

Improving tourist arrival prediction: A big data and artificial neural network approach

Abstract

Due to high fluctuations of tourism demand, accurate predictions of tourist arrivals are of high importance for tourism organisations. The study at hand presents an approach to enhance autoregressive prediction models by including travellers' web search traffic as external input attribute for tourist arrival prediction. The study proposes a novel method to identify relevant search terms and to aggregate them into a compound web-search index, used as additional input of an autoregressive prediction approach. As methods to predict tourism arrivals, the study compares autoregressive moving average (ARIMA) models with the machine learning-based technique artificial neural network (ANN). Study results show that (1) Google Trends data, mirroring traveller's online search behavior (i.e. big data information source), significantly increase the performance of tourist arrival prediction compared to autoregressive approaches using past arrivals alone, and (2) the machine learning technique ANN has the capacity to outperform ARIMA models.

Keywords

Big data; web search traffic; Google Trends data; tourist arrival forecasting; ARIMA; artificial neural networks

1 Introduction

Understanding, explaining and forecasting tourism demand has been an important area of scientific enquiry and practice for a long time (cf. Witt & Witt 1995), and is also a relevant domain in contemporary tourism research (Frechtling 2002; 2011). Forecasting tourism demand is a way of reducing risk and is imperative in tourism, since the tourism service product is perishable and the process of production and consumption is inseparable, depends on complementary services and is extremely sensitive to crises (Frechtling 2011; Vanhove 2011). Forecasting in tourism implies indicating the direction of future demand and, as such, it provides meaningful information for destination management and the various tourism suppliers (Vanhove 2011). Tourism demand modelling and forecasting use a variety of measures. However, a commonly used measure is tourist arrivals (Song & Li 2010). Managerial interest in accurate forecasts of future events and changes in tourism demand is related to the extent to which actions can be implemented to influence the demand, monitor fluctuations for planning or resource capacities as well as manage the (e.g. socio-economic) effects of demand. Hence, the success of tourism businesses largely depends on the ability to predict tourism demand most accurately. In turn, the consequences of poor forecasting lead to poor marketing effectiveness, inefficient use of resources and decay of sustainability (Frechtling 2011).

Accurate forecasting of tourism demand is a thought-provoking science and art (Li, Chen, Wang & Ming 2018). As a key finding by Song & Li (2008) and Song, Qiu & Park (2019), the methods used in modelling and forecasting tourism demand are highly diverse, and, in fact, there is no single model that consistently outperforms other models in all situations. A further challenge in demand forecasting is the access to timely and cost-effective data (Önder 2017). Other essential challenges include the lack of historical time series data and

demand volatility (Frechtling 2011; Song, Li, Witt & Fei 2010). Therefore, new data sources and modeling techniques are crucial issues of contemporary tourism research and big data sources (e.g. web traffic or search engine traffic), apparently, have shown promising potentials to overcome these issues.

Even at an early stage, the use of big data in tourism research has achieved, even at an early stage, amazing improvements, both in terms of new theoretical and methodological insights (Li, Xu, Tang, Wang & Li 2018; Mariani, Baggio, Fuchs & Höpken 2018). In recent literature, big data has already been used for tourism demand predictions (Yang, Pan, & Song 2014; Önder & Gunter, 2016; Höpken et al. 2017b; Höpken et al. 2018; Liu, Tseng & Tseng 2018). A growing group of international tourism scholars has concluded that, in particular, search engine traffic has the potential to significantly increase the accuracy and robustness of tourism demand forecasts (cf. Antolini & Grassini 2019; Bangwayo-Skeete & Skeete 2015; Bokelmann & Lessmann 2019; Camacho & Pacce 2018; Dergiades, Mavragani & Pan 2018; Höpken et al. 2017b; Jackman & Naitram 2015; Kim & Malek 2018; Li, Pan, Law & Huang 2017; Li & Law 2020; Pan & Yang 2017; Park, Lee & Song 2017; Rivera 2016; Siliverstovs & Wochner 2018; Sun, Wang, Wei & Yang 2017; Sun, Wei, Tsui & Wang 2019; Wu, Song & Shen 2017; Yang, Pan, Evans & Lv 2015; Zhang, Huang, Li & Law 2017; Önder & Günter 2016; Önder 2017). However, further efforts to improve the performance of forecasting models by incorporating tourist online behavioural data are advocated (Wu et al. 2017). More precisely, some researchers point out, that there are particular challenges in using Google Trends data, such as the constant changes in the search result ranking-algorithm and changes to the functionality of searches (Rivera 2016).

In past literature, non-causal time series models and causal econometric models are the two dominant approaches used for quantitative demand modeling (Song & Li 2008; Song &

Turner 2006). More concretely, integrated autoregressive moving average (ARIMA) models (Box & Jenkins 1970) appear most frequently in the tourism literature (Song & Li 2008, 210; Song, Qiu & Park 2019). Findings by Moro and Rita (2016) show that time series analyses are still the most strongly adopted methods and that in particular the seasonality phenomena in tourism continues to justify this use (Kulendran & Wong 2005; Chen, Li, Wu & Shen 2019). Yet, a major advantage of econometric approaches is their ability to identify economic factors influencing tourism demand (Peng, Song & Crouch 2014; Athanasopoulos, Song & Sun 2018). Most interestingly, recent research has sought to find explanations of tourism demand beyond neo-classic economic theory by including additional explanatory variables, such as tourists' online behaviour (Wu, Song & Shen 2017).

It has even been proposed that the emerging big data paradigm is set to transform the landscape for socio-economic policy and research along with management and decision-making (Blazquez & Domenech 2018). In fact, big data sources, like web search traffic, web usage data, or customer online feedback, are naturally related to tourism demand and, therefore, turn out to be valuable inputs for prognosticating tourism demand.

Furthermore, from a methodological perspective, machine-learning approaches, like k-nearest neighbor (k-NN), support vector machines (SVM) or artificial neural networks (ANNs), are increasingly used for the purpose of tourism demand modeling and forecasting (Moro & Rita 2016). These artificial intelligence-based approaches offer the advantage of not depending on specific statistical characteristics of the data set at hand, such as (normal) distribution, linearity, and non-collinearity (Song & Li 2008, 212), and being more robust against biased, incomplete, redundant and noisy data (Li, Wu, Peng & Lv 2016). Thus, machine learning-based approaches typically yield superior results, not only for non-causal,

but in particular also for causal approaches, showing a high-dimensional input space (Kon & Turner 2005; Song & Li 2008; Lin, Chen, & Lee 2011; Ricardo, Goncavales & Costa 2018).

The objective of this paper is to enhance tourism demand forecasting based on past arrivals alone (i.e. autoregressive approaches), by including travellers' induced web search traffic as an additional explanatory input variable. More concretely, the study presents a novel approach to identify and aggregate tourism-relevant search terms, and evaluates whether the corresponding web search volume, represented by Google Trends data as additional input to forecast tourist arrivals, can increase prediction accuracy compared to models using past arrivals alone (*research proposition 1*). As a second objective, the study uses machine learning techniques for predicting tourist arrivals (i.e. artificial neural networks) additionally to an ARIMA model as a comparative statistical approach. As said, machine learning techniques are not limited to linear models and are typically more robust against data not conforming to specific statistical characteristics, like non-collinearity of input attributes or biased data. These characteristics of machine learning approaches make them a promising methodological alternative, especially in the case of more complex and diverse input data, which holds true when adding big data sources as additional input for arrival prediction. Thus, in this study we further evaluate whether the machine learning-based method of artificial neural networks (ANNs) achieves better prediction accuracies than ARIMA models as a statistical approach, also in case of a big data enriched prediction (*research proposition 2*). The tourist destination of Sweden serves as a case for this study, using arrival data of major sending countries (i.e. Denmark, Finland, Norway, Russian Federation, and United Kingdom) and Google Trends data for the period 2008-2016, including text and video search, respectively.

The structure of the paper is as follows: Section 2 gives an overview of related literature dealing with the task of tourism demand prediction when including travellers' online search traffic. Section 3 discusses the study design and the methodology by presenting techniques of collecting and preparing data as well as the process of model building. Section 4 presents major findings. Finally, section 5 summarizes and discusses insights and results, while section 6 considers limitations of the study and points at possible future extensions and consecutive research activities.

2 Literature review

2.1 Demand modeling in tourism

As outlined, literature on quantitative demand modeling consists of two major sub-domains: (non-causal) time series modeling and (causal) econometric methods. Time-series models explain a target variable based on its own past values. In the course of the past decades, tourism research has used the integrated autoregressive moving average (ARIMA) approach (Box & Jenkins 1970) heavily (Song & Li 2008, 210; Song, Qiu & Park 2019). Moreover, exponential smoothing models (Geurts & Ibrahim 1975; Lim & McAleer 2001; Cho 2003) as well as shift-share techniques (Fuchs, Rijken, Peters, & Weiermair 2000) are found in the literature for forecasting and modeling tourism demand. Econometric approaches, by contrast, offer the advantage to enable the analysis of causal relationships between tourism demand (i.e. the dependent variable) and its explanatory variables (Peng, Song, & Crouch 2014; Peng et al. 2014). Recent econometric studies proposed a broad range of possible determinants of tourism demand (Khaidi, Abu & Muhammad 2019), for instance, consumer price index in a destination, substitute prices, Gross Domestic Product, currency exchange rates, but also interest and unemployment rates, as well as export/import rates (Cho 2001; Song & Li 2008, 211; Athanasopoulos, Song & Sun 2018). Additionally, events (especially mega-events), advertising investments (Divisekera & Kulendran 2006; Kronenberg et al. 2016), but also financial crises or terrorist attacks (Smeral 2009; Smeral 2017; Song & Lin 2010) and disasters (e.g. SARS, H1N1 pandemy, etc.) (Chen, Kang, & Yang 2007) have shown to significantly influence tourism demand. From a statistical point of view, the error correction model (EDM), autoregressive distributed lag models (ADLM), the time varying parameter (TVP) model, and the vector autoregressive (VAR) model emerged as the main

econometric models (Peng et al. 2014). In addition, tourism demand modelling also used the linear structural equation model (SEM (Turner & Witt 2001)).

2.2 Using Google Trends data to predict tourism arrivals

In recent years, researchers all over the globe used web search traffic for predicting economic indicators (Wu & Brynjolfsson 2015). For instance, Vosen & Schmidt (2011) constructed a model to forecast private consumption using Google Trends data and revealed that the addition of web search data outperforms the majority of survey-based factors. Carrière-Swallow & Labbé (2013) could improve the prediction accuracy for automobile purchases by using online search traffic as additional input, while Hand & Guy (2012) reached the same result for cinema admissions.

Information search is a fundamental feature of consumer decision-making behavior (Xiang, Magnini & Fesenmaier 2015). In a tourism context, travel information search is defined as the stage of the decision-making process, wherein travelers actively collect and integrate information from numerous sources *prior* to making their travel decision and destination choice (Vogt & Fesenmaier 1998; Fodness & Murray 1999). In fact, travel information search serves a variety of travelers' goals, from simple ones, addressing basic functional information needs, such as knowledge about the price of a hotel, to highly emotional ones, such as understanding the symbolic meaning of destination places (Xiang, Magnini & Fesenmaier 2015; Fuchs & Baggio 2017).

Since the late 1990s, the Internet has fundamentally changed the way tourism-related information is distributed and the way travelers search for travel products. Travelers particularly use search engines to find relevant information for all aspects of the trip, including accommodations, attractions, activities, and dining (Pan & Fesenmaier 2006; Steen

Jacobsen & Munar 2012; Xiang, Wang, Leary & Fesenmaier 2015; Choe et al. 2017).

Technically, every time a tourist interacts with the Internet, be it through a search engine, a website, a mobile phone, or a social media platform, electronic traces of this interaction can be captured, stored, and analyzed, later on (Fuchs et al. 2014; Höpken et al. 2015).

Researchers use these online data, such as search engine query volumes, amount and types of tweets, website traffic, and social media posts, for various analytical purposes, such as online customer segmentation (Pitman et al. 2010) and tourism sentiment analysis (Schmunk et al. 2014; Höpken et al. 2017a). Most importantly, search engine query volumes are successfully used for tourism demand forecasting and modelling (Artola, Pinto & de Pedraza García 2015; Yang et al. 2015; Padhi & Pati 2017; Siliverstovs & Wochner 2018; Höpken et al. 2019).

Literature on travelers' search behavior builds upon different theoretical foundations, like economic approaches (Stabler et al. 2010; Kronenberg et al. 2016), information processing theories (Chung et al. 2015), and theory of planned behavior - TPB (Ajzen 2005; Erawan et al. 2011). In fact, the *best* predictor of tourist behavior in selecting travel destinations when searching online seems to be information-processing (Padhi & Pati 2017, 36). More precisely, the TPB framework suggests that factors, like attitude and behavioral control, are strongly associated with tourist behavior and intention for online search and subsequent holiday bookings (Kim et al. 2016). Indeed, with the current high penetration of the internet, online search engines are gratifying a wide spread of individuals' needs compared with traditional information sources (Padhi & Pathi 2017, 36). Thus, recent studies have impressively demonstrated that Google Trends data reflect crucial aspects of tourists' keyword-based queries which, in turn, provide vast opportunities to investigate and predict travelers' planned behavior (ibid 2017; Yang et al. 2015; Kim et al. 2016; Höpken et al. 2019).

A number of researchers in the travel and tourism domain have begun to use web search data to predict tourism demand. The latter variable is typically expressed in terms of tourist arrivals. For example, Bangwayo-Skeete and Skeete (2015) underline that web search traffic increases the quality of predicting tourism demand, based on autoregressive mixed-data sampling (AR-MIDAS) models. Likewise, Önder and Gunter (2016) demonstrate that Google Trends data for text and image search enhances the quality of tourism demand forecasts, as compared to simple exponential smoothing time-series models (e.g. Holt-Winters) or autoregressive models. Li et al. (2016) make use of search engine data for tourism forecasting with noise processing. The work by Yang et al. (2015) demonstrates that employing the volume of web search traffic for tourist arrivals prediction helps to improve forecasting accuracy significantly as compared to auto-regressive moving average (ARMA) models. Pan, Wu and Song (2012), Yang, Pan and Song (2014) and Pan and Yang (2017) use web search traffic to improve the forecast accuracy of hotel demand. More recently, the study by Li, Pan, Law and Huang (2017) presents an approach for a composite search index integrated into a generalized dynamic factor model (GDFM) to forecast tourist demand. Findings show that the approach increases prediction accuracy compared to a time series model and a search index-based model created by principal component analysis. Moreover, Camacho and Páez (2018) show that Google's search volume indices improve predictions of overnight stays in Spain, thereby outperforming models that exclude these leading indicators. More recently, Antolini and Grassini (2019) use Google Trends data to predict foreign arrivals in Italy. Thereby, assessing the contribution of lagged Google Trends variables in a standard ARIMA model and in a time series regression model with seasonal dummies and autoregressive components. In a similar way, Gunter, Önder and Gindl (2019) integrate Google Trends data in autoregressive distributed lag (ADL) models to predict tourist arrivals

in four Austrian cities. Most recently, Li and Law (2020) demonstrate that a decomposition-based approach using Google Trends data is particularly superior in forecasting turning points.

2.3 Machine learning and demand prediction

Fairly recently, non-statistical, i.e. machine learning, methods have been applied to tourism demand prediction. The main advantage of machine learning techniques over statistical approaches is that they do not make any preliminary assumptions about the data, such as normal distribution, linearity, and non-collinearity (Yu & Schwartz 2006; Song & Li 2008, 212; Song, Qiu & Park 2019). Concretely, the machine learning methods artificial neural network (ANN), rough set theory, fuzzy time-series method, genetic algorithms (GAs), support vector machines (SVM) and, most recently, deep learning approaches (Law, Li and Feng 2019) are commonly used for tourism demand forecasting. While emulating the human brain, artificial neural networks (ANN) became apparent as a superior model to predict tourism demand compared to ARIMA and multiple regression models, respectively (Kon & Turner 2005; Song & Li 2008, 212). Applications of ANNs for tourism demand prediction are presented by Law and Au (1998), Palmer, Montano and Sese (2006), Lin et al. (2011), Claveria and Torra (2014), Çuhadar, Cogurcu and Kukrer (2014), Claveria, Monte and Torra (2015) and Silva, Hassani, Heravi and Huang (2019). The approach of a decision rule induction is the basis of the rough set theory focusing on analyzing the classification of imprecise, uncertain or incomplete data (Song & Li 2008, 213). Law and Au (1998) and Goh, Law, and Mok (2008) have used the rough set theory for tourism demand modeling. The fuzzy time-series method has shown to work particularly well in analysing short time series with a limited amount of observations in the past (Hadavandi, Ghanbari, Shahanaghi, & Abbasian-Naghneh 2011; Tsaur & Kuo 2011). Genetic algorithms (GAs), i.e. optimization algorithms applying the fundamental principles of

evolution (Song & Li 2008, 213), are useful in recognizing changes in the structure of tourism demand (Hernández-López & Cáceres-Hernández 2007; Hong, Dong, Chen, & Wei 2011). Finally, support vector machines (SVM) are used in solving the non-linear estimation and prediction problem and have been successfully used for tourism demand analysis by Pai, Hong, Chang, and Chen (2006), Hong (2006), Chen and Wang (2007) and Pai, Hung, and Lin (2015). Importantly, empirical evidence shows that SVM can outperform ARIMA models and traditional counterparts in predicting tourism demand (Song & Li 2008). A study by Zhang, Huang, Li and Law (2017) hybridizes support vector regression (SVR) with the Bat algorithm to forecast tourist volume by incorporating search engine data, where the Bat algorithm is used to adjust the SVR parameters (ibid, 245). Most recently, Assaf, Li, Song and Tsionas (2019) have shown the advantages of Bayesian global vector autoregressive (BGVAR) models to capture spillover-effects of international tourism demand, typically accruing in touristically strongly interlinked countries, such as Southeast Asia.

However, when looking specifically at the comparison of ARIMA and ANN, the two approaches used in the study at hand, their relative superiority in predicting tourism demand is judged differently in research studies. In a study of Lin, Chen and Lee (2011), comparing ARIMA, ANN and MARS (multivariate adaptive regression splines) for predicting visitors to Taiwan, ARIMA outperformed ANN and MARS. Similarly, in a study of Claveria and Torra (2014) ARIMA outperformed ANN (as well as a self-exciting threshold auto-regression) when forecasting overnight stays to Catalonia, especially for shorter forecasting horizons. In contrast, Aslanargun et al. (2007) could show that models with non-linear components, like ANNs with a non-linear activation function, can outperform linear models like ARIMA, demonstrated in a study, which is forecasting tourist arrivals in Turkey. Burger et al. (2001) compare a variety of classical forecasting methods, like moving average, decomposition,

exponential smoothing, ARIMA, and multiple regression with non-traditional approaches, like genetic regression and ANN, applied to forecasting the US demand to Durban, South Africa, and could show that ANN reaches the overall best performance. Chun-Fu Chen, Lai, and Yeh (2012) combine a decomposition model with an ANN model and can show, based on forecasting international visitors to Taiwan, that the combined approach outperforms both an ANN, directly learned on the original time series data, as well as a traditional ARIMA model.

2.4 Preprocessing web search data

Preprocessing web search data typically does not follow a standard approach. Nevertheless, three main tasks for preparing search engine traffic as input to demand prediction emerged: keyword selection, identification of time lags and construction of search indices. *Keyword selection* aims at identifying potential keywords, either by making use of domain specific knowledge (e.g. by domain-specific ontologies), by employing web information extraction and text mining methods, or by keyword recommendations offered by search engine providers (Liu, Lv, Peng, & Yuan 2012). The second task, *identification of time lags*, intends to identify those time lags with the strongest correlation between tourist arrivals and corresponding search queries. The third task, *construction of search indices*, intends to combine a multitude of different search requests into a compound search index with a high predictive power in order to avoid both, multi-collinearity and the overfitting phenomenon, often caused by high-dimensional time series data (Varian 2014). Liu et al. (2012) used search engine requests to predict the Chinese stock market and proved that such a compound search index, which is aggregating lagged search queries, significantly increased forecasting performance. Yang et al. (2015) recently adopted this approach considering online search requests to forecast tourism demand.

Related to the second task, *identification of time lags*, the literature shows different approaches for measuring lagged correlations between predictor and target time series attributes. Liu et al. (2012) combined the Pearson correlation with the Kullback-Leibler divergences, while other authors use the Pearson correlation alone (Yang et al. 2015; Xiaoxuan, Qi, Geng & Benfu 2016; Pan, Li, Law & Huang 2017).

A final step of preprocessing search query data is the elimination of useless information by methods of noise reduction. Yang et al. (2015) and Xiaoxuan et al. (2016) point out the importance of noise reduction when using Google Trends data as forecasting input and propose the 'Hilbert-Huang-Transformation' (HHT), which shows the capacity to reduce prediction errors significantly. Another methodological advancement has been proposed by Peng, Liu, Wang, and Gu (2017) by using the Hurst exponent (Hurst, Black & Simaika 1965) to remove search queries with a low predictive power.

3 Methods

3.1 Data collection and preparation

Specification of data set

The dataset used in this study contains monthly data concerning inbound tourist arrivals to Sweden from major sending countries (Denmark, Finland, Norway, the Russian Federation and the United Kingdom) for the period between January 2008 and December 2016, resulting in 9 years of monthly time series data for each sending country. More precisely, the resulting dataset consists of 108 entries, reflecting past tourist arrivals to Sweden (Statistics Sweden 2017). Fig. 1 shows the time series of tourist arrivals per sending country for the complete observation period and clearly demonstrates a strong seasonality of tourist arrivals on a monthly basis.

--- Fig. 1 to be placed here ---

Besides data on tourist arrivals, the data set includes search traffic on web search engines for the different sending countries. Analogous to Bangwayo-Skeete and Skeete (2015), Yang et al. (2015), Önder and Gunter (2016), and Höpken et al. (2018), search traffic on the Google search engine has been selected, as Google is the predominant search engine in most sending countries (c.f. Pearson CMG 2017). Google's search volume is available via the Google Trends service, providing the relative search volume for specific keywords over time and geographical regions. Although Google's world dominance is undisputable, in specific sending countries, like China, Japan or Russia, local favourites are more popular than Google. In Russia, for instance, the majority of internet users make use of Yandex for web searches, while Google search plays only a subordinate role for most Russian internet users (Return On Now 2017). In the literature, this is known as platform bias, as users from certain countries

tend to perform searches on local search engines (Dergiades, Mavragani, & Pan 2018). Since Yandex does not provide search queries older than one year, we used Google for Russian travellers as well, although a low search intensity was expected.

Collection of web search data

We used Google's Keyword Planner for collecting the keywords that tourists tend to use when planning their trip to Sweden. Suggested keywords were used as input for retrieving appropriate time series from *Google Trends*, reflecting tourists' planning behaviour before visiting Sweden. Accordingly, we chose the search category "Travel and Tourism" to restrict results to search queries related to tourism. As the study's aim is the prediction of Swedish inbound tourist arrivals with respect to the above-mentioned sending countries, the search term "Sweden" was spelled in English as well as the official language of each sending country, i.e. "sverige" for travellers from Denmark and Norway, "ruotsi" and "sverige" for Finnish travellers, as well as "швеция" and "shvetsiya" for travellers from Russia, respectively. For Finland and Russia, two terms for "Sweden" were used in order to account for all the languages spoken in the respective sending country, thus, preventing the data to be biased by language (Dergiades, Mavragani, & Pan 2018). As a result, Google suggested a maximum number of 700 keywords for the sending countries Denmark and Norway, followed by 144 keyword suggestions for the sending country Finland. For Russia and the UK, Google suggested 29 and 52 keywords, respectively. Additionally, as a specific extension, each of the keyword lists were extended by adding 290 Swedish tourist destinations (i.e. regions, cities, and villages).

For the extraction of query series (defined as the daily search volume for a given search term over a specific time-period), an automatic crawling algorithm was individually developed.

The algorithm starts by iterating over a list of seed keywords (the keywords described above)

and extracts corresponding query series. If no query series exists for a given keyword, the algorithm skips the keyword. If the iteration finds a query series, the list of keywords is expanded by keywords the users are likely to use within the same search session, as suggested by Google. Besides obtaining time series reflecting search behaviour regarding Google Search (i.e. text search), as done by previous research, in this study, search queries are extended to tourism-related video content as well, in order to optimally reflect search behavior of potential tourists. Table 1 shows the number of search query series that correspond to the number of keywords used within tourism-related search queries for both text and video search queries collected for each sending country.

--- Table 1 to be placed here ---

Normalization of search terms

Compared to existing literature, a step of *normalization* enhances the process of identifying relevant search terms, intending to improve predictive power and to reduce any kind of redundancy by removing synonymous or interchangeable search terms (Liu 2008). More specifically, search term normalization deals with identifying linguistic variations, synonyms or even meaningless variations caused by misspellings. In this study, we handled the following types of variations concerning query names: First, queries containing the same terms but arranged differently (e.g. 'sweden skiing' vs. 'skiing sweden'); second, variations only caused by the existence of stop-words (e.g. 'skiing sweden' vs. 'skiing in sweden '); third, different keywords belonging to the same word stem, e.g. 'ski' and 'skiing'; additionally, the same search term appearing in different languages (e.g. 'sverige sää' and 'sweden weather'). Finally, queries may also differ by the usage of special characters, in this case by nordic special characters {Å, å, Ä, ä, Æ, æ} and {Ø, ø, Ö, ö}, which were substituted by {a} and {o}, respectively.

To cope with search term variations, as mentioned above, first, typical text preprocessing techniques were executed, i.e. tokenization, character substitution, stop-word elimination, and stemming (Liu 2008). Second, for each query we created a word vector, based on the contained search terms, and we calculated similarities based on cosine similarity (Liu 2008, 190). The cosine similarity $\cos(\theta)$ between two vectors A and B is defined as follows:

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|_2 \|\mathbf{B}\|_2} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

with A_i and B_i being word occurrences of word vectors A and B , respectively. Search queries, containing the same words in principle, but having a different order of words, other stopwords, or other word forms, will still have a cosine similarity of one and, thus, were merged.

3.2 Statistical evaluation of the arrival series

Statistical techniques for time series prediction usually require the time series to comply with the characteristics of *stationarity* (Mukherjee, White & Wuyts 1998, 335, Song, et al. 2010; Frechtling 2011), i.e. having a constant mean and variance (weak stationarity) and auto-correlations between two values being independent of the point in time within the series (strong stationarity) (Frechtling 2002).

To achieve stationarity, the time series were first analyzed for the existence of seasonal patterns, one of the main reasons for non-stationarity, especially in case of tourism arrival series. Yearly seasonality (i.e. a seasonal frequency of 12 months), has been tested by Maravall's QS test (Maravall 2011) and the Kruskal-Wallis (KW) test. The QS test is a variant of the Ljung-Box test executed on seasonal lags, where only positive auto-correlations are considered (Maravall 2011; Nwogu, Iwueze and Nlebedim 2016). By contrast, the KW test

(Kruskal et al. 1952) is similar to the Friedman test (Friedman 1937), where the observations are checked for significant variances in their period-specific mean ranks, with the difference that period-specific values of observations are assigned to ranks over the entire observation period. According to Webel et al. (2018), the KW-test can be understood as a one-way ANOVA without repeated measures.

Following the methodology of Webel and Oellech (Webel et al. 2018), the QS test was applied twice. First to the original arrival series and second to the fitted residuals of a non-seasonal ARIMA model, estimated with the Hyndman-Khandakar algorithm (Hyndman et al. 2008). Similarly, we applied the Kruskal-Wallis test to the residuals of the non-seasonal ARIMA model, as well. If the p-values of the QS tests are below 0.01 or the p-value of the Kruskal-Wallis test is below 0.002, the Webel-Oellech test will classify the corresponding time series as seasonal (Webel et al. 2018).

--- Table 2 to be placed here ---

Test results in Table 2 clearly show that seasonal patterns exist in the arrival time series as the p-value of all QS tests for all arrival series is below 0.01.

For that reason, we further analyzed the time series for the strength of seasonal patterns with the method of Yang et al. (2006), and for the existence of seasonal unit roots using the Osbourn-Chui-Smith-Birchenhall (OCSB) test (Osborn et al. 1988).

In a second step, the time series were analyzed for level- and trend-stationarity by the *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS) test, in contrast to other tests having stationarity as the null hypothesis (Hill, Griffith, & Lim 2011), and for covariance stationarity by the *Augmented Dickey-Fuller* (ADF) test, identifying non-seasonal unit-roots (Baddeley & Barrowclough 2009).

--- Table 3 to be placed here ---

Test results in Table 3 show that all time series either contain seasonal unit roots (Finland and Norway), were found to be not trend or level stationary (Denmark and Russian Federation) or both (United Kingdom). While the arrival series for the Russian Federation and the United Kingdom are neither trend nor level stationary, the KPSS test applied to the arrival series for Denmark suggests the time series to be at least trend stationary at a 90 per cent significance level. In addition to the seasonal unit roots found, the seasonal strength of all time series is close to unity, thus, confirming that the time series have strong seasonal patterns. According to Yang et al. (2006), time series should be differentiated when the measure of seasonal strength exceeds 0.64. Therefore, we first differentiated all time series with lag 12 to remove strong seasonality. Second, we again differentiated the arrival time series for Denmark, the Russian Federation and the United Kingdom with lag 1 in order to eliminate both, level- and trend non-stationarity as identified by the KPSS tests.

--- Table 4 to be placed here ---

Table 4 shows that all time series, except the arrival series for the Russian Federation, are clearly stationary after transforming the time series as described above. As the p-value of the KPSS test for Russia is only slightly above the threshold ($0.110 > 0.100$), no further action has been taken. The differentiation with lag 12 clearly eliminated seasonal unit roots and the strength of seasonality declined significantly. Therefore, no additional transformations were necessary.

3.3 Construction of aggregated web search indices

Both Song et al. (2010) and Liu et al. (2012) see search queries as a manifestation of tourists' preferences and needs, accounting for important trends that are relevant for the

development of tourism destinations. As dimension reduction techniques show the capability to increase the predictive power in case of a high-dimensional input space, different single search queries from the same sending country are aggregated into a compound search index (c.f. Yang et al. 2015). As specific information needs of tourists differ in the various phases of the information and decision-making process, different search queries, satisfying different needs, will have a different time lag with the corresponding tourist arrivals. Thus, before aggregating multiple search queries into a compound search index, the predominant time lag has to be identified for each search query, and each single search query is shifted by the appropriate time lag, before being incorporated into the search index. The described approach proved its capability to increase forecasting performance significantly (Liu et al. 2012).

Previous literature, often use the *Pearson correlation* coefficient to identify the time lag, leading to the highest correlation between a search query series and the corresponding arrival series (Yang et al. 2015; Xiaoxuan, Qi, Geng & Benfu 2016; Pan, Li, Law & Huang 2017). Thus, the first step of the index aggregation procedure is the calculation of the Pearson correlations between each search query and the corresponding arrival series for a time lag of zero to six months in order to cope with both short- and mid-term travel planning behavior (Fesenmaier, Xiang, Pan, & Law 2010). Based on such correlation coefficients, search query series were shifted towards the arrival series by their most dominant time lag (i.e. the time lag with the highest correlation), and weighted by the corresponding squared correlation coefficient. The approach of shifting each single search term series by its most dominant time lag before aggregating all search term series is justified by the fact that search terms are mostly used in one dominant month ahead of arrival and, thus, the corresponding time series shows a strong peak in this dominating month. According to

literature (Yang et al. 2015), online search activities executed immediately prior to arrival have typically no predictive power in forecasting tourism demand. Thus, queries with a dominant time lag of zero are excluded from the index building process.

Peng et al. (2017) have shown that an input series has a higher predictive power, if it follows the same auto-correlative patterns as the target series, measured by the *Hurst exponent* of both series. Hence, as an attempt to further increase the predictive power of the search index, we excluded query series with a significantly different *Hurst exponent* than the corresponding arrival series from the search index as well. However, we could show that the exclusion of the affected time series could not improve the forecasting accuracy in this specific case.

As a final step of input parameter selection, a backward-stepwise regression has been executed, to successively remove query series from the search index which do not lead to a significant increase of the correlation with the corresponding arrival series (Roecker 1991), thus, in order to increase model parsimony and generalizability (Liu 2008).

To ensure stationarity of the search index series, analogous to the arrival series, the (seasonal) unit root and stationarity tests were applied to the search index time series as well. Test results in Table 5 clearly show the absence of (seasonal) unit roots and, therefore, confirm stationarity. Thus, no further transformations had to be applied to the search index time series.

--- Table 5 to be placed here ---

Fig. 2 shows the stationary tourist arrival series and the corresponding aggregated search index series. The high conformity of the time series evidently demonstrate a potentially high predictive power of the aggregated search index related to tourist arrivals.

--- Fig. 2 to be placed here ---

The Pearson correlations between the arrival series and the corresponding search index series, shown in Table 6, confirms this.

--- Table 6 to be placed here ---

Importantly, the exact overlap of the main peaks of the arrival series and the corresponding search index series is simply caused by the fact that for each search term the corresponding search query series is shifted towards the arrival series by their dominant time lag, as explained before.

3.4 Model building

While typical autoregressive (i.e. univariate) forecasting approaches use past arrivals as the only prediction input, the study at hand makes use of web search traffic as an additional input variable (i.e. bivariate approach) (Hill et al. 2011; Frechtling et al. 2011). The aim is to evaluate whether the web search volume (for relevant search terms in form of Google Trends data), as additional input to predicting tourist arrivals can increase predictive accuracy, compared to using past arrivals alone (*research proposition 1*).

This study first makes use of a traditional ARIMA (i.e. autoregressive integrated moving average) model, proposed by Box and Jenkins (1970), to predict tourist arrivals autoregressively (i.e. univariate) and a regression model with ARIMA errors for the prediction with the bivariate datasets consisting of the tourist arrivals series and the corresponding aggregated search indices. Regression models with ARIMA errors are equal to linear regression models with

$$y_t = \beta_0 + \beta_1 x_t + \eta_t$$

where η_t describes the ARIMA error for a given model (Hyndman & Athanasopoulos 2018). In both cases (i.e. univariate and bivariate), we chose appropriate ARIMA models by using the Hyndman-Khandakar algorithm which selects the model with the lowest AICc (Akaike Information Criteria) by estimating models with different combinations of the p and q parameters (i.e. the order of the autoregressive part and the order of the moving average part). We determined the optimal parameter combinations by a stepwise search, traversing the model space (Hyndman et al. 2008).

Due to the fact, that the degree of differentiation involved is zero for all time series (because the time series were made stationary before), the fitted models are equal to ARMA models for the autoregressive prediction and regression models with ARMA errors for the bivariate datasets, respectively. Table 7 shows the models fitted to the univariate as well as the bivariate data (i.e. the arrivals series with corresponding aggregated search indices).

--- Table 7 to be placed here ---

In addition to ARIMA models, we applied *artificial neural network (ANN)* models in this study, representing a modern machine-learning approach for time series prediction (Kamel et al. 2008). Artificial neural networks (ANNs) represent a well-known type of machine learning technique, used for both, supervised learning (i.e. classification, estimation or prediction) as well as unsupervised learning (i.e. clustering) (Du & Swamy 2019). ANNs imitate the human brain and consist of neurons (called nodes) responsible for identifying certain patterns, i.e. correlations of attributes (McCulloch & Pitts 1943). A specific form of ANN is a feed-forward, fully connected network, called multilayer perceptron (MLP) (Rosenblatt 1961). An MLP consists of an input layer, one or more hidden layers and an output layer, where each node at one layer passes on information to all nodes at the consecutive layer (feed-forward), thus, each node is connected with all nodes of the

subsequent layer (fully connected). In each node, an activation function takes the sum of all weighted inputs, provided by the nodes of the preceding layer, and calculates a corresponding output, passed on to all nodes of the subsequent layer. Non-linear activation functions (like SIGMOID, tanh, Gaussian, etc.) enable the MLP to learn complex and non-linear patterns. While the nodes of the input layer simply represent the input (i.e. independent) attributes and the node of the output layer the target (i.e. dependent) attribute, the nodes of the hidden layers represent patterns meaningful to most precisely identify the correct target attribute values. Most importantly, these patterns are not predefined, but the system learns them automatically, thus, constituting the high flexibility and predictive power of MLPs. In fact, the process of learning meaningful patterns means adapting the weights of the edges between two nodes in order to reduce the overall prediction error (typically the sum of squared errors (SSE) of all training data examples). We use the gradient descent optimization algorithm in order to iteratively find the optimal weights minimizing the SSE. We also use the back-propagation algorithm to percolate the necessary weight adaptations, identified by the gradient descent method, back to the preceding layers within the network (Rumelhart, Hinton, & Williams 1986). The hyper-parameter *learning rate* and *momentum* allow specifying, how fast the weights are adapted and how fast the direction of weight adaptations can be changed in order to avoid oscillating behavior. The well-known problem of the vanishing gradient constitutes an upper limit for the number of hidden layers of an MLP. Previous research has shown that in most cases already one hidden layer leads to optimal results (Hochreiter 1998).

In this study, the ANN model represents a *multilayer perceptron neural network* (MLP), using *back-propagation* for optimization (Rumelhart, Hinton, & Williams 1986) and a single-hidden layer as a common architecture for ANNs in research practice (Kamel et al. 2008). More

precisely, the number of neurons in the hidden layer is set to the number of input variables divided by 2 plus 1. A *sigmoid activation function* is employed within the hidden layer. The hyperparameter error epsilon was set to 1.0E-5 for all prediction models. The hyperparameters *learning rate*, *momentum* and *training cycles* have been set differently for each prediction model (cf. Table 8).

--- Table 8 to be placed here ---

To summarize, the study at hand evaluates whether the machine learning technique artificial neural networks (ANNs) achieves superior prediction accuracies than the statistical approach of ARIMA especially in case of making use of Google trends data as an additional '*exogenous*' variable (*research proposition 2*).

Finally, we used a six-month forecasting horizon to validate the forecasting performance for a mid-term prediction scenario. The remaining prediction error is usually used to measure forecast accuracy (Frechtling 2002; Song et al. 2010) and typically operationalized by the root mean square error (RMSE).

In addition to the performance measure itself, an appropriate validation method has to be chosen (Frechtling 2002; Kennedy 2010). As a traditionally used method for estimating forecasting performance in time series data, we evaluated the prediction accuracy based on an out of sample validation (Mozetic et al. 2019). The training data set consisted of 83 samples from August 2009 until June 2016 and the test data set of 6 samples, from July 2016 until December 2016 (Chatfield 2000).

As a final step of validating the quality of the learned prediction model, the *Shapiro-Wilk* test is used to test whether the residuals are normally distributed and, thus, do not contain any remaining patterns or information (Hill et al. 2011).

4 Results and discussion

In this study, we evaluated four different approaches for predicting tourist arrivals. An *autoregressive* approach based solely on tourist arrivals itself, and an approach based on web search traffic as additional model input in order to validate research proposition 1. Additionally, for both approaches, we employed an ARIMA model and a neural network multilayer perceptron (MLP) in order to answer research proposition 2. All four prediction models were learned for all five sending countries separately. For evaluating mid-term forecasting capabilities, we evaluated the prediction models with a forecasting horizon of 6 months. Exemplarily, Figure 3 shows monthly tourist arrivals and the corresponding predicted tourist arrivals based on the forecasting approach artificial neural network (ANN) with Google Trends data. The high prediction accuracy of the presented approach becomes evident.

--- Fig. 3 to be placed here ---

Table 9 shows the overall results for all four prediction models and Sweden's major sending countries. For the ARIMA and the artificial neural network approach, we calculated the relative difference between the purely autoregressive models and the approach including Google Trends data (where negative values indicate an error reduction by adding Google Trends data), in order to assess research proposition 1. Furthermore, we used a *Shapiro-Wilk* test for both models using *Google Trends* data to check if residuals (error terms) comply with a normal distribution. The latter is considered as an important indicator that the residuals do not include any kind of information or patterns that have not been recognized by the regression model (Kennedy 2010; Hill et al. 2011). Finally, the two right-most columns calculate the difference between the ARIMA model and the artificial neural network

regression for the autoregressive approach as well as the approach adding Google Trends data (i.e. assessment of research proposition 2).

--- Table 9 to be placed here ---

The *root mean squared error (RMSE)* values for the four different prediction models in Table 9 show satisfactory results. When forecasting autoregressively, the ARIMA models achieved better results for the sending countries Denmark, Finland and Russian Federation, while the autoregressive forecasts for Norway and the United Kingdom using ANNs were more accurate. On average, the prediction errors for the autoregressive ANN models were slightly higher than the RMSE achieved by the autoregressive ARIMA models (+ 2.4 percent). In contrast, nearly all the Neural Network based models led to significantly more accurate forecasting results compared to the ARIMA models, when including Google Trends data into the prediction. The inclusion of Google Trends data to the ANN models led to a RMSE reduction for all sending countries (except United Kingdom) of at least 18.61 percent and a maximum RMSE reduction of 78.91 percent. On average, the inclusion of Google Trends data achieved an RMSE reduction of 47.75 percent. In this case, the average RMSE for the Neural Network models was slightly lower compared to the corresponding ARIMA models. The results of the Shapiro-Wilk test underpin the superior results of the ANN models with Google Trends data, where for all sending countries the hypothesis of normal distribution is not rejected (p -values clearly above 0.05). Regarding the ARIMA models, the results of the Shapiro-Wilk test indicate that at least for the sending country Denmark, the hypothesis of normal distribution of the residuals has to be rejected (p -values below 0.05), and, thus, ARIMA could not fit the data for the Denmark data set optimally. To summarize, we view the model fit and forecasting accuracy of all proposed approaches as sufficient to meaningfully substantiate the research propositions of this study.

Empirical findings in Table 9 impressively demonstrate that extending a purely autoregressive approach by adding Google Trends data reduces the prediction error, both for the ARIMA and the ANN approach. On average, the prediction errors for both models (i.e. ARIMA and Neural Net) declined nearly 50 percent compared to the corresponding autoregressive prediction errors. Thus, adding Google Trends data to an autoregressive approach to predict tourist arrivals clearly reduces the prediction error and, thus, confirms research proposition 1. This (1) demonstrates that tourists make extensive use of internet search engines, like Google, during their information gathering and travel planning phase, and (2) that different search terms can be assigned to different travel planning phases as search terms have a dominant time lag, which is a prerequisite to aggregate them to a compound search index.

When comparing the two different estimation methods, on average the ANN models slightly outperform the ARIMA models when including Google Trends data. Overall, the RMSE obtained by the ANN models with Google Trends data is 4.2 per cent lower compared to the corresponding ARIMA models. Thus, ANN-based estimation models tend to outperform the ARIMA models when adding Google Trends data. Consequently, we can confirm research proposition 2 as well. This demonstrates the ability of the machine learning technique artificial neural networks to learn complex and non-linear patterns effectively, which turns out as a clear advantage relative to linear models, especially when adding multiple and diverse input data, like Google Trends data (Kamel et al. 2008). While ARIMA delivers competitive or even slightly superior results in a standard autoregressive setting, ARIMA cannot benefit from adding Google Trends data to the same extent than ANNs do. The robustness of ANNs against correlated, irrelevant or even biased input data enables ANNs to fully benefit from the predictive power of the relevant search volume on a search engine, like Google.

However, Google Trends data is not the only new potential input to be added to tourism demand prediction. Other big data sources, such as e-reviews (user-generated content), social media interactions, network data, etc., constitute promising data input as well (Mariani et al. 2018). Therefore, we expect machine learning-based techniques, such as artificial neural networks, to gain more attention in tourism demand forecasting and in big data analytics in the future.

5 Conclusions

This study presented a novel approach to extend the autoregressive time-series forecasting method by adding the web-search traffic of tourists as external input for predicting tourist arrivals. More concretely, the study introduced a new method to identify and aggregate relevant Google search terms into a compound web-search index, serving as an additional input variable to an autoregressive forecasting approach. As prediction methods, the study compared the forecasting performance (i.e. accuracy in terms of root mean squared error) gained by the statistical approach of ARIMA with those gained by the machine learning-based technique artificial neural network (ANN). The forecasting study was conducted for Sweden, using arrival data of Sweden's major sending countries (i.e. Denmark, Finland, Norway, Russian Federation, and the United Kingdom) for the period 2008-2016, and corresponding Google Trends data (i.e. text and video search) as a big data source. All statistical computations used R-Statistics, while the algorithm for retrieving search queries used Spyder® ('Scientific Python Development Environment'). Rapid Miner Studio® was used in the study for the machine learning processes for both learning and evaluating the prediction models.

Study results clearly show that extending purely autoregressive forecasting approaches by Google Trends data significantly reduces the prediction error. Google Trends data can be used as an effective data source to increase forecasting accuracy for the mid-term range of future tourism demand and demand fluctuations. Findings of this study constitute a remarkable insight, as the web search volume in the form of Google Trends data is only one out of many available big data sources (e.g. web navigation behaviour, social media interactions, e-reviews, etc.), likely to have a similar potential to increase the accuracy of tourism demand prediction (Mariani et al. 2018). Therefore, we envisage that adding

further big data sources can increase the precision of tourist arrival predictions in the future (Kamel et al. 2008).

Second, the study could demonstrate that when comparing statistical approaches, such as ARIMA models, with machine learning-based approaches, like artificial neural networks (ANNs), ANNs tend to outperform the ARIMA models when using the big data-based approach. Interestingly enough, the big data approach benefits relatively more from applying ANNs than the autoregressive approach, which can be explained by the high flexibility and robustness of ANNs, especially suitable for a multi-dimensional input space (Haykin 2008). Thus, for future scenarios of considering the inclusion of additional big data sources as input to tourism demand prediction, ANNs, or similar machine learning techniques, like deep learning, will most likely play a dominant role for reliably predicting tourist arrivals.

To summarize, as the main scientific contribution, the study proved that (1) big data sources, like Google Trends data, show strong potentials to increase the prediction performance of tourist arrivals compared to autoregressive approaches. Furthermore, (2) machine learning techniques, like ANN, have the potential to outperform statistical approaches, such as an ARIMA when adding search query indices obtained by Google Trends.

Additionally, the study results are of high practical relevance. Due to the perishable nature of tourism services, precise and reliable demand predictions are of utmost importance for tourism stakeholders and decision makers (Grönroos 2008; Edgell, Del Mastro Allen, Smith, & Swanson 2008). The extension of traditional (e.g. autoregressive) approaches by integrating big data sources does not only increase the forecasting precision, but specifically enables the prediction of demand fluctuations in extraordinary, or even singular circumstances (e.g. shocks, such as financial and economic crises, natural disasters, mega

events, epidemics, etc.), especially when autoregressive approaches fail systematically (Chen et al. 2007; Song & Lin 2010; Song, Qiu & Park 2019).

When looking at major study limitations, first, the used cosine similarity definitely has limited capabilities for matching semantically identical search terms when constructing the aggregated search index. Thus, in order to automatically and more powerfully detect semantically identical search terms, text mining approaches are recommended for future studies (see: Schmunk, Höpken, Fuchs, & Lexhagen 2014; Menner, Höpken, Fuchs, & Lexhagen 2016; Höpken, Fuchs, Menner & Lexhagen 2017a). Second, while the current study makes use of web search traffic in terms of Google Trends data, certainly, other big data sources, e.g. web navigation behaviour, social media interactions, e-reviews, etc., show a similarly strong potential to serve as input for tourism demand modeling and prediction (Kamel et al. 2008; Mariani et al. 2008). Therefore, future studies in the domain of tourism demand forecasting should extend the presented approach to other big data sources. Third, the current study is potentially limited by using a feed-forward multi-layer perceptron (MLP), learned by back propagation with gradient descent as machine learning approach. MLPs are typically limited to a relatively small number of hidden layers (i.e. only one in the case of this study) due to the vanishing gradient problem typically occurring within back-propagation over multiple hidden layers (Hochreiter, Bengio, Frasconi & Schmidhuber 2001). Deep learning approaches might best overcome this limitation, as the use of a high number of hidden layers can cope with large model complexity and flexibility and, ultimately, increase explanation power (Schmidhuber 2015). Therefore, deep learning might constitute a promising future forecasting approach, especially when adding multiple big data sources as additional input to the prediction of tourism demand. Finally, the identified search queries and their corresponding time lags with a high correlation with future tourist arrivals are an

excellent input to analyse tourists' online search behavior (Höpken et al. 2019). Thus, identifying concrete search terms, actually preceding tourism demand would constitute valuable input to destination marketing generally and search engine marketing and optimization specifically.

References

Ajzen, I. 2005. Attitudes, personality and behaviour, 2nd ed., Maidenhead, Open University Press.

Antolini, F. and L. Grassini. 2019. "Foreign arrivals now casting in Italy with Google Trends data." *Quality & Quantity* 53:2385–2401.

Artola, C., F. Pinto, and P. de Pedraza García. 2015. "Can internet searches forecast tourism inflows?" *International Journal of Manpower* 36 (1):103-116.

Aslanargun, A., M. Mammadov, B. Yazici, & S. Yolacan. 2007. "Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting" *Journal of Statistical Computation and Simulation* 77 (1): 29-53.

Assaf, A.G., G. Li, H. Song, and M. Tsionas. 2019. "Modeling and forecasting regional tourism demand using the Bayesian Global Vector Autoregressive (BGVAR) Model". *Journal of Travel Research*, 58 (3):383–397.

Athanasopoulos, G., H. Song, and J. A. Sun. 2018. "Bagging in tourism demand modelling and forecasting." *Journal of Travel Research* 57 (1):52-68.

Baddeley, M. C., and D. Barrowclough. 2009. *Running regressions – A practical guide to quantitative research in economics, finance and development studies*. Cambridge: University Press.

Bangwayo-Skeete, P. F., and R. W. Skeete. 2015. "Can Google data improve the forecasting performance of tourist arrivals? A mixed-data sampling approach." *Tourism Management* 46:454-464.

- Blazquez, D., and J. Domenech. 2018. "Big data sources and methods for social and economic analyses." *Technological Forecasting & Social Change* 130:99-113.
- Bokelmann, B. and S. Lessmann. 2019. "Spurious patterns in Google Trends data - An analysis of the effects on tourism demand forecasting in Germany." *Tourism Management*, 75(Dec):1-12.
- Box, G. E., and G. M. Jenkins. 1970. *Time series analysis, forecasting and control*. San Francisco: Holden Day.
- Burger, C. J.S.C., M. Dohnal, M. Kathrada, & R. Law. 2001. "A practitioners guide to time-series methods for tourism demand forecasting - A case study of Durban, South Africa." *Tourism Management* 22:403-409.
- Camacho, M., and M. J. Pacce. 2018. "Forecasting travellers in Spain with Google's search volume indices." *Tourism Economics* 24 (4):434-448.
- Carrière-Swallow, Y., and F. Labbé. 2013. "Nowcasting with Google Trends in an emerging market." *Journal of Forecasting* 32 (4):289-298.
- Chatfield, C. 2000. "Time-Series Forecasting." CRC Press, 238-239.
- Chen, Y.-C., H.-H. Kang, and T.-Ch. Yang. 2007. „A study on the impact of SARS on the forecast of visitor arrivals to China." *Journal of Asia Pacific Business* 8 (1):31-50.
- Chen, K. Y., and C. H. Wang. 2007. "Support vector regression with genetic algorithms in forecasting tourism demand." *Tourism Management* 28:215-226.
- Chen, J. L., G. Li, D. C. Wu, and S. Shen. 2019. „Forecasting seasonal tourism demand using a multi-series structural time series method". *Journal of Travel Research*, 58 (1): 92-103.

Cho, V. 2001. "Tourism forecasting and its relationship with leading economic indicators." *Journal of Hospitality and Tourism Research* 25:399-420.

Cho, V. 2003. "A comparison of three different approaches of tourist arrivals." *Tourism Management* 24:323-330.

Choe, Y., C. Vogt, and D. R. Fesenmaier. 2017. "Twenty-five years since Vogt: Assessing the changing information needs of American travelers", in *Information and communications technologies in Tourism 2017*, edited by Schegg, R. & Stangl, B., New York, Springer.

Chung, N., H. Lee, S. J., Lee, S. J., and C. Koo. 2015. „The influence of tourism website on tourist's behavior to determine destination selection". *Technology Forecasting and Social Change*, 96: 130–143.

Chun-Fu Chen, C-F., M-C. Lai, & C-C. Yeh. 2012. "Forecasting tourism demand based on empirical mode decomposition and neural network." *Knowledge-Based Systems* 26:281–287.

Claveria, O., and S. Torra. 2014. "Forecasting tourism demand to Catalonia - Neural networks vs. time series models." *Economic Modelling* 36:220-228.

Claveria, O., E. Monte, and S. Torra. 2015. "Tourism demand forecasting with neural networks models: Different ways of treating information." *International Journal of Tourism Research* 17:492-500.

Çuhadar, M., I. Cogurcu, and C. Kukrer. 2014. "Modelling and forecasting cruise tourism demand to Zmir by different artificial neural network architectures." *International Journal of Business and Social Research* 4:12-28.

Dergiades, Th., E. Mavragani, and B. Pan. 2018. "Google Trends and tourists' arrivals: Emerging biases and proposed corrections." *Tourism Management* 66:108-120.

Divisekera, S., and N. Kulendran. 2006. "Economic effects of advertising on tourism demand." *Tourism Economics* 12:187–205.

Du, K-L and M.N. Swamy. 2019. "Neural networks and statistical learning." 2nd edition, London: Springer.

Edgell, D. L. Sr., M. Del Mastro Allen, G. Smith, and J. R. Swanson. 2008. *Tourism policy and planning – Yesterday, today and tomorrow*. New York: Routledge.

Erawan, T., D. Krairit, and D. B. Khang. 2011. "Tourists' external information search behavior model: The case of Thailand". *Journal of Modelling in Management*, 6 (3): 297–316.

Fesenmaier, D. R., Z. Xiang, B. Pan, and R. Law. 2010. "An analysis of search engine use for travel planning." In *Information and Communication Technologies in Tourism 2010*, edited by U. Gretzel, R. Law, and M. Fuchs, 381-392. New York: Springer.

Fodness, D., and B. Murray. 1999. "A model of tourist information search behavior". *Journal of Travel Research*, 37: 220-30.

Frechtling, D. C. 2002. *Forecasting tourism demand*. Oxford: Butherworth-Heinemann.

Frechtling, D. C. 2011. *Forecasting tourism demand: methods and strategies*. Abingdon, UK: Routledge.

Friedman, M. 1937. „The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance.“ *Journal of the American Statistical Association* 32(200): 675–701.

Fuchs, M., L. Rijken, M. Peters, and K. Weiermair. 2000. "Modelling Asian incoming tourism - A shift-share approach." *Asia Pacific Journal of Tourism Research* 5 (2):1-10.

Fuchs, M., W. Höpken, and M. Lexhagen. 2014. „Big data analytics for knowledge generation in tourism destinations – A case from Sweden”. *Journal of Destination Marketing and Management*, 3 (4):198-209.

Fuchs, M. and R. Baggio. 2017. “Creativity and tourism networks: A contribution to a post-mechanistic economic theory”. *Critical Tourism Studies Conference*, Palma de Mallorca, Spain, 25–9 June.

Geurts, M. D., and I. B. Ibrahim. 1975. “Comparing the Box Jenkins approach with the exponentially smoothed forecasting model application to Hawaii tourists.” *Journal of Marketing Research* 7:182-188.

Goh, C., R. Law, and H. M. K. Mok. 2008. “Analyzing and forecasting tourism demand: a rough sets approach.” *Journal of Travel Research* 46 (3):327-338.

Granger, C. W. 1969. “Investigating causal relations by econometric models and cross-spectral methods.” *Econometrica* 37 (3):424-438.

Granger, C. W. 1988. “Some recent developments in a concept of causality.” *Journal of Econometrics* 39 (1-2):199-211.

Granger, C. W., and P. Newbold. 1974. “Spurious regressions in econometrics.” *Journal of Econometrics* 2:111-120.

Grönroos, C. 2008. “Service logic revisited - Who creates value? And who co-creates?” *European Business Review* 20 (4):298–314.

Gunter, U., and I. Önder. 2016. “Forecasting city arrivals with Google Analytics.” *Annals of Tourism Research* 61:199-212.

Gunter, U., I. Önder, and S. Gindl. 2019. „Exploring the predictive ability of LIKES of posts on the Facebook pages of four major city DMOs in Austria.” *Tourism Economics* 25 (3):375–401.

Hadavandi, E., A. Ghanbari, K. Shahanaghi, and S. Abbasian-Naghneh. 2011. “Tourist arrival forecasting by evolutionary fuzzy systems.” *Tourism Management* 32 (5):1196-1203.

Hand, C., and J. Guy. 2012. “Searching for the picture: forecasting UK cinema admissions using Google Trends data.” *Applied Economics Letters* 19 (11):1051-1055.

Hernández-López, M., and J. J. Cáceres-Hernández. 2007. “Forecasting tourists’ characteristics by a genetic algorithm with a transition matrix.” *Tourism Management* 28:290-297.

Hill, R. C., W. E. Griffith, and G. C. Lim. 2011. *Principles of econometrics*. 4th ed. New York: Wiley and sons.

Hochreiter, S. 1998. “The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions.” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6(2):107-116.

Hochreiter, S., Y. Bengio, P. Frasconi, and J. Schmidhuber. 2001. “Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.” In *A Field Guide to Dynamical Recurrent Neural Networks*, edited by S. C. Kremer and J. F. Kolen. IEEE Press.

Höpken, W., M. Fuchs, D. Keil, D. and M. Lexhagen. 2015. Business intelligence for cross-process knowledge extraction at tourism destination, *Journal of Information Technology and Tourism*, 15 (2): 101-130.

Höpken, W., M. Fuchs, Th. Menner, and M. Lexhagen. 2017a. "Sensing the online social sphere – the sentiment analytical approach." In *Analytics in Smart Tourism Design – Concepts and Methods*, edited by Z. Xiang, A. Alzua, and D. Fesenmaier, 129-146. Springer.

Höpken, W., D. Ernesti, M. Fuchs, K. Kronenberg, and M. Lexhagen. 2017b. „Big data as input for predicting tourist arrivals." In *Information and Communication Technologies in Tourism 2017*, edited by R. Schegg and B. Stangl, 187-200. New York: Springer.

Höpken, W., T. Eberle, M. Fuchs, and M. Lexhagen. 2018. „Search engine traffic as input for predicting tourist arrivals." In *Information and Communication Technologies in Tourism 2018*, edited by B. Stangl and J. Pesonen, 381-393. New York: Springer.

Höpken, W., T. Eberle, M. Fuchs, and M. Lexhagen. 2019. „Google trends data for analysing tourists' online search behavior and improving demand forecasting: The case of Åre, Sweden. *Information Technology & Tourism*, 21 (1): 45-62.

Hong, W. C. 2006. "The application of support vector machines to forecast tourist arrivals in Barbados - An empirical study." *International Journal of Management* 23:375-385.

Hong, W., Y. Dong, L. Chen, and S. Wei. 2011. "SVR with hybrid chaotic genetic algorithms for tourism demand forecasting." *Applied Soft Computing* 11 (2):1881-1890.

Hyndman, R. J., and Y. Khandakar (2008). "Automatic Time Series Forecasting: The forecast Package for R". *Journal of Statistical Software* 27 (3), 1-22.

Hyndman, R. J., and G. Athanasopoulos. 2018. "Forecasting: principles and practice." 276-278.

Hurst, H. E., R. P. Black, and Y. M. Simaika. 1965. *Long-term storage: an experimental study*. London: Constable.

Jackman, M., and S. Naitram. 2015. "Nowcasting tourist arrivals to Barbados. Just Google it!" *Tourism Economics* 21 (6):1309-1313.

Kamel, N., A. F. Atiya, N. El Gayar, and H. El-Shishiny. 2008. "Tourism demand forecasting using machine learning methods." *International Journal on Artificial Intelligence and Machine Learning* 8:1-7.

Kennedy, P. 2010. *A guide to econometrics*. 6th ed. Oxford: Wiley-Blackwell.

Khaidi, S. M., N. Abu, and N. Muhammad. 2019. "Tourism demand forecasting - a review on the variables and models". *Journal of Physics, Conference Series*, 1366 (2019) 012111) doi:10.1088/1742-6596/1366/1/012111.

Kim, S., and A. Kim. 2016. "A new metric of absolute percentage error for intermittent demand forecasts." *International Journal of Forecasting* 32 (3):669-679.

Kim, E., J.-H. Lee, Y. Sung, and S. M. Choi. 2016. "Predicting selfie-posting behavior on social networking sites: An extension of theory of planned behavior". *Computers in Human Behavior*, 62: 116–123.

Kim, W-H., and K. Malek. 2018. "Forecasting casino revenue by incorporating Google trends." *International Journal of Tourism Research* 20:424–432.

Kon, S. C., and W. L. Turner. 2005. "Neural Network forecasting of tourism demand." *Tourism Economics* 11:308-328.

Kronenberg, K., M. Fuchs, K. Salman, M. Lexhagen, and W. Höpken. 2016. "Economic effects of advertising expenditures – A Swedish destination study of international tourists." *Scandinavian Journal of Hospitality & Tourism Research* 16 (4):352-374.

Kruskal, W. H., and W. A. Wallis (1952). Use of Ranks in One-Criterion Variance Analysis.

Journal of the American Statistical Association 47 (260): 583-621.

Kulendran, N., and K. Wong. 2005. "Modelling seasonality in tourism forecasting." Journal of

Travel Research 44:163-170.

Law, R., and N. Au. 1998. "A rough set approach to hotel expenditure decision rules

induction." Journal of Hospitality & Tourism Research 22 (4):359-375.

Law, R., G. Li, and L. Feng. 2019. Tourism demand forecasting: a deep learning approach.

Annals of Tourism Research 75:410-423.

Li, S., T. Chen, L. Wang, and C. Ming. 2018. "Effective tourist volume forecasting supported

by PCA and improved BPNN using Baidu index." Tourism Management 68:116-126

Li, J., L. Xu, L. Tang, S. Wang, and L. Li. 2018. "Big data in tourism research: a literature

review." Tourism Management 68:301-323

Li, X., Q. Wu, G. Peng, and B. Lv. 2016. « Tourism forecasting by search engine data with

noise processing." African Journal of Business Management 10 (6):114-130.

Li, X., B. Pan, R. Law, and X. Huang. 2017. "Forecasting tourism demand with composite

search index." Tourism Management 59:57-66.

Li, X., and R. Law, R. 2020. "Forecasting tourism demand with decomposed search cycles."

Journal of Travel Research 59 (1):52-68.

Lim, Ch., and M. McAleer. 2001. "Forecasting tourist arrivals." Annals of Tourism Research

28 (4):965-977.

Lin, Ch.-J., H.-F. Chen, and T.-S. Lee. 2011. „Forecasting tourism demand using time series, artificial neural networks and multivariate adaptive regression splines.” *International Journal of Business Administration* 2 (2):14-24.

Liu, B. 2008. *Web data mining – Exploring hyperlinks, contents, and usage data*. Heidelberg: Springer.

Liu, Y., B. Lv, G. Peng, and Q. Yuan. 2012. “A pre-processing method of Internet search data for prediction improvement.” In *Proceedings of the Data Mining and Intelligent Knowledge Management Workshop*, 3:1-3:7. New York: ACM.

Liu, Y. Y., F. M. Tseng, and Y. H. Tseng. 2018. „Big data analytics for forecasting tourism destination arrivals with the applied Vector Autoregression model.” *Technological Forecasting & Social Change* 130:123-134.

Maravall, A. 2011. “Seasonality Tests and Automatic Model Identification in TRAMO-SEATS.” Bank of Spain.

Mariani, M., R. Baggio, M. Fuchs, and W. Höpken. 2018. “Business intelligence and big data in hospitality and tourism: A systematic literature review.” *International Journal of Contemporary Hospitality Management* 30 (10):3514-3554.

McCulloch, W., and W. Pitts. 1943. "A Logical Calculus of Ideas Immanent in Nervous Activity." *Bulletin of Mathematical Biophysics* 5(4):115-133.

Menner, Th., W. Höpken, M. Fuchs, and M. Lexhagen. 2016. “Topic detection – Identifying relevant topics in tourism reviews.” In *Information and Communication Technologies in Tourism 2016*, edited by A. Inversini and R. Schegg, 411-423. New York: Springer.

Mozetic, I., V. Cerqueira, and L. Torgo. 2019. „Evaluating time series forecasting models: An empirical study on performance estimation methods.“ arXiv:1905.11744:2.

Moro, S., and P. Rita. 2016. “Forecasting tomorrow’s tourist.” *Worldwide Hospitality and Tourism Themes* 8 (6):643-653.

Mukherjee, C., H. White, and M. Wuyts. 1998. *Econometrics and data analysis for developing countries*. New York: Routledge.

Osborn, D., Chui, A., Smith, J., Nwogu, E.C., I. S. Iwueze, and Birchenhall, C. 1988. "Seasonality and the order of integration". *Journal of Applied Econometrics* 3(2):101-113.
U. Nlebedim. 2016. “Some tests for consumption.”, *Oxford Bulletin of Economics and Statistics* 50(4):361-377.

Önder, I., and U. Gunter. 2016. “Forecasting tourism demand with Google Trends for a major European city destination.” *Tourism Analysis* 21:203-220.

Önder, I. 2017. “Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities.” *International Journal of Tourism Research* 19:648-660.

Padhi, S. & R. K. Pati. 2017. “Quantifying potential tourist behavior in choice of destination using Google Trends”. *Tourism Management Perspectives*, 24:34–47.

Pai, P.-F., W. C. Hong, P. T. Chang, and C. T. Chen. 2006. “The application of support vector machines to forecast tourism arrivals in Barbados: An empirical study.” *International Journal of Management* 23:375-385.

Pai, P.-F., K.-Ch. Hung, and K.-P. Lin. 2015. “Tourism demand forecasting using novel hybrid system.” *Expert Systems with Applications* 41 (8):3691-3702.

Palmer, A., J. J. Jose Montano, and A. Sese. 2006. “Designing an artificial neural network for forecasting tourism time series.” *Tourism Management* 27:781-790.

- Pan, B., and D. R. Fesenmaier. 2006. Online information search: Vacation planning process. *Annals of Tourism Research*, 33 (3):809-832.
- Pan, B., X. Li, R. Law, and X. Huang. 2017. "Forecasting tourism demand with composite search index." *Tourism Management* 59 (1):57–66.
- Pan, B., C. Wu, and H. Song. 2012. "Forecasting hotel room demand using search engine data." *Journal of Hospitality and Tourism Technology* 3 (3):196-210.
- Pan, B., and Y. Yang. 2017. "Forecasting destination weekly hotel occupancy with big data." *Journal of Travel Research* 56 (6):957-970.
- Park, S., J. Lee, and W. Song. 2017. "Short-term forecasting of Japanese tourist inflow to South Korea using Google trends data." *Journal of Travel & Tourism Marketing* 34 (3):357–368.
- Pearson CMG. 2017. Internet and search engine use by country: Global search engine marketing.
http://ptgmedia.pearsoncmg.com/images/9780789747884/supplements/9780789747884_a_ppC.pdf
- Peng, B., H. Song, and G. Crouch. 2014. "A meta-analysis of international tourism demand forecasting and implications for practice." *Tourism Management* 45:181-193.
- Peng, G., Y. Liu, J. Wang, and J. Gu. 2017. "Analysis of the prediction capability of web search data based on the HE-TDC method - Prediction of the volume of daily tourism visitors." *Journal of Systems Science and Systems Engineering* 26 (2):163–182.
- Pitman A., M. Zanker, M. Fuchs M., and M. Lexhagen. 2010. Web usage mining in tourism: A query term analysis and clustering approach. In *Information and Communication*

Technologies in Tourism 2010, edited by Gretzel U., R. Law, and M. Fuchs, Springer, New York, 393-403.

Return on now. 2017. "Search engine market share by country."

<http://returnonnow.com/internet-marketing-resources/2015-search-engine-market-share-by-country/>.

Ricardo, H., I. Gonçalves, and A. C. Costa. 2018. "Forecasting tourism demand for Lisbon's region through a data mining approach." In Proceedings of the 11th IADIS International conference on Information Systems 2018, 58-66. IS2018.

Rivera, R. 2016. "A dynamic linear model to forecast hotel registrations in Puerto Rico using Google trends data." *Tourism Management* 57:12-20.

Roecker, E. B. 1991. "Prediction error and its estimation for subset—selected models." *Technometrics* 33:459–468.

Rosenblatt, F. 1961. "Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms." Washington DC: Spartan Books.

Rumelhart, E. D., G. E. Hinton, and R. J. Williams. 1986. "Learning representations by back-propagating errors." *Nature* 323 (6088):533-536.

Schmidhuber, J. 2015. "Deep learning in neural networks: An overview." *Neural Networks* 61:85–117.

Schmunk, S., W. Höpken, M. Fuchs, and M. Lexhagen. 2014. "Sentiment analysis – Implementation and evaluation of methods for sentiment analysis with Rapid-Miner®." In *Information and Communication Technologies in Tourism 2014*, edited by Ph. Xiang and I. Tussyadiah, 253-265. New York: Springer.

Silverstovs, B. and D. S. Wochner. 2018. Google trends and reality: Do the proportions match? Appraising the informational value of online search behavior - Evidence from Swiss tourism regions." *Journal of Economic Behavior & Organization* 145:1–23.

Silva, S.E., H. Hassani, S. Heravi, and X. Huang. 2019. "Forecasting tourism demand with denoised neural networks". *Annals of Tourism Research* 74(Jan):134-154.

Smeral, E. 2009. "The impact of the financial and economic crisis on European tourism." *Journal of Travel Research* 48 (1):3-13.

Smeral, E. 2017. "Tourism forecasting performance considering the instability of demand elasticities." *Journal of Travel Research* 56 (7):913-926.

Song, H., and G. Li. 2008. "Tourism demand modelling and forecasting: A review of recent research." *Tourism Management* 29:203-220.

Song, H., G. Li, St. F. Witt, and B. Fei. 2010. "Tourism demand modelling and forecasting - How should demand be measured?" *Tourism Economics* 16 (1):63-81.

Song, H., and S. Lin. 2010. "Impact of the financial and economic crisis on tourism in Asia." *Journal of Travel Research* 49 (1):16-30.

Song, H., and L. Turner. 2006. "Tourism demand forecasting." In *International handbook on the economics of tourism*, edited by L. Dwyer and P. Forsyth. Cheltenham: Edward Elgar.

Song, H., R. T R. Qiu, and J. Park. 2019. "A review of research on tourism demand forecasting methods". *Annals of Tourism Research*, 75:338–362.

Stabler, M. J., A. Papatheodorou, and M. T. Sinclair. 2010. *The economics of tourism*. 2nd ed. Routledge, London.

Statistics Sweden. 2017. "Statistiska centralbyråns (SCB)." <http://www.scb.se/en/> (assessed February 2, 2017).

Steen Jacobsen, J. K. and A. M. Munar. 2012. "Tourist information search and destination choice in a digital age". *Tourism Management Perspectives*, 1:39–47.

Sun, S., W. Wang, Y. Wei, and X. Yang. 2017. „Forecasting tourist arrivals with machine learning and internet search index." In *2017 IEEE International conference on big data*, 4165-4169.

Sun, S., Y. Wei, K.-L. Tsui, and S. Wang. 2019. "Forecasting tourist arrivals with machine learning and internet search index." *Tourism Management* 70:1-10.

Tsaur, R., and T. Kuo. 2011. "The adaptive fuzzy time series model with an application to Taiwan's tourism demand." *Expert Systems with Applications* 38:9164-9171.

Turner, L. W., and S. F. Witt. 2001. "Factors influencing demand for international tourism: tourism demand analysis using structural equation modelling." *Tourism Economics* 7:21-38.

Vanhove, N. 2011. *The economics of tourism destinations*. London: Elsevier.

Varian, H. 2014. "Big data - New tricks for econometrics." *Journal of Econometric Perspectives* 28 (2):3-28.

Vogt, C., and D. R. Fesenmaier. 1998 "Expanding the functional information search model". *Annals of Tourism Research*, 25 (3):551-578.

Vosen, S., and T. Schmidt. 2011. "Forecasting private consumption: survey-based indicators vs. Google trends." *Journal of Forecasting* 30 (6):565-578.

Wang, X., K. Smith, R. Hyndman. 2006. "Characteristic-based clustering for time series data", *Data Mining and Knowledge Discovery* 13 (3):335-364.

- Witt, S. F., and C. A. Witt. 1995. "Forecasting tourism demand: a review of empirical research." *International Journal of Forecasting* 11 (3):447-475.
- Wu, L., and E. Brynjolfsson. 2015. "The future of prediction: How Google searches foreshadow housing prices and sales." In *Economics of Digitalisation*, 89-118. Chicago: University of Chicago Press.
- Wu, D. C., H. Song, and S. Shen. 2017. "New developments in tourism and hotel demand modelling and forecasting." *International Journal of Contemporary Hospitality Management* 29 (1):507-529.
- Xiaoxuan, L., W. Qi, P. Geng, and L. Benfu. 2016. "Tourism forecasting by search engine data with noise processing." *African Journal of Business Manager* 10 (6):114-130.
- Xiang, Z., D. Wang, J.T., O'Leary, and D.R. Fesenmaier. 2015. "Adapting to the internet: Trends in Travelers' Use of the Web for Trip Planning". *Journal of Travel Research*, 54 (4): 511 – 527.
- Xiang, Z., V. P. Magnini, and D. R. Fesenmaier. 2015. "Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet". *Journal of Retailing and Consumer Studies*, 22: 244- 249.
- Yang, Y., B. Pan, and H. Song. 2014. "Predicting hotel demand using destination marketing organizations' web traffic data." *Journal of Travel Research* 53 (4):433-447.
- Yang, X., B. Pan, J. A. Evans, and B. Lv. 2015. "Forecasting Chinese tourist volumes with search engine data." *Tourism Management* 46 (3):386-397.
- Yu, G. & Z. Schwartz. 2006. „Forecasting short time-series tourism demand with artificial intelligence models". *Journal of Travel Research* 45(Nov):194-203.

Webel, K., and D. Ollech (2018). "An overall seasonality test based on recursive feature elimination in conditional random forests". Proceedings of the 5th International Conference on Time Series and Forecasting, 20-31.

Zhang, B., X. Huang, N. Li, and R. Law. 2017. "A novel hybrid model for tourist volume forecasting incorporating search engine data." Asia Pacific Journal of Tourism Research 22 (3):245-254.

Table 1: Number of search query series per sending country

	<i>Denmark</i>	<i>Finland</i>	<i>Norway</i>	<i>Russian Federation</i>	<i>United Kingdom</i>
<i>Google Search</i>	677	345	729	170	464
<i>Google Video</i>	183	10	106	3	47
Sum	860	355	835	173	511

Table 2: Seasonality tests on the arrival time series

	Test	Test Statistic	p-value
Denmark	QS-Test	123.74	0
	QS-Test on residuals	125.87	0
	KW-Test on residuals	79.71	1.682654e-12
Finland	QS-Test	148.61	0
	QS-Test on residuals	146.24	0
	KW-Test on residuals	87.26	5.717649e-14
Norway	QS-Test	153.91	0
	QS-Test on residuals	162.35	0
	KW-Test on residuals	85.77	1.116884e-13
Russian Federation	QS-Test	157.61	0
	QS-Test on residuals	104.14	0
	KW-Test on residuals	54.38	1.006473e-07
United Kingdom	QS-Test	105.33	0
	QS-Test on residuals	92.13	0
	KW-Test on residuals	76.2	7.988055e-12

Table 3: Stationarity tests for the original arrival data sets

Data set	KPSS-Test				ADF-Test		OCSB-Test		Seasonal strength	
	Level		Trend		Covariance		Saisonal unit roots		-	
	Test statistic	p-value	Test statistic	p-value	Test statistic	p-value	Test statistic	Critical value	Test statistic	Limit
Arrivals <i>(DK)</i>	0.585	0.024	0.052	0.100	5.953	0.010	-2.153	-1.803	0.951	0.64
Arrivals <i>(FI)</i>	0.087	0.100	0.015	0.100	6.092	0.010	0.309	-1.803	0.988	0.64
Arrivals <i>(NO)</i>	0.041	0.100	0.016	0.100	6.395	0.010	1.820	-1.803	0.993	0.64
Arrivals <i>(RU)</i>	0.899	0.010	0.273	0.010	4.859	0.010	-2.257	-1.803	0.960	0.64
Arrivals <i>(UK)</i>	0.576	0.025	0.168	0.032	4.804	0.010	-0.093	-1.803	0.943	0.64

Table 4: Stationarity tests for the transformed arrival data sets

Data set	KPSS-Test				ADF-Test		OCSB-Test		Seasonal strength	
	Level		Trend		Covariance		Saisonal unit roots		-	
	Test statistic	p-value	Test statistic	p-value	Test statistic	p-value	Test statistic	Critical value	Test statistic	Limit
Arrivals''_(DK)	0.025	0.100	0.020	0.100	7.032	0.010	-7.317	-1.803	0.025	0.64
Arrivals'_(FI)	0.022	0.100	0.022	0.100	7.120	0.010	-11.399	-1.803	0.067	0.64
Arrivals'_(NO)	0.021	0.100	0.020	0.100	6.647	0.010	-5.581	-1.803	0.152	0.64
Arrivals''_(RU)	0.110	0.100	0.057	0.100	5.565	0.010	-4.658	-1.803	0.115	0.64
Arrivals''_(UK)	0.028	0.100	0.024	0.100	7.042	0.010	-4.706	-1.803	0.106	0.64

Table 5: Stationarity tests for the search index data sets

Data set	KPSS-Test				ADF-Test		OCSB-Test		Seasonal strength	
	Level		Trend		Covariance		Saisonal unit roots		-	
	Test statistic	p-value	Test statistic	p-value	Test statistic	p-value	Test statistic	Critical value	Test statistic	Limit
Index (DK)	0.089	0.10	0.097	0.10	5.580	0.01	-6.599	-1.803	0.072	0.64
Index (FI)	0.062	0.10	0.027	0.10	6.644	0.01	-10.798	-1.803	0.068	0.64
Index (NO)	0.085	0.10	0.044	0.10	6.695	0.01	-7.577	-1.803	0.068	0.64
Index (RU)	0.021	0.10	0.021	0.10	8.161	0.01	-6.021	-1.803	0.052	0.64
Index (UK)	0.074	0.10	0.034	0.10	6.400	0.01	-4.903	-1.803	0.073	0.64

Table 6: Correlation between arrival series and corresponding indices

<i>Dataset</i>	<i>Pearson R</i>
Denmark	0.848
Finland	0.823
Norway	0.801
Russian Federation	0.615
United Kingdom	0.809

Table 7: ARIMA models

<i>Dataset</i>		<i>Model</i>	<i>Coefficients</i>	<i>AICc</i>
Denmark	Auto-regressive	ARIMA(1,0,2) with zero mean	AR1: -0.6924 MA1: 0.1541 MA2: -0.6558	142.25
	With Google Trends	Regression with ARIMA(0,0,2) errors	MA1: -0.2636 MA2: -0.3389 Intercept: -86471.02 Xreg: 1473.874	124.77
Finland	Auto-regressive	ARIMA(0,0,1) with zero mean	MA1: -0.6885	119.42
	With Google Trends	Regression with ARIMA(0,0,0) errors	Intercept: -10506.55 Xreg: 260.3937	107.29
Norway	Auto-regressive	ARIMA(0,0,1) with zero mean	MA1: -0.9574	141.96
	With Google Trends	Regression with ARIMA(0,0,0) errors	Intercept: 161450.3 Xreg: -2790.873	135.96
Russian Federation	Auto-regressive	ARIMA(0,0,1) with zero mean	MA1: -0.5753	115.5
	With Google Trends	Regression with ARIMA(0,0,0) errors	Intercept: -7417.555 Xreg: 145.1758	111.2
United Kingdom	Auto-regressive	ARIMA(0,0,3) with non-zero mean	MA1: -0.5941 MA2: -0.0664 MA3: -0.1984 Mean: 171.59	123.05
	With Google Trends	Regression with ARIMA(0,0,3) errors	MA1: -0.1038 MA2: -0.1714 MA3: -0.4629 Intercept: -12688.35 Xreg: 219.4587	114.94

Table 8: Neural network models

<i>Dataset</i>		<i>Training cycles</i>	<i>Learning rate</i>	<i>Momentum</i>
Denmark	Autoregressive	848	0.848	0.195
	With Google Trends	2348	0.154	0.720
Finland	Autoregressive	325	0.1	0.768
	With Google Trends	1950	0.417	0.562
Norway	Autoregressive	480	0.892	0.287
	With Google Trends	674	0.201	0.009
Russian Federation	Autoregressive	1796	0.717	0.076
	With Google Trends	470	0.292	0.464
United Kingdom	Autoregressive	750	0.501	0.092
	With Google Trends	2345	0.660	0.173

Table 9: Results for all four prediction models and sending countries

Sending Country	Autoregressive Moving Average (ARIMA)				Artificial Neural Network (ANN)				Difference ANN-ARIMA	
	Auto-regressive only (RMSE)	With Google Trends (RMSE)	Relative Diff. (%)	Shapiro-Wilk	Auto-regressive only (RMSE)	With Google Trends (RMSE)	Relative Diff. (%)	Shapiro - Wilk	Auto-regressive only	With Google Trends
Denmark	24705.940	6489.383	-73.734	0.043	30071.500	6342.256	-78.909	0.700	+ 21.72 %	- 2.27 %
Finland	4076.076	1564.450	-61.619	0.872	4929.336	1441.386	-70.759	0.939	+ 20.93 %	- 7.87 %
Norway	22458.494	15690.907	-30.134	0.708	14782.616	12031.290	-18.612	0.902	- 34.18 %	- 23.32 %
Russia	2759.682	1994.371	-27.732	0.633	2297.836	1769.573	-22.990	0.974	- 16.74 %	- 11.27 %
UK	4963.617	2534.586	-48.937	0.919	5968.317	3135.668	-47.461	0.939	+ 20.24 %	+ 23.72 %
Average	11792.76	5654.74	-48,431	-	11609.921	4944.035	-47.746	-	+ 2.40 %	- 4.20 %

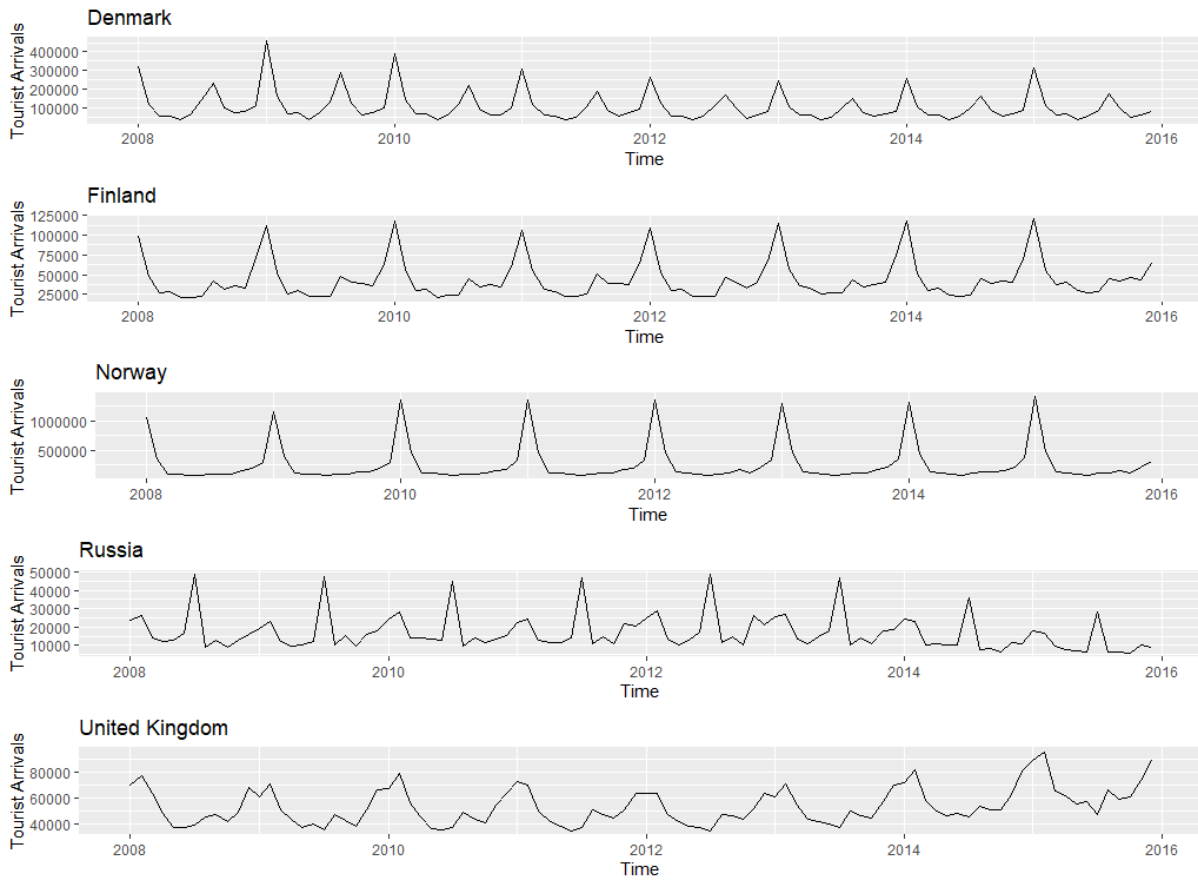


Fig. 1: Monthly tourist arrivals to Sweden per sending country between 2008 and 2016

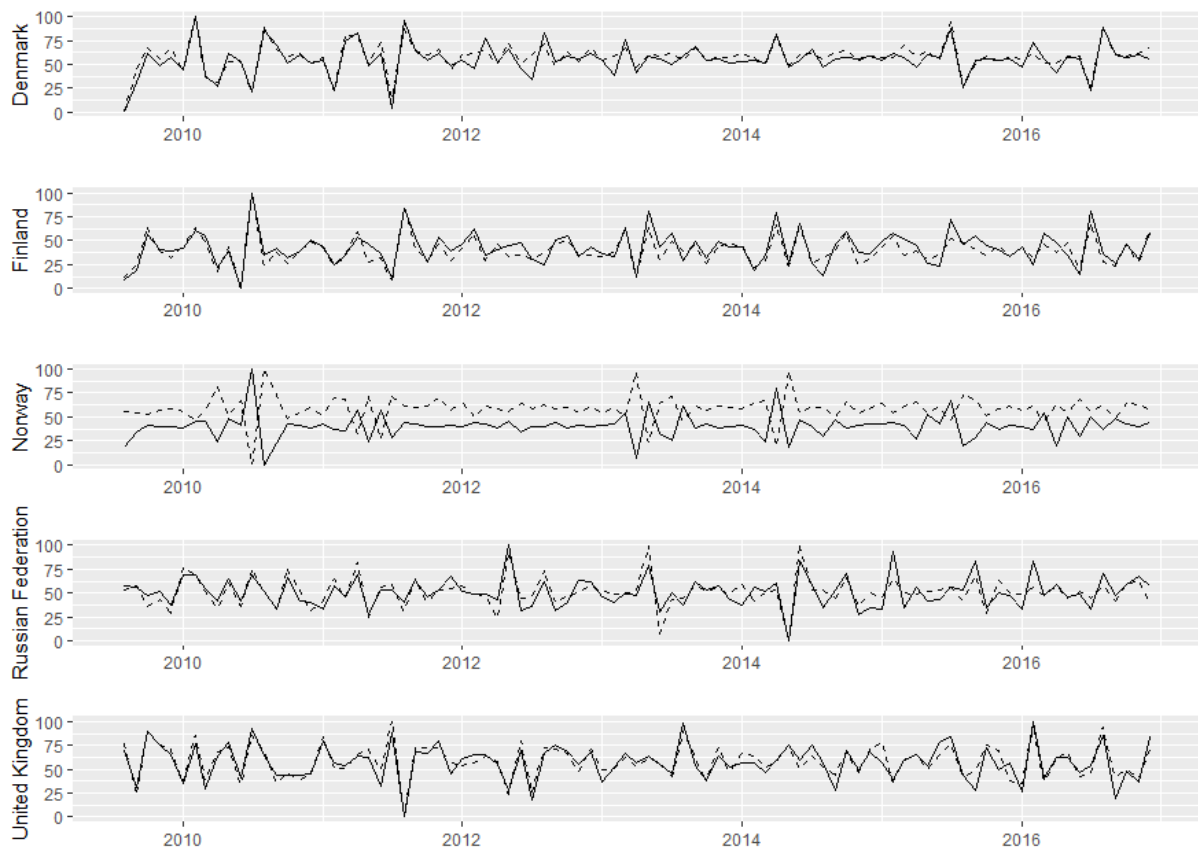


Fig. 2: Monthly tourist arrivals (solid line) and corresponding search indices (dashed line) normalized with a range between 0 and 100

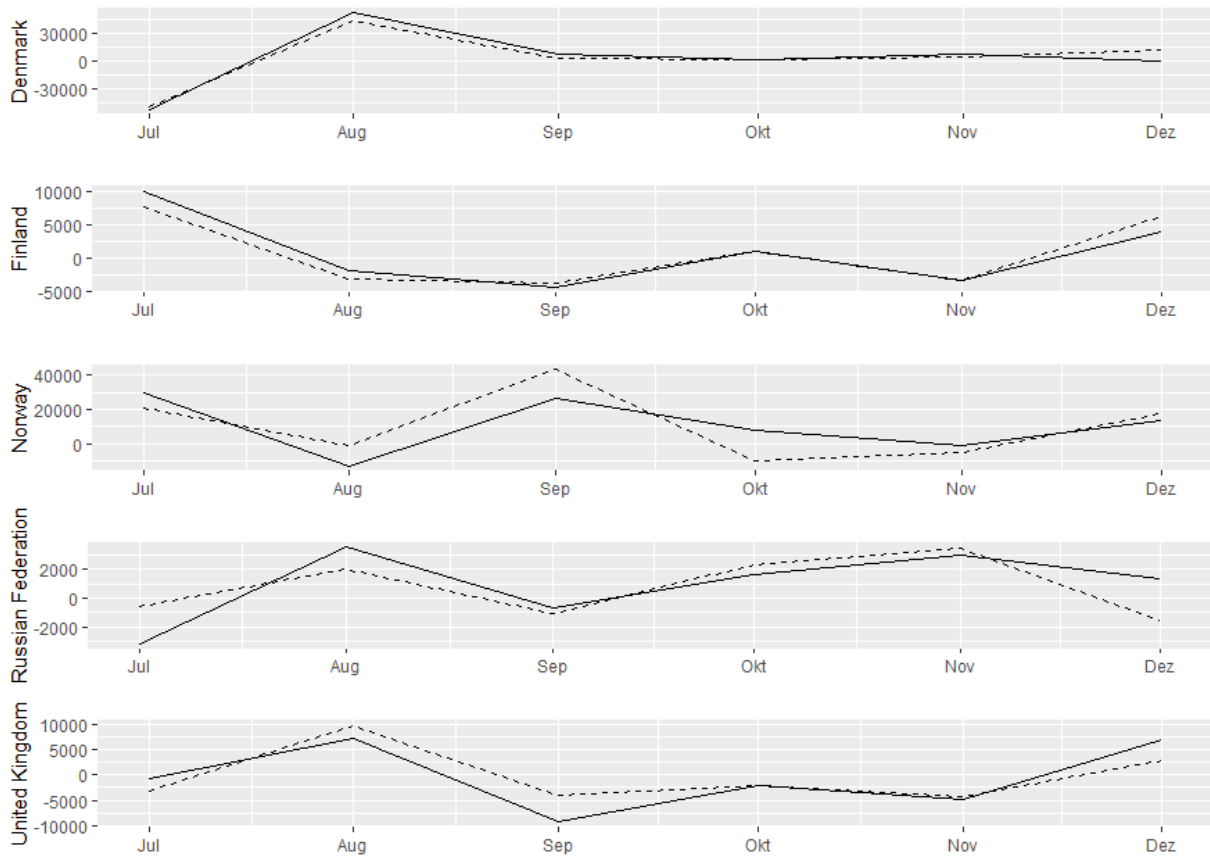


Fig. 3: Predicted values (dashed line) versus stationary arrival series (solid line) from July 2016 to December 2016 based on an artificial neural network (ANN) with Google Trends data