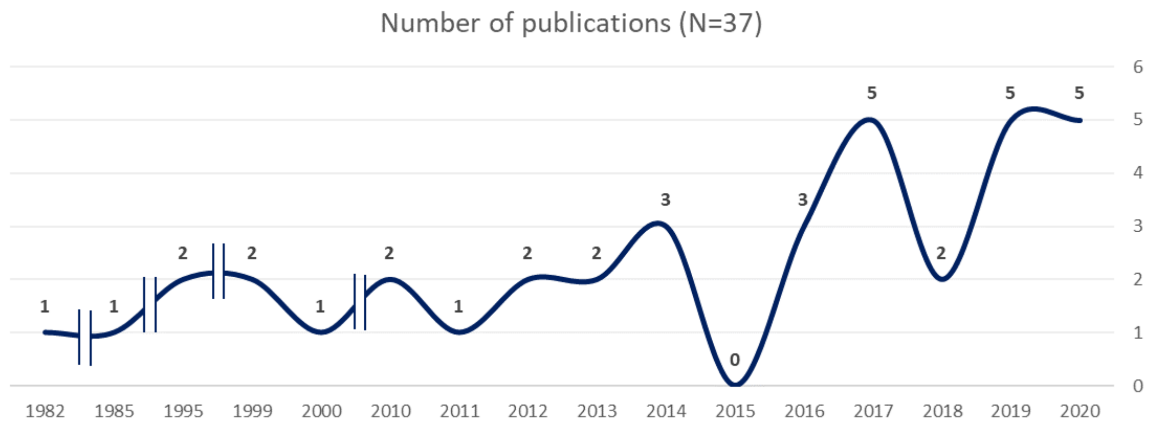

Figure 12. Number of research papers by the year of publication

Table 2 draws a picture of publication channels in which the articles included in this SLR have been published. By subject, unsurprisingly, the typical journal classifications are related to transportation. Other frequent classes include economics, management, management science, and operations research. Most of the publication channels that met the basic requirements of the Publication Forum quality assessment were basic level journals and conferences (n=30). No publications have so far been published in the highest-rated journals (JUFO 3).

Table 2. Publication channels

JUFO 2	
Journal of Transport Geography	**2**
Transportation Research Part A: Policy and Practice	**1**
Transport Policy	**1**
Transportation Research Part E: Logistics and Transportation Review	**1**
Automation in Construction	**1**
Tourism Management	**1**
JUFO 1	
Journal of Air Transport Management	**9**
Transportation Research Record	**5**
Communications in Computer and Information Science	**3**
* IOP Conference Series: Materials Science and Engineering	**3**
* Journal of Physics: Conference Series	**2**
** Digital Marketplaces Unleashed - Springer	**1**
Expert Systems with Applications	**1**
* In Proceedings of 5th the International Conference on Operations Research and Enterprise Systems	**1**
Tourism Economics	**1**
Aviation	**1**
Transportmetrica A: Transport Science	**1**
Case Studies on Transport Policy	**1**
Journal of Forecasting	**1**
Total	**37**

* denotes conference proceedings
** denotes book

The interest towards research on airport passenger traffic forecasting is global. Most of the research has been conducted in China (n=10). Asia and Pacific as an area cover half (n=19, 51 %) of the total research. The area also had the most individual airports (n=30, 63 %) covered by the research (see table 3). Suryani et al. (2010, 2325) has noted that the rapid growth of aviation in Asia Pacific regions is why the topic has attracted increased attention from academia. While most of the research is conducted in China, also its airports are widely covered. Nearly a third (n=13, 27 %) of the airports covered by research are located in mainland China. The airports of China have started to appear in scientific work published from 2019 onwards, which may be explained by the boom of aviation in China and the need for infrastructure development (Zhang 2020, 1). Aviation in China is expected to continue its growth, which justifies the importance of constructing accurate forecasting models to support capacity expansion, for example (ibid., 1). The most popular airport in scholarly research is Hong Kong International (n=4, 8 %), which has been used as the case airport by Tsui, Balli, Gilbey, and Gow (2014), Xiao, Liu, Liu, Xiao, and Gu (2014), Xiao et al. (2016), and Xie, Wang, and Lai (2016).

Table 3. Airports covered by the research

Airport	n	Airport	n	Airport	n	Airport	n
Australia		**China**		**Indonesia**		**Taiwan**	
Sydney	1	Sanya Phoenix International Airpor	3	Samarinda	1	Taoyuan International Airpor	1
Melbourne	1	Beijing Capital	3	Djalaluddin Gorontalo Airport	1	**Turkey**	
Gold Coast	1	Shanghai Pudong	2	**Portugal**		Istanbul Ataturk Airport	1
Brisbane	1	Guangzhou	2	Lisbon Airport	1	Ankara Esenboga Airport	1
Perth	1	Xianyang International Airport	1	**Saudi Arabia**			
Cairns	1	Chengdu	1	Jeddah Airport	1		
Adelaide	1	Mianyang	1	**South Korea**			
Darwin	1	**Colombia**		Seoul Incheon International Airport	1		
Bangladesh		Bogotá-El Dorado International Airport	1	**Spain**			
Shahjalal International Airport	1			Adolfo Suarez Madrid-Barajas airport	1		
Brazil		**Germany**		**The Neatherlands**			
Sao Jose dos Campos	1	Frankfurt Airport	1	Amsterdam Schiphol Airport	1		
Barreiras	1	**Greece**		**The United States of America**			
São Paulo International Airpor	1	Rhodes Airport	2	Hartsfield–Jackson Atlanta International Airport	1		
Chapeco	1	**Hong Kong**					
Lages	1	Hong Kong International Airport	4	Robert Mueller Municipal Airport	1		
				Honolulu International Airport	1		

The most cited publication is the article by Suryani (2010), who discussed airport passenger forecasting at Taiwanese Taoyuan International Airport by applying a system dynamics approach. It is followed by the article by Tsui et al. (2014), who applied time series methods to forecasting passenger volumes at Hong Kong International airport. A complete list of publications included in the SLR and the number of citations are presented in table 4 next page. Both Google Scholar and Microsoft Academic were used to count citations. The latter is supposed to return a cleaner set of search results (Harzing 2020, 5), and, therefore, it controls the results from Google Scholar, which tends to cover a broader range of non-scholarly works.

Table 4. Publications included in the SLR (sorted by the number of citations)

No	Authors (year) title	Type of publication	Google Scholar citations	Microsoft Academics citations
1	Suryani (2010) Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework	Journal article	197	186
2	Tsui et al. (2014) Forecasting of Hong Kong airport's passenger throughput	Journal article	108	107
3	Profillidis (2000) Econometric and fuzzy models for the forecast of demand in the airport of Rhodes	Journal article	84	76
4	Xiao et al. (2014) A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting	Journal article	73	64
5	Xie et al. (2014) Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches.	Journal article	53	31
6	Kim & Shin (2016) Forecasting short-term air passenger demand using big data from search engine queries	Journal article	45	27
7	Samagaio & Wolters (2010) Comparative analysis of government forecasts for the Lisbon Airport	Journal article	37	23
8	Graham (1999) Airport-specific traffic forecasts: a critical perspective	Journal article	38	20
9	Scarpel (2013) Forecasting air passengers at Sao Paulo International Airport using a mixture of local experts model	Journal article	20	12
10	Tsui et al. (2017) International arrivals forecasting for Australian airports and the impact of tourism marketing expenditure	Journal article	19	10
11	Wadud (2013) Simultaneous modeling of passenger and cargo demand at an airport	Journal article	18	9
12	Strand (1999) Airport-specific traffic forecasts: the resultant of local and nonlocal forces	Journal article	14	12
13	Wadud (2011) Modeling and forecasting passenger demand for a new domestic airport with limited data	Journal article	15	8
14	Kressner & Garrow (2012) Lifestyle segmentation variables as predictors of home-based trips for Atlanta, Georgia, airport	Journal article	11	7
15	Ashley et al. (1995) A policy-sensitive traffic forecasting model for Schiphol Airport	Journal article	12	6
16	Uddin et al. (1985) Methodology for forecasting air travel and airport expansion needs	Journal article	13	5
17	Sun et al. (2019) Nonlinear vector auto-regression neural network for forecasting air passenger flow	Journal article	8	8
18	Sismanidou & Tarradellas (2017) Traffic demand forecasting and flexible planning in airport capacity expansions: Lessons from the Madrid-Barajas new terminal area master plan	Journal article	9	3
19	Xiao et al. (2016) Oscillations extracting for the management of passenger flows in the airport of Hong Kong	Journal article	7	5
20	Kawad & Prevedouros (1995) Forecasting air travel arrivals: model development and application at the Honolulu international airport	Journal article	7	5
21	Profillidis (2012) An ex-post assessment of a passenger demand forecast of an airport	Journal article	6	4
22	Jin et al. (2020) Forecasting air passenger demand with a new hybrid ensemble approach	Journal article	4	5
23	Liu et al. (2017b) Prediction of passenger flow at sanya airport based on combined methods	Conference paper	5	3
24	Hofer et al. (2018) Socio-economic mobility and air passenger demand in the U.S.	Journal article	4	3
25	Ferhatosmanoglu & Macit (2016) Incorporating explanatory effects of neighbour airports in forecasting models for airline passenger volumes	Conference paper	4	3
26	Suh & Ryerson (2019) Forecast to grow: Aviation demand forecasting in an era of demand uncertainty and optimism bias	Journal article	2	2
27	Li et al. (2018) Passenger flow forecast of Sanya airport based on ARIMA Model	Conference paper	3	1
28	Karasek (1982) Forecasting and Planning the Jeddah Air Traffic with a Mini Model	Journal article	2	2
29	Liu et al. (2017a) Prediction for passenger flow at the airport based on different models	Conference paper	2	1
30	de Paula et al. (2019) Forecasting passenger movement for Brazilian airports network based on the segregation of primary and secondary demand applied to Brazilian civil aviation policies planning	Journal article	1	-
31	Li & Jiang (2020) Airport Passenger Throughput Forecast Based on PSO-SVR Model	Conference paper	-	-
32	Ramadiani et al. (2020) Forecasting the number of airplane passengers uses the double and the triple exponential smoothing method	Conference paper	-	-
33	Rodriguez et al. (2020) Air traffic forecasting in post-liberalization contect: a dynamic linear models approach	Journal article	-	-
34	Zhang (2020) Research on forecasting method of aviation traffic based on social and economic indicators	Conference paper	-	-
35	Djakaria (2019) Djalaluddin Gorontalo Airport Passenger Data Forecasting with Holt's-Winters' Exponential Smoothing Multiplicative Event-Based Method	Conference paper	-	-
36	Lei et al. (2019) Aviation Business Volume Forecast of Xianyang International Airport Based on Multiple Prediction Models	Conference paper	-	-
37	Felkel et al. (2017) Hub airport 4.0 - How frankfurt airport uses predictive analytics to enhance customer experience and drive operational excellence	Book chapter	-	-

citations retrieved 25.11.2020

3.4 Forecasting methods

All of the publications empirically testing forecasting methods (n=35) applied quantitative approach. Of those, Suryani (2010) adopted a mixed model approach with the system dynamics model. Along with the quantitative methods, only Samagaio and Wolters (2010) also adopted opinion-based methods (Gardner & McKenzie model and Grubb & Mason model), classified as qualitative judgmental methods. No single article focused solely on qualitative methods, which indicates the quantitative approach is the preferred approach to forecasting airport passenger traffic in scholarly research. The methods were categorized according to Banerjee et al. (2020) (see figure 6, pp. 21) with minor modifications. Figure 13 aims to illustrate the development of methods used over the years.

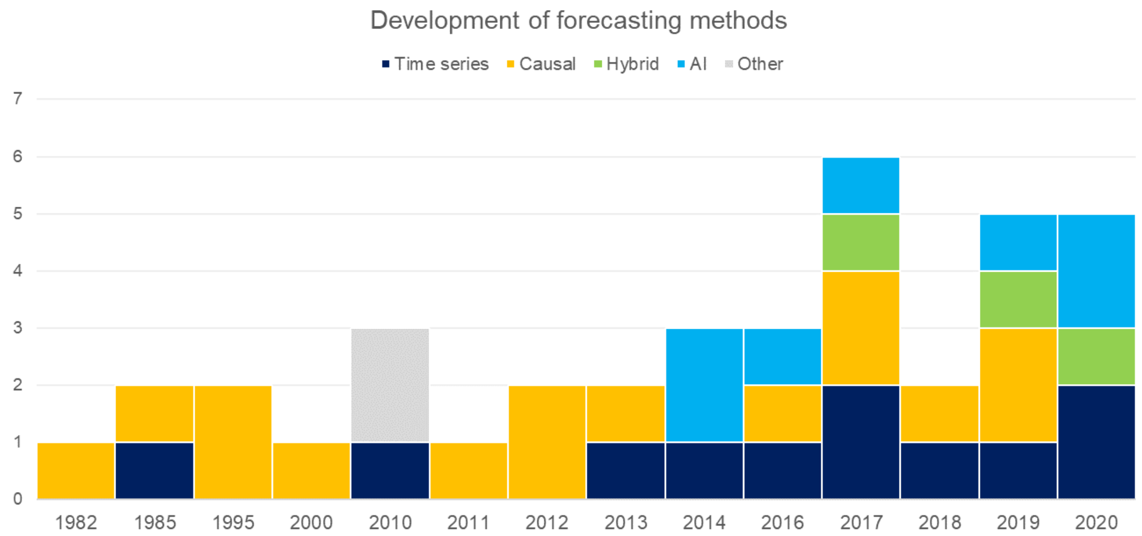

Figure 13. Development of forecasting methods

Time series models, which were applied in 11 publications, and causal methods (n=15) were the most applied methods. It is worth noting that publications that were considered adopting AI-based methods (n=7) may incorporate a time-series approach, but the model was not classified as a time-series or hybrid (n=3) if the model was using machine learning techniques, such as neural networks for time-series forecasting. Banerjee et al. (2020) did not have a category for hybrid quantitative methods, although they were recognized in the paper. However, such a category was added for better categorization since hybrid models incorporating different quantitative forecasting methods can be expected to become commonplace. Examples of hybrid approaches include Zhang (2020), who applied a combination of trend extrapolation and econometric model. Lei, Chong, and Long (2019) added a market share method in the model with trend extrapolation and an econometric model. Liu et al. (2017a) combined Holt-Winters seasonal prediction model with ARMA and unary linear regression. In addition to combining traditional methods, such as time series and causal methods, each AI-based forecasting model applied more than one method or approach, too. However, despite their hybrid approach, AI tools were categorized as their own.

Causal methods

Causal methods have traditionally dominated research and are still playing a significant role in estimating airports' future air passenger volumes. Karasek (1982) adopted an econometric regression model to forecast long-term annual passenger volumes of Jeddah Airport, Saudi Arabia. For predicting long-term annual passenger numbers, Ashley, Hanson, and Weldhuis (1995) introduced an econometric model that was in use at Amsterdam Schiphol Airport to support government decision-making. In addition to forecasting passenger volumes, the Competition Model was also used to forecast cargo demand, aircraft movements, airport costs and revenues, and compute the airport's economic impacts (ibid. 91). Using both time-series and causal methods, Uddin, McCullough, and Crawford (1985) used regression models to forecast long-term passenger volumes at the Robert Mueller Municipal Airport in Austin, Texas. The forecasts were used to assess airport expansion needs, a common reason for producing forecasts in the airport business.

It is typical to use causal models for long-term forecasting. In his first article, Wadud (2011) applied the gravity model and panel regression to forecast passenger demand for a proposed new airport near Khulna, Bangladesh, by using the aggregate national level and peer-airport data, as well as data obtained by a survey. In his other article, Wadud (2013) applied OLS and seemingly unrelated regressions (SUR) to forecast both long-term annual passenger and cargo demand at Hazrat Shahjalal International Airport in Bangladesh, finding SUR more suitable for forecasting future passenger demand for capacity expansion purposes.

Causal methods have also been under examination in ex-post assessments of demand forecasts. Sismanidou and Tarradellas (2017) assessed the past traffic forecasts in Madrid-Barajas Airport’s capacity expansion master plan and criticized the too simplistic approach of using GDP as the only predictor of passenger demand in their linear model. They advise against using too complex models, but too simple ones, too. They suggest using additional analysis of other demand-driving factors and including expert opinions in the forecasts. In construction planning, real options methodology may be helpful as well, Sismanidou and Tarradellas (ibid., 197) argue. Profillidis (2012) conducted an ex-post assessment on demand forecasts of Rhodes Airport in Greece. He compared the prediction accuracy of a linear regression model and polynomial second-degree calibration, of which the latter was found superior (ibid., 48). Previously, Profillidis (2000, 100) found the econometric and fuzzy linear regression models satisfactory in terms of predicting power when the exchange rate of Greek currency (drachma) compared to the currencies of origin countries of the passengers were used as the regressor. In his concluding remarks, despite using only one predictor (parity exchange rate), Profillidis (2012, 49) noted the issue with econometric models is the difficulty of predicting the future values of independent regressors.

Kawad and Prevedouros (1995) build country-specific regression models to forecast arriving passengers at Honolulu International Airport in short- to medium-term, which in their study translates to a forecasting horizon of 1-18 years. Similar to the previously introduced studies, annual passenger numbers were used. Public data was used for short-term forecasting (approximately two years), but for longer-term forecasts, Kawad & Prevedouros (1995, 23–24) used trend extrapolation with ARIMA and educated estimates to predict the future values of independent regressors. Thus, the model holds attributes of a hybrid model, although it was classified as causal models in this SLR. Kim & Shin (2016) were the others who adopted a short-term perspective in their forecasts with causal methods. They relied on regression models and big data from search engine queries to forecast monthly passenger volumes for Seoul Incheon International Airport. The findings of Kim & Shin (ibid., 107) indicate that key search engine queries, which there were 51, can be successfully used to forecasting air passengers eight-month away with a mean error of 5,30 percent.

De Paula, Silva, Vilela, and Cruz (2019, 25) consider two types of demand in their passenger demand forecasts at Brazilian airports: Primary demand, which is based on the needs of inhabitants living in the area of their closest airport, and; Secondary demand, which arises from the unmet demand of inhabitants living in another airport's area. De Paula et al. (ibid., 25–26) used OLS regression to estimate the primary demand of a specific airport and gravitational model to estimate secondary demand. They argue the model, which produced satisfactory forecasts, is suitable for deciding the location and timing of a new airport, analyzing potential routes, and estimating the potential impact on rival airports when a new airport is being planned (ibid., 28). There are several more demand-affecting aspects to consider when using causal methods. For example, socio-economic mobility has been found to negatively affect air passenger demand at US airports (Hofer, Kali & Mendez 2018, 93). Kressner and Garrow (2012) applied a least-square regression model to predict the number of trips for Hartsfield–Jackson International Airport in Atlanta, Georgia, that originated or terminated at the Atlanta Metropolitan residences area. Their findings indicate that using lifestyle-related variables, in addition to income, improved forecasting accuracy measured by adjusted R-squared. The lifestyle-related data was collected by a survey and using credit-reporting agency data, and, thus, they also concluded that nontraditional data sources are also suitable for accurate forecasts using causal methods (Ibid., 29).

Suh & Ryerson (2019) applied a method called reference class forecasting to estimate long-term passenger demand at multiple airports in the United States (US) and eliminate optimism bias in traffic forecasts. The basic idea of reference class forecasting lies in identifying peer airports by using econometric data and including their forecast errors to forecast future passenger volumes of an individual airport (ibid., 414). Such an approach has also been adopted by Ferhatosmanoglu and Macit (2016, 184), who found that forecasting accuracy can be improved by considering neighbor airports' traffic data in the models.

Time series

Ferhatosmanoglu and Macit (2016) applied traditional time series models for predicting passengers at Istanbul Ataturk and Ankara Esenboga airports in Turkey. Unlike in any other publication presented in this SLR that used monthly or annual data, Ferhatosmanoglu and Macit (ibid., 181) used hourly data in six-hour intervals. In addition to applying the ARIMA model, they applied TBATS-model (Trigonometric, Box-Cox transform, ARMA errors, Trend, and Seasonal components) to forecast an individual airport’s passengers without considering interactions. In addition to using traditional time series models, regression with ARMA errors was used to take into account neighbor airport traffic in a model. According to Ferhatosmanoglu and Macit (ibid., 178-179), TBATS can handle multiple seasonalities in data, making it often superior compared to ARIMA. The findings of the study indicated how TBATS was often able to outperform ARIMA and the dynamic linear model in terms of lower MAPE (ibid., 182–184).

ARIMA-based models are popular in air passenger traffic forecasting. Out of the 12 publications adopting ARIMA or one of its variations, four (Tsui et al. 2014; Tsui & Balli 2017; Li et al. 2018; Rodriguez et al. 2020) focused solely on them, four applied ARIMA models in comparison to others (Uddin et al. 1985; Samagaio & Wolters 2010; Ferhatosmanogly &

Macit 2016; Liu et al. 2017b), and four benefitted from ARIMA models as part of the final model (Kawad & Prevedouros 1995; Xie et al. 2014; Liu et al. 2017a; Jin et al. 2020).

Further developments of ARIMA include SARIMA (seasonal ARIMA), which is able to handle seasonal time series (see, e.g., Li et al. 2018), and ARIMAX (ARIMA with exogenous variables), which is also sometimes referred to as intervention model, due to its ability to consider intervention events or factors (such as SARS effect, oil price) in a time-series model (see, e.g., Tsui et al. 2014). The combination of this is called SARIMAX (seasonal ARIMA with exogenous variables), which has been implemented by, for example, Tsui & Balli (2017) in their attempts to accurately forecasting monthly international arrivals to Australian airports.

Like Ferhatosmanoglu and Macit (2016), Rodriguez et al. (2020) applied a dynamic linear model approach to forecasting medium-term passenger volumes for Bogotá-El Dorado International Airport, Colombia. Dynamic linear models are regression models, in which some of the regressors are defined as a function of time (ibid., 12). The reason for the model by Rodriguez et al. (2020) not being classified as the causal model is the chosen approach of defining all predictors by using individual ARIMA models. Thus, the model can be deemed as time-series based similar way as ARIMAX models.

Another common approach is to apply exponential smoothing methods. Such examples include Djakaria (2019), who, in his conference paper, applied multiplicative Holt-Winter's exponential smoothing to forecast monthly passengers for Djalaluddin Gorontalo Airport in Indonesia. Ramadiani, Syahrani, Astuti, and Azainil (2020) applied double and triple exponential smoothing methods to model monthly short-term passenger demand at Samarinda Airport, Indonesia. A multiplicative Holt-Winters was used as an independent model and part of the IMLEM (the integrated mixture of local experts model) model introduced by Scarpel (2013), which was used to predict the number of air passengers at São Paulo International Airport, Brazil. Samagaio and Wolters (2010) compared Holt-Winter's model to an ARIMA model and two judgmental opinion-based methods but did not find significant differences between Holt-Winter's and the ARIMA model. However, the judgmental methods overcame the forecast results of both time-series models (ibid., 216).

Hybrid methods

Classification into categories was not an easy task since multiple methods had features suitable for different classes. For example, it was typical to use time series models as independent regressors or refine time series models with exogenous variables. Thus, it was not always clear to define whether the models should be classified as causal or time series. In search of superior forecasting methods, scholars have lately come up with the idea of combining different statistical methods to produce more accurate forecasts.

Liu et al. (2017) combined Holt-Winter's exponential smoothing method with ARMA and unary linear regression to forecast monthly passengers at Sanya Phoenix International Airport, China. The final 1-year short-term forecast was produced as a weighted combination of the three independent models (ibid., 738). Lei et al. (2019) constructed a model combining trend extrapolation, econometric model, and market share method to produce medium-term forecasts of annual passengers for Xianyang International Airport in China. Zhang (2020) combined trend extrapolation and an econometric model to predict long-term passenger volumes for the Chinese Mianyang Airport.

AI-based methods

AI tools are becoming commonplace in research on airport passenger forecasting. Seven publications applied this modern approach. Xiao et al. (2014) introduced a neuro-fuzzy combination model, which combines singular spectrum analysis (SSA), adaptive-network-based fuzzy inference system (ANFIS), and improved particle swarm optimization (IPSO). The model was harnessed to produce 1-1,5 -year short-term forecasts of monthly air passengers for Hong Kong International Airport. The results indicated the model can successfully handle irregularities, high volatility, and seasonality in the data and was found superior compared to seven benchmark models (e.g., ARIMA and six other AI-based tools), which adopted a time-series approach (Xiao et al. 2014, 9–11). To illustrate this, an ARIMA model had a MAPE of 7,65 percent, while the proposed model had a MAPE of 1,53 percent. Hong Kong International Airport was also used as a case airport when Xiao et al. (2016) applied hybrid oscillations analysis model to forecast monthly passengers in a short-term horizon (one year). The model was based on SSA and fruit fly algorithm optimized generalized regressions neural networks (GRNN). The model reached an in-sample MAPE of 2,38 percent, while a stand-alone feedforward neural network had a MAPE of 8,39 percent (ibid., 76).

Once again, short-term forecasts were produced for Hong Kong International airport. This time two hybrid approaches were implemented by Xie et al. (2014): X12-ARIMA-LSSVR and TRAMO/SEATS-LSSVR. The models combine a least squares support vector machine (LSSVR) with two different seasonal decomposition methods, X12-ARIMA and TRAMO/SEATS (see Xie et al. 2014, 21–22 for more details and definitions). Xie et al. (ibid., 25–26 claim how hybrid AI methods can overcome the challenges of traditional time series and single AI-based tools. This assumption is supported by comparing models, in which the proposed model achieved MAPE of 3,4-4,6 percent and ARIMA 23,1-25,1 percent.

The other three AI-based hybrid methods were applied at Chinese airports to forecast air passengers in the short term. Sun, Lu, Tsui, and Wang (2019) introduced the MIV-NVARNN method to forecast air passengers at Beijing Capital Airport in a 1-6 months forecasting horizon. As a result, the model, which combines the mean impact value method with a non-linear vector auto-regressive neural network, outperformed the benchmark models (SARIMA, MLPNN (multi-layer perceptron neural network), VAR NVARNN, MIV-VAR) with high accuracy in each forecasting horizon: one, three and six months. In comparison, MAPE for SARIMA was 9,05 percent in one-month-ahead forecasting, while the proposed method produced forecasts with MAPE of 0,28 percent only. They wrap up their findings by describing how the multivariate forecasting approach improves forecasting accuracy compared to univariate methods and how AI-based methods outperform traditional methods, namely SARIMA, due to their ability to handle complex data. (Sun et al. 2019, 59.).

Li and Jiang (2020) applied their model, a combination of support vector regression (SVR) and particle swarm optimization (PSO), for Beijing Capital, Shanghai Pudong, Guangzhou Baiyun, and Chengdu Shuangliu airports. PSO-SVR model produced more accurate forecasts than traditional time series models ARIMA, exponential smoothing, and moving average methods (Li & Jiang 2020, 5). However, the performance of PSO-SVR was only slightly better than ARIMA. In their recent article, Jin et al. (2020) applied VMD-ARMA/KELM-KELM model at Beijing Capital, Guangzhou, and Shangai Pudong airports. The model is a combination of variational mode decomposition, ARMA, and kernel extreme learning machine. Once again, the model was able to outperform its peers, which there were 12 in total (ARIMA, ELM, PSO-SVM, to mention a few) (Jin et al. 2020, 9–12).

Felkel et al. (2017, 447) introduced a method used in Frankfurt Airport in Germany that combines multiple statistical methods such as decision trees, linear regression, and multiple imputation. The method is used to predict, for example, passengers onboard a specific flight number and at specific process points (ibid., 447–448). Although the method could also have been classified as a decision support tool (in the mixed models' category defined by

Banerjee et al. 2020, 799), the method was classified as an AI-based tool since it approaches passenger volume forecasting from the machine learning perspective. In the most cited article by Suryani et al. (2010), the system dynamics model is an example of a decision support tool providing a promising approach to passenger volumes forecasting and airport decision-making. The model considers internal and external factors and their causal relationships to forecast future passenger volumes for Taiwan Taoyuan International Airport (ibid., 2324; 2326). Although the model considers economic factors and their causality to passenger demand, the model itself is not considered a quantitative method per se, which are the focus of this research. However, due to visual illustrations and relatively easy interpretation, the article is suggested for those interested in understanding causal relationships of certain internal and external factors to passenger demand and airport capacity requirements. Table 5 on the next page has listed all the methods included in this SLR.

Table 5. Methods used in airport passenger traffic forecasting

Classification	Authors	Method(s)
AI	Li & Jiang (2020)	PSO-SVR (support vector regression machine based on particle swarm optimization)
	Jin et al. (2020)	VMD-ARMA/KELM-KELM (variational mode decomposition, ARMA, kernel extreme learning machine)
	Sun et al. (2019)	MIV-NVARNN (mean impact value based on nonlinear vector auto-regressive neural network)
	Felkel et al. (2017)	Combination of multiple statistical methods such as decision-trees, linear regression and multiple imputation
	Xiao et al. (2016)	Hybrid oscillations analysis model (singular spectrum analysis (SSA) and generalized regression neural network (GRNN))
	Xie et al. (2014)	Two hybrid approaches: X12-ARIMA-LSSVR and TRAMO/SEATS-LSSVR
	Xiao et al. (2014)	Neuro-fuzzy combination model (based on singular spectrum analysis (SSA), adaptive-network-based fuzzy inference system (ANFIS) and improved particle swarm optimization (IPSO))
Causal	Suh & Ryerson (2019)	Reference class forecasting
	de Paula et al. (2019)	OLS, gravitational model
	Hofer et al. (2018)	Three-stage least squares regression
	Sismanidou & Tarradellas (2017)	Linear regression
	Kim & Shin (2016)	Regression model
	Wadud (2013)	OLS regression, seemingly unrelated regression (SUR)
	Profillidis (2012)	Linear model, polynomial second-degree model
	Kressner & Garrow (2012)	Least-square regression
	Wadud (2011)	Gravity model, panel regression
	Profillidis (2000)	Fuzzy regression
	Kawad & Prevedouros (1995)	Cochrane-Orcutt regression procedure
	Ashley et al. (1995)	Econometric model
	Karasek (1982)	Econometric model
Decision support tool	Suryani (2010)	System dynamics
Hybrid	Zhang (2020)	Combination of trend extrapolation and econometric model
	Lei et al. (2019)	Trend extrapolation-econometric-market share method
	Liu et al. (2017a)	Combination of Holt-Winters seasonal prediction model, ARMA and unary linear regression
Time series	Rodriguez et al. (2020)	Dynamic linear model
	Ramadiani et al. (2020)	Exponential smoothing (double and triple)
	Djakaria (2019)	Holt-Winters' exponential smoothing multiplicative event-based method
	Li et al. (2018)	ARIMA
	Tsui & Balli (2017)	SARIMA, SARIMAX, SARIMAX/EGARCH volatility model
	Ferhatosmanoglu & Macit (2016)	ARIMA, TBATS, regression with ARIMA errors
	Tsui et al. (2014)	SARIMA, ARIMAX
	Scarpel (2013)	Integrated mixture of local experts model
Time series & causal	Liu et al. (2017b)	ARMA, Grey prediction model, regression model based on ARMA (RE-ARMA)
	Uddin et al. (1985)	ARIMA, multiple linear regression
Time series & judgement	Samagaio & Wolters (2010)	Holt-Winters, ARIMA, Gardner & McKenzie model, Grubb & Mason model

3.5 Forecasting horizons and the number of observations

Different methods were applied for different purposes and time-spans. According to Suh and Ryerson (2019, 402), machine learning algorithms, classified as AI tools, are more suitable for short-term forecasts, even as short as daily forecasting (Suh & Ryearson, 402). Such very short-term forecasting is not limited to AI tools only. Ferhatosmanoglu and Macit (2016, 181) applied ARIMA and TBATS to predictions as short as six hours. Figure 14 illustrates how the publications with a time-series approach (n=19) have typically been used for short-term forecasting, while the models not applying the time-series approach (n=14, i.e., causal econometric models) typically were used for long-term forecasting.

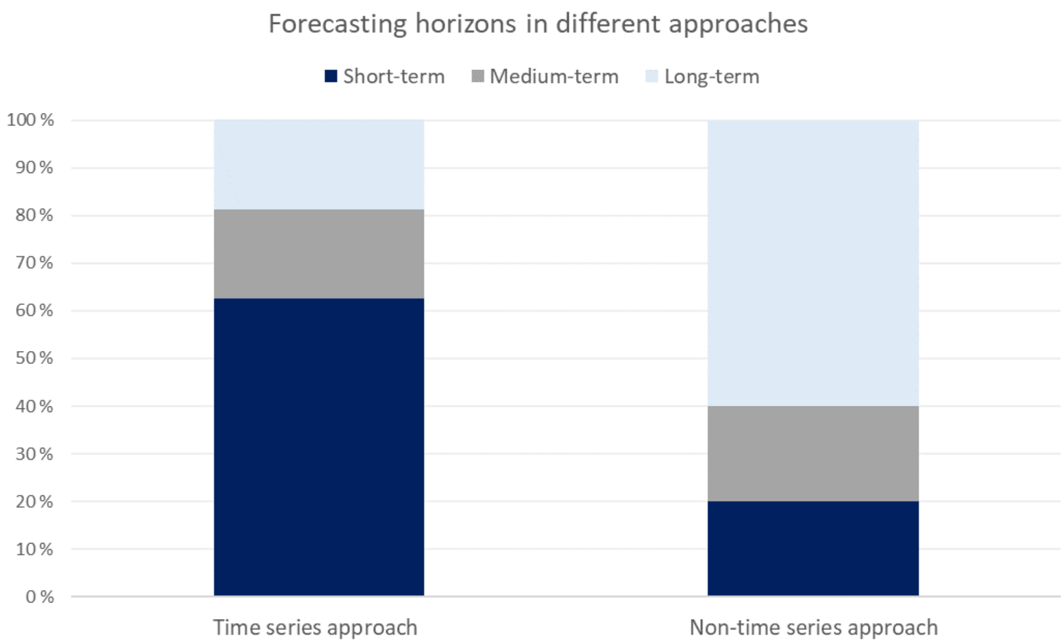

Figure 14. Forecasting horizons of methods applying time series and non-time series approaches

There is significant variation between forecasting horizons. Figure 15 illustrates forecasting horizons in years categorized by author-defined classes: short-, medium-, and long-term. The forecasting horizon was categorized in 15 publications. Some of the authors produced forecasts for more than one horizon. In such cases, for analysis purposes, the horizon was defined as the maximum forecasted. For example, the short- and long-term forecasting horizons of two and 3–7 years of Scarpel (2013) were combined and defined as long-term forecasting horizon of seven years. The short- to medium-term horizon of 1-18 years applied by Kawad and Prevedouros (1995) was translated as the medium-term horizon of 18 years.

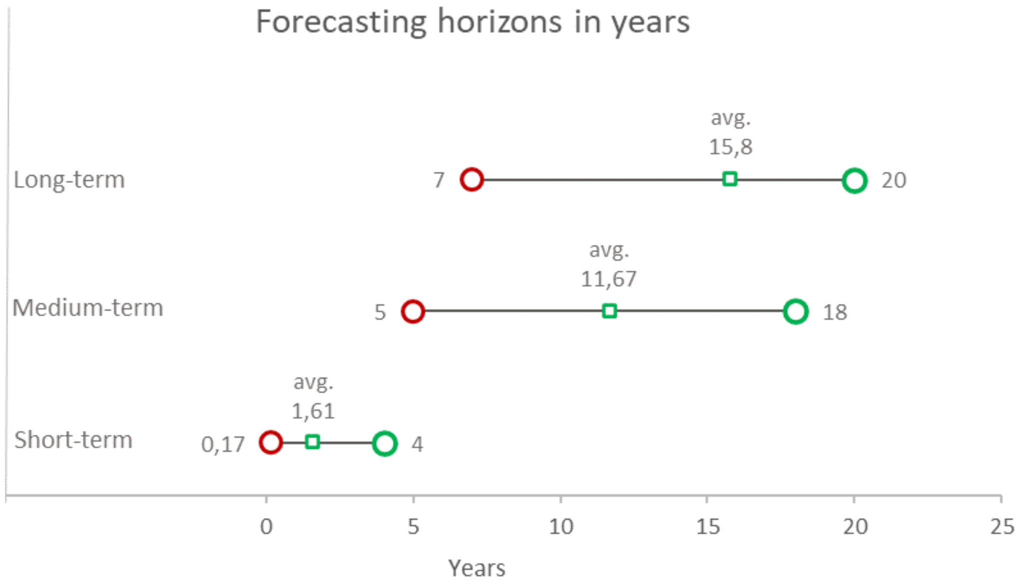

Figure 15. Forecast horizons in years, as defined by the authors (N=15)

As the figure illustrates, short-term forecasts (n=7) were produced for a minimum of one to two months (Jin et al. 2020) and a maximum of four years (Xie et al. 2014). The average horizon for short-term forecasts was 19-20 months. The average horizon for the author-defined medium-term forecast (n=3) was approximately 11,5 years, and close to 16 years in long-term forecasts (n=5). As the results show, it is often unclear whether a forecast horizon should be classified as long- or medium-term. The remaining 20 publications that were either empirically testing or discussing a model were classified according to the theory and practice, where able. The longest forecasting horizon, 27 years, was discussed by Sismanidou & Tarradellas 2017, who conducted a post-assessment on passenger demand forecasts included in the master plan of Madrid Barajas International Airport.

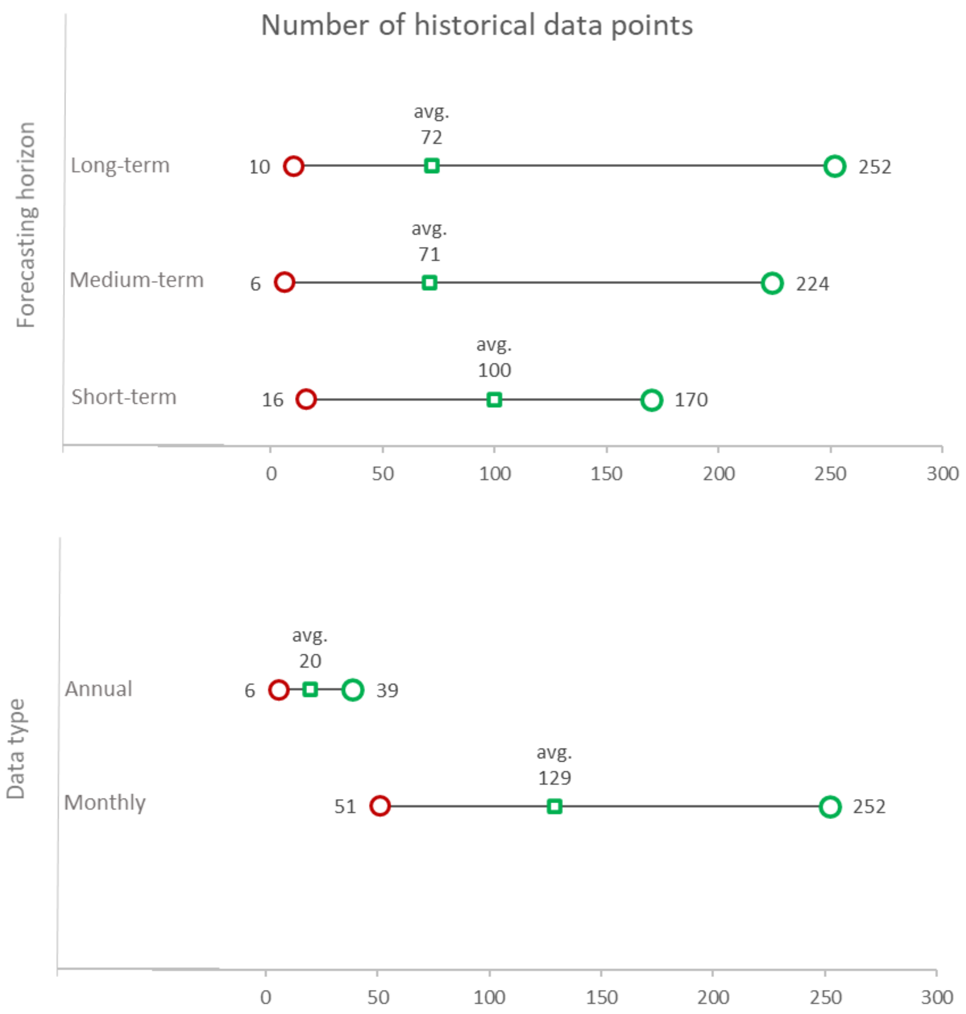

Figure 16. Number of historical data points of monthly and annual data

To support empirical analysis, the number of observations for each forecasting horizon category was examined (see figure 16). Excluding the study by Ferhatosmanoglu and Macit (2016), who used hourly data in six-hour intervals in their short-term forecasting model, the rest of the publications (n=31) used monthly or annual data. Those studies that did not use time-series or panel data were not included. On average, data from 133 points of time was used. However, the maximum number of historical observations was 1460 (an outlier), which was applied in the study of Ferhatosmanoglu and Macit (2016). The 1460 six-hour observations translate to one year of data. When the study with hourly data was excluded, the average number of point-of-time observations dropped to 86, the minimum being six and the maximum 252. Figure 16 also illustrates how it is typical to include more historical data points, 129 on average, into analysis when monthly data is used. Models with annual data tend to use significantly fewer observations to forecast the future. In addition to these, it was found out that the models applying the time-series approach (n=19) tend to use a greater number of historical observations, 109 on average, than the models not applying the time-series approach (38 point-of-time obs.).

3.6 *Variables and performance measures*

In addition to considering historical passenger numbers to produce forecasts, several other variables were used as well. The variables that were used more than once are shown in table 6. Out of the 76 unique variables detected, GDP was the most used one, applied in the quarter (24 %) of the publications. The findings indicate that GDP, population, GDP per capita, and currency exchange rates would be the most suitable predictors for air passenger demand. The four most used variables were also the most popular in the models that incorporated a time-series approach. Notably, in ten of the publications which used a time-series approach, no exogenous variables were used in addition to historical air passenger volumes.

Table 6. Variables used in the models

Variable (N=76)	Frequency
GDP	9
Population	7
Currency exchange rate	4
GDP per capita	3
Dummy variable for Gulf War	2
Crude oil price	2
Herfindahl index	2
…	…
…	…
Total	**98**

One of the interest areas was to understand how market shocks or significant intervention events have previously been handled in the models. Surprisingly, only five publications (out of 37) had taken into account exogenous shocks and interventions. In their article, Tsui and Balli (2017, 423) used exponential general autoregressive conditional heteroscedasticity (EGARCH) methodology to conclude that both positive and negative exogenous shocks affect air travel demand at Australian airports. In their other article, Tsui et al. (2014) applied the ARIMAX model to include shocks and interventions when forecasting passenger volumes for Hong Kong International Airport. They included SARS[9] dummy to note the outbreak during a period from November 2000 until July 2003, a dummy variable to account for the effects of cross-strait agreement in and after 2008, an oil dummy to get value "1" when the crude oil price was more than 80 US dollars, zero otherwise, and a dummy variable to account for the effects of Individual Visit Scheme (IVS) after 2003 (ibid., 70). Without further description, Xiao et al. (2014, 10) briefly note that their AI-based model, which also predicts passengers for Hong Kong International Airport, is able to control the intervention of the financial crisis in 2008.

In his fuzzy regression model, Profillidis (2000) included a dummy variable to take into account the 1991 Persian Gulf War. The Gulf War was also noted in a regression model of Kawad and Prevedouros (1995), who attempted to forecast arriving passengers for Honolulu International Airport from Japan and the USA. Similar to the model of Profillidis (2000), The Gulf War was accounted for by a dummy variable, getting a value of 1 for 1991 and 0 otherwise. The recessions of 1983 and 1991 were noted similarly in the model, predicting volumes of arriving passengers from Japan. The model predicting passengers from the USA included dummy variables for the United Airlines strike in 1985 and Hurricane Iniki in 1992. (Kawad & Prevedouros 1996, 20-21).

A wide variety of different performance measures have been used. Surprisingly, approximately half (n=14, 52 %) of the publications in which the forecasting performance was evaluated (N=27) used only a single error measure. On the contrary, the majority (n=9, 69 %) of the remaining publications used more than two error measures. Based on findings, it seems like a rule that more than one error measure is used when AI-based methods are used. Out of the seven publications classified using AI-based methods, six introduced their error measures, and out of those, five relied on more than two measures. When only a

---

[9] Severe Acute Respiratory Syndrome

single variable was used, it was typically MAPE (n=6, 46%), R-squared (n=2, 15%), or percentage error rate (n=2, 15%). The greatest number of error measures was used by Sun et al. (2019), who applied six different error measures to evaluate the performance of their AI-based method. This was followed by Xiao et al. (2014), Xiao et al. (2016), and Jin et al. (2020), who used four measures for their AI-based tools. Suh and Ryerson (2019), with their causal model, also evaluated performance through four measures. The complete list of error of different error measures is presented in table 7, which shows MAPE, MAE, and RMSE being the most used ones.

Table 7. List of error measures

Error measure (N=27)	Frequency
MAPE	15
MAE	6
RMSE	6
R^2	3
Error rate (percentage)	2
Adjusted R^2	2
Directional change (DC)	2
Absolute percentage error (APE)	1
MSE	1
Mean growth-based forecast error (MGBFE)	1
Enhanced Mean Peer-Based Forecast Error (EMPBFE)	1
Directional Symmetry (DS)	1
Mean forecast error (MFE)	1
Mean Peer-Based forecast Error (MPBFE)	1
D_{stat}	1
Mean Square Deviation (MSD)	1
Mean square prediction error (MSPE)	1
Deviation from actual	1
Normalized root mean squared error (NRMSE)	1
Percent error rate (deviation from actual value)	1
Pesaran-Timmermann (PT) statistic	1
Diebold-Mariano (DM) statistic	1
Root mean square percentage error (RMSPE)	1
MASE	1
Standard errors	1
Mean Absolute Deviation (MAD)	1
Mean error	1
Total	**56**

The SLR provided valuable insights which influenced the decisions made in the empirical part of this research. For example, time-series were considered the most suitable methods methodologically to produce short-term forecasts. On the other hand, due to the varying performance of different methods, different tools, from traditional to more sophisticated AI-based tools, were considered. Evaluation of the amount of data used suggested that five years of daily data (1825 observations) would be sufficient to train the models and validate the performance. On average, 129 historical data points were used with monthly data, and thus, 11 years of monthly data (132 observations) was considered a suitable choice for projections with monthly data. Moreover, the literature suggests using more than one error measure. The practice was adopted in this thesis, too.

4. Data and methodology

4.1 Data overview

The airport operator Finavia Corporation has provided daily and monthly traffic data for this non-commissioned research. Daily data was collected from January 2016 until December 2020, monthly data from January 2010 until December 2020. The dataset includes arriving, departing, and transfer passengers for international and domestic routes from commercial flight operations (scheduled, charter, air taxi). Roser, Ritchie, Ortiz-Ospina and Hasell (2020) have maintained publicly available datasets of policy responses to the coronavirus pandemic on OurWorldInData.org. Datasets of interest, which are updated daily, are International Travel Controls and the Government Response Stringency Index, based on nine individual measures. Data of COVID-19 cases, along with the demographic information, will be retrieved from the same data source.

Figure 17 on the next page illustrates monthly passenger time-series data, including and excluding the COVID-19 crisis. We can see how the data contains evident trend and seasonality, thus indicating the data is non-stationary. In addition to visually interpreting the pre-covid data (1/2010-12/2019), it can be confirmed by applying Augmented Dickey-Fuller Test (ADCF), where the null hypothesis is that the data is stationary. ADCF returns a DF test statistic of -0,94 with lag order 12 and a p-value of 0,94, which confirms our interpretation that the time-series is non-stationary.

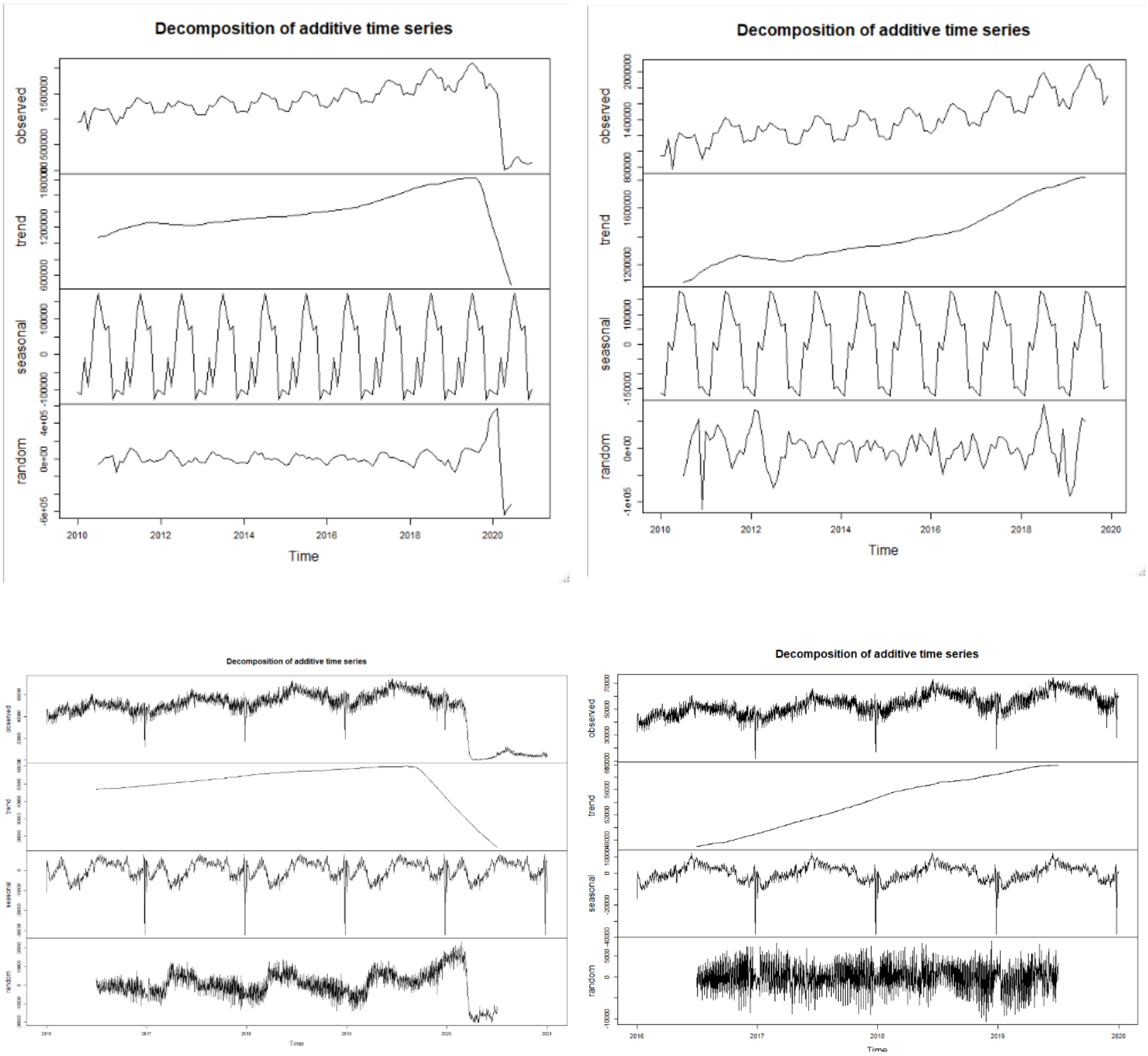

Figure 17. Decomposition of monthly (upper graphs) and daily time series (bottom)

Before the crisis, the passenger numbers at Helsinki Airport had a relatively constant upward trend with annual seasonality. Passenger numbers tend to peak during summer and lower during winter months. The figure also illustrates the development of daily passengers at Helsinki Airport. Excluding the fall in passenger demand at the beginning of the year 2020, an upward trend with annual and weekly seasonality can be identified (see figure 18). ADCF test returns with a test statistic 2,69 with lag order 365 and with a p-value over 0.99, confirming our assumption of non-stationary time series. The findings indicate non-constant variance and mean.

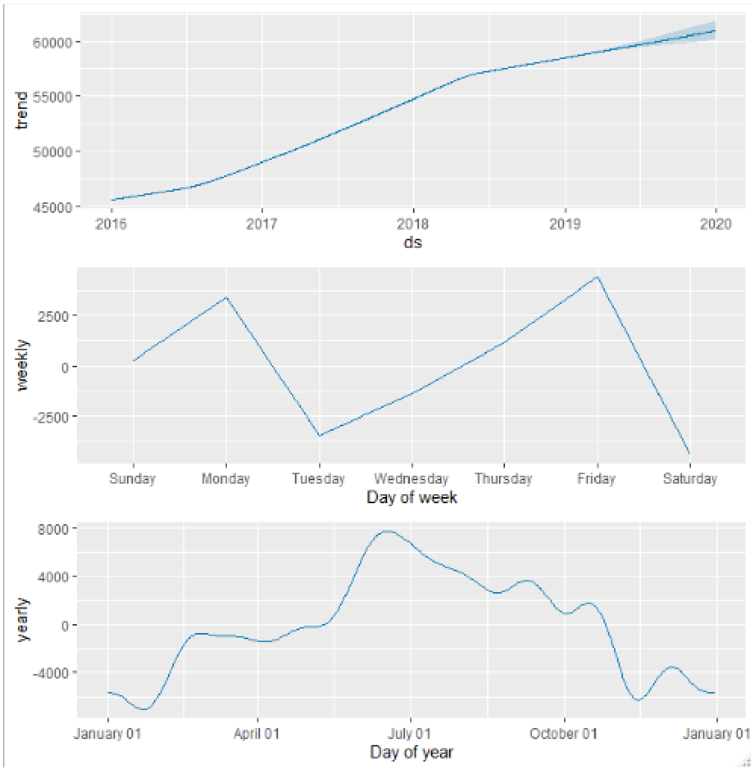

Figure 18. Seasonalities in daily data (2016-2020, pre-covid) recognized by Prophet

The number of COVID-19 cases and other pandemic-related variables were sourced from the dataset maintained by the University of Oxford on OurWorldInData.org. Data from Europe was filtered and included in this research. Limiting the data to Finland would have been troublesome because a significant share of passengers, 38,6 % in 2019, were international transfer passengers (Finavia 2020). Therefore, the COVID-19 cases in Finland and government measures restricting entry to Finland may not be considered suitable drivers for evaluating traffic development at Helsinki Airport. On the other hand, the worldwide context was not deemed suitable since the policy responses and COVID-19 situation varies between continents, and not all world events are equally impacting traffic at Helsinki Airport. Since the airport serves as a popular transfer airport for traffic between Europe and Asia (Finavia 2020), European events were considered the most significant factor for influencing passenger demand at the Nordic airport

Daily data of COVID-19 cases was aggregated from 46 European countries[10] , and the total daily population of those countries was used to calculate cases per 100 000 inhabitants. 14-days incidence rate, the metric determining entry conditions to Finland, was calculated for Europe as the moving sum (14 days) of daily cases per 100 000 inhabitants. Figure 19 illustrates the daily development of both the incidence rate and passengers at Helsinki Airport. The figure, together with the correlation analysis (see figure 23, pp. 65), illustrates how the number of COVID-19 cases has little to do with passenger numbers per se. Instead, the crisis as a whole is the evident driving force for traffic development. Therefore, two measures to reflect governments' policy responses were considered instead.

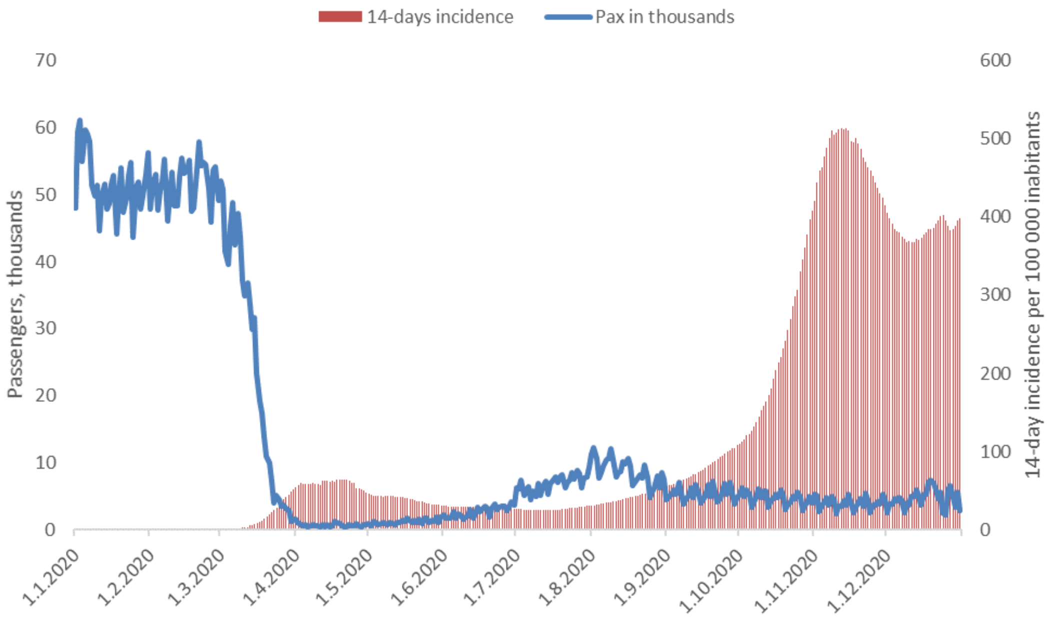

Figure 19. Daily Helsinki Airport passengers and daily 14-days COVID-19 incidence per 100 000 inhabitants mapped.

International Travel Controls (travel controls, TC) and the Government Stringency Index (GSI) are the two measures among 17 others developed to reflect government policy responses to the COVID-19 pandemic. TC is an ordinal indicator with five stages: 0 indicating

[10] Listed as European in the dataset: Albania, Andorra, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Kosovo, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Moldova, Monaco, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Russia, San Marino, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine, United Kingdom, Vatican

no measures, 1 indicating screening, 2 indicating quarantine from high-risk regions, 3 for a ban on arrivals from some regions, 4 indicating the ban on all regions or total border closure, and blank for cases where no data is available (Hale et al. 2020, 22). GSI, on the other hand, combines nine ordinal policy indicators: school closings, workplace closings, cancellation of public events, restrictions on gathering size, closure of public transport, stay at home requirements, restrictions on internal movements, restrictions on international travel (TC), and public information campaign (ibid., 4). The value varies between 0 and 100, the latter indicating the strictest response. Figure 20 illustrates the correlation between the two policy metrics, which is 0,97 for 1st January – 31st December 2020 and 0,90 for the period 1st April – 31st December 2020, excluding the first three months of the crisis. The latter eliminates the strong effect when the government lockdowns spread across Europe at the beginning of the crisis.

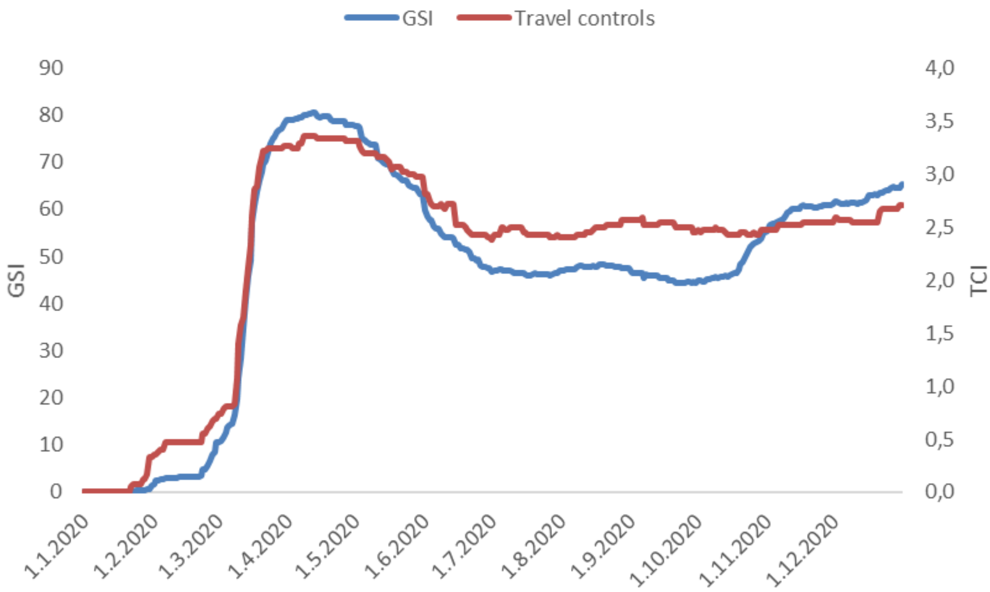

Figure 20. Correlation between GSI and TC

Since the dataset included missing values, especially at the beginning of 2020, few countries were influencing the GSI and TC index for the whole of Europe, while in reality, the majority of the European countries had not yet introduced harsh measures against the virus. Therefore, missing values were coded to 0 (no measures) until the measures were identified by the team (Hale et al. 2020) responsible for data collection. In those cases, where a value for an indicator existed for a day but was missing for the upcoming days, the missing values were replaced by previously observed values. European TC was calculated from the average of 42 countries[11] that had observations in the dataset. Thus, all countries had equal weight in the aggregated indicator. There was a strong correlation between both GSI and TC and passenger numbers. GSI had a -0,92 correlation with passenger numbers (pax) with full-year data, while TC had an even stronger correlation of -0,96. When the first three months of 2020 were eliminated, GSI was still correlating strongly with pax (-0,74). TC once again had a slightly better correlation (-0,75). The correlation between pax and TC is illustrated in figure 21.

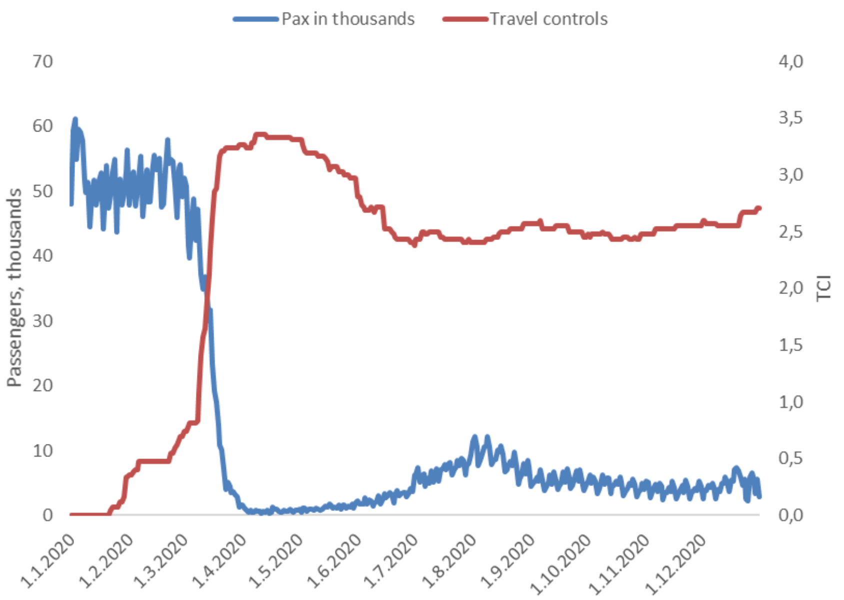

Figure 21. Correlation between Helsinki Airport passengers and TC

With monthly data, GSI and TC were averaged for the month since only the end-of-month values have not affected the passenger numbers for the entire month. Thus, average values take into account also development within the month. With the full-year (2020) dataset, passenger numbers strongly correlated with GSI (-0,92) and TC (-0,96). When the first three

[11] All the countries included in the COVID-dataset except Liechtenstein, Montenegro, North Macedonia, and Vatican.

months were excluded, the correlation was still strong and significant (at 0.95 confidence level) between pax and GSI (-0,79) and TC (-0,80).

4.2 Empirical research methodology

In order to thoroughly understand the performance of the selected models, monthly and daily data were used. The intention was to first test the performance during regular times, before the COVID-19 crisis. For this purpose, monthly data from January 2010 until December 2018 was assigned as a training set and data from January 2019 until December 2019 as a testing set to evaluate forecasting performance. No in-sample performances were evaluated in this thesis. Pre-covid performance with daily data was tested with the same period of time, from 1st January to 31st December in 2019. The training set was composed of three years of daily data from January 2016 until December 2018.

The other aim was to evaluate performance during the COVID-19 crisis. For this purpose, daily data from 1st April 2020 until 30th September 2020 was used for training and the remaining months of 2020 for validating the performance. The reason for selecting this period was to exclude historical development prior to the crisis and the major fall in passenger numbers in March, and, instead, consider only the period of “new normal”. Since enough observations could not be collected to train the models with monthly data from 2020 only, it was decided to extend the training set from January 2010 until September 2020 and include the major fall.

To determine *the accuracy of selected forecasting methods during the COVID-19 pandemic,* five automated forecasting methods were initially considered: ARIMA, TBATS, Facebook’s Prophet, multilayer perceptron (MLP), and extreme learning machine (ELM). ARIMA models and TBATS represent more traditional forecasting methods, former commonly used in air passenger demand forecasting. ARIMA and its variations are generally unable to handle data with multiple seasonalities (Taylor & Letham 2018, 38). However, to test this claim, ARIMA models were also introduced to daily data. The most suitable ARIMA model was selected by the **auto.arima** function with default settings in the ‘forecast’ package in statistical program system R (Hyndman et al. 2020). The selection procedure of this automated function is described in the article by Hyndman and Khandakar (2008). Following

Ferhatosmanoglu and Macit (2016), TBATS was selected to deal with both weekly and annual seasonalities in daily data. The model fitting was performed with default settings of the **tbats** function included in the 'forecast' package, which follows the selection process described in De Livera, Hyndman, and Snyder (2011). TBATS was also introduced to monthly data.

Facebook's Prophet is a relatively new automatic forecasting method introduced to the public in 2017. According to Taylor & Letham (2017), it is optimized for business forecasting and can handle multiple seasonalities (see figure 19), data with outliers and missing values, and automatically detect changes even in non-linear trends. The forecast was fitted in R by using package 'prophet' and its default settings. Although the method allows its user to modify, for example, trend change points manually, significant modifications were not done. Prophet has three main components: trend, seasonality, and holidays and it can be presented mathematically with the formula 6. A detailed walkthrough of the model is presented in the paper of Taylor & Letham (2018).

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{6}$$

where,	g(t)	is the trend function,
	s(t)	represents seasonality,
	h(t)	represents irregularly occurring events (user defined), and
	ϵ_t	is the error term.

In addition to Prophet, two neural networks were included in the comparison due to their increasing popularity in forecasting: multilayer perceptron (MLP) and extreme learning machine (ELM), both having one hidden layer. The methods were implemented using the 'nnfor' package in R, which allows automatic time series modeling with neural networks (Kourentzes 2019a). The standard number of repetitions is 20, but the models were trained 200 times each to improve forecasting accuracy. The forecasts were combined by using the median operator, the standard argument in the R function. For MLP, the standard setting is five neurons in a hidden layer, which was not changed. According to Kourentzes (2019b), ELM automatically specifies the hidden layer. However, it applies a shrinkage estimator to estimate weights, which means only certain nodes are connected to the output layer to contribute to the forecast (see figure 22, where black lines denote connected nodes). The grey input nodes in figure 22 represent autoregressions, and the red ones are seasonalities. Additional regressors would be shown in blue. (Kourentzes 2019a.) Huang, Zhu, and Siew (2006, 499) argue how ELM can learn faster than those using the back-propagation learning algorithm (MLP, for example) and can produce more accurate forecasts by being able to deal with several issues such as overfitting.

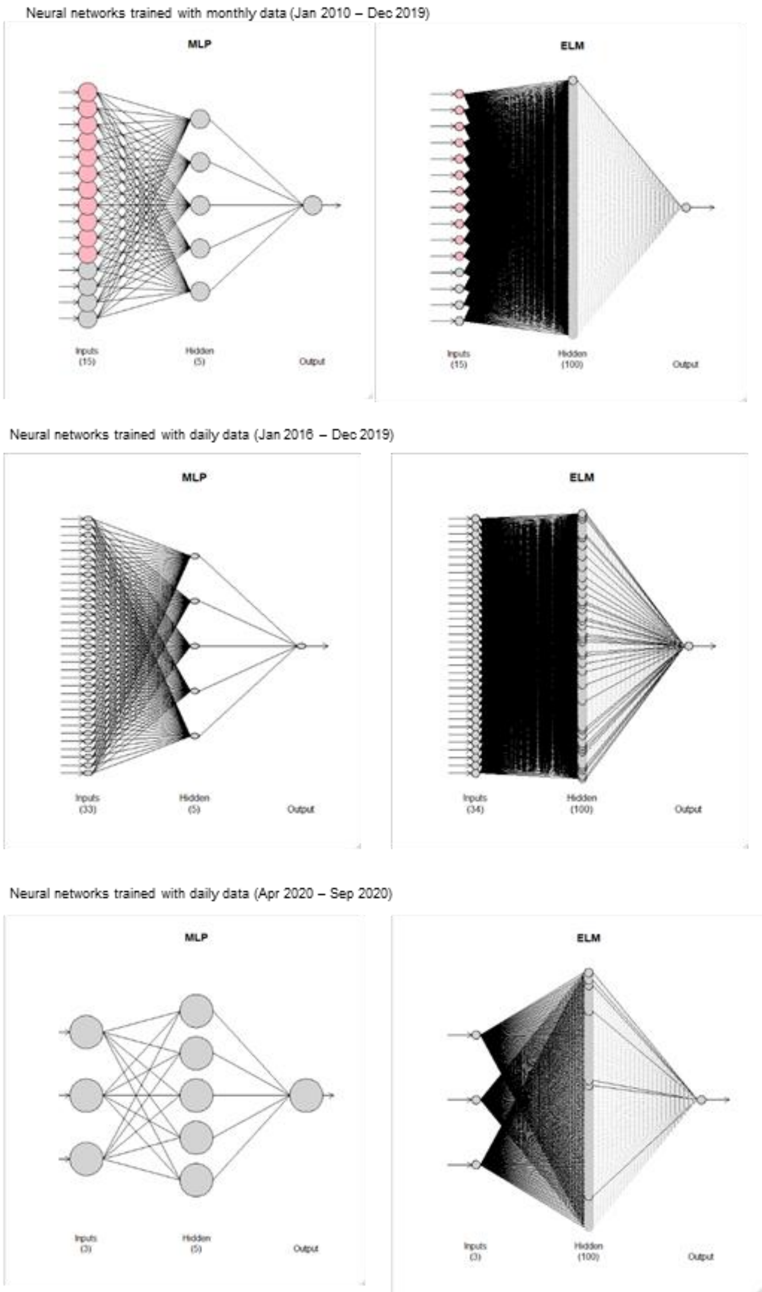

Figure 22. Neural networks trained with different training sets

Although the COVID-19 14-day incidence per 100 000 inhabitants correlated slightly (-0,37) with Helsinki Airport passenger numbers (PAX) in 2020, the variable was excluded since the number of cases, and thus, the incidence rate, are highly dependent on the testing capacity. As figure 23 illustrates, the correlation was non-existent when data from April until the end of December 2020 was used. That is, there is no evidence for the correlation between the number of COVID-19 cases and air passenger volume per se. Instead, GSI and TC were considered more suitable measures. Both GSI and TC correlated strongly (GSI -0,92, TC -0,96) with daily passenger numbers. During the “stable phase” of the crisis from April to December 2020, the correlation was still strong and significant (-0,74 for GSI/PAX, -0,75 for TC/PAX).

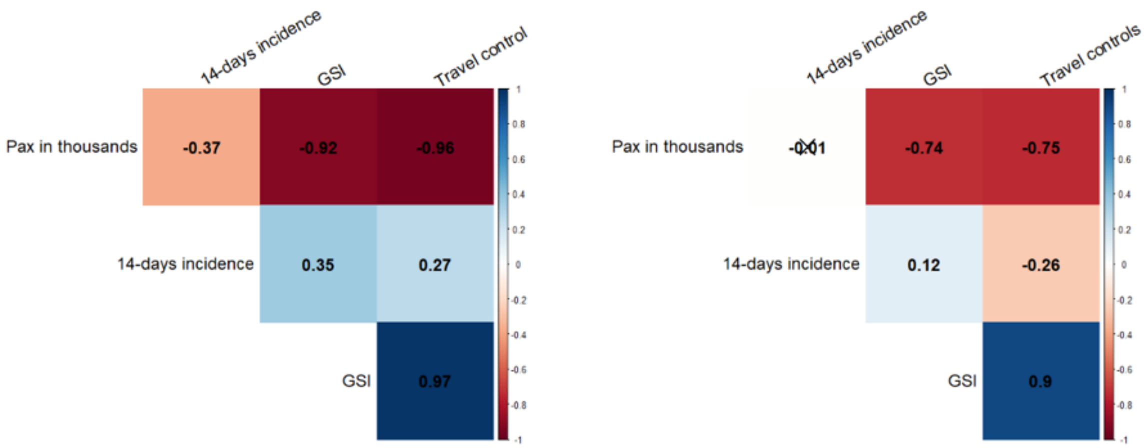

Figure 23. Correlation matrices for COVID-19 related variables (1-12/2020 data on the left, 4-12/2020 data on the right, “x” denotes non-significant correlation)

Figure 23 also illustrates how the 14-day incidence per 100 000 inhabitants has little to do with GSI and TC. There was only a low positive correlation between the incidence rate and the two policy measures when the whole-year data was examined. Even a lower positive correlation between the incidence rate and GSI was recognized when data from April to December 2020 was explored. The incidence rate turned into a negative correlation with TC. Meanwhile, a very strong correlation between GSI and TC was observed in both datasets. Since GSI is an aggregate measure of nine different measures, it is not easy for the analyst to estimate it in the forecasts. Therefore, more intuitive TC was selected as the exogenous variable representing the pandemic-related variable in the models.

5. Forecasting airport passenger volumes during the pandemic

5.1 Accuracy of the models

One of the aims of this research was to evaluate the predictive power of selected forecasting methods during the pandemic. The models were first run with daily and monthly datasets prior to the COVID-19 crisis to understand their predicting power during regular times. Since the pandemic started to show its first signs at the end of 2020, the selected forecasting periods were the first three, six, and twelve months of 2019. The training period was then extended until September 2020, and forecasts were produced for the last three months of 2020. The monthly forecast results are shown in Table 8, where the most accurate method for each forecasting horizon is bolded.

Table 8. Forecasting accuracies with monthly data

		3 MONTHS					6 MONTHS					12 MONTHS				
	Error measure	ARIMA	PROPHET	TBATS	MLP	ELM	ARIMA	PROPHET	TBATS	MLP	ELM	ARIMA	PROPHET	TBATS	MLP	ELM
Training set 01/2010 - 12/2018	**MAPE**	.0446	.0753	**.0332**	.0479	.0700	**.0242**	.0401	.0430	.0296	.0452	**.0294**	.0437	.0459	.0306	.0424
	MAE	71 370	120 453	**52 976**	76 478	111 699	**39 391**	64 901	77 650	49 288	75 942	**49 346**	73 635	86 217	52 645	73 340
	RMSE	71 642	122 253	**54 076**	77 620	113 633	**50 966**	86 763	82 571	58 250	86 072	66 453	94 736	95 079	**64 279**	86 303
Training set 01/2010 - 9/2020	**MAPE**	N/A	2.8563	N/A	N/A	**.0998**	-	-	-	-	-	-	-	-	-	-
	MAE	N/A	394 709	N/A	N/A	**12 809**	-	-	-	-	-	-	-	-	-	-
	RMSE	N/A	422 911	N/A	N/A	**15 007**	-	-	-	-	-	-	-	-	-	-

Based on MAPE, all of the methods performed comparably well before the crisis. TBATS seemed to perform the best in the short term, while Prophet's performance was the worst compared to others. However, surprisingly, the longer the forecasting period, the lower the prediction error. SARIMA (*ARIMA (0,1,1)(1,1,0)[12]*) outperformed other methods in predictions both six months and one year ahead (see figure 24). The lower prediction errors during the summer months may explain the lower overall prediction error for the forecasts of six and twelve months for all other methods except TBATS, which performed better at the beginning and end of the year (see appendices 2 and 3 for detailed outputs).

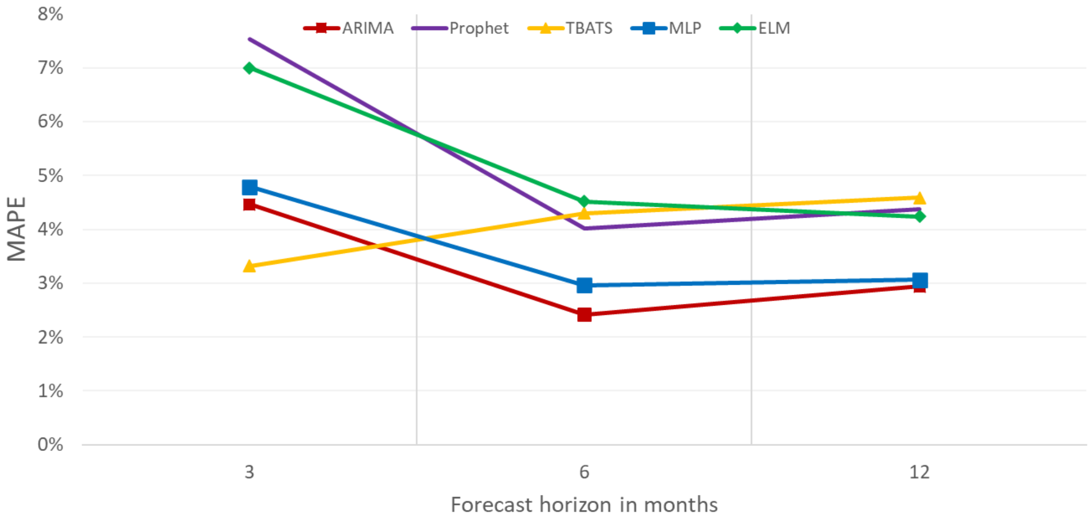

Figure 24. Forecasting errors (MAPE) before COVID-19 in 2019 (monthly data)

Unsurprisingly, more traditional methods, SARIMA and TBATS, were unable to produce usable forecasts when the complex data was introduced, and forecasts were produced for October until December 2020, the months amid the COVID-19 crisis. Both methods returned negative values. Since Prophet first returned negative values, its trend flexibility was tuned from 0.05 to 0.5 following an example on Prophet's documentation (Prophet n.d.). This resulted in Prophet producing too optimistic forecasts and thus resulted in MAPE 2.86 (286 %). However, it is to be noted that other settings were not modified, and thus, the real potential of Prophet was not fully covered – the method allows its user to manually adjust and add points where the trend changes occur, allowing more flexibility to account for significant events. While MLP produced forecasts with negative values, ELM returned surprisingly accurate results: a three-month forecast with 10% MAPE.

Figure 25 visually shows forecasts in both scenarios, prior and during COVID-19. The results indicate how traditional methods are powerful enough to produce high-quality forecasts during regular, stable times. However, they heavily underperform when data starts containing complexities. While MLP with five hidden nodes in one layer seemed too simple to forecast with complex data, an increase in the number of nodes increased forecasting accuracy. ELM provided promising results by its comparably high predicting power during the crisis. It should be noted that the results should be considered with caution since the models were not validated with other data than described above. Thus, with different training or testing sets, the results may be different.

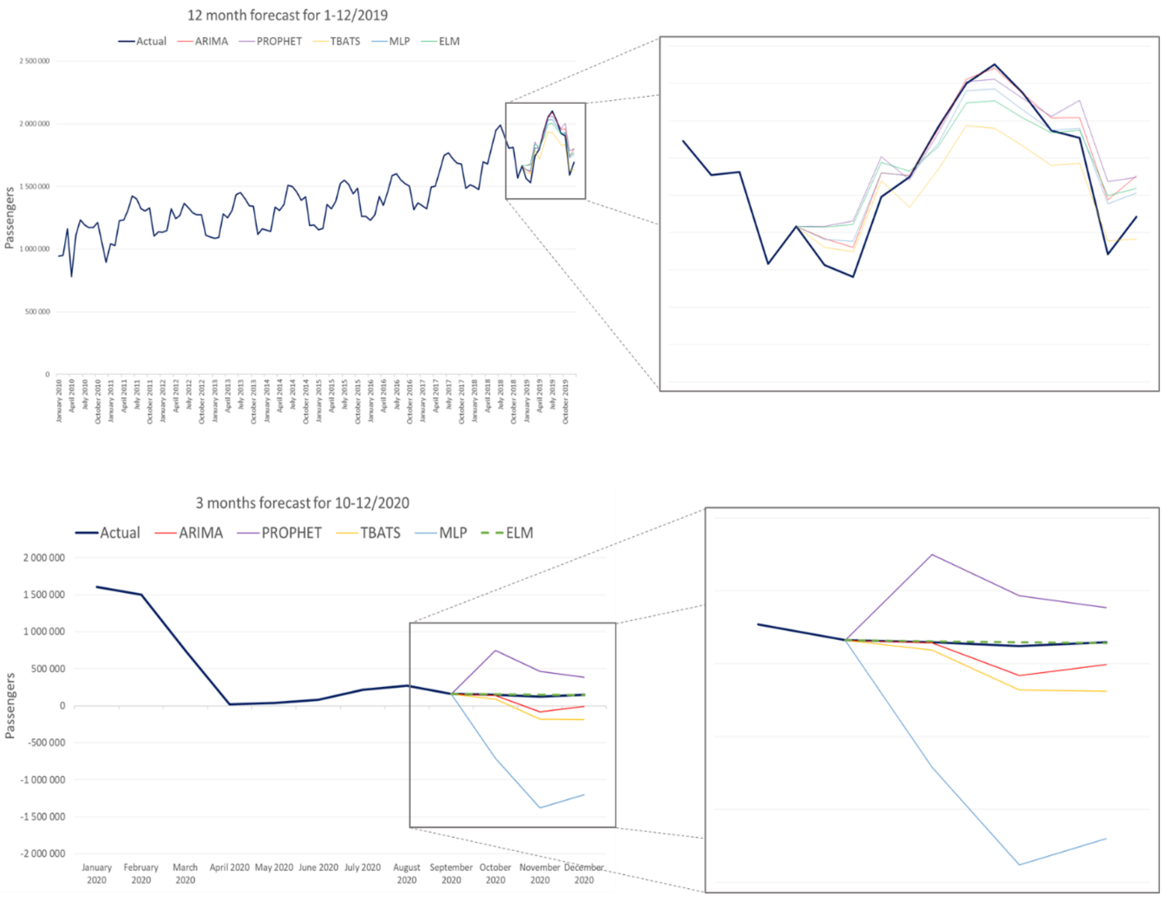

Figure 25. Forecasts before and during COVID-19

In daily forecasts, the accuracies were more drifted away than with monthly values. Figure 26 illustrates how Prophet outperformed other methods with the lowest MAPE of 5,3 – 6,4 percent prior to the COVID-19. TBATS, which is able to handle multiple seasonalities, performed nearly as well as Prophet. As expected, SARIMA (ARIMA (5,1,2)(0,1,0)[365]) did not stand out in the results, perhaps since the data contains multiple seasonalities. However, surprisingly, more sophisticated neural networks did perform even worse than the SARIMA model. The forecasting error of MLP increased significantly from 18,7 to 30,0 percent when the forecasting period was extended from 90 to 365 days.

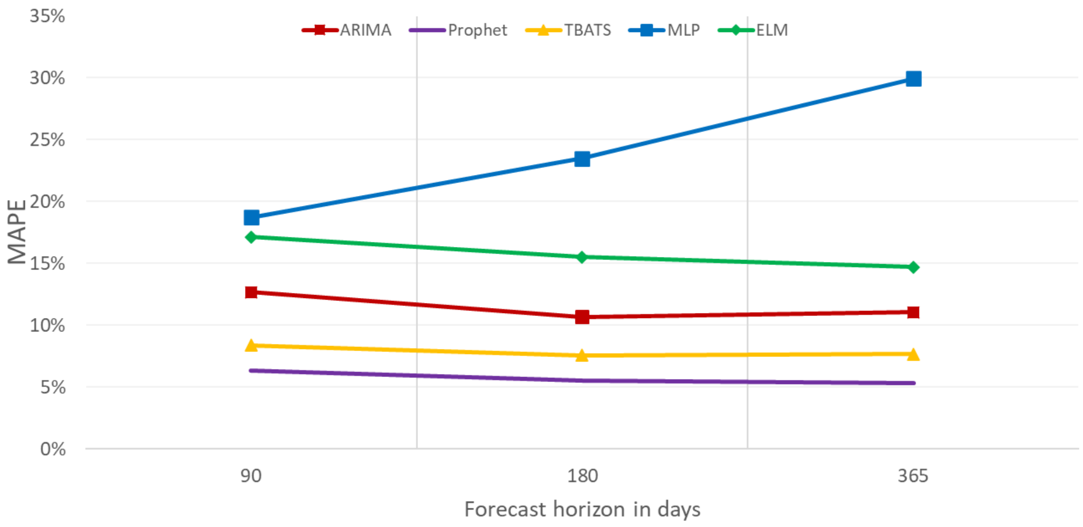

Figure 26. Forecasting errors (MAPE) prior to COVID-19 crisis in 2019 (daily data)

When the training period was extended until the end of September, and forecasts were produced for the last three months of 2020, only traditional methods could produce non-negative forecasts. However, the performance of ARIMA (5,1,1) was poor (MAPE 34,4 %), and so was the performance of TBATS (MAPE 28,3 %). The results with Prophet, MLP, and ELM were recorded as “not available” since they all produced forecasts with negative values. Unlike ARIMA and TBATS -models, which were introduced nearly five-year data from January 2016 until September 2020, MLP and ELM models were trained with data from January 2018 until the end of September 2020 with 20 repetitions each. This was done due to limited computing power.

Table 9. Forecasting accuracies with daily data during COVID-19

		90 days					180 days					365 days				
	Error measure	ARIMA	PROPHET	TBATS	MLP	ELM	ARIMA	PROPHET	TBATS	MLP	ELM	ARIMA	PROPHET	TBATS	MLP	ELM
Training set 01/2016 - 12/2018	MAPE	.1266	**.0636**	.0839	.1873	.1714	.1067	**.0553**	.0755	.2348	.1550	.1105	**.0532**	.0766	.2995	.1471
	MAE	6 502	**3 302**	4 364	9 500	8 704	5 977	**3 137**	4 342	13 560	8 584	6 197	**2 929**	4 401	17 350	8 195
	RMSE	7 631	**3 926**	5 212	11 557	10 850	6 991	**3 846**	5 139	15 668	10 537	7 397	**4 073**	5 402	19 322	10 374
Training set 01/2016 - 09/2020	MAPE	.3439	N/A	**.2830**	N/A	N/A	-	-	-	-	-	-		-	-	-
	MAE	1 222	N/A	**1 049**	N/A	N/A	-	-	-	-	-	-		-	-	-
	RMSE	1 493	N/A	**1 286**	N/A	N/A	-	-	-	-	-	-		-	-	-
Training set 04/2020 - 09/2020	MAPE	.3580	.3140	**.2959**	.6524	.3637	-	-	-	-	-	-		-	-	-
	MAE	1 275	1 480	**1 079**	2 422	1 295	-	-	-	-	-	-		-	-	-
	RMSE	1 543	2 035	**1 314**	2 689	1 566	-	-	-	-	-	-		-	-	-

Table 9 combines results for the models mentioned above and the models trained with data only during the crisis. During the crisis (training set 4-9/2020), daily forecasting errors significantly increased among all methods compared to pre-covid performance, resulting in high MAPEs of 29,6 – 65,2 percent. However, no negative forecast vales were produced when the historical development and the fall in passenger numbers were eliminated. TBATS was the best performer with MAPE 29,6 percent and was followed by PROPHET (MAPE 31,4 percent). The results suggest that neural networks are not performing any better than traditional methods, even when data contains more complex patterns than multiple seasonalities and linear trends. However, the results should be considered with caution since the training period during the crisis was somewhat short (185 observations).

5.2 Effects of the pandemic-related variables on forecasting accuracy

In the attempt to evaluate the relevance of including pandemic-related variables in the models, monthly data was examined first. Based on the arguments made in the methodological part, TC was selected as the sole exogenous regressor to reflect the overall development of the pandemic. Monthly data from January 2010 until September 2020 was used as a training set, and the remaining three months as testing set for validation. This out-of-sample validation provided poor results when univariate time-series data was applied, as described in the previous section: Out of the five methods, only ELM could produce usable results, with a surprisingly low MAPE of 9,98. Similar to ARIMA, TBATS, and MLP, Prophet first returned negative values with univariate models and highly optimistic values (MAPE 285,6 %) when its trend flexibility was adjusted.

In this part, TBATS was not applied since the method is not capable of considering exogenous regressors. For the remaining methods, ARIMA and MLP returned negative values despite the highly correlated variable was introduced to the model. ELM, which performed surprisingly well, improved its predicting power slightly from 9,98 percent to 9,86 percent, measured as MAPE. The most significant improvement was with Prophet, which improved from MAPE 285,6 percent to 71,1 percent when similar trend flexibility (0,5) was applied for the univariate model. With default settings, MAPE was 79,4 percent, which is still a significant improvement compared to the univariate model that returned negative values.

ARIMAX (5,1,2) was the only out of four applied methods in this phase of research returning non-negative forecasts for 90 days of daily data (training set January 2016 to September 2020). The accuracy of the ARIMA model improved by 8,7 percentage points when TC was introduced to the model. While MAPE for 90 days forecast with univariate data was 34,4 percent, it was 25,7 percent with the multivariate one. Although the improvement was significant, the performance of the ARIMA model was still somewhat unsatisfactory. The minimum daily prediction error was 0,83 percent, while the maximum was 101,2 percent. The median absolute percentage error was 18,02 percent, while the MAPE was 25,7 percent. Prophet, MLP, and ELM produced forecasts below-zero values with both univariate and multivariate data. However, it is to be noted that MLP and ELM were trained with fewer data from January 2018 onwards with 20 repetitions each due to limited computing power.

When the models were trained with data during the “stable phase” of the COVID-19 crisis only (4-9/2020), three out of four methods produced meaningful results. Since Prophet returned some negative values with both default settings and trend flexibility adjusted, its results were not considered. Thus, refining the model with TC did not improve the forecasting accuracy of Prophet. The rest of the models, ARIMAX (0,1,2), MLP, and ELM, slightly improved their predicting power when TC was included as an exogenous regressor. The overall results are presented in figure 27.

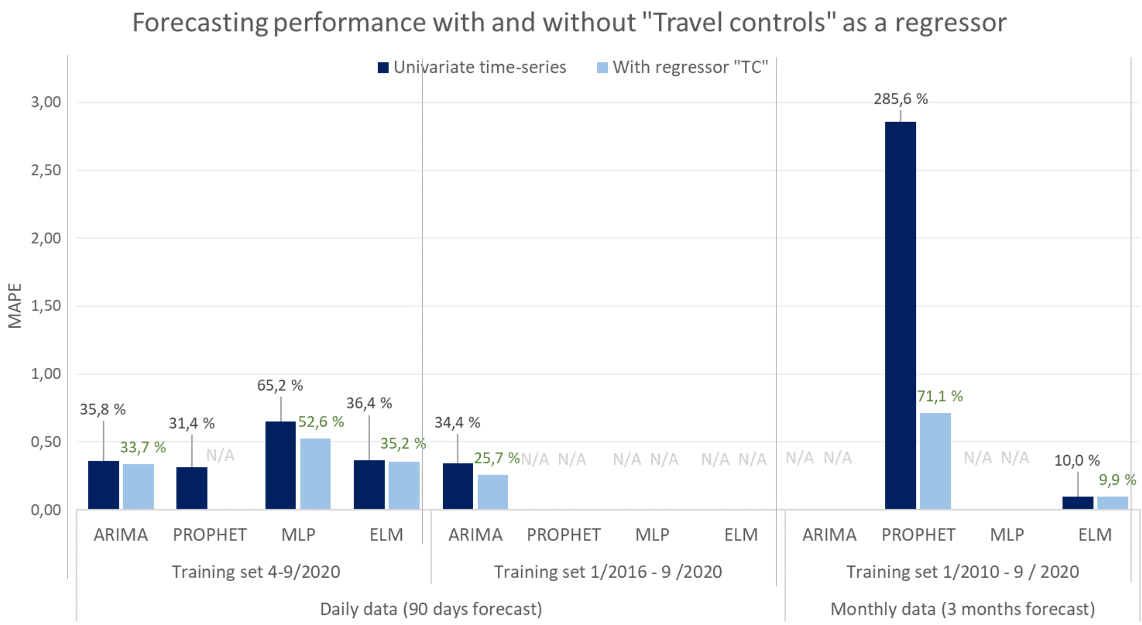

Figure 27. Comparison of results

6. Discussion and conclusions

The section is dedicated to presenting the results of this research comprehensively and discussing them in light of the theory. In the first subsection, research questions are answered, and findings discussed. The second subsection is dedicated to discussing the findings from an academic perspective. The third subsection discusses the findings from the practical point standpoint. Finally, in the final subsection, limitations are discussed and future research avenues are suggested.

6.1 Conclusions

In addition to empirically evaluating the performance of selected quantitative forecasting models in the context of airport passengers, the thesis aimed to draw a picture of the current status of scholarly research on airport passenger traffic forecasting. In general, the results indicate growing interest towards the topic, especially within the Asia Pacific region, where the air traffic has been growing (Suryani 2010, 2325) and is expected to continue its growth (Zhang 2020, 1). An increase in computing power and easier access to more sophisticated forecasting tools through free, open-source statistical programs suggest the airport passenger forecasting may gain popularity in upcoming years.

The first objective was to seek an answer for the sub RQ1: "*What quantitative methods have been used in passenger volumes forecasting at airports, and how past industry shocks have been handled in the models?"* A systematic literature review was conducted, and it was found out that historically causal methods have dominated research, but lately, hybrid and AI-based models have shown their potential to take over the positions as the most used methods. The results also showed how the time-series approach is more common in short-term forecasting, translating to the forecasting horizon of 0-4 years. Both time-series and causal methods are used in medium-term forecasting (5-18 years), but it is much more common to apply non-time-series approach when long-term forecasts up to 27 years are produced. The (different types of) ARIMA models have been popular as primary models, part of other ones, and independent benchmarking models due to their ability to handle seasonality and consider exogenous variables. However, hybrid models, especially AI-based methods, have often shown their superiority compared to traditional methods, which lack the ability to consider nonlinearities and other complexities in data. Also, it was found out that the more granular the data is, the greater the number of historical observations in the model. The number of observations ranged from six (annual data) up to 1460 (hourly data).

In addition to historical passenger numbers, a total of 76 different predictors have been applied in research. The typical ones are GDP and GDP per capita, population, and currency exchange rate. However, less attention has been given to industry shocks as only five publications had at least noted shocks and interventions. Historically, shocks have been dealt with dummy variables. There was a surprisingly broad variation with error measures. A total of 27 different measures were recognized. The three most used measures – MAPE, MAE, and RMSE – represent approximately half of the observations. The use of more than one measure is a relatively new practice – the study by Wadud (2013) was the first out of the studies included in this SLR, applying more than one error measure. Since then, it has become more common to consider more measures.

The purpose of the exhaustive literature search was to select suitable methods and to understand the sufficient amount of data needed to train the models. In the quest of searching for an answer to the sub RQ2 – *what is the accuracy of selected forecasting methods during the COVID-19 pandemic* – the following evidence was found: performance of the models was expectedly less-optimal during than before the pandemic. The overall forecasting performance during the pandemic, both including and excluding the major fall in the first months of 2020, was poor with all other methods except ELM, which produced surprisingly accurate forecasts with monthly data. Since the results were not cross-validated with other data, the results should be considered with caution. More research is needed on whether ELM would be suitable in predicting univariate monthly time series during the pandemic. The cautious approach is also supported by the finding with daily data, where ELM failed to produce meaningful forecasts when the model was trained with data from January 2018 until September 2020. The performance was also poor if only data during the crisis was introduced: MAPEs varied between 30 and 65 percent. Thus, I conclude that standalone univariate models are not the best solution to accurately forecast air passengers during a crisis such as COVID-19, but they all seem suitable when the forecasting environment is stable.

In the third sub-research question – *how does including COVID-19 related variables in the models affect forecasting accuracy* – the aim was to understand whether it would be meaningful to include pandemic-related variables in the models to improve predictions. It was understood that, in general, yes it is. Theory suggests incorporating an exogenous shock, even just by including a dummy variable in the models (see Tsui & Balli 2017, Tsui et al.

2017, Profillidis 2000, Kawad & Prevedouros 1996), is an appropriate approach to exclude the effect of an exogenous shock. However, the reasons for including dummy variables to account for crises or other intervention events were to eliminate the effect of occasional historical events when forecasts during “normal times” were produced. Thus, the theory could not guide how to deal with the shocks when forecasts during the pandemic are produced. Therefore, an exploratory approach was selected, and a variable that highly correlated with Helsinki Airport passengers was used to experiment this: international travel control index (TC) for the European continent. Based on the findings, introducing the additional variable to the models improved forecasting accuracy. Therefore, including external variables in models can be suggested, although further research is needed with more data.

Finally, the ultimate objective was to reach an understanding by answering the main research question of this thesis: *How quantitative forecasting methods are able to predict airport passenger volumes in the era of COVID-19 pandemic?* Although the pandemic-related variable was able to improve forecasting accuracy, and ELM provided promising results as a standalone model, it can be concluded that the possibilities of successfully adopting quantitative methods to forecast airport passenger traffic during the crisis are weak. At least more research is needed by conducting a more profound analysis of the individual automated methods and the opportunities they provide in terms of manual adjustments. Spitz and Golaszewski (2007, 20) pointed out how successful use of time-series requires, for example, the environment being stable. This may be one reason why the models introduced in this thesis do not perform well during the COVID-19 crisis.

6.2 Theoretical contribution

To the best of my knowledge, the thesis was the first to empirically test the predicting power of quantitative forecasting models during a crisis such as COVID-19 and first to evaluate the forecasting performance of quantitative methods at a Finnish airport. In addition to the empirical contribution, the thesis shed light on the current status of research by conducting presumably the first systematic literature review on airport passenger traffic forecasting.

No forecasting theory for a crisis environment was identified so that the research findings could be directly compared to the previous empirical evidence. If only the findings before the crisis are discussed, and the crisis is excluded from the discussion, the findings are in line with the theory built around the stable business environment. The reason for discussing the findings excluding the effects of COVID-19 allows this research to compare the findings with the current body of knowledge. In general, excluding the COVID-19 factor, it was found out how the performance of ARIMA was not any worse than its rival models. Unlike the literature suggests (see e.g., Jin et al., 2020; Xiao et al., 2014; Xie et al. 2014), no evidence was found that AI-based methods would systematically beat the traditional methods. Instead, the ARIMA model outperformed the other models with monthly data. The ARIMA model's performance was comparably good also with daily data, although Prophet and TBATS, both able to handle multiple seasonalities, outperformed the ARIMA model.

Jin et al. (2020, 9–12) applied ELM among other methods as a benchmark model against a VMD-ARMA/KELM-KELM. Although standalone ELM significantly underperformed compared to its more sophisticated counterparts, it outperformed ARIMA. However, the accuracy of both ARIMA and ELM were comparably poor in different data sets compared to the hybrid approach, which combined multiple different methods as one. In the paper of Xiao et al. (2016, 7), a stand-alone feed-forward neural network performed significantly worse than the proposed model that was combining multiple methods. While MLP, in the study by Xiao et al. (2014, 10), achieved 7,06 percent in-sample performance, the developed model combining multiple methods achieved MAPE of 1,53 percent. No single paper included in the SLR focused on a single AI method. Thus, I can conclude that a stand-alone AI-based method, namely neural networks, may not be a solution for accurately forecasting the number of future air passengers during stable times. In this research, hybrid models were not introduced.

Makridakis et al. (1981, 127) argued how the models that do not take into account a trend are inferior compared to those considering them. Kourentzes (2016) has discussed that neural networks are not good at dealing with trends. Since air passenger data typically contain a strong linear trend or changing trends, this may explain why the stand-alone neural networks failed at this research. Therefore, applying MLP and ELM with their default settings only should be avoided with trended time series. Kourentzes (ibid.) does not have off-the-self solutions available to best deal with trends, but one option would be differencing, which is done for ARIMA models to transform data stationary. Forcing a neural network to model ever-increasing or decreasing trends may be a too bold assumption in an environment full of uncertainties.

Similar to the findings of Ferhatosmanogly and Macit (2016, 184), TBATS performed better than ARIMA when data with multiple seasonalities (daily data) was introduced. They (ibid.) argued that ARIMA and TBATS produced comparable results when single seasonality was included. The good performance of ARIMA with single seasonality was supported, but no evidence was found in our monthly dataset to support the assumption of TBATS producing comparable results with ARIMA with single seasonality.

The results highlight that no single method is useful on all occasions – it depends on data, its granularity, and complexity. While some method performs exceptionally well during regular times with monthly data, the same method may not be useful in predicting daily passengers. While more sophisticated methods, ELM namely, may not perform any better than easier-to-understand traditional methods, their performance may stand up during extraordinary times when the data patterns turn from simple to complex.

Based on the findings of the SLR, an improved taxonomy for air passenger demand forecasting at airports is proposed. General, although comprehensive, classifications by Banerjee et al. (2020) and ICAO (2006) are not specific enough to meet the needs of classifying methods for airport passenger demand forecasting. The development of the models has improved significantly along with an increasing interest towards air passenger demand forecasting. An increasing number of scholars have started to combine methods to improve forecasting accuracy and consider AI tools to handle non-linearities and exogenous events better. It no longer works for the purpose of classifying the methods only to time series and causal, or quantitative and qualitative. That is, it has become commonplace to consider both time series and causal methods in one predictive model and perhaps apply them by using modern machine learning approaches. Figure 28 illustrates an improved taxonomy for quantitative methods based on the classification by Banerjee et al. (2020). The proposed classification considers hybrid models as their own, separate from stand-alone causal, time-series, and AI methods.

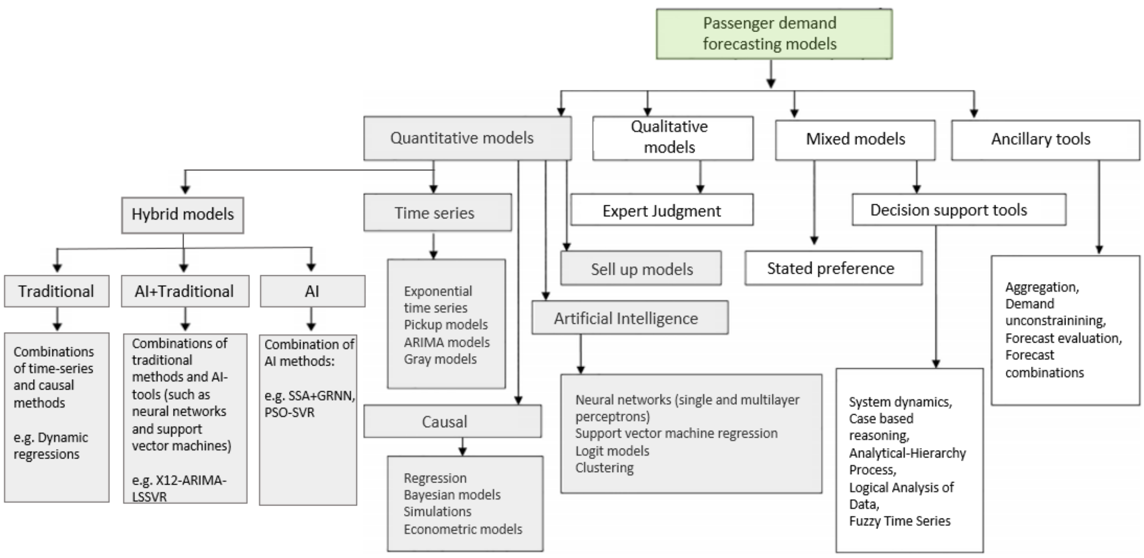

Figure 28. Improved taxonomy for air passenger traffic forecasting (adapted from Banerjee et al. (2020, 799)

6.3 *Practical implications*

In addition to contributing to the theory of air passenger forecasting from the academic perspective, it is essential to discuss the results from a practical standpoint, too. The results suggest how international travel controls form a significant driver for air passenger demand development at the Nordic airport. Although adding variables of international travel controls slightly improves forecasting performance, the difficulty for an analyst stems from estimating the future of external regressors. Profillidis (2012, 49) also claimed this difficulty in his research. The aggregated measure requires estimating the future travel controls for each European country, which may be troublesome for an analyst. One option to deal with this is to forecast the values using a separate time series model for international travel controls. Another way is to evaluate the development of international travel controls through different scenarios – one scenario for a complete lock-down, one scenario for the present situation, one scenario for significantly eased international travel controls. By adopting this approach, the method would not be considered pure quantitative any longer.

In a constantly evolving environment, where there are multiple factors affecting air travel demand, a purely quantitative approach may not be the best option. Judgmental methods are underrepresented in scientific research. Applying a qualitative approach together with quantitative methods, in the form of scenario planning, for example, could be an exciting way to understand the fuzzy future better. Relying only upon crisp numbers may lead to wrong decisions if the space of possible future outcomes is left unexplored. The fuzzy approach, as implemented by, for example, Profillidis (2000), could provide a broader perspective for the management.

Airport operators have access to even more granular data, which they could benefit from. The variable used in this research may be criticized for being too general. For example, each country equally contributed to the European international travel control index when country data was aggregated and averaged. While this may sound fair, it should be noted that the government restrictions made in different countries do not necessarily equally affect air passenger volumes at a particular airport. Therefore, practitioners could consider forecasting passenger volumes at route-level (country or airport) and aggregate these results. Another option would be to assign different weights to different countries by, for example, the size of GDP or the number of passengers transported to that country.

The results of the research are not generalizable due to airport-specific factors. This research did not separate international and domestic passengers nor arriving, transfer, and departing passengers. The traffic at certain airports may be highly influenced by domestic traffic, which may not be affected by international travel controls similar to international traffic. However, domestic traffic may also sometimes be strongly dependent on international transfer traffic. Therefore, the driving factors of passenger demand during the crisis and the impact of pandemic-related variables should be carefully evaluated before applying the methods presented in this research. Perhaps, point-to-point and transfer passengers should also be separated if travel restrictions are applied differently to point-to-point passengers and passengers only passing through an airport without entering the country.

Once the effects of the pandemic disappear and the development of air travel demand returns to regular, quantitative models introduced in this thesis provide promising opportunities in forecasting air passenger traffic at airports. Compared to other stand-alone methods, ARIMA models and MLP seem to provide the most accurate results with monthly data. When data complexity increases (daily data with weekly and annual seasonality), Prophet and TBATS seem to deliver the most accurate results. Prophet and its manually adjustable trend changepoints, trend flexibility, and other adjustable aspects are worth being explored further, also during the pandemic. Neural networks may not necessarily be the best choices.

Besides being slow and requiring enough computing power, the “black box” nature is also troublesome for the analysts and management since it may be impossible to understand the logic behind the outcomes.

The performance of the models trained with the data during the pandemic (training set 4/2020 onwards) should be re-evaluated when more data points become available. The data was now divided into 67/33 testing and training set. The results may be more meaningful when more historical data is introduced to the models. This is also supported by the literature, in which an average of 129 data points was applied with monthly data. 1460 data points were used with the hourly data of 6-hour intervals. Based on these, 185 daily observations introduced to the models can be deemed too little to produce meaningful results and accurately compare forecasting methods.

Browne and Kline (2020) criticized airline management teams for not being prepared for the risk of a pandemic after a string of financially successful years. A similar critique may be addressed to airport operators, too: none of the four major Nordic operators[12] recognized pandemic as a risk in their 2018 annual reports. Perhaps, after this ordeal, the business risks arising from the threat of pandemics will not be dismissed with a shrug. Although the thesis did not provide off-the-shelf solutions to the present problem, I genuinely wish that the findings of this research, which provides a novel perspective to the theory of air passenger demand forecasting, will help the industry make better predictions of the fuzzy future.

6.4 Limitations and suggestions for future research

Although the aim was to thoroughly understand the theory of air passenger traffic forecasting at airports and contribute to the body of knowledge by bringing new insights from a novel perspective, it was quickly understood how much further research is needed to comprehensively understand the opportunities to forecast airport passengers during the pandemic. Thus, the thesis instead touched upon the topic and left several questions open worth exploring in the future. The pandemic will most certainly not be the last one. One limitation of

[12] Keywords ”pandemic”, ”epidemic”, and ”outbreak” searched from 2018 annual reports of Avinor, Copenhagen Airports A/S, Finavia, and Swedavia.

this research was that it relied only on one pandemic-related variable. Therefore, it would be essential to answering an important question: what are the determinants of a well-performing quantitative air passenger demand forecasting model during a pandemic?

This research focused on the perspective of airports. To the best of my knowledge, no such research has been done either from the airlines' perspective. Airlines have access to different kinds of data, forward-looking booking data, for example. Thus, it would be fascinating to explore the determinants of a well-performing model from the perspective of airlines. Also, this thesis did not take into consideration the effects of country-level restrictions. To better understand the effects of government policy responses, it would be interesting to examine the effects on passenger volumes between Finland and a country that either falls in the list of restricted countries or is released from it. By understanding this aspect, we could learn how the government restrictions affect air passenger traffic and how soon the impact is realized as reduced or increased passenger volumes.

This thesis aimed to form the base for future research on airport passenger forecasting during an exogenous shock. Therefore, the automated methods were compared with their default settings only, which is also a limitation of this study. Artificial intelligence tools provide exciting opportunities for airport passenger volume forecasting. Thus, the research could be replicated by focusing on modern AI-based tools only.

Following Ferhatosmanogly and Macit (2016), neighbor-dependent models, in a slightly different form called reference class forecasting (see Suh & Ryerson 2019), could be examined to estimate future passenger volumes during a pandemic. In addition to considering the passenger development of a single airport, the model considers the traffic development of its rival airports. The practice is claimed to reduce optimism bias in forecasts (Suh & Ryerson 2019) and, thus, assists in preparing more realistic forecasts.

Finally, besides focusing on quantitative methods only, the mixed-methods approach would bring additional value to the scientific and expert discussion. Although this thesis considered quantitative studies only, it was evident how little weight has been given to qualitative or semi-qualitative air passenger forecasting approaches. Shifting research focus from quantitative to mixed methods approach is recommended since this approach remains still heavily unexamined. Amid the crisis, expert judgments and other qualitative approaches, such as scenario planning, could bring valuable and richer insights for the management teams who try to navigate through the present rough seas.

REFERENCES

*denotes the publication was included in the SLR

ACI (2016) ACI guide to world airport traffic forecasts. Airport Council International. [web resource]. [cited 2.12.2020]. Retrieved from: https://store.aci.aero/wp-content/uploads/2018/05/ACI_Guide_to_World_Airport_Traffic_Forecasts_2016-2-1.pdf.

ACI (2020) European airports revise recovery projection to 2024 whilst reporting only marginal traffic increase for June. [web article]. [cited 20.9.2020]. Retrieved from: https://www.aci-europe.org/media-room/263-european-airports-revise-recovery-projection-to-2024-whilst-reporting-only-marginal-increases-in-passenger-traffic-for-june.html.

Archer, B. (1994) Demand forecasting and estimation. In: Ritchie, J.R.B. & Goeldner, C.S (ed.). Travel, tourism, and hospitality research: A handbook for managers and researchers. 2nd ed. New York: Wiley.

Armstrong, J.S. (2001a) Selecting forecasting methods: In: Armstrong, J.S. (ed.) Principles of forecasting: a handbook for researchers and practitioners. MA: Kluwer Academic Publishers. [web resource]. [cited 7.12.2020]. Retrieved from: https://www.researchgate.net/profile/J-Armstrong/publication/228255479_Selecting_Forecasting_Methods/links/5b6478fa458515298ce42f28/Selecting-Forecasting-Methods.pdf.

Armstrong, J.S. (2001b) Evaluating forecasting methods. In: Armstrong, J.S. (ed.) Principles of forecasting: a handbook for researchers and practitioners. MA: Kluwer Academic Publishers. [web resource]. [cited 7.12.2020]. Retrieved from: http://repository.upenn.edu/marketing_papers/146.

Ashford, N. (1985) Problems with long term air transport forecasting. Journal of Advanced Transportation, 19, 2, 101–113.

*Ashley, D. J., Hanson, P. & Veldhuis, J. (1995) A policy-sensitive traffic forecasting model for Schiphol Airport. Journal of Air Transport Management, 2, 2, 89–97.

ATAG (2018) Aviation: benefits beyond borders. [web document]. [cited 21.9.2020]. Retrieved from: https://aviationbenefits.org/downloads/aviation-benefits-beyond-borders/.

ATAG (2020) Aviation: benefits beyond borders. [web document]. [cited 3.10.2020]. Retrieved from: https://aviationbenefits.org/media/167143/abbb20_full.pdf.

Banerjee, N., Morton. A. & Akartunali, K. (2020) Passenger Demand Forecasting in Scheduled Transportation. European Journal of Operational Research, 286, 3, 797–810.

Blalock, G., Kadiyali, V. & Simon, D.H. The impact of post-9/11 airport security measures on the demand for air travel. The Journal of Law and Economics, 50, 4, 731–755.

Box. G.E.P. & Jenkins, G.M. (1976) Time series analysis: forecasting and control. Rev. ed. San Francisco: Holden-Day.

Browne, R.S. & Kline, W.A. (2020) Exogenous shocks and managerial preparedness: A study of U.S. airlines' environmental scanning before the onset of the COVID-19 pandemic. Journal of Air Transport Management, 89, 1–9.

Chambers, J.C., Satinder, K.M., Smith, D.D. (1971) How to choose the right forecasting technique. Harward Business Review, July issue.

Cheng, L. & Mengting, X. A review of research on airline passenger volume forecasting. 4th International Conference on Machinery, Materials and Computer. Advances in Engineering Research, 150.

Dantas, T.M., Oliveira, F.L.C. & Repolho, H.M.V. (2016) Air transportation demand forecast through Bagging Holt Winters methods. Journal of Air Transport Management, 59, 116–123.

De Livera, A. M., Hyndman, R. J. & Snyder, R.D. (2011) Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American Statistical Association, 106, 496, 1513–1527.

*de Paula, R. O., Silva, L. R., Vilela, M. L. & Cruz, R. O. M. (2019) Forecasting passenger movement for Brazilian airports network based on the segregation of primary and secondary demand applied to Brazilian civil aviation policies planning. Transport Policy, 77, 23–29.

* Djakaria, I. (2019) Djalaluddin Gorontalo Airport Passenger Data Forecasting with Holt's-Winters' Exponential Smoothing Multiplicative Event-Based Method. Journal of Physics: Conference Series, 1320, 1.

Do, Q.H., Lo, S-K., Chen, J.-F., Le, C.-L. & Anh, L.H. Forecasting Air Passenger Demand: A comparison of LSTM and SARIMA. Journal of Computer Science, 16, 7, 1063–1084.

EU 2018/1136. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018R1139.

*Felkel, R., Steinmann, D. & Follert, F. (2017) Hub airport 4.0 - How frankfurt airport uses predictive analytics to enhance customer experience and drive operational excellence. In: Linnhoff-Popien, C., Schneider, R. & Zaddach, M. (ed). Digital Marketplaces Unleashed - Springer, 443–453.

*Ferhatosmanoglu, N. & Macit, B. (2016) Incorporating explanatory effects of neighbour airports in forecasting models for airline passenger volumes. In Proceedings of 5th the International Conference on Operations Research and Enterprise Systems, 178–185.

Finavia (2021) Financial statements release January-December 2021: The COVID-19 crisis had a significant impact on Finavia's result. Press release 24.3.2021. [web article]. [cited 27.3.2021]. Retrieved from: https://www.finavia.fi/en/newsroom/2021/financial-statements-release-january-december-2021-covid-19-crisis-had-significant.

Finavia (2020) Finavia airports had 26 million passengers in 2019 – a year of moderate growth in air traffic. [web article]. [cited 22.9.2020]. Retrieved from: https://www.finavia.fi/en/newsroom/2020/finavia-airports-had-26-million-passengers-2019-year-moderate-growth-air-traffic.

Flyvbjerg, B., Skamris Holm, M.K. & Buhl, S.L. (2005) How (in)accurate are demand forecast in public works projects. Journal of American Planning Association, 71, 2, 131–146.

Forsyth, P., Guiomard, C., & Niemeier, H.-M. (2020) Covid-19, the collapse in passenger demand and airport charges. Journal of Air Transport Management, 89, 1–5.

Gallego, I. & Font, X. (2020) Changes in air passenger demand as a result of the COVID-19 crisis: using Big Data to inform tourism policy. Journal of sustainable tourism. Journal of sustainable tourism, ahead-of-print, 1–20. [web document]. [cited 3.10.] Retrieved from: https://www.tandfonline.com/doi/epub/10.1080/09669582.2020.1773476?needAccess=true.

Gelhausen, M.C., Berster, P., Wilken, D. (2018) A new direct demand model of long-term forecasting air passengers and air transport movements at German airports. Journal of Air Transport Management, 71, 140–152.

Gerrish, R.J. & Baggaley, P.A. (2020) From bad to worse: global air traffic to drop 60%-70& in 2020. S&P Global. [web article]. [cited 19.9.2020]. Retrieved from: https://www.spglobal.com/ratings/en/research/articles/200812-from-bad-to-worse-global-air-traffic-to-drop-60-70-in-2020-11610389?utm_campaign=corporatepro&utm_medium=contentdigest&utm_source=Airlines.

Ghalehkhondabi, I., Ardjmand, E., Young, W.A. & Weckman, G.R. A review of demand forecasting models and methodological developments within tourism and passenger transportation industry. Journal of Tourism Futures, 5, 1, 75–93.

Gudmundsson, S.V., Cattaneo, M. & Redondi, R. (2020) Forecasting recovery time in air transport markets in the presence of large economic shocks: COVID-19. SSRN.

*Graham, B. (1999) Airport-specific traffic forecasts: a critical perspective. Journal of Transport Geography, 7, 4, 285–289.

Hale, T., Angrist, N., Boby, T., Cameron-Blake, E., Hallas, L., Kira, B., Majumdar, S., Petrherick, A., Phillips, T., Tatlow, H. & Webster, H. (2020) Variation in government responses to COVID-19. Version 10.0. Blavatnik School of Government Working Paper, 10 December 2020. [web resource]. [cited 13.3.2021]. Retrieved from: https://www.bsg.ox.ac.uk/covidtracker.

Harzing, A.-W. (2020) Everything you always wanted to know about research impact… Version April 2019. Accepted for the 2nd edition of Clark, T. & Wright, M. (2020) How to get published in the best management journals. [web resource]. [cited 25.11.2020]. Retrieved from: https://harzing.com/download/impact.pdf.

Hemmingway, P. (2009) What is a systematic review? Evidence-based medicine, 1–8.

*Hofer, C., Kali, R. & Mendez, F. (2018) Socio-economic mobility and air passenger demand in the U.S.. Transportation Research Part A: Policy and Practice, 112, 85–94.

Huang, G.-B., Zhu, Q.-Y. & Siew C.-K. (2006) Extreme learning machine: Theory and applications. Neurocomputing, 70, 489–501.

Hyndman, R., Athanasopoulos, F., Bergmeier, C., Caceres, G., Chhay, L., O-Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E. & Yasmeen, F. (2020). Package ‘forecast’.

Manual 12.9.2020. [web document]. [cited 27.2.2020]. Retrieved from: https://cran.r-project.org/web/packages/forecast/forecast.pdf.

Hyndman, R.J. & Athanasopoulos, G. (2019) Forecasting: principles and practice, 3rd edition. [online book]. [cited 3.10.2020]. Retrieved from: OTexts.com/fpp3.

Hyndman, R.J. & Khandakar, Y. (2008) Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software, 27, 3.

Hyndman, R.J. & Koehler, A.B. (2006) Another look at measures of forecast accuracy. International Journal of Forecasting, 22, 679–688.

Iacus, M.S., Natale, F., Santamaria, C., Spyratos, S. & Vespe, M. (2020) Estimating and projecting air passenger traffic during the COVID-19 coronavirus outbreak and its socio-economic impact. Safety Science, 129, 1–11.

IATA (2021) 2020 Worst year in history for air travel demand. Press release 3.2.2021. [web article]. [cited 27.3.2021]. Retrieved from: https://www.iata.org/en/pressroom/pr/2021-02-03-02/.

IATA (2020a) Don’t make a slow recovery more difficult with quarantine measures. [web article]. [cited 21.9.2020]. Retrieved from: https://www.iata.org/en/pressroom/pr/2020-05-13-03/.

IATA (2020b) Recovery delayed as international travel remains locked down. [web article]. [cited 20.9.2020]. Retrieved from: https://www.iata.org/en/pressroom/pr/2020-07-28-02/.

IATA (2020c) What can we learn from past pandemic episodes. IATA economics’ chart of the week. [web article]. [cited 21.9.2020] retrieved from: https://www.iata.org/en/iata-repository/publications/economic-reports/what-can-we-learn-from-past-pandemic-episodes/.

ICAO (2006) Manual on air traffic forecasting. Third edition. International Civil Aviation Organization. [web document]. [cited 3.10.2020]. Retrieved from: https://www.icao.int/MID/Documents/2014/Aviation%20Data%20Analyses%20Seminar/8991_Forecasting_en.pdf.

ICAO (2009) Review of the classification and definitions used for civil aviation activities. Working paper. International Civil Aviation Organization. [web document]. [cited 4.10.2020]. retrieved from: https://www.icao.int/meetings/sta10/documents/sta10_wp007_en.pdf.

ICAO (2020) Effects of novel coronavirus (COVID-19) on civil aviation: economic impact analysis. Analysis 16.9.2020. [web document]. [cited 20.9.2020]. Retrieved from: https://www.icao.int/sustainability/Pages/Economic-Impacts-of-COVID-19.aspx.

*Jin, F., Li Y. Sun S. & Li H. (2020) Forecasting air passenger demand with a new hybrid ensemble approach. Journal of Air Transport Management, 83.

Kagan Albayrak, M. B., Özkan, I. C., Can. R. & Dobruszkes F. (2020) The determinants of air passenger traffic at Turkish airports. Journal of Air Transport Management, 86.

*Karasek, M. (1982) Forecasting and Planning the Jeddah Air Traffic with a Mini Model. Journal of Forecasting, 1, 4, 409–417.

*Kawad, S. & Prevedouros, P. D. (1995) Forecasting air travel arrivals: model development and application at the Honolulu international airport. Transportation Research Record, 1506, 18–25.

Khurshid, W. & Chandrasekhar, A. (2020) Revamping passenger demand models for a post-COVID aviation world. Lufthansa Consulting. [web article]. [cited 22.9.2020]. Retrieved from: https://www.lhconsulting.com/fileadmin/dam/downloads/studies/20200512_Article_Covid_demand_forecasting_Lufthansa_Consulting.pdf.

Kim, S. & Kim, H. (2016) A new metric of absolute percentage error for intermitted demand forecasts. International Journal of Forecasting, 32, 3, 669–679.

*Kim, S. & Shin, D. H. (2016) Forecasting short-term air passenger demand using big data from search engine queries. Automation in Construction, 70, 98–108.

Kourentzes, N. (2019a) Package ‘nnfor’. Documentation. [web resource]. [cited 28.2.2021]. Retrieved from: https://cran.r-project.org/web/packages/nnfor/nnfor.pdf.

Kourentzes, N. (2019b) Tutorial for the nnfor R package. [Web resource]. [cited 21.2.2021]. Retrieved from: https://kourentzes.com/forecasting/2019/01/16/tutorial-for-the-nnfor-r-package/.

Kourentzes, N. (2016) Can neural networks predict trended time series? Blog post 26.12.2016. [web article]. [cited 5.4.2021]. Retrieved from: https://kourentzes.com/forecasting/2016/12/28/can-neural-networks-predict-trended-time-series/.

*Kressner, J. D. & Garrow, L. A. (2012) Lifestyle segmentation variables as predictors of home-based trips for Atlanta, Georgia, airport. Transportation Research Record, 2266, 20–30.

Lamb, T.L., Winter, S.R., Rice, S., Ruskin, K.J. & Vaughn, A. (2020) Factors that predict passengers willingness to fly during and after the COVID-19 pandemic. Journal of Air Transport Management, 89, 1–10.

*Lei, J., Chong, X. & Long, X. (2019) Aviation Business Volume Forecast of Xianyang International Airport Based on Multiple Prediction Models. IOP Conf. Series: Materials Science and Engineering, 688, 2.

*Li, Y. & Jiang, X. (2020) Airport Passenger Throughput Forecast Based on PSO-SVR Model. IOP Conference Series: Materials Science and Engineering, 780, 6.

*Li, Y.-H., Han, H.-Y., Liu, X. & Li, C. (2018) Passenger flow forecast of Sanya airport based on ARIMA Model. Communications in Computer and Information Science, 902, 442–454.

Liu, J., Liu, B., Liu, Y., Chen, H., Feng, L., Xiong, H. & Huang, Y. (2017) Personalized air travel prediction: A multi-factor perspective. ACM Transactions on Intelligent Systems and Technology. 9, 3.

*Liu, X., Huang, X., Chen, L., Qiu, Z. & Chen, M.-R. (2017a) Prediction for passenger flow at the airport based on different models. Communications in Computer and Information Science, 729, 25–40.

*Liu, X., Huang, X., Chen, L., Qiu, Z. & Chen, M.-R. (2017b) Prediction of passenger flow at Sanya airport based on combined methods. Communications in Computer and Information Science, 727, 729–740

Liu, X., Li, L., Liu, X., Zhang, T., Rong, X., Yang, L. & Xiong, D. (2018) Field investigation on characteristics of passenger flow in a Chinese hub airport terminal. Building and Environment, 133, 51–61.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J. Parzen, E. & Winkler. R. (1982) The accuracy of extrapolation (Time series) methods: Results of a forecasting competition. Journal of Forecasting, 1, 111–153.

Miller, J.P. & Clarke, J.-P. (2007) The hidden value of air transportation infrastructure. Technological Forecasting & Social Change, 74, 18–35.

Ministry of Foreign Affairs (2020) Restrictions on entry into the country to be amended due to COVID-19. [web article]. [cited 4.10.2020]. Retrieved from: https://valtioneuvosto.fi/-/1410869/maahantulon-rajoituksia-muutetaan-koronatilanteen-perusteella?languageId=en_US.

Næss, P. & Strand, A. (2015) Traffic Forecasting at 'Strategic', 'Tactical' and 'Operational' level: A differentiated methodology is necessary. disP -The Planning Review, 51, 2, 41–48.

Nielsen, S. (2018) Reflections on the applicability of business analytics for management accounting – and future perspectives for the accountant. Journal of Accounting & Organizational Change, 14, 2, 167–187.

Njegovan, N. (2006) Are shocks to air passenger traffic permanent of transitory – Implications for long-term air passenger forecasts for the UK. Journal of Transport Economics and Policy, 40, 2, 315–328.

Okoli, C. (2015) A Guide to Conducting a Standalone Systematic Literature Review. Communications of the Association for Information Systems, 37, 879–910.

Peters, M.D.J., Godfrey, C.M., Khalil, H., McInerney, P., Parker, D. & Soares, C.B. (2015) Guidance for conducting systematic scoping reviews. International Journals of Evidence-Based Healthcare, 13, 141–146.

*Profillidis, V. A. (2012) An ex-post assessment of a passenger demand forecast of an airport. Journal of Air Transport Management, 25, 47–49.

*Profillidis, V.A. (2000) Econometric and fuzzy models for the forecast of demand in the airport of Rhodes. Journal of Air Transport Management, 6, 95–100.

Prophet (n.d.) Documentation. [web resource]. [cited 27.3.2021]. Retrieved from: https://facebook.github.io/prophet/docs/trend_changepoints.html.

Publication forum (no date) Evaluations. [web resource]. [accessed 15.11.2020]. Retrieved from: https://julkaisufoorumi.fi/en/evaluations.

*Ramadiani, Syahrani, R., Astuti, I. F. & Azainil (2020) Forecasting the number of airplane passengers uses the double and the triple exponential smoothing method. Journal of Physics: Conference Series, 1524, 1.

*Rodriguez, Y., Pineda, W. & Diaz Olariaga, O. (2020) Air traffic forecasting in post-liberalization contect: a dynamic linear models approach. Aviation, 24, 1, 10–19.

Roser, M. & Ritchie, H. Ortiz-Ospina, E. & Hasell, J. (2020) Coronavirus Pandemic (COVID-19). [Online resource]. [cited 20.9.2020]. Retrieved from: https://ourworldindata.org/coronavirus.

*Samagaio, A. & Wolters, M. (2010) Comparative analysis of government forecasts for the Lisbon Airport. Journal of Air Transport Management, 16, 4, 213–217.

*Scarpel, R.A. (2013) Forecasting air passengers at Sao Paulo International Airport using a mixture of local experts model. Journal of Air Transport Management, 26, 35–39.

Schmueli, G. (2010) To explain or to predict? Statistical Science, 2010, 25, 3, 289–310.

Schmueli, G. and Koppius, O. (2011) Predictive analytics in information systems research. MIS Quarterly, 35, 553-572.

*Sismanidou, A. & Tarradellas, J. (2017) Traffic demand forecasting and flexible planning in airport capacity expansions: Lessons from the Madrid-Barajas new terminal area master plan. Case Studies on Transport Policy, 5, 2, 188–199.

Snyder, H. (2019) Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 140, 333–339.

Spitz, W. & Golaszewski, R. (2007) Airport aviation activity forecasting: a synthesis of airport practice. National Academies of Sciences, Engineering, and Medicine. Washington: The National Academies Press.

*Strand, S. (1999) Airport-specific traffic forecasts: the resultant of local and nonlocal forces. Journal of Transport Geography, 7, 1, 17–29.

*Suh, D. Y. & Ryerson, M. S. (2019) Forecast to grow: Aviation demand forecasting in an era of demand uncertainty and optimism bias. Transportation Research Part E: Logistics and Transportation Review, 128, 400–416.

*Sun, S., Lu, H., Tsui, K.-L. & Wang, S. (2019) Non-linear vector auto-regression neural network for forecasting air passenger flow. Journal of Air Transport Management, 78, 54–62.

Sung, J. & Monschauer, Y. (2020) Changes in transport behaviour during the Covid-19 crisis. International Energy agency IEA. [web article]. [cited 4.10.2020]. Retrieved from: https://www.iea.org/articles/changes-in-transport-behaviour-during-the-covid-19-crisis.

*Suryani, E., Chou, S.-Y. & Chen, C.-H. (2010) Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework. Expert Systems with Applications, 37, 3, 2324–2339.

Taylor, S.J. & Letham, B. (2018) Forecasting at Scale. The American Statistician, 72,1,37–45.

Taylor, S.J. & Letham, B. (2017) Prophet: forecasting at scale. Facebook research blog 23.2.2017. [web article]. [cited 27.2.2021]. Retrieved from: https://research.fb.com/prophet-forecasting-at-scale/.

The World Bank (2020) Air transport, passengers carried. [Database]. [accessed 20.9.2020. Retrieved from: https://data.worldbank.org/indicator/IS.AIR.PSGR?end=2019&start=1970.

THL (2020) Traffic light model to help in the assessment of risks associated with foreign travel. Finnish institute for health and welfare. [web resource]. [cited 21.10.2020]. Retrieved from: https://thl.fi/en/web/infectious-diseases-and-vaccinations/what-s-new/coronavirus-covid-19-latest-updates/travel-and-the-coronavirus-pandemic/traffic-light-model-to-help-in-the-assessment-of-risks-associated-with-foreign-travel.

THL (2021) Traffic light model to help in the assessment of risks associated with foreign travel. Finnish institute for health and welfare. [web resource]. [cited 13.3.2021]. Retrieved from: https://thl.fi/en/web/infectious-diseases-and-vaccinations/what-s-new/coronavirus-covid-19-latest-updates/travel-and-the-coronavirus-pandemic/traffic-light-model-to-help-in-the-assessment-of-risks-associated-with-foreign-travel.

Tranfield, D., Denyer, D. & Smart, P. (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. British Journal of Management, 4, 207–222.

*Tsui, W.H.K., Balli, H.O., Gilbey, A. & Gow, H. (2014) Forecasting of Hong Kong airport's passenger throughput. Tourism Management, 42, 62–76.

*Tsui, W. H. K. & Balli, F. (2017) International arrivals forecasting for Australian airports and the impact of tourism marketing expenditure. Tourism Economics, 23, 2, 403–428.

Twinn, I., Qureshi, N., Rojas, D.S.P. & Conde, M.L. (2020) The impact of COVID-19 on airports: an analysis. International Finance Corporation (IFC), a member of the World Bank Group. [web article]. [cited 19.9.2020]. Retrieved from: https://www.ifc.org/wps/wcm/connect/26d83b55-4f7d-47b1-bcf3-01eb996df35a/IFC-Covid19-Airport-FINAL_web3.pdf?MOD=AJPERES&CVID=n8lgpkG.

*Uddin, W., McCullough, B. F. & Crawford, M. M. (1985) METHODOLOGY FOR FORECASTING AIR TRAVEL AND AIRPORT EXPANSION NEEDS.. Transportation Research Record, 1025, 7–14

*Wadud, Z. (2011) Modeling and forecasting passenger demand for a new domestic airport with limited data. Transportation Research Record, 2214, 59–68.

*Wadud, Z. (2013) Simultaneous modeling of passenger and cargo demand at an airport. Transportation Research Record, 2336, 63–74.

Wang, M. & Song, H. (2010) Air travel demand studies: A review. Journal of China Tourism Research, 6,1,29–49.

WHO (2020) Naming the coronavirus disease (COVID-19) and the virus that causes it. World Health Organization. [web resource]. [cited 11.10.2020]. Retrieved from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it.

Wohlin, C. (2014) Guidelines for snowballing in systematic literature studies and replication in software engineering. Proceedings of the 18th International Conference on evaluation and assessment in software engineering 2014-05-13, 1–11.

Wu, X., Xiang, Y., Mao, G., Du, M., Yang, X & Zhou, X. (2020) Forecasting air passenger traffic flow based on the two-phase learning model. Journal of Supercomputing.

*Xiao, Y., Liu, J.J., Hu, Y., Wang, Y., Lai, K.K. & Wang, S. (2014) A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting. Journal of Air Transport Management, 39, 1–11.

*Xiao, Y., Liu, Y., Liu, J. J., Xiao, J. & Hu, Y. (2016) Oscillations extracting for the management of passenger flows in the airport of Hong Kong. Transportmetrica A: Transport Science, 12, 1, 65–79.

*Xie, G., Wang, S. & Lai, K.K. (2014) Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches. Journal of Air Transport Management, 37, 20–26.

Yokota, J., Tsoneva, T., Iyer, P., Lu, G., Forsgren, K.U. & Sperling-Tyler, B . (2020) Airports face a long haul to recovery. S&P Global. [web article]. [cited 19.9.2020]. Retrieved from: https://www.spglobal.com/ratings/en/research/articles/200528-airports-face-a-long-haul-to-recovery-11506553

Zhang, G.P. (2004) Business forecasting with artificial neural networks: An overview. In: Zhang, G.P. (ed). Neural networks in business forecasting. Hershey: Idea Group Publishing.

*Zhang, X. (2020) Research on forecasting method of aviation traffic based on social and economic indicators. IOP Conference Series: Materials Science and Engineering, 780, 6.

APPENDICES

Appendix 1: SLR summary table

Authors (year) title	Type of publication	GS citations	MA citations	Airport(s) covered	Forecasting Method	Proposed classification	Forecast horizon	Data	Number of time-units (hours, months, years)	Error measures	Regressors (passenger/trend not mentioned if time-series in question)
Jin et al. (2020) Forecasting air passenger demand with a new hybrid ensemble approach	Journal article	4	5	Beijing, Guangzhou, Shanghai Pudong	VMD-ARMA/KELM-KELM (variational mode decomposition, ARMA, kernel extreme learning machine)	AI	Short-term	Monthly	143	MAE,MAPE,RMSE,Dstat	-
Li & Jiang (2020) Airport Passenger Throughput Forecast Based on PSO-SVR Model	Conference paper	-	-	Beijing Capital, Shangai Pudong, Guangzhou, Chengdu	PSO-SVR (support vector regression machine based on particle swarm optimization)	AI	Short-term[1]	Annual	16	MAPE	-
Ramadiani et al. (2020) Forecasting the number of airplane passengers uses the double and the triple exponential smoothing method	Conference paper	-	-	Samarinda	Exponential smoothing (double and triple)	Time series	Short-term[1]	Monthly	72	MSE	-
Rodriguez et al. (2020) Air traffic forecasting in post-liberalization contect: a dynamic linear models approach	Journal article	-	-	Bogotá-El Dorado International Airport	Dynamic linear model	Time series	Medium-term	Annual	39	MAPE	**National passengers:** Consumer Price Index, GDP per capita, national passengers with delay -1 **International passengers:** GPD, population, Currency Exchange Rate, international passengers with delay t-1.
Zhang (2020) Research on forecasting method of aviation traffic based on social and economic indicators	Conference paper	-	-	Mianyang	Combination of trend extrapolation and econometric model	Hybrid	Long-term	Annual	10	-	GDP, Fixed asset investment, Industrial enterprises main business income, Local public revenue, Investment promotion, Urban per capita disposable income, Number of tourists
de Paula et al. (2019) Forecasting passenger movement for Brazilian airports network based on the segregation of primary and secondary demand applied to Brazilian civil aviation policies planning	Journal article	1	-	Barreiras, Lages, Chapeco, Sao Jose dos Campos	OLS, gravitational model	Causal	-	Monthly	144	-	**Primary demand model:** regional population, GDP, Regions dummy **Secondary demand model:** The proportion of the population of each municipality inrelation to the Brazilian population, The Brazilian global demand estimate
Djakaria (2019) Djalaluddin Gorontalo Airport Passenger Data Forecasting with Holt's-Winters' Exponential Smoothing Multiplicative Event-Based Method	Conference paper	-	-	Djalaluddin Gorontalo Airport	Holt-Winters' exponential smoothing multiplicative event-based method	Time series	-	Monthly	180	MAPE, Mean Absolute Deviation (MAD), Mean Square Deviarion (MSD)	-
Lei et al. (2019) Aviation Business Volume Forecast of Xianyang International Airport Based on Multiple Prediction Models	Conference paper	-	-	Xianyang International Airport	Trend extrapolation-econometric-market share method	Hybrid	Medium-term[1]	Annual	6	Deviation from actual	GDP, population, social average wage, international tourism income, passenger thoughput
Suh & Ryerson (2019) Forecast to grow: Aviation demand forecasting in an era of demand uncertainty and optimism bias	Journal article	2	2	Multiple US airports	Reference class forecasting	Causal	Long-term[1]	Annual	11	Mean forecast error (MFE), Mean growth-based forecast error (MGBFE), Mean Peer-Based forecast Error (MPBFE), Enhanced Mean Peer-Based Forecast Error (EMPBFE)	Airport competition % change (5-year average annual percentage change (5AAC)), connecting passenger share, Connecting passenger share % change (5AAC), Population % change (5AAC), per capita income, service sector employment Variables for cluster analysis: passengers, connecting passenger chare, Avg. ticket price, HHI, Per capita income, Service sector employment, 5-year avg. annual % change up to base year, passengers change (5AAC), Airport competition (5AAC), Connecting passenger share (5AAC), Avg number of seats per aircraft (5AAC), Avg. ticket price (5AAC), HHI (5AAC), Population (5AAC), Per capita income (5AAC), Service Sector Employment (5AAC)
Sun et al. (2019) Nonlinear vector auto-regression neural network for forecasting air passenger flow	Journal article	8	8	Beijing Capital	MIV-NVARNN (mean impact value based nonlinear vector auto-regressive neural network)	AI	Short-term[1]	Monthly	123	Normalized root mean squared error (NRMSE), MAPE, Directional Symmetry (DS), Diebold-Mariano (DM) statistic, Pesaran-Timmermann (PT) statistic, Mean square prediction error (MSPE)	Beijing International Airport (BJIA) air passenger flows, Tianjin air passenger flows, Beijing railway passenger flows, The air passenger market index of China, The flight movements of BJIA, The flight release rate of BJIA

[1] in "Forecast horizon" column indicates that the forecast horizon was not defined by the author of the publication. Instead, it was defined by the author of this thesis. Table continues on the next page…

continued (appendix 1)

Authors (year) title	**Type of publication**	**GS citations**	**MA citations**	**Airport(s) covered**	**Forecasting Method**	**Proposed classification**	**Forecast horizon**	**Data**	**Number of time-units (hours, months, years)**	**Error measures**	**Regressors** (passenger/trend not mentioned if time-series in question)
Hofer et al. (2018) Socio-economic mobility and air passenger demand in the U.S.	Journal article	4	3	-	Three-stage least squares regression	Causal	-	-	-	-	Socio-economic mobility (absolute mobility and relative mobility), Income and population, number of alternate airports, yield, number of destinations, HUB, Herfindahl Index (HHI) of aggregated routes
Li et al. (2018) Passenger flow forecast of Sanya airport based on ARIMA Model	Conference paper	3	1	Sanya Phoenix International Airport	ARIMA	Time series	Short-term	Monthly	108	Absolute percentage error (APE), R^2	-
Felkel et al. (2017) Hub airport 4.0 - How frankfurt airport uses predictive analytics to enhance customer experience and drive operational excellence	Book chapter	-	-	Frankfurt Airport	Combination of multiple statistical methods such as decision-trees, linear regression and multiple imputation	AI	Short-term[1]	-	-	-	The model takes into account various external influencing factors such as school holidays, public holidays, seasonal effects, and fairs
Liu et al. (2017b) Prediction of passenger flow at sanya airport based on combined methods	Conference paper	5	3	Sanya Phoenix International Airport	ARMA, Grey prediction model, regression model based on ARMA (RE-ARMA)	Time series, causal	Short-term	Monthly	108	MAPE	aircraft movements, luggage, passenger load factor FOR RE-ARMA only
Liu et al. (2017a) Prediction for passenger flow at the airport based on different models	Conference paper	2	1	Sanya Phoenix International Airport	Combination of Holt-Winters seasonal prediction model, ARMA and unary linear regression	Hybrid	Medium-term[1]	Monthly	108	MAPE	-
Sismanidou & Tarradellas (2017) Traffic demand forecasting and flexible planning in airport capacity expansions: Lessons from the Madrid-Barajas new terminal area master plan	Journal article	9	3	Adolfo Suarez Madrid-Barajas airport	Linear regression	Causal	Long-term[1]	Annual	12	-	GDP
Tsui & Balli (2017) International arrivals forecasting for Australian airports and the impact of tourism marketing expenditure	Journal article	19	10	Adelaide, Brisbane, Cairns, Darwin, Gold Coast, Melbourne, Perth,	Box-Jenkins SARIMA, SARIMAX, SARIMAX/EGARCH volatility model	Time series	-	Monthly	81	MAPE, MAE, RMSE	GDP (t, t-1, t-2), tourism marketing expenditure, total scheduled international flight seats
Ferhatosmanoglu & Macit (2016) Incorporating explanatory effects of neighbour airports in forecasting models for airline passenger volumes	Conference paper	4	3	Ankara Esenboga Airport & Istanbul Ataturk Airport	ARIMA, TBATS, regression with ARIMA errors	Time series	Short-term[1]	Hourly (6-hour interval)	1460	MAE, MAPE, MASE	-
Kim & Shin (2016) Forecasting short-term air passenger demand using big data from search engine queries	Journal article	45	27	Incheon International Airport	Regression model	Causal	Short-term	Monthly	51	Percent error rate (deviation from actual value), mean error	51 key queries from a search engine with time shift of 8 months
Xiao et al. (2016) Oscillations extracting for the management of passenger flows in the airport of Hong Kong	Journal article	7	5	Hong Kong International Airport	Hybrid oscillations analysis model (based on singular spectrum analysis (SSA) and generalized regression neural network (GRNN))	AI	Short-term	Monthly	102	MAE, MAPE, RMSE, Directional change (DC)	-
Tsui et al. (2014) Forecasting of Hong Kong airport's passenger throughput	Journal article	108	107	Hong Kong International Airport	Box-Jenkins SARIMA, ARIMAX	Time series	Short-term, medium-term[1]	Monthly	224	MAPE, RMSE	GDP per capita, lagged GDP per capita by 1-4, connecting traffic, visitors by air transport, SARS, Cross - straight agreement, Fuel prices, Individual visit scheme
Xiao et al. (2014) A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting	Journal article	73	64	Hong Kong International Airport	Neuro-fuzzy combination model (based on singular spectrum analysis (SSA), adaptive-network-based fuzzy inference system (ANFIS) and improved particle swarm optimization (IPSO))	AI	Short-term	Monthly	108	MAE, MAPE, RMSE, DC	-

continues on the next page…

continued (appendix 1)

Authors (year) title	Type of publication	GS citations	MA citations	Airport(s) covered	Forecasting Method	Proposed classification	Forecast horizon	Data	Number of time-units (hours, months, years)	Error measures	Regressors (passenger/trend not mentioned if time-series in question)
Xie et al. (2014) Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches.	Journal article	53	31	Hong Kong International Airport	Two hybrid approaches: X12-ARIMA-LSSVR and TRAMO/SEATS-LSSVR (based on seasonal decomposition and least squares support vector regression (LSSVR) model)	AI	Short-term	Monthly	170	MAE, MAPE,RMSE	-
Scarpel (2013) Forecasting air passengers at Sao Paulo International Airport using a mixture of local experts model	Journal article	20	12	São Paulo International Airport	Integrated mixture of local experts model (applying multiplicative Holt Winters model + Monte Carlo simulation)	Time series	Short-term, long-term	Monthly	252	MAPE	GDP
Wadud (2013) Simultaneous modeling of passenger and cargo demand at an airport	Journal article	18	9	Shahjalal International Airport	OLS regression, seemingly unrelated regression (SUR)	Causal	Long-term	Annual	28	Standard errors, adjusted R^2	GDP, crude oil price national price level
Kressner & Garrow (2012) Lifestyle segmentation variables as predictors of home-based trips for atlanta, Georgia, airport	Journal article	11	7	Hartsfield–Jackson Atlanta International Airport	Least-square regression	Causal	-	-	-	Adjusted R^2	Income groups, demographic profiles: Travel clusters and non-travel clusters
Profillidis (2012) An ex-post assessment of a passenger demand forecast of an airport	Journal article	6	4	Rhodes Airport	Linear model, polynomial second-degree model	Causal	-	Annual	-	Mean divergence (D_{mean}) = MAPE	-
Wadud (2011) Modeling and forecasting passenger demand for a new domestic airport with limited data	Journal article	15	8	The proposed airport in Khulna	Gravity model, panel regression	Causal	Long-term[1]	Annual	36	-	GDP per capita, population, travel time ratio (between air and road travel)
Samagaio & Wolters (2010) Comparative analysis of government forecasts for the Lisbon Airport	Journal article	37	23	Lisbon Airport	Holt-Winters, ARIMA, Gardner & McKenzie model, Grubb & Mason model	Time series, judgement	Long-term	Monthly	149	Root mean square percentage error (RMSPE)	-
Suryani (2010) Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework	Journal article	197	186	Taiwan Taoyuan International Airport	System dynamics	Decision support tool	-	Annual	12	Error rate (percentage)	airfare, service level, GDP, population, number of daily flights, dwell time
Profillidis (2000) Econometric and fuzzy models for the forecast of demand in the airport of Rhodes	Journal article	84	76	Rhodes Airport	Fuzzy regression	Causal	Medium-term	Annual	32	R^2	1) the exchange rate of Greek currency per year compared to the currencies of origin countries passengers of the airport 2) the dummy variable (1 for 1991 Gulf War), 0 for the other years
Graham (1999) Airport-specific traffic forecasts: a critical perspective	Journal article	38	20	-	-	-	-	-	-	-	-
Strand (1999) Airport-specific traffic forecasts: the resultant of local and nonlocal forces	Journal article	14	12	-	-	-	-	-	-	-	-
Ashley et al. (1995) A policy-sensitive traffic forecasting model for Schiphol Airport	Journal article	12	6	Amsterdam Schiphol Airport	Econometric model	Causal	Long-term[1]	Annual	-	-	Set of economic factors
Kawad & Prevedouros (1995) Forecasting air travel arrivals: model development and application at the Honolulu international airport	Journal article	7	5	Honolulu International Airport	Cochrane-Orcutt regression procedure	Causal	Short-term, medium-term	Annual	19	Error rate (percentage)	**USA model:** Annual change of GDP, Annual change of the consumer price index, strike-dummy (1985 United Airlines), Nature-dummy (Hurricane Iniki), Dummy-variable on DJIA (Dow Jones Industrial Average index) **Japan model:** Annual change of GNP, YEN-dummy for large increases in the strenght of Yen, War-dummy for Persian Gulf War 1991, Recession dummy (value 1 during 1983-1991), Dummy variable to account for the "cultural habit" of little vacation and foreign travel before the 1980s
Uddin et al. (1985) Methodology for forecasting air travel and airport expansion needs	Journal article	13	5	Robert Mueller Municipal Airport, Austin, Texas	Box-Jenkins ARIMA, multiple linear regression	Time series, causal	Medium- to long-term[1]	Monthly	77	R^2	Sales tax revenue, population
Karasek (1982) Forecasting and Planning the Jeddah Air Traffic with a Mini Model	Journal article	2	2	Jeddah Airport	Econometric model	Causal	Long-term	Annual	-	-	**Economic factors:** Government spending, imports, GDP Socio-economic factors standard of living (Public consumption), **pre-determined and exogenous variables:** Exchange rate, average realized price (per barrel of crude oil), Stability; **Policy:** Government spending, Crude oil production

Appendix 2. Forecasting results with monthly data prior to COVID-19 pandemic

Training set 01/2010 - 12/2018

ARIMA (0,1,1)(1,1,0)[12]	3 steps ahead prediction	6 steps ahead prediction	12 steps ahead prediction	actual	error	error-%	Absolute error	Absolute percentage error	Standard error
January 2019	1 633 883	1 633 883	1 633 883	1 563 007	70 876	4,53 %	70876	0,0453	5023407376
February 2019	1 610 510	1 610 510	1 610 510	1 531 267	79 243	5,17 %	79243	0,0517	6279453049
March 2019	1 809 255	1 809 255	1 809 255	1 745 263	63 992	3,67 %	63992	0,0367	4094976064
April 2019	-	1 803 581	1 803 581	1 798 639	4 942	0,27 %	4942	0,0027	24423364
May 2019	-	1 926 170	1 926 170	1 932 232	-6 062	-0,31 %	6062	0,0031	36747844
June 2019	-	2 061 290	2 061 290	2 050 062	11 228	0,55 %	11228	0,0055	126067984
July 2019	-	-	2 090 207	2 101 303	-11 096	-0,53 %	11096	0,0053	123121216
August 2019	-	-	2 022 516	2 023 868	-1 352	-0,07 %	1352	0,0007	1827904
September 2019	-	-	1 957 026	1 923 122	33 904	1,76 %	33904	0,0176	1149481216
October 2019	-	-	1 958 219	1 903 679	54 540	2,86 %	54540	0,0286	2974611600
November 2019	-	-	1 737 954	1 592 224	145 730	9,15 %	145730	0,0915	21237232900
December 2019	-	-	1 801 796	1 692 612	109 184	6,45 %	109184	0,0645	11921145856
MAPE	.0446	.0242	.0294						
MAE	71370	39391	49346						
RMSE	71642	50966	66453						

Training set 01/2010 - 12/2018

PROPHET	3 steps ahead prediction	6 steps ahead prediction	12 steps ahead prediction	actual	error	error-%	Absolute error	Absolute percentage error	Standard error
January 2019	1666606	1666606	1666606	1 563 007	103 599	6,63 %	103599	0,0663	10732752801
February 2019	1681177	1681177	1681177	1 531 267	149 910	9,79 %	149910	0,0979	22473008100
March 2019	1853112	1853112	1853112	1 745 263	107 849	6,18 %	107849	0,0618	11631406801
April 2019	-	1791175	1791175	1 798 639	-7 464	-0,41 %	7464	0,0041	55711296
May 2019	-	1916367	1916367	1 932 232	-15 865	-0,82 %	15865	0,0082	251698225
June 2019	-	2054779	2054779	2 050 062	4 717	0,23 %	4717	0,0023	22250089
July 2019	-	-	2060701	2 101 303	-40 602	-1,93 %	40602	0,0193	1648522404
August 2019	-	-	2009758	2 023 868	-14 110	-0,70 %	14110	0,0070	199092100
September 2019	-	-	1961323	1 923 122	38 201	1,99 %	38201	0,0199	1459316401
October 2019	-	-	2004760	1 903 679	101 081	5,31 %	101081	0,0531	10217368561
November 2019	-	-	1786719	1 592 224	194 495	12,22 %	194495	0,1222	37828305025
December 2019	-	-	1798342	1 692 612	105 730	6,25 %	105730	0,0625	11178832900
MAPE	.0753	.0401	.0437						
MAE	120453	64901	73635						
RMSE	122253	86763	94736						

Training set 01/2010 - 12/2018

TBATS	3 steps ahead prediction	6 steps ahead prediction	12 steps ahead prediction	actual	error	error-%	Absolute error	Absolute percentage error	Standard error
January 2019	1610330	1610330	1610330	1 563 007	47 323	3,03 %	47323	0,0303	2239466329
February 2019	1599421	1599421	1599421	1 531 267	68 154	4,45 %	68154	0,0445	4644967716
March 2019	1788715	1788715	1788715	1 745 263	43 452	2,49 %	43452	0,0249	1888076304
April 2019	-	1718312	1718312	1 798 639	-80 327	-4,47 %	80327	0,0447	6452426929
May 2019	-	1818973	1818973	1 932 232	-113 259	-5,86 %	113259	0,0586	12827601081
June 2019	-	1936680	1936680	2 050 062	-113 382	-5,53 %	113382	0,0553	12855477924
July 2019	-	-	1929706	2 101 303	-171 597	-8,17 %	171597	0,0817	29445530409
August 2019	-	-	1881911	2 023 868	-141 957	-7,01 %	141957	0,0701	20151789849
September 2019	-	-	1830318	1 923 122	-92 804	-4,83 %	92804	0,0483	8612582416
October 2019	-	-	1835724	1 903 679	-67 955	-3,57 %	67955	0,0357	4617882025
November 2019	-	-	1627388	1 592 224	35 164	2,21 %	35164	0,0221	1236506896
December 2019	-	-	1633383	1 692 612	-59 229	-3,50 %	59229	0,0350	3508074441
MAPE	.0332	.0430	.0459						
MAE	52976	77650	86217						
RMSE	54076	82571	95079						

continues on the next page

continued (Appendix 2)

Training set 01/2010 - 12/2018 **MLP**	**3 steps ahead prediction**	**6 steps ahead prediction**	**12 steps ahead prediction**	**actual**	**error**	**error-%**	**Absolute error**	**Absolute percentage error**	**Standard error**
January 2019	1632285	1632285	1632285	1 563 007	69 278	4,43 %	69278	0,0443	4799441284
February 2019	1626354	1626354	1626354	1 531 267	95 087	6,21 %	95087	0,0621	9041537569
March 2019	1810331	1810331	1810331	1 745 263	65 068	3,73 %	65068	0,0373	4233844624
April 2019	-	1801736	1801736	1 798 639	3 097	0,17 %	3097	0,0017	9591409
May 2019	-	1888863	1888863	1 932 232	-43 369	-2,24 %	43369	0,0224	1880870161
June 2019	-	2030231	2030231	2 050 062	-19 831	-0,97 %	19831	0,0097	393268561
July 2019	-	-	2035675	2 101 303	-65 628	-3,12 %	65628	0,0312	4307034384
August 2019	-	-	1979083	2 023 868	-44 785	-2,21 %	44785	0,0221	2005696225
September 2019	-	-	1923914	1 923 122	792	0,04 %	792	0,0004	627264
October 2019	-	-	1929923	1 903 679	26 244	1,38 %	26244	0,0138	688747536
November 2019	-	-	1726909	1 592 224	134 685	8,46 %	134685	0,0846	18140049225
December 2019	-	-	1756491	1 692 612	63 879	3,77 %	63879	0,0377	4080526641
MAPE	.0479	.0296	.0306						
MAE	76478	49288	52645						
RMSE	77620	58250	64279						

Training set 01/2010 - 12/2018 **ELM**	**3 steps ahead prediction**	**6 steps ahead prediction**	**12 steps ahead prediction**	**actual**	**error**	**error-%**	**Absolute error**	**Absolute percentage error**	**Standard error**
January 2019	1664926	1664926	1664926	1 563 007	101 919	6,52 %	101919	0,0652	10387482561
February 2019	1671979	1671979	1671979	1 531 267	140 712	9,19 %	140712	0,0919	19799866944
March 2019	1837728	1837728	1837728	1 745 263	92 465	5,30 %	92465	0,0530	8549776225
April 2019	-	1814773	1814773	1 798 639	16 134	0,90 %	16134	0,0090	260305956
May 2019	-	1880677	1880677	1 932 232	-51 555	-2,67 %	51555	0,0267	2657918025
June 2019	-	1997194	1997194	2 050 062	-52 868	-2,58 %	52868	0,0258	2795025424
July 2019	-	-	2003106	2 101 303	-98 197	-4,67 %	98197	0,0467	9642650809
August 2019	-	-	1957177	2 023 868	-66 691	-3,30 %	66691	0,0330	4447689481
September 2019	-	-	1917259	1 923 122	-5 863	-0,30 %	5863	0,0030	34374769
October 2019	-	-	1924741	1 903 679	21 062	1,11 %	21062	0,0111	443607844
November 2019	-	-	1749184	1 592 224	156 960	9,86 %	156960	0,0986	24636441600
December 2019	-	-	1768268	1 692 612	75 656	4,47 %	75656	0,0447	5723830336
MAPE	.0700	.0452	.0424						
MAE	111699	75942	73340						
RMSE	113633	86072	86303						

Appendix 3: Forecasting results with monthly data during the COVID-19 pandemic

training set 01/2010 - 9/2020 **ARIMA (1,1,0)(2,0,0)[12]**	**3 steps ahead prediction**	**actual**	**error**	**error-%**	**Absolute error**	**Absolute percentage error**	**Standard error**
October 2020	141 741	146571	-4 830	-3,30 %	4830	0,0330	23328900
November 2020	-81 946	121610	-203 556	-167,38 %	203556	1,6738	41435045136
December 2020	-6 389	145956	-152 345	-104,38 %	152345	1,0438	23208999025
MAPE	0,9169						
MAE	120244						
RMSE	146819						

training set 01/2010 - 9/2020 **PROPHET**	**3 steps ahead prediction**	**actual**	**error**	**error-%**	**Absolute error**	**Absolute percentage error**	**Standard error**
October 2020	747 083	146571	600 512	409,71 %	600512	4,0971	360614662144
November 2020	466 543	121610	344 933	283,64 %	344933	2,8364	118978774489
December 2020	384 637	145956	238 681	163,53 %	238681	1,6353	56968619761
MAPE	2,856						
MAE	394709						
RMSE	422911						

training set 01/2010 - 9/2020 **TBATS**	**3 steps ahead prediction**	**actual**	**error**	**error-%**	**Absolute error**	**Absolute percentage error**	**Standard error**
October 2020	92 830	146571	-53 741	-36,67 %	53741	0,3667	2888095081
November 2020	-180 249	121610	-301 859	-248,22 %	301859	2,4822	91118855881
December 2020	-188 151	145956	-334 107	-228,91 %	334107	2,2891	111627487449
MAPE	1.7126						
MAE	229 902						
RMSE	261 811						

training set 01/2010 - 9/2020 **MLP**	**3 steps ahead prediction**	**actual**	**error**	**error-%**	**Absolute error**	**Absolute percentage error**	**Standard error**
October 2020	-710 357	146571	-856 928	-584,65 %	856928	5,8465	734325597184
November 2020	-1 381 464	121610	-1 503 074	-1235,98 %	1503074	12,3598	2259231449476
December 2020	-1 200 642	145956	-1 346 598	-922,61 %	1346598	9,2261	1813326173604
MAPE	9.1441						
MAE	1235533						
RMSE	1265818						

training set 01/2010 - 9/2020 **ELM**	**3 steps ahead prediction**	**actual**	**error**	**error-%**	**Absolute error**	**Absolute percentage error**	**Standard error**
October 2020	154 354	146571	7 783	5,31 %	7783	0,0531	60575089
November 2020	148 057	121610	26 447	21,75 %	26447	0,2175	699443809
December 2020	141 761	145956	-4 195	-2,87 %	4195	0,0287	17598025
MAPE	.0998						
MAE	12808						
RMSE	16100						