Transportation Research Part E 128 (2019) 400–416

Contents lists available at ScienceDirect

Transportation Research Part E

journal homepage: www.elsevier.com/locate/tre

Forecast to grow: Aviation demand forecasting in an era of demand uncertainty and optimism bias

Daniel Y. Suh[a], Megan S. Ryerson[a,b,*]

[a] Department of City & Regional Planning, University of Pennsylvania, 210 S 34th Street, Philadelphia, PA 19104, United States
[b] Department of Electrical & Systems Engineering, University of Pennsylvania, 200 S 33rd Street, Philadelphia, PA 19104, United States

ABSTRACT

Errors in forecasting airport passenger demand arise from uncertain economic climates and planners' optimism, leading airport planners to make misinformed infrastructure investments. We use publicly available data to develop and test methodologies that enable airport planners to (1) predict the probability of a severe contraction in passenger volumes and (2) improve forecast accuracy by systematically incorporating past forecast errors of airport peers thus "grounding" optimistic forecasts. By incorporating past forecast errors from like airports into airport forecasting models, we build a methodology that is grounded in established demand forecasting practices and is able to significantly improve the accuracy of aviation demand forecasting models.

1. Introduction

One of the most pressing issues facing the busiest U.S. commercial airports in the past decades has been the growth in passenger demand translating into more flights and congestion. Airport planners, both at the federal and local levels, have responded to the problem of congestion overwhelmingly through airport expansions (Ryerson and Woodburn, 2014). Airport planners use two broad categories of methods to estimate future use of the airport; (1) peer group learning (as in, considering the experiences of "peer" airports), in which airports engage in an ad-hoc and unstructured way (Suh and Ryerson, 2017) and (2) forecasting future passenger demand with statistical models based on historical data, which airports do as a part of the federally mandated planning process for airport master planning. A decision to plan and expand a runway, which typically spans a period of 10 years from planning to completion, predicates on the accuracy of the projected future passenger demand. Make a wrong decision and an airport planner could be harming the future growth of their airport and city. An airport might expand its airfield based on a robust passenger forecast only to have passenger forecasts not materialize; the airport manager is now saddled with a large debt and a small revenue stream (through fees levied on airlines per operation). As airline fees at an airport are a function of the cost to maintain the airfield and the number of flights using the airfield, an airport with a relatively large airfield to revenue stream ratio will charge airlines high fees to operate, discouraging new flight demand (Ryerson, 2016). An airport that decides against an expansion due to a modest demand forecast only to see passenger demand surge is also hindering growth, as each new flight option has the potential to bring substantial direct and indirect economic development benefits to the airport and their city (Mosbah and Ryerson, 2016; Brueckner, 2003).

Despite the criticality of getting the forecasts right, forecasting passenger demand at an airport is quite challenging. Consider that Lambert-St. Louis International Airport (STL), after forecasting large increases in passenger demand in the 1990s, built a $1.1 billion third runway. This runway is now rarely used after passenger traffic declined by more than half between 2000 and 2004 because a major airline declared bankruptcy (ACRP Report 76, 2012). In the 1990s, Denver International Airport (DIA) experienced a protracted planning process which at one point estimated that the airport would have passenger demand to justify 12 runways; almost

* Corresponding author.
E-mail addresses: dansuh@design.upenn.edu (D.Y. Suh), mryerson@design.upenn.edu (M.S. Ryerson).

https://doi.org/10.1016/j.tre.2019.06.016
Received 31 December 2018; Received in revised form 19 June 2019; Accepted 26 June 2019
1366-5545/

30 years later this demand has not materialized (Goetz and Szyliowicz, 1997). Earlier in history, planners built airports for seaplanes and small wind sensitive aircraft to fly routes that were heavily regulated by the U.S. government; the advent of jets and the plummeting of airfare due to deregulation left airport managers with runways that were inadequate (Bednarek, 2001). Indeed, in the early days of aviation during the 1930s, the nascent aviation technology was changing so unpredictably fast that no meaningful group of experts on the issue of long-term airport planning existed (Barrett and Rose, 1999).

For the airports of Denver and St. Louis, among other airports, the process of forecasting passenger demand is inadequately static in the face of increasing uncertainty. Uncertainties such as changes in global, regional, or local economic conditions or policies could lead to an airline experiencing a severe contraction of demand (De Neufville et al., 2013; ACRP Report 76, 2012). Moreover, an uncertain forecasting environment allows for more of the planners' bias, or optimism to taint forecast results (Wachs, 1989; Flyvbjerg et al., 2005). The concepts of demand uncertainty and optimism bias (explored more thoroughly in Section 2) create large forecast errors; these forecasting errors could lead airport planners to make unwise infrastructure investment decisions.

In the following study, we develop methodologies to address two major shortcomings of the current practice of passenger demand forecasting: first, that of predicting precipitous decline in passenger demand which forecasters currently do a poor job of anticipating and yet has a significant impact on infrastructure planning, and second, reducing optimism bias or inflationary tendency in forecasts. Towards improving forecasts in a way that systematically addresses these two issues, we develop methodologies which blend the tenets of peer group learning and statistical forecasting models yet are built to directly solve the issue of incorporating demand uncertainty and reducing optimism bias into the method of estimating passenger demand. Specifically, we harness large-scale historical publicly available aviation data and census data to develop and test new methodologies that enable airport planners to (1) predict the probability of a severe contraction in passenger volumes in the next 10 years and (2) improve forecast accuracy by systematically incorporating past forecast errors of airport peers into the current forecast and "ground" optimistic forecasts. By grounding forecasts with information from airport peers, we utilize a novel forecasting methodology termed reference class forecasting; we modify the method and adapt it to the airport context by incorporating the airport peer group concept. Our study constitutes a first serious attempt at moving away from pinpointing passenger demand accurately and instead predicting demand contraction and applying and adapting reference classes to aviation demand forecasting.

Our methods and validation of these methods indicate that identifying a reference class of airports that share similar socioeconomic and airport trends improves the forecast accuracy. Given that our methodology is rooted in classic aviation demand forecasting methods – identifying airport peer groups and developing a statistical model based on historical data to estimate passenger demand – our new method leverages and innovates on existing frameworks and therefore has a higher possibility of acceptance in the planning community. In the following study we begin with a review of the sources of uncertainty in airport passenger demand forecasting and the presence of optimism bias, as well as the practice of forecasting demand (Section 2). We next develop a model to predict a severe contraction in passenger volumes over a 10 year period at an airport by incorporating operational and socioeconomic variables (Section 3). Finally, we introduce the reference class forecasting model and validate it with a dataset of the forecasted and realized demand at the top 64 airports in the US by passenger volume (Section 4). We conclude (Section 5) with the discovery that the combination of peer-group learning framework and reference class forecasting results in a powerful forecasting methodology for producing more accurate predictions of passenger.

2. Airport planning in a new era of planning

2.1. Methods for forecasting passenger demand

The Federal Aviation Administration (FAA) requires all U.S. airports receiving federal grants (of which the top 100 airports in the U.S. do) to produce airport master plans (De Neufville et al., 2013). As a blueprint for long-term airport development, airport master plans are impactful documents. New runways typically cost more than $1 billion; they require long-term planning and investment and a broad view of economic and environmental impacts (Li et al., 2018). Given the significant impact of airport master plans on both airports and surrounding regions, airport master plans are designed to be "a comprehensive study of an airport" to "meet future aviation demand" while considering "potential environmental and socioeconomic impacts" (FAA, 2015). Addressing demand growth often becomes the dominant consideration in airport master planning (May and Hill, 2006). The practical apparatus of airport master planning heavily relies on a forecast of future demand; it involves aviation demand forecasts, the elaboration of alternatives and their analysis according to their costs and benefits, and the selection of that alternative which best addresses the forecast (Ryerson and Woodburn, 2014). The FAA sets strict guidelines for an Airport Master Planning process, a linear process that hinges on the key step of the passenger demand forecast (Wijnen et al., 2008).

Preparing a passenger demand forecast is multidimensional. There are largely four types of statistical, data driven forecasting models used for airport demand forecasting: market share forecasting, time series model forecasting, simulation, and econometric model forecasting (ACRP Synthesis 2, 2007). Market share forecasting measures airport traffic as a share of a larger aggregate measure and assumes the relationship to extend into the future (Corsi et al., 1997). A time series model is a relatively simple method in which the existing data trend is extrapolated into the future (Wadud, 2011). Simulation provides more disaggregate information such as how a passenger might travel through an airport terminal (Parker et al., 2012). The most widely used forecasting method for aviation demand forecasts is econometric modeling (ACRP Synthesis 2, 2007). Econometric modeling involves statistical estimation of a regression model that assumes a causal relationship between dependent and independent variables. In its simplest form (which many airports adopt), the relationship between the dependent variable (e.g., aviation passenger demand) and the independent variables (e.g., socioeconomic and airport specific metrics) is assumed to be linear. The model is estimated on historical data (i.e., realized airport demand, socioeconomic, and airport specific metrics).

The current approach to forecasting, i.e., econometric modeling, is a classic example of what Ascher (1979) calls the “insider's” approach where the relevant concerns are limited to “the basic scientific information and techniques at the forecaster's disposal” (Ascher, 1979). Embedded in the econometric modeling process is a set of assumptions about the relationship between the dependent and independent variables as well as the presumed trajectories of the independent variables (i.e., what the socioeconomic and airport metrics will be in the future). This is both the strength and the weakness of the econometric models. One can test the underlying assumptions by fitting the model and evaluating the model parameters and develop a more informed set of assumptions. On the other hand, because the outcome is determined by its core assumptions, there is a danger in using a set of incorrect assumptions that may lead to wildly inaccurate estimation of the reality. Incorrect assumptions of exogenous inputs or independent variables, can result in over- or under-estimation of the dependent variable (Pickrell, 1992). Furthermore, assumptions do not hold well for long-term forecasts (Makridakis and Bakas, 2016); as the forecast target year gets farther into the future, there are more uncertainties and risks that play into the dynamics of the forecasts.

Recent developments in aviation demand forecasting literature apply more complex techniques using machine learning algorithms. However, these techniques leverage machine learning for shorter term demand forecasts, which are more useful for planning airport's daily operations. For example, a popular area of inquiry seems to be improving short term demand forecast accuracy with variances on decomposition time series modeling in which disparate components of the time series are predicted separately using algorithms such as Support Vector Machines (Bao et al., 2012; Xie et al., 2014; Xu et al., 2019) and Holt Winters (Dantas et al., 2017). For even shorter term demand forecasts with heavy focus on predicting short-term fluctuations in air passenger demand, models using “big data” such as search engine queries are proposed (Kim and Shin, 2016). In fact, most applications of deep learning and machine learning techniques in the context of demand forecasting in transportation in general tend to focus on short term demand forecasting (Wei and Chen, 2012; Moreira-Matias et al., 2013; Ke et al., 2017).

While the statistical models are the key part of an aviation forecast, they are not the only practice an airport will uptake in the aviation demand estimation process. Airports engage in a combination of peer group learning and econometric modeling. Suh and Ryerson (2017) explore how airports utilize peer group learning to do forecasting: they review how airports identify their own peer groups based on their aspirations and not the actual position of the airport, and then present a data driven method to identify airport peer groups. Peer group learning, despite its drawbacks, is an indication of how airports prefer to learn from their “peers” in addition to learning from their own history.

Peer group learning is reminiscent of the concept of reference class forecasting, developed by behavioral economists Kahneman and Tversky (1977). Their work showed that there is often a type of cognitive bias in decision-making process and the judgement errors are more systematic and predictable than random. Reference class forecasting builds on this idea and formalizes the process of identifying the predictable cognitive errors in forecasts and removing or “de-biasing” them from the forecasts. It is essentially a formalization of the idea that one needs to learn from the past mistakes, which has been proposed as a simple yet powerful safeguard against optimism bias in forecasting by the likes of Ascher (1979). Specifically, reference class forecasting involves the following three steps (Flyvbjerg et al., 2005):

(1) Identification of a relevant reference class of past, similar projects. The class must be broad enough to be statistically meaningful but narrow enough to be truly comparable with the specific project.
(2) Establishing a probability distribution for the selected reference class. This requires access to credible, empirical data for a sufficient number of projects within the reference class to make statistically meaningful conclusions.
(3) Comparing the specific project with the reference class distribution, in order to establish the most likely outcome for the specific project.

The efficacy of reference class forecasting was first demonstrated by Lovallo and Kahneman (2003) citing an example of curriculum planning. It was not until 2005 that reference class forecasting was implemented in practice in the field of planning. Flyvbjerg (2008) used reference class forecasting for the first time for cost estimates for large transportation infrastructure investments in the UK with generally positive outcomes. His work led the American Planning Association to recommend the use of reference class forecasting for large infrastructure projects (American Planning Association, 2005). Researchers have subsequently applied reference class forecasting to cost estimates of hydroelectric dams (Awojobi and Jenkins, 2016), project management (Batselier and Vanhoucke, 2017), and public school building costs (Bayram and Al-Jibouri, 2017). As of this writing, there is no practical application or research on reference class forecasting in aviation demand forecasts.

2.2. Uncertainty and optimism in planning

Following the deregulation of the airline industry in the U.S., airlines and airports became highly susceptible to changing economic and market conditions; the number of airlines increased dramatically, routes were expanded, fares declined, and airlines adopted hub-and-spoke system in which airlines concentrated service on key hub airports (Moore, 1986). The 21st century bore witness to large spikes in fuel prices and a number of airline mergers and consolidations that brought uneven changes across airports in the U.S. (Fuellhart et al., 2016; Kim and Ryerson, 2018; Ryerson and Kim, 2013). During the 2000s, seven major U.S. airlines merged into three; these newly merged airlines consolidated their networks and hub operations and established fewer, more concentrated airline hubs (Goetz and Vowles, 2009; Ryerson, 2016). Some major airports situated in the largest cities saw their air service strengthen while airports in smaller metropolitan areas lost significant service, leaving their customers with fewer flight options and higher airfares due to reduced competition (Brueckner et al., 2003). Also during the 2000s, concern for the environment and climate change led to a wave of protests against airport capacity expansions (Dewey and Davis, 2013). Precisely predicting demand with so much uncertainty – and so many sources of uncertainty – is challenging at best and impossible at worst (Frechtling, 2012).

Further complicating the practice of forecasting is optimism; optimism that aviation demand will grow, and optimism that investment in the airport will spur development itself. Since the beginning of airport development in the U.S., the mindset in the aviation industry has pro-growth based underlying assumptions of economic growth and global consumer capitalism (Bednarek, 2001; May and Hill, 2006). At the same time, local governments tend to overgeneralize the indisputable benefits of air service to the local economy (e.g., direct and indirect creation of jobs and facilitation of business travel) and believe that providing higher quantity of air service via airport expansions will increase the economic benefits (Mosbah and Ryerson, 2016). There are indeed clear and measurable economic impacts from airports (Brueckner, 2003; Green, 2007) as well as intangible benefits such as civic pride (Ryerson and Woodburn, 2014). There is even some evidence in the literature suggesting that optimism itself can also manifest real benefits in a competitive market (Jiang and Liu, 2019). However, the urban boosterism, i.e., promoting a city or a region, has become one of the dominating drives in airport planning over actual need. Ryerson and Woodburn (2014) show that airport infrastructure planning documents that build on the master plans put significant focus on growing operations to preserve their hub status.

Federal funding mechanisms may also provide perverse incentives for airports to inflate their forecasts. The federal government is strongly committed to the maintenance, development, and expansion of the National Airspace System (NAS) and, more so than any other modes of transportation, the federal government plays an outsized role in regulation of aviation (U.S. Department of Transportation, 2009). As an entity in charge of the NAS, FAA funds airport capital projects through their Airport Improvement Program (AIP) designed to improve, upgrade, and expand infrastructure in support of the NAS. The funding allocation is based on the estimated future activity at airport (i.e., aviation demand forecasts) (FAA, 2017). Because airports are competing with each other for limited AIP funding, it creates incentives for airports to inflate their forecasts. Moreover, local governments have the most detailed knowledge of their infrastructure needs and this information asymmetry can incentivize airport sponsors to "overstate future activity demand" (ACRP Synthesis 2, 2007).

Towards alleviating the symptoms of optimism bias, FAA also produces the official forecasts of aviation demand called the Terminal Area Forecast (TAF) for airports in the US. FAA produces TAFs annually for all US airports to help federal, state, and local authorities plan in regards to airport and air traffic operations (Federal Aviation Administration, 2017). FAA requires the 5-year forecast prepared by a specific airport to remain within 10% of the TAF and within 15% of the 10-year forecast (FAA, 2008). However, FAA gives airport sponsors leeway to negotiate (i.e., work with the FAA to update the TAF) in the case their forecasts fall outside the ranges of the TAF (FAA, 2008).

The type of inflationary incentives in the federal airport funding structure is by no means unique to airport planning. Pickrell (1992) found that local officials in eight US cities showed bias towards high-capital transit investments due to the structure of the federal transit grant programs that levies "financial risk of forecasting errors" to the federal treasury rather than local government. In other words, because the federal government provides a majority of the funding for transit projects and local governments prepare projections of the ridership to justify the investments, local governments have incentives to inflate their forecasts and the federal government ends up taking on the risk.

While there is much suspicion about similar types of bias in aviation demand forecasts, literature on aviation demand forecast accuracy, let alone forecast bias, is very scant. Maldonado (1990) represents by far the most comprehensive analysis on aviation demand forecasts; his research evaluated aviation demand forecast accuracy for 22 master plans in the FAA New England region and found that forecasts in general perform poorly. At the same time, as airports have experienced drastically different and disproportionate changes in the 21st century it is clear that these dynamic changes and uncertainty require a more nuanced understanding of forecast accuracy. Specifically, we need to understand different patterns of demand growth/contraction in order to justify airport infrastructure investments instead of relying solely on the forecast accuracy (precisely because we know the forecasts are inaccurate).

3. Predicting the probability of a severe contraction in passenger volumes

Towards understanding indicators of a future contraction in passenger demand, we build a predictive model using binary logistic regression that estimates the probability that an airport will experience a dramatic contraction in passenger demand in the next 10 years. Overall, airport planners use the 10-year aviation demand forecasts to prepare the airport master plans and plan for infrastructure investments such as building new runways. Therefore, this time period represents an important planning horizon as well as a time horizon long enough to be impacted by external socio-economic forces and a possible systematic bias to overestimate. Because the baseline assumption for airport planning is growth in demand, detecting the signs of a future passenger contraction is an important insight into airport planning that may prevent wasteful investments.

We build a predictive model using the FAA's official 10-year demand forecasts, the TAF, for the 64 large, medium, and small hub airports (as defined by the FAA based on the share of total passengers moved) that are located within the top 50 metropolitan statistical areas (MSAs) by population. These airports served about 90% of total passengers in the US in 2016 (FAA, 2016). We dichotomize the outcome by categorizing these patterns into that of a dramatic contraction and of cyclical changes; two categories that we define in this section. We then build a binary logistic regression model using several airport and MSA explanatory variables to predict the binary outcome. Our results indicate that the regional socioeconomic trends can be robust predictors of a dramatic contraction in demand at an airport.

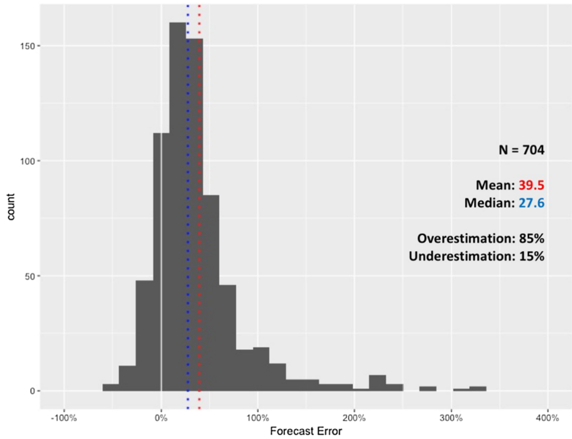

Fig. 1. Histogram of the forecast errors for 10-year aviation demand forecasts.

3.1. Exploratory analysis of 10-year passenger demand

We begin looking at the forecast errors (percent difference between the predicted passenger demand and the actual passenger boardings) of the 704 10-year aviation demand forecasts for the 64 study airports from 1995 to 2005 (see Fig. 1). The majority (85%) of these forecast errors are positive errors, meaning the forecasts were higher than the actual passenger volumes. Only a small portion (15%) of the forecasts underestimated the passenger volumes. On average, these forecasts overestimated the passenger volumes by almost 40%, however, there are also several extremely large and positive forecast errors that are skewing this average. The median forecast error, which is less affected by these outliers, is close to 30%. Because these major airports handle hundreds of millions of passengers per year, the forecast error of 30% means that there are millions fewer passengers that are actually using the airports than estimated.

We next deep dive into the forecast errors of four airports: Miami (MIA), San Francisco (SFO), St. Louis (STL), and Pittsburgh (PIT). We do so as MIA and SFO are airports that have achieved relatively stable growth during the study period and STL and PIT experienced dramatic declines in their passenger demand. We investigate their forecast errors to understand the value of identifying a severe contraction in passenger demand.

Fig. 2 shows the annual passenger demand (in the form of realized passenger boardings) at MIA and SFO along with the 10-year TAF forecasts with a base year 1995 and a target year 2005 (shown in red). As suspected, the 10-year forecasts for both MIA and SFO (red bars) overestimated by substantial margins. At the same time, the annual demand for both MIA and SFO (grey bars) show a pattern of growth subsequently after the forecast target year (2005) to 2015; both airports ultimately achieve the passenger boarding equal to their 2005 projected value in 2015. These airports, while they did not meet their forecasts in the forecast year, did not experience a severe, sustained contraction of demand and thus grew to their forecasted level, albeit later than planned. Fig. 2 also shows the 10-year TAF forecasts (red bars) and annual demand (grey bars) for STL and PIT. The margins of error are even more pronounced for these forecasts. In fact, the forecasts overestimated by more than twice the actual demand in 2005. Additionally, the annual demand for both STL and PIT show a pattern of a dramatic contraction; the annual demand in 2015 was almost half of the demand in 1995 both airports. These airports for St. Louis and Pittsburgh, formerly prosperous industrial cities, have lost tremendous amounts of passengers along with population since their peak from decades ago. Airline strategies contributed to the major contractions in passengers at these airports because the airports' major hub airlines experienced financial difficulties and declared bankruptcies during this period (Redondi et al., 2012). The difference between airports like SFO and MIA and those similar to STL and PIT is drastic in terms of the future growth and planning needs.

3.2. Identifying 10-year passenger demand

A binary logistic regression model describes the relationship between the explanatory variables and the binary outcome variable. As our outcome variable of 10-year change in passenger volumes is continuous rather than binary, we first need to dichotomize the outcome variable. We dichotomize the 10-year passenger demand into that of severe contraction and otherwise stable patterns. The literature is relatively scarce on the empirical definition of a severe contraction in passenger volumes. Instead, a data-driven approach can distinguish distinct patterns in the 10-year percent changes in passenger volumes. Specifically, a model-based clustering algorithm such as Gaussian mixture model can tease out these distinct patterns and enable dichotomization of the outcome variable.

The left histogram in Fig. 3 shows the distribution of the 10-year percent changes in the passenger volumes for the 64 airports for all base years from 1995 to 2005 (i.e., total of 704 10-year percent changes). The distribution almost seems normal (i.e., Gaussian) but there are a few spikes in the distribution towards the left tail as well as a very long right tail. The mean and the median are 14.39 and 13.11, respectively (i.e., 14.39% growth in passenger volumes and 13.11% growth in passenger volumes) (Table 1, first column).

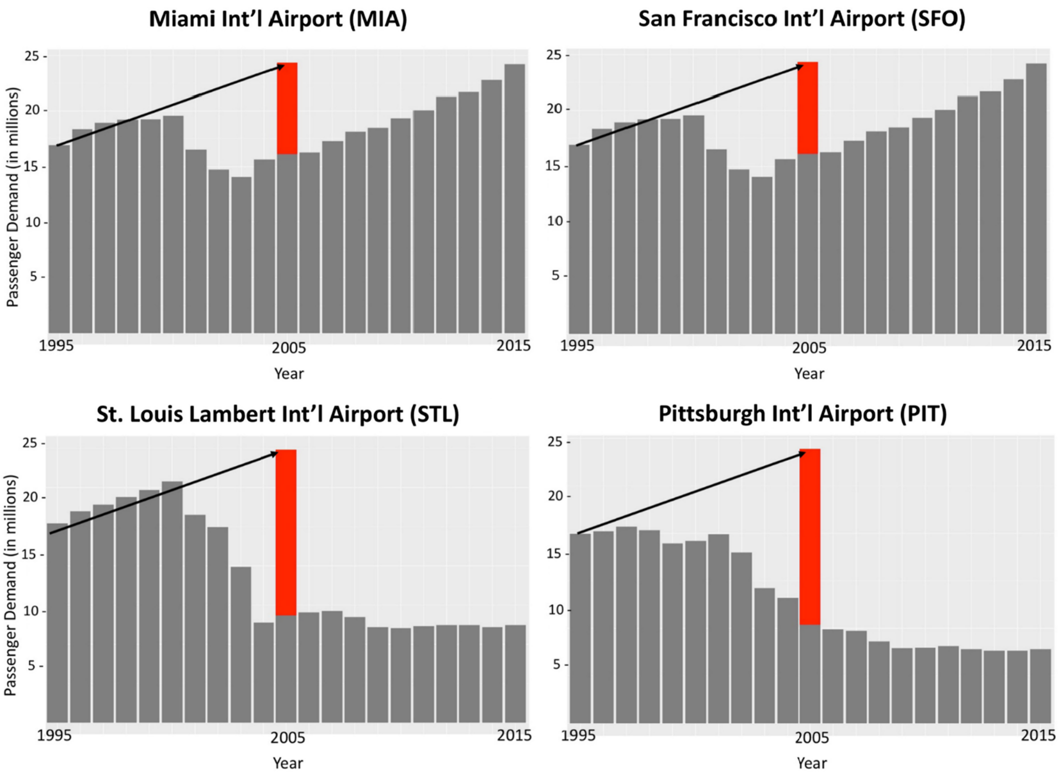

Fig. 2. Annual passenger demand and 10-year TAF forecast (base year 1995, target year 2005) for sample airports.

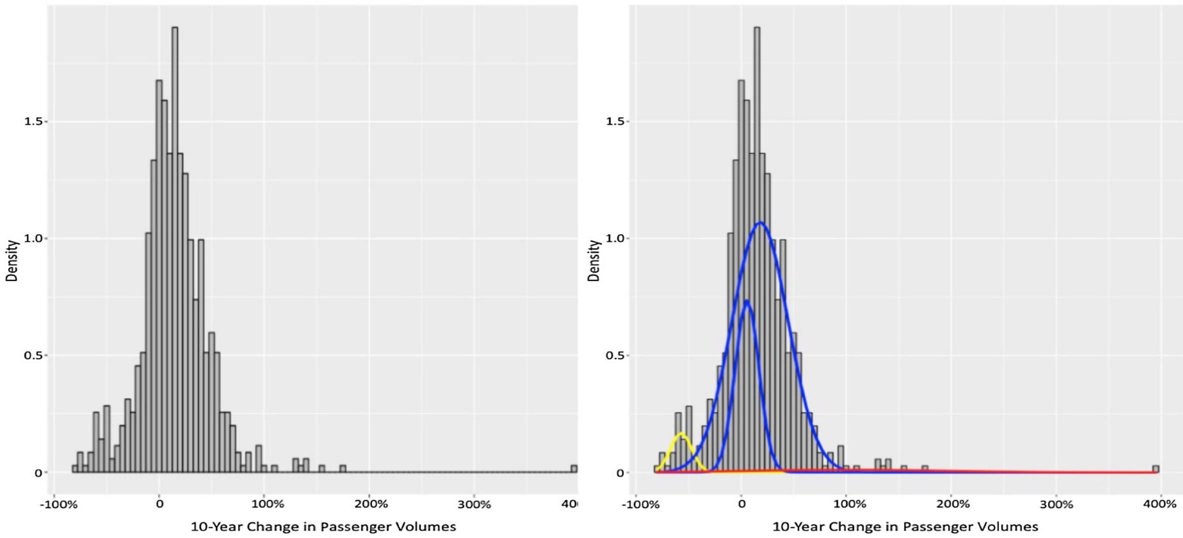

Fig. 3. Histograms of 10-year percent changes in passenger volumes.

However, there seems to be a wide spread in the data as indicated by the standard deviation of 34.91. There are also some extreme values as big as 395.70 and as small as −80.79.

One approach to clustering such data points is by using a model-based algorithm such as Gaussian Mixture Model, which assumes the data points are generated from a mixture of Gaussian distributions with unknown parameters. This method uses the Expectation-Maximization (EM) algorithm, an iterative algorithm that estimates the parameters of the distributions, and produces the posterior probabilities, the probabilities of each data point generated from a particular distribution. This method essentially allows for clustering of similar data points together by assigning each data point to a distribution with the highest posterior probability.

Table 1
Summary statistics for 10-year change clusters.

Statistic	All Data (N = 704)	Growth Cluster (n = 9)	Cyclical Cluster (n = 559)	Contraction Cluster (n = 136)
Mean	14.39	167.20	22.40	−28.61
Median	13.11	138.70	18.09	−22.00
Std. Dev.	34.91	87.51	23.08	18.26
Max	395.70	395.70	99.15	−11.23
Min	−80.79	110.00	−10.74	−80.79

The histogram on the right in Fig. 3 shows the resulting mixture of four Gaussian distributions ($k = 4$). The two blue distributions in the middle include the majority of the data points and the flat red distribution encompasses the large positive outliers. The yellow distribution near the left tail includes the data points of interest, the 10-year percent changes that are negative (i.e., the passenger volumes declined) and extreme (i.e., distinctively larger in magnitude than other negative data points). Based on these distributions, we further cluster or group the data points into 3 distinct clusters (Table 1, right three columns). The first cluster only contains 9 data points that have more than 167% growth on average in passenger volumes in the 10-year period (growth cluster). The second cluster contains the majority of the data points (559) that show a moderate average growth of 18.09% (cyclical cluster). The last cluster contains 136 data points that showed a distinct pattern of negative growth (contraction cluster); in the 10-year period, the passenger volumes for these airports declined by 29% on average with some airports losing more than 50% of their passengers.

We code the 559 data points in the cyclical cluster as 0 (non-event) and the 136 data points in the contraction cluster as 1 (event) for the total of 695 data points for the binary outcome variable. The data points in the growth cluster are extreme outliers that may skew the data; we do not include them and we focus on the dichotomy between cyclical demand and demand contraction.

3.3. Predicting ten-year contraction in passenger demand

With the dichotomized outcome variable, we use both point-in-time metrics and change-over-time metrics of airport operations and socioeconomic conditions as explanatory variables in a binary logistic regression model.

Binary logistic regression estimates the probability that an event will occur given the values of explanatory variables. For a binary outcome variable Y and a set of explanatory variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_k)$, the model takes on the expression: $\pi_i = \Pr(Y_i{=}1|\boldsymbol{X_i} = \boldsymbol{x}_i) = \frac{\exp(\boldsymbol{\beta x})}{1-\exp(\boldsymbol{\beta x})}$. Once the parameters ($\boldsymbol{\beta}$) are estimated, the coefficients are typically interpreted in exponents (exp (β_i)), i.e., odds ratio. An odds ratio indicates that for every unit increase in X_1, the odds of an event happening is multiplied by exp(β_1).

Literature on air travel demand shows that there is intrinsic relationship between airport passenger demand as well as airline strategy and sociodemographic characteristics of its host city/region (Brueckner, 2003; Alkaabi and Debbage, 2007; Green, 2007; Bel and Fageda, 2008). Therefore, we selected explanatory variables pertaining to both airports and the MSAs in which they reside. We use 9 such variables for the base year figures (i.e., point-in-time numbers in base years) as well as 5-year average annual percentage change (5AAC) in these 9 variables up to the base year in order to capture dynamic changes within the past 5 years (Suh and Ryerson, 2017). For example, population of 2 million in the base year 1995 for an MSA is a point-in-time figure while 5AAC would be 5% (averaged over 2% change during ‘90-‘91, 3% during ’91-’92, 3% during ’92-’93, 4% during ’93-’94, 3% during ’94-’95). In total, we start with 18 explanatory variables (9 point-in-time and 9 5AACs) (Table 2). Data sources include the U.S. Census, the FAA’s databases of airports and passengers carried; the U.S. Bureau of Economic Analysis (BEA) measure of competition; and the Bureau of Transportation Statistics (BTS) database of passenger fares (DB1B, a 10% sample of all tickets purchased on U.S. Carriers) and database of flight schedules and details (T100).

We highlight here some of these variables that may require further explanations.

Airport competition: In multiple-airport regions, the main airport faces competition from the secondary airports which may offer different types of accessibility, charges, and quality of service to compete for passenger traffic. Passengers may also consider tradeoffs between the quality of service offered at competing airports and distance to travel to the airports (Johnson et al., 2014). We estimate the potential airport competition for airport a_i by summing annual passenger enplanements (in 100,000 s) for airports within 100 miles divided by distance from airport a_i: Airport competition $(AC_{ai}) = \sum_n \frac{\left(\frac{E_{ni}}{100000}\right)}{d_{na_i}}$, where $n \in N$(set of airports within 100 miles from airport a_i),

E_{ni} = annual passenger enplanements at airport a_{ni} and
d_{na_i} = distance in miles of airport a_{ni} from airport a_i.

HHI: When a few airlines have a large share of airport operations, the impacts of the airlines’ decisions are more powerful than if a large number of airlines share the airport. A hub airline’s decision to de-hub from an airport, for example, has a long-lasting impact on the airport, leaving it with excess capacity and overbuilt infrastructure (Redondi et al., 2012). The HHI is calculated as: $HHI_{a_i} = \sum_l m_{li}^2$, where m_{li} is market share of an airline l at airport a_i as estimated by proportion of seats provided by airline l over total seats by all airlines. A higher HHI indicates a higher concentration, while a lower HHI means greater competition among airlines.

In order to prevent overfitting, we split the data 80–20 into a training set (n = 556) and a test set (n = 139). We fit the binary logistic regression model on the training set and evaluate its performance on the test set. We selected the final model in Table 3 using backward-stepwise regression. We started with a model with all the explanatory variables (minus base population due to collinearity,

Table 2

Summary of analysis variables (5-year avg. annual % change up to base year corresponding information in parenthesis).

Variables in base year numbers	Notes	Unit	Mean	Std. Dev.	Data Source
Passengers	Reflects overall volume of traffic at an airport	Persons (millions)	8.42 (4.00)	7.95 (7.76)	FAA
Airport competition	Function of the share of demand at the airport with respect to all airports within 100 miles	Unitless	3.74 (1.97)	5.57 (3.49)	FAA
Connecting passenger share	A high proportion of connecting passengers indicates high facility needs, such as more gates and terminal space	Proportion	0.47 (0.53)	0.11 (4.51)	BTS DB1B
Avg. number of seats per aircraft	Fleet mix of aircraft at an airport indicate types of destinations and travel demand at the airport. We estimate the fleet mix by using the average number of seats per aircraft at each airport as a proxy.	Seats	118. 40 (−1.68)	26.87 (3.02)	BTS T-100
Avg. ticket price	Average direct flight ticket prices and percentage of seats flown by low-cost airlines; indicate the level of competition among airlines	Dollars	227.70 (−2.94)	53.01 (3.39)	BTS DB1B
HHI	Airline concentration is measured here by using the Herfindahl–Hirschman index (HHI), a frequently applied economic concept that measures the amount of competition among firms in an industry. HHI is computed as a sum of squared market shares of companies,	Unitless	0.35 (0.44)	0.20 (7.52)	BTS T-100
Population	Air traffic has historically been correlated with economic conditions. The economic variables are population, income, employment in service sectors. Service sector employment as opposed to total employment was used because the literature indicated a stronger systematic relationship between employment in that sector and air passenger demand.	Persons (millions)	3.56 (1.12)	3.44 (0.90)	Census
Per capita income		Dollars (thousands)	45.87 (1.77)	7.91 (1.27)	BEA
Service sector employment		Persons (millions)	0.92 (4.76)	0.91 (3.45)	Census

Table 3
Summary table for the binary logistic regression.

	Odds ratio	$P > \|z\|$
(Intercept)	0.120	0.000***
Airport competition % change (5AAC)	0.612	0.000***
Connecting passenger share	1.555	0.000***
Connecting passenger share % change (5AAC)	0.962	0.005**
Avg. number of seats per aircraft	0.709	0.000***
Avg. ticket price	0.612	0.000***
HHI	2.234	0.004**
HHI % change (5AAC)	1.346	0.003**
Population % change (5AAC)	0.201	0.000***
Per capita income	1.539	0.001**
Service sector employment	0.406	0.001**
	n = 556	
	AIC = 422.66	

**p < 0.01, ***p < 0.001.

and standardized) and eliminated a variable at a time based on its significance (p-value) until the optimal number of variables was decided (when the Akaike information criterion, AIC, is minimized). We tested the selected model on the test data set and achieved 86% accuracy, 84% true positive rate, and 23% false positive rate. The purpose of this research is to identify the operational and socioeconomic metrics that are predictive of the demand uncertainty and therefore, the predictive performance is a secondary issue although it is reflected in the model selection. We report the odds ratios ($\exp(\beta)$) instead of log-odds (β) for the ease of interpretation.

3.4. Discussion of results

As all the explanatory variables were standardized before fitting the model, therefore, the unit for each selected variable reported in Table 3 is one standard-deviation. The following discussion of each of the selected variables assumes that all other variables are held constant (*ceteris paribus*).

Airport competition 5-year avg. annual % change (5AAC) (odds ratio = 0.612): One unit increase in the 5AAC of airport competition reduces the likelihood of experiencing a severe contraction in passenger volumes by almost a half. In other words, if the airlines in the neighboring airports have been offering more seats or services in the past 5 years, the airport of interest is less likely to experience a severe contraction in passenger volumes. The airlines' decision to provide more service in the region may indicate that the region as a whole is a growing market and airports in this region are less likely to experience a sudden disruption in the passenger trends.

Connecting passenger share (odds ratio = 1.555) and *5-year avg. annual % change (5AAC)* (odds ratio = 0.965): Our model indicates that one unit increase in the connecting passenger shares will increase the likelihood of experiencing a severe contraction in passenger demand by a factor of 1.5. This falls in line with the fact that the airports with high connectivity are typically the hub airports and the hub airports have historically experienced a sudden disruption in passenger volumes due to de-hubbing. On the other hand, one unit increase in the 5-year average annual % change (5AAC) in the connecting passenger shares slightly reduces the likelihood of a severe contraction in passenger volumes. The gain in the connectivity can be interpreted as gaining more passengers in general because the connectivity indicates that the airlines are pooling passengers at the airport.

Average number of seats per aircraft (odds ratio = 0.709): One unit increase in the average number of seats per aircraft at an airport reduces the likelihood of a severe contraction in passenger volumes by a factor of 0.71. The literature indicates that the small aircraft size is related to the high frequency routes (i.e., short domestic routes) (Wei and Hansen, 2005; Ryerson and Hansen, 2013) and conversely, the large aircraft size can indicate longer routes including international destinations. Our model indicates that airports with larger aircraft potentially serving international routes are less likely to experience a severe contraction in passenger volumes.

Average ticket price (odds ratio = 0.612): One unit increase in the average ticket price decreases the likelihood of a severe contraction in passenger volumes by a factor of 0.61. This is in line with the findings just discussed above. Higher ticket prices often indicate longer routes including international destinations. In addition, the literature indicates that the higher ticket prices may be explained by the mix of leisure and business passengers (Lee and Luengo-Prado, 2005).

HHI (odds ratio = 2.234) and *5-year avg. annual % change (5AAC)* (odd ratio = 1.346): One unit increase in HHI, a measure of market concentration, increases the likelihood of a severe contraction in passenger volumes by a factor of 2.2. This result finds support from many case studies in which the dominant airline with a large share of the market at the airport discontinues their service at the airport resulting in a sharp contraction in passenger volumes. Similarly, one unit increase in the 5-year average annual % change (5AAC) in HHI also increases the likelihood by a factor of 1.3. This indicates that as fewer and fewer airlines start gaining larger shares of the market at the airport, the airport is more likely to experience a severe contraction in passenger volumes.

Population 5-year avg. annual % change (5AAC) (odd ratio = 0.201): One unit increase in 5-year average annual % change (5AAC) in MSA population reduces the likelihood of a dramatic contraction in passenger demand in the next 10 years by more than a half. This result supports the existing literature that links air travel demand and socio-demographic conditions of airport host cities/regions (Alkaabi and Debbage, 2007).

Per capita income (odds ratio = 1.539): One unit increase in the per capita income increases the odds of a severe demand contraction by a factor of 1.5. This result is somewhat counterintuitive as it indicates that airports in the MSAs with higher per capita income are more likely to experience a sudden disruption in passenger volumes. This may be a case where the MSAs with higher per capita income tend to host hub airports and by definition, hub airports are at a greater risk for "de-hubbing". The literature, however, is not conclusive on this point.

Service sector employment (odds ratio = 0.406): One unit increase in the service sector employment reduces the likelihood of a severe contraction in passenger volumes by a factor of 0.41. This result supports the findings in the literature that show a strong positive relationship between employment in service sector and air passenger volumes (Alkaabi and Debbage, 2007). In other words, airports located in the MSAs with strong service sector employment base have stable passenger volumes bolstered by the service sector employment.

4. Grounding optimistic forecasts: Testing the efficacy of reference class forecasting

Demand forecasts in general are prone to optimism bias especially when the use cases of forecasts create incentives to inflate benefits and downplay costs (Wachs, 1989; Kain, 1990; Pickrell, 1992; Button and Lall, 1999; Flyvbjerg et al., 2005). Because aviation demand forecasts are used to justify and fund infrastructure investments, optimism bias is a highly relevant issue to address. In literature, a methodological framework known as reference class forecasting has been introduced and applied to real world forecasts to reduce optimism bias in demand forecasts with very promising results (Flyvbjerg, 2008).

We pose the following research questions in order to evaluate the feasibility of applying reference class forecasting to the aviation demand forecasts:

(1) Does reference class forecasting produce statistically significant reductions in the forecast errors for the aviation demand forecasts compared to the traditional forecasts?
(2) What is the relevant and effective definition of a reference class of the forecast errors?

The two questions above are inherently related because the effective implementation of reference class forecasting hinges on the identification of a relevant reference class. Towards addressing these two questions, we develop the following four methodologies through which we construct different sets of reference class forecast errors and test the forecast accuracy of each of these methods compared to that of the original forecasts. These methods differ in their approach to identifying a reference class, a set of relevant past forecasts that could provide useful information for the forecast of interest.

Mean Forecast Error (MFE): We use each airport's own past empirical forecast errors. For instance, for a 10-year forecast for Philadelphia International Airport (PHL) in 2005, we use the forecast errors (i.e., how much the forecast was off by) of PHL's past 10-year forecasts (e.g., 10-year forecast in 1995) and calculate the average or MFE. Then we adjust (i.e., reduce) the current 10-year forecast by the calculated MFE in order to remove the systematic optimism.

Mean Growth-Based Forecast Error (MGBFE): To reflect the observation that there may exist a correlation between forecasted growth percentage and forecast error, we use the empirical forecast errors of the past forecasts (of any airport) with forecasted growth percentage that is within a range of the forecasted growth percentage of interest. In other words, for a given forecasted growth percentage (e.g., 30% growth for PHL), we find past forecasts of any airport that forecasted a similar range of growth (e.g., 27.5 − 32.5%) and use the mean of their forecast errors (MGBFE) to adjust the forecast of interest.

Mean Peer-Based Forecast Error (MPBFE): Using the peer identification methodology developed in Suh and Ryerson (2017), we find airports with similar socioeconomic and airport characteristics and use their past forecast errors to adjust the current forecast. The main logic behind this approach is that there have been dynamic socioeconomic changes with disproportionate impacts on airports and airports that have gone through similar changes in the socioeconomic trends may also share similar forecast errors. In this approach, we first identify peer groups of airports and calculate the mean forecast errors for each group (Mean Peer-Based Forecast Error or MPBFE) and adjust the forecast of interest by its peer group's MPBFE.

Enhanced Mean Peer-Based Forecast Error (EMPBFE): We incorporate the predicted probabilities of a severe contraction in passenger volumes developed in Section 3 to adjust the MPBFE in the previous method. Because the predicted probabilities provide the information on how likely it is for an airport to experience a dramatic drop in the passenger volumes (and thus, a potentially larger forecast error), we use this additional information to calibrate the MPBFE and name this approach Enhanced Peer-Based Forecast Error (EPBFE).

We evaluate the performance of each of these approaches (MFE, MGBFE, MPBFE, EPBFE) by comparing the forecast errors between the actual forecasts and the adjusted forecasts (i.e., forecast adjusted by the mean forecast errors from our methods). The actual forecasts and their observed errors constitute the baseline performance of current forecasts against which our four methods will be measured. Specifically, we use the paired Wilcoxon signed rank test to evaluate the null hypothesis that the median of the absolute values of the adjusted forecast errors are less than those of the original forecast errors. By using the absolute values, the rejection of the null hypothesis means that the adjusted forecast errors are closer to zero (i.e., more accurate). We use the Wilcoxon test instead of the *t*-test for equal means because the Wilcoxon test is a nonparametric test that does not assume normal distributions of the samples and thus suitable for these forecast errors containing several outliers.

Table 4
Summary statistics for the adjusted forecast errors and actual forecast errors.

Statistics	Actual (n = 64)	Actual (**n = 52**)	MFE (n = 64)	MGBFE (**n = 52**)	MPBFE (n = 64)	EMPBFE (n = 64)
Mean	43.71	39.18	14.66	8.86	3.64	26.95
Median	27.38	26.58	2.19	7.53	0.10	17.28
MAPE	48.48	43.63	53.59	30.16	31.05	36.41
Proportion above 0	83%	85%	55%	60%	50%	72%
Proportion below 0	17%	15%	45%	40%	50%	28%
% Change in MAPE	–	–	+10.5%	−30.9%	−36.0%	−25%
Paired Wilcoxon Signed rank test (p-value)	–	–	0.5584	0.0243	0.0001	0.0000

In addition we use the Mean Absolute Percentage Error ($MAPE = \frac{\sum_{t=1}^{n}\left[\frac{|X_t - \widehat{X}_t|}{X_t}\right]}{n} \times 100\%$), as an additional measure of forecast accuracy, where X_t is the observed value at time t and $\widehat{X}_t$ is the predicted value at time t, with smaller MAPE indicating smaller forecast errors (i.e., higher forecast accuracy). This is a popular method among forecasters to compare the effectiveness of competing forecast models if the only criteria of evaluation is forecast accuracy. MAPE is a popular measurement of forecast errors because it is scale-independent and easy to implement. One of the major drawbacks of MAPE is that it is undefined if $X_t = 0$, that is, the observed value is zero (Hyndman, 2006); however, this situation is irrelevant for aviation demand forecasting and for this research because all of the passenger demand for the major airports in our dataset is non-zero.

We continue to use data from the FAA's 10-year TAF for the top 64 busiest airports in the top 50 MSAs from 1995 to 2015 as the forecasts ($\widehat{X}_t$). We also use the enplanement or boardings data (also from FAA) as the actual passenger demand (X_t). Due to the availability of data, we set 2005 as the base year when we assume the forecast of interest is being prepared and evaluate the forecast errors of the 10-year demand forecasts of the target year 2015. This arrangement arises from the fact that access to past 10-year demand forecasts is needed and because the dataset goes back only as far as 1995, the year 2005 represents the only base year feasible for this research.

Table 4 presents both the summary statistics of the actual forecast errors and the results of the adjusted forecast errors. The actual forecast errors on the left columns are the actual observed errors from historical Terminal Area Forecasts (TAFs) and constitute the baseline benchmark. Each method's performance is measured using Mean Absolute Percentage Error (MAPE) as well as paired Wilcoxon signed rank test, which tests whether the forecast errors under the proposed methods are statistically lower than the baseline errors. The following sections will go through each method in detail and reference this table for summary.

4.1. *Mean forecast errors (MFE)*

In this approach, we take Ascher (1979)'s advice in its simplest form and use each airport's own past forecast errors to adjust the current forecast. First, we identify all available historic 10-year demand forecasts and calculate the Mean Forecast Error (MFE) for each airport. That is,

$$MFE_\alpha = \frac{\sum_{t=1}^{n}\left[\frac{A_{\alpha t} - F_{\alpha t}}{A_{\alpha t}}\right]}{n} \tag{1}$$

where $F_{\alpha t}$ is the 10-year forecast of demand for airport α ($\alpha \in \{1, 2, ..., 64\}$) and $A_{\alpha t}$ is the actual demand in the target year for airport α. As mentioned, due to data availability, we only have access to one set of historic 10-year demand forecasts available, namely, 1995 TAFs for target year 2005. Therefore, $t = 1$ for this analysis. We still keep the notation t for the purpose of use in the future research when more data becomes available.

After identifying and calculating MFE_α for all available historic forecasts, we adjust the current 10-year demand forecast (i.e., 2005 TAF for target year 2015) by MFE_α,

$$\widehat{F}_{MFE_\alpha} = \frac{F_\alpha}{1 + MFE_\alpha} \tag{2}$$

where $\widehat{F}_{MFE_\alpha}$ is the MFE-adjusted 10-year demand forecast. Then forecast error can be recalculated for the MFE-adjusted forecast.

$$\hat{e}_\alpha = \frac{A_\alpha - \widehat{F}_{MFE_\alpha}}{A_\alpha} \tag{3}$$

Now we apply this process to the study airports and evaluate whether the adjusted forecasts improve forecast accuracy. First, the summary statistics of the MFE-adjusted forecast errors and the actual errors indicate the MFE method may have produced better results on average (Table 4). There were significant reductions both in the mean and the median of the forecast errors using the MFE method. However, the MFE method also resulted in a larger proportion of the forecast errors that now underestimate (45% vs 18%) and the mean and the median might not give a full picture of whether the MFE method produced substantial and statistically significant reduction in forecast errors. For example, a large negative forecast error could distort the mean forecast error for the MFE-

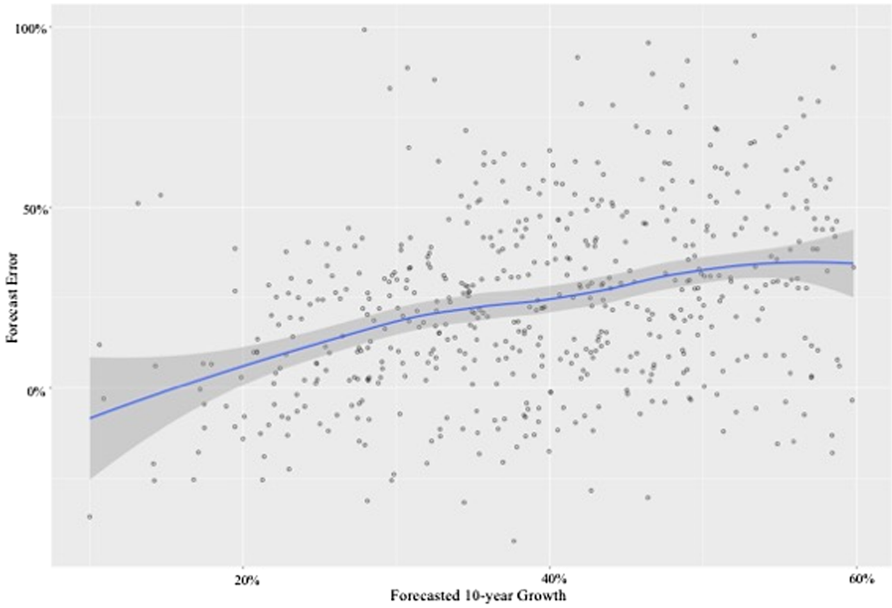

Fig. 4. Forecasted 10-year growth vs. forecast errors.

adjusted forecasts. In addition, the Mean Absolute Percent Error (MAPE) is higher for the MFE-adjusted forecast errors than that for the actual forecast errors, indicating that the forecast accuracy has declined using the MFE method.

Indeed, a paired Wilcoxon signed ranked test indicates that the median forecast error for the absolute MFE-adjusted forecasts, abs (median) = 29.16, is statistically significantly higher than the median for the absolute actual forecast errors, abs(median) = 28.48 (P-value = 0.5584). In other words, the MFE-adjusted forecasts in fact produced larger forecast errors than the actual forecasts.

This result is confirmed by looking at the bar plots of the forecast errors. The actual forecast errors (shown in orange bars) are ranked from the largest positive errors to the largest negative errors along with the corresponding MFE-adjusted forecast errors (shown in blue bars). The plot shows that in some cases, the MFE method over-compensated and resulted in significant under-estimations while in others it actually increased the forecast errors. While the net effect on the average is a reduction in the errors (as evidenced by the reduction in the mean), the MFE method performs poorly because it induces extremely large errors. The bar plots in Fig. 5 are also informative in terms of understanding the Mean Absolute Percent Error (MAPE) in Table 4; MAPE measures the average length of the bars. For example, the MAPE of 48.48 for the actual forecast errors means that the average length of the orange bars is 48.48. The relatively higher MAPE of 53.39 for the MFE-adjusted forecasts can be visually confirmed in Fig. 5 by noting that there are some extreme lengths for the blue bars.

4.2. Mean Growth-Based Forecast Errors (MGBFE)

In this approach that builds on the previous method, we use growth-based forecast errors of any airport to calculate the Mean Growth-Based Forecast Errors (MGBFE) by which we adjust the forecast of interest. The logic behind this method stems from an observation that there may exist some correlation between the forecasted growth percentage (how much growth is forecasted in the next 10 years) and the forecast errors. In Fig. 4, the forecasted growth percentage and forecast errors (%) for all 10-year demand forecasts from 1995 to 2005 for the 64 airports (n = 704) are plotted and a LOWESS line (i.e., smooth line through the scatter plot) is fitted. The LOWESS line indicates that there is a hint of a positive correlation between how much one forecasts to grow and how (in)accurate the forecasts are. In other words, it seems that the more growth is forecasted, the larger the forecast error becomes.

We leverage this information and select a set of past 10-year demand forecasts that share similar forecasted growth percentages and forecast errors. First, we calculate the forecasted percentage growth (g_α) for the forecast of interest for airport α. For instance, if the 10-year demand forecast (base year 2005) for Boston Logan International Airport (BOS) forecasts 35 million passengers in 2015 and their base year passenger demand is 25 million, g_α for the BOS' forecast is 40%. Then, we identify all available historic 10-years demand forecasts of any of the 64 airports (not just BOS' forecasts) within a range of 5 percentage points (±2.5%) of the forecasted growth percentage (g_α) (e.g., $37.5-42.5\%$ for $g_\alpha = 40\%$) and calculate the Mean Growth-Based Forecast Error, $MGBFE_\alpha = \frac{\sum_{i=1}^{n}\left[\frac{A_i - F_{ig_\alpha}}{A_i}\right]}{n}$, where F_{ig_α} is the 10-year demand forecast of any airport i whose forecasted growth percentage lies in the range of ($g_\alpha \pm 2.5$) and A_i is the actual passenger demand for airport i.

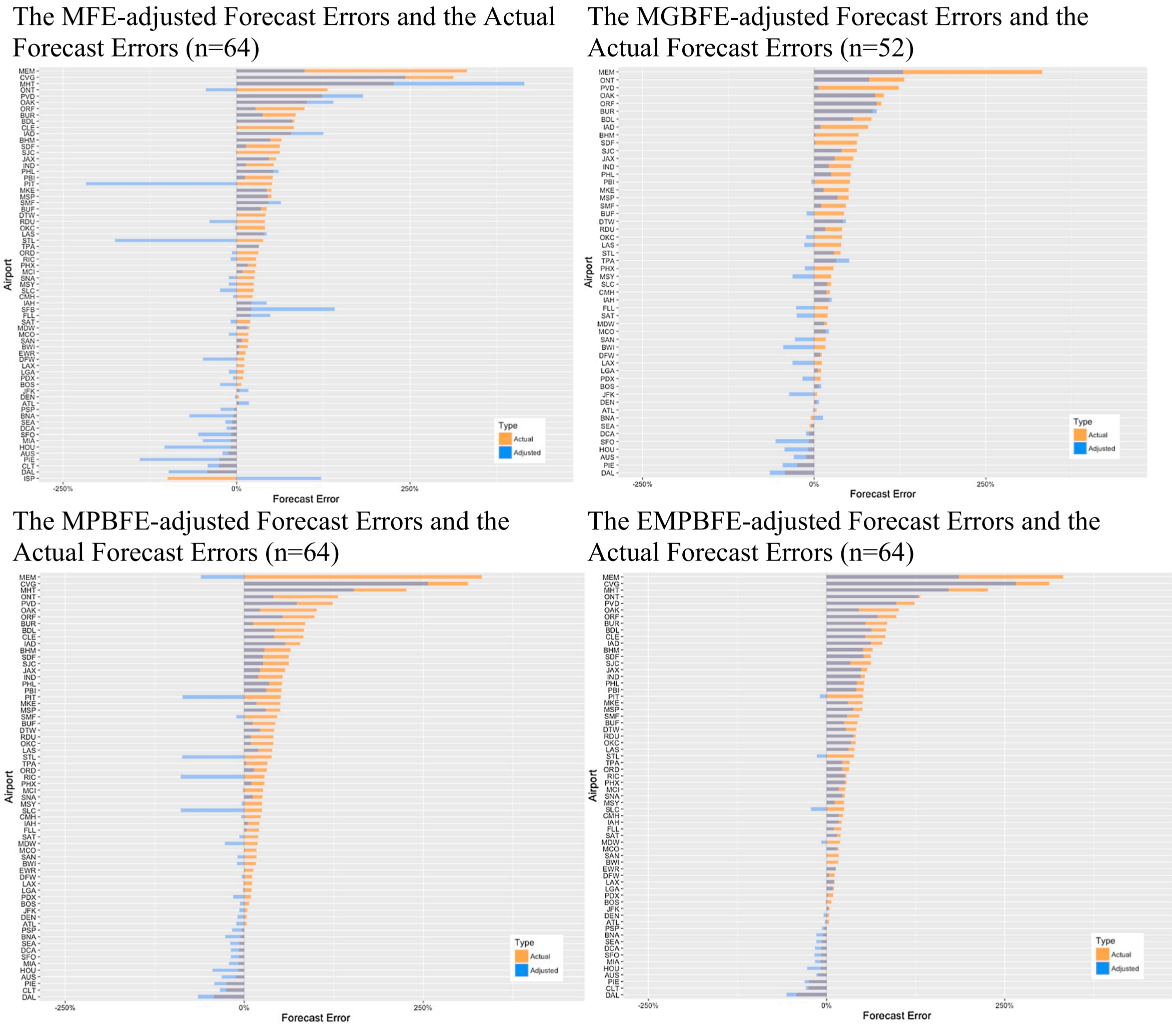

Fig. 5. Adjusted-forecast errors vs. actual errors.

After identifying and calculating $MGBFE_\alpha$ for all available historic forecasts, we adjust the current 10-year demand forecast (i.e., 2005 TAF for target year 2015) by $MGBFE_\alpha$, $\widehat{F}_{MGBFE_\alpha} = \frac{F_\alpha}{1+MGBFE_\alpha}$, where $\widehat{F}_{MGBFE_\alpha}$ is the MGBFE-adjusted 10-year demand forecast. Then we recalculate forecast error for the MGBFE-adjusted forecast, $\hat{e}_\alpha = \frac{A_\alpha - \widehat{F}_{MGBFE_\alpha}}{A_\alpha}$.

We apply this method to the top 64 airports in the top 50 MSAs. Unlike the previous method, MGBFE narrowed the number of airports down to 52 airports because some of the forecast growth percentages for the 10-year demand forecasts could not be matched to any available historic 10-year demand forecasts. That is, $MGBFE_\alpha$ for some airports could not be calculated because there was no historic demand forecast F_{ig_α} whose forecasted growth percentage was outside the range of ($g_\alpha \pm 2.5$).

As before, the mean and the median for the MGBFE-adjusted forecast errors showed significant reduction (Table 4). However, the MGBFE method also resulted in a larger proportion of the forecast errors that now underestimate (40% vs 15%). But in this case, the Mean Absolute Percentage Error (MAPE) is lower for the MGBFE-adjusted forecast errors than that for the actual forecast errors, indicating that the forecast accuracy may have improved using the MGBFE method.

The improvement is statistically significant with the paired Wilcoxon signed rank test indicating that at the 5% significance level, we can reject the null hypothesis that the median of the differences between the absolute actual errors and the absolute MGBFE-adjusted forecast errors is greater than 0 (P-value = 0.0243). In other words, the MGBFE-adjusted forecast errors are statistically significantly closer to 0 than the actual forecast errors.

This can be confirmed visually looking at the bar plots (Fig. 5, top left). The MGBFE-adjusted forecast errors (again shown in blue bars) do not contain any extreme values like the ones for the MFE-adjusted forecast errors and they are closer to zero than the actual forecast errors (shown in orange bars). The MAPE for the MGBFE-adjusted forecast errors (i.e., the average bar length for the blue bars) is 30.16, which is lower than that for the actual forecast errors (43.63).

Table 5
Variables for cluster analysis.

	Unit	Data source
Variables in base year numbers		
Passengers	Persons (millions)	FAA
Connecting passenger share	Proportion	BTS DB1B
Avg. ticket price	Dollars	BTS DB1B
HHI	Unitless	BTS T-100
Per capita income	Dollars (thousands)	BEA
Service sector employment	Persons (millions)	Census
5-year avg. annual % change up to base year		
Passengers (5AAC)	% change	FAA
Airport competition (5AAC)	% change	FAA
Connecting passenger share (5AAC)	% change	BTS DB1B
Avg. number of seats per aircraft (5AAC)	% change	BTS T-100
Avg. ticket price (5AAC)	% change	BTS DB1B
HHI (5AAC)	% change	BTS T-100
Population (5AAC)	% change	Census
Per capita income (5AAC)	% change	BEA
Service sector employment (5AAC)	% change	Census

4.3. Mean Peer-Based Forecast Errors (MPBFE)

In this method, we test the proposition that airports with similar characteristics ("peers") may produce similar forecast errors. We begin by identifying airport peers, and then incorporate the errors of these peers into our passenger prediction models.

We follow the work of Suh and Ryerson (2017) which pioneered a new method to identify airport peers. They use the k-means clustering algorithm to create groups of airports that are the most alike based on socioeconomic and airport variables. More specifically, the k-means algorithm determines the k-number of clusters and assigns membership to each of the observations on the basis of their Euclidean distance to the nearest mean. It determines, for a given set of airports, the airports that are the most similar and the optimal number of airport groups, or clusters, that contain those similar entities. Suh and Ryerson (2017) use a unique set of variables that include static socioeconomic and airport variables as well as the changes in those variables over time, to group airports both based on their current state and their history.

We begin the cluster analysis with the full set of variables shown in Table 5 with values in 2005 (2000–2005 values for 5-year annual average change or 5AAC) and select the final set of 14 variables after removing highly correlated variables (these variables are crossed out in Table 5). We cluster the 64 study airports using k-means clustering algorithm; this algorithm identifies that there are seven distinct airport groups. While Suh and Ryerson's study focuses on the meaning of each airport peer group, our study is more focused on utilizing the information of each airport's peer to improve forecasting.

For each airport, we identify its peers (airports in the same cluster) and use the available historic 10-year demand forecasts of the peer airports as the reference class. Specifically, for each airport α in cluster c ($c \in \{1, 2, \ldots, 7\}$), we calculate the Mean Peer-Group Forecast Error, $MPGFE_{\alpha} = \frac{\sum_{j=1}^{n}\left[\frac{A_{c\alpha j} - F_{c\alpha j}}{A_{c\alpha j}}\right]}{n}$, using the forecasts in the same cluster c, where $F_{c_{\alpha}j}$ is the 10-year demand forecast of any airport j in cluster c_{α} (cluster to which airport α belongs) and $A_{c_{\alpha}j}$ is the actual passenger demand for airport j in cluster c_{α}. Then, we adjust the forecast of interest by MPGFE: $\widehat{F}_{MPGFE_{\alpha}} = \frac{F_{\alpha}}{1 + MPGFE_{\alpha}}$.

We apply the MPGFE method to the top 64 airports in the top 50 MSAs. As before, the mean and the median for the MPBFE-adjusted forecast errors showed significant reduction with the median becoming close to 0 (Table 4). The MPBFE method also resulted in an even split of overestimation and underestimation (50–50%) from the actual forecast errors' positive error bias (83–17% split). In this case, the Mean Absolute Percentage Error (MAPE) is also lower for the MPGFE-adjusted forecast errors than that for the actual forecast errors, indicating that the forecast accuracy may have improved using the MPBFE method.

The improvement is tested statistically significant with the paired Wilcoxon signed rank test indicating that at the 5% significance level, we can reject the null hypothesis that the median of the differences between the absolute actual errors and the absolute MPBFE-adjusted forecast errors is greater than 0 (P-value = 0.0001).

Again, this can be confirmed visually looking at the bar plots. While the MGBFE-adjusted forecast errors (again shown in blue bars) contain a few extreme negative errors, most of them are tighter around zero. The MAPE for the MPBFE-adjusted forecast errors (i.e., the average bar length for the blue bars) is 31.05, which is lower than that for the actual forecast errors (48.48). This represents about 36% reduction in the average absolute forecast errors.

4.4. Enhanced Mean Peer-Based Forecast Errors (EMPBFE)

In this last method, we leverage the probabilistic information on the risk of a dramatic decline in passenger volumes in Section 3 in applying reference class forecasting. Specifically, we use the Mean Peer-Based Forecast Errors (MPBFE), which have shown the best improvement in forecast accuracy so far among the tested methods, and adjust them by the predicted probabilities of the decline. The rationale is that the airports with higher probability of experiencing this decline in passenger volumes can use the MPBFE more liberally than those airports with lower probabilities. In other words, this Enhanced MPBFE method (EMPBFE) may prevent over-correcting the forecasts for those airports that are not likely to see their passenger volumes decline in the next 10 years. We apply the predicted probabilities in the following way; using the model built in Section 3 and the predictors for the study airports in the base year 2005, we predict the probability of a dramatic contraction p_α for each study airport α. Then we reduce the MPBFE by this predicted probability and recalculate the EMPGE-adjusted forecasts $\widehat{F}_{EMPBFE_\alpha} = \frac{F_\alpha}{1 + MPBFE_\alpha \times (1 - p_\alpha)}$.

We apply the EMPBFE method to the top 64 airports in the top 50 MSAs. While the mean and the median for the EMPBFE-adjusted forecast errors declined compared to those for the actual forecast errors (Table 4), the magnitudes of reduction in the mean and the median are relatively small compared to the other methods. The MPBFE method, for example, reduced both the mean and the median close to zero. Likewise, the proportion of negative forecast errors (i.e., underestimations) for the EMPBFE has increased very little (17% to 28%) compared to all other methods (45%, 40%, and 50% for MFE, MGBFE, and MPBFE, respectively). The Mean Absolute Percent Error (MAPE) for the EMPBFE-adjusted forecast errors showed reduction, indicating improvement in forecast accuracy. Together, this indicates that as expected, the EMPBFE method induced relatively conservative corrections.

Again the improvement in forecast accuracy is tested statistically significant with a paired Wilcoxon signed rank test at the 5% significance level (P-value = 0.00000). The bar plots of the EMPBFE-adjusted forecast errors (Fig. 5) tell the above statistics in a very compelling manner. Unlike the other methods, the EMPBFE method produced little to no extreme negative errors. On the other hand, some of the corrections for the large positive forecast errors (the bars in the top portion of the plot) are more conservative than the other methods as well. But this represents the overall improvement in forecast accuracy as indicated by the MAPE of 36.41 for the EMPBFE-adjusted forecast errors (i.e., the average bar length for the blue bars), compared to the MAPE of 48.48 for the actual forecast errors. This represents about 25% reduction in the average absolute forecast errors.

We developed the four methods to identify a relevant reference class to implement reference class forecasting for the aviation demand forecasts and answer the following research questions:

(1) Does reference class forecasting produce statistically significant reduction in the forecast errors for the aviation demand forecasts compared to the traditional forecasts?
(2) What is the relevant and effective definition of a reference class of the forecast errors?

The summary statistics for the four reference class forecasting methodologies in Table 4 show that for three of the methods, there is a statistically significant reduction in the forecast errors. The only method that did not show a statistically significant result is also the simplest form of reference class forecasting tested in this chapter. Specifically, the Mean Forecast Error (MFE) method simply took each airport's own past forecast errors and applied them to adjust the current forecast. On one hand, the MFE method represents the purest form of learning from the airport's past but it also runs into the same question of whether an airport's own past trajectories of passenger volumes can be a good indicator of its future passenger volumes, a tenuous assumption in most current aviation demand forecasting techniques. As more nuanced approaches for identifying a relevant reference class were employed in the MGBFE, MPBFE, and EMPBFE methods, these methods all produced statistically significant and substantial amount of reduction in the forecast errors.

One of the key takeaways from this research is the validation of the peer group learning framework in demand forecasting. Among the four reference class forecasting methods proposed above (MFE, MGBFE, MPBFE, and EMPBFE), the only method that did not yield accuracy improvement happens to be the only method that does not incorporate any peer group information, namely, the Mean Forecast Error (MFE) method. This suggests that when information is limited to only airport's own past history, its demand forecasts may suffer. Leveraging information gained through peer group learning, on the other hand, can moderate any idiosyncratic demand trends for one particular airport and provide more well balanced picture of past trends that can help produce more accurate predictions.

The definition of a relevant class for the aviation demand forecast errors depends on the desired goal of error reduction. For example, if the goal of error reduction is simply to minimize the overall forecast errors, the use of the Mean Absolute Forecast Error (MAPE) can distinguish the best method. Using this criterion, it seems that the Mean Peer-Based Forecast Error (MPBFE) produced the most significant reduction in the forecast errors (-36% change in MAPE). If, on the other hand, the goal of error reduction is to minimize the forecast error without over-correcting the forecasts, the probabilistic information incorporated in the Enhanced Peer-Based Forecast Error (EPBFE) method can reduce the forecast errors without over-correcting. This type of a situation may arise if airport planners would like to ground their forecasts but they also do not want to underestimate the passenger volumes for the fear of providing inadequate amount of infrastructure based on the forecasts.

5. Discussion

In this study, we developed methodologies to address two major shortcomings of the current practice of passenger demand forecasting: first, that of predicting precipitous decline in passenger demand which forecasters currently do a poor job of anticipating and yet has a significant impact on infrastructure planning, and second, reducing optimism bias or inflationary tendency in forecasts. We developed methodologies that incorporate experiences of peer airports directly into statistical forecasting models. To do so, we used publicly available aviation demand data and census data and developed a two part methodology that: (1) predicts the probability of a severe contraction in passenger volumes in the next 10 years and (2) incorporates past forecast errors of airport peers – a reference class – into the current forecast and "ground" optimistic forecasts. Our study constitutes a first serious attempt at moving away from pinpointing passenger demand accurately and instead predicting demand contraction and applying and adapting reference classes to aviation demand forecasting.

Our approach to predicting a contraction in demand, rather than focusing on pinpointing demand exactly is novel in that we introduce a forecasting model not to predict demand with accuracy, but rather to focus on the likelihood of a severe contraction. Our approach is motivated by the fact that the most risk for developing a new runway is in the possibility of contraction. The insight our model provides carries a significant relevance to airport master planning and aviation demand forecasting processes for runway expansions, as it could help airport planners reconsider unwise investment decisions. The primary purpose of this research is to identify and understand the explanatory variables that impact the outcome of a severe contraction in passenger volumes, rather than to achieve the highest predictive performance.

The research presented herein indicates that reference class forecasting could be not only a viable option but an effective one for grounding optimistic aviation demand forecasts. The relevance and appeal of this framework in aviation demand forecasting lies in the already established culture of peer group learning in airport planning. Reference class forecasting quantifies the identification of peers and incorporates more rich past data into forecasting. It's worth noting that our research is only the starting point for incorporating reference class forecasting into airport demand estimation. The four methods of reference class identification in this research are by no means the only ways to apply reference class forecasting to aviation demand forecasting. For instance, the use of median instead of the mean of the forecast errors of a reference class, can result in much more conservative corrections on the forecast errors. Because there are several outliers in the data (i.e., extremely large forecast errors), the mean of the forecast errors of a reference class may be skewed while the median would represent a more balanced central tendency.

Indeed our research was limited in scope because of the data availability. We therefore end on a suggestion that airport managers be more focused on collecting rich datasets over multiple timescales and variables. We encourage the research community and airports to consider how more fine-grained data can be collected and incorporated into our models (which are based on yearly aggregate demand data). These data will not be difficult to collect in the future as airport managers, airlines, and related stakeholders look to transition to full connectivity through the Internet of Things (IoT) (Fischer, 2011). All the operational elements at the airport as well as the smartphones and wearables of passengers and personnel will be connected with one another, allowing for full capture of fine-grained operational data. Airport managers could build on our foundational, baseline passenger demand models based on yearly data and update the models on a monthly, daily, or more frequent basis with flight information, passenger information, concessions sales information, external environmental and economic information, and so on. These data could be used in a Bayesian approach, such that passenger demand forecasts are estimated first on the yearly data and then updated based on more frequent passenger, operational, and environmental data. This approach would allow airport managers to have a detailed understanding of their airport's demand both in the short and long term, based in history, based in the experiences of their peers, and based on the daily happenings at and around their own airport.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tre.2019.06.016.

References

ACRP Report 76, 2012. Addressing Uncertainty about Future Airport Activity Levels in Airport Decision Making. Transportation Research Board, Washington, D.C.. https://doi.org/10.17226/22704.

ACRP Synthesis 2, 2007. Airport Aviation Activity Forecasting. Transportation Research Board of the National Academies, Washington, D.C.

Alkaabi, K.A., Debbage, K.G., 2007. Air passenger demand and skilled labor markets by US metropolitan area. J. Air Transport Manage. 13 (3), 121–130. https://doi.org/10.1016/j.jairtraman.2006.11.006.

American Planning Association, 2005. Planners Called to Help End Inaccuracies in Public Project Revenue Forecasting. Retrieved from https://www.planning.org/newsreleases/2005/apr07.htm.

Ascher, W., 1979. Forecasting: An Appraisal for Policy-Makers and Planners. Johns Hopkins University Press, Baltimore, Md.

Awojobi, O., Jenkins, G.P., 2016. Managing the cost overrun risks of hydroelectric dams: An application of reference class forecasting techniques. Renew. Sustain. Energy Rev. 63, 19–32. https://doi.org/10.1016/j.rser.2016.05.006.

Bao, Y., Xiong, T., Hu, Z., 2012. Forecasting air passenger traffic by support vector machines with ensemble empirical mode decomposition and slope-based method. Discrete Dyn. Nature Soc. 2012. https://doi.org/10.1155/2012/431512.

Barrett, P., Rose, M.H., 1999. Street smarts: the politics of transportation statistics in the American City, 1900–1990. J. Urban History 25 (3). https://doi.org/10.1177/009614429902500305.

Batselier, J., Vanhoucke, M., 2017. Improving project forecast accuracy by integrating earned value management with exponential smoothing and reference class forecasting. Int. J. Project Manage. 35 (1), 28–43. https://doi.org/10.1016/j.ijproman.2016.10.003.

Bayram, S., Al-Jibouri, S., 2017. Cost forecasting using RCF: a case study for planning public building projects costs in Turkey. Int. J. Construct. Manage. 1–13. https://doi.org/10.1080/15623599.2017.1333399.

Bednarek, J.R.D., 2001. America's Airports: Airfield Development, 1918–1947, 1st ed. Texas A&M University Press.

Bel, G., Fageda, X., 2008. Getting there fast: globalization, intercontinental flights and location of headquarters. J. Econ. Geography 8 (4), 471–495. https://doi.org/10.1093/jeg/lbn017.

Brueckner, J.K., 2003. Airline traffic and urban economic development. Urban Stud. 40, 1455–1469. https://doi.org/10.1080/0042098032000094388.

Brueckner, J.K., Lee, D., Singer, E.S., 2013. Airline competition and domestic US airfares: A comprehensive reappraisal. Econ. Transp. 2 (1), 1–17.

Button, K., Lall, S., 1999. The economics of being an airport hub city. Res. Transport. Econ. 5, 75–105. https://doi.org/10.1016/S0739-8859(99)80005-5.

Corsi, T., Dresner, M., Windle, R., 1997. Air passenger forecasts: Principles and practices. J. Transport. Res. Forum 36 (2), 42–62.

Dantas, T.M., Oliveira, F.L.C., Repolho, H.M.V., 2017. Air transportation demand forecast through Bagging Holt Winters methods. J. Air Transport Manage. 59, 116–123. https://doi.org/10.1016/j.jairtraman.2016.12.006.

De Neufville, R., Odoni, A., Belobaba, P., Reynolds, T., 2013. Airport Systems, Second Edition: Planning, Design and Management, 2nd ed. McGraw-Hill Education, New York.

Dewey, O.F., Davis, D.E., 2013. Planning, politics, and urban mega-projects in developmental context: Lessons from Mexico City's Airport Controversy. J. Urban Affairs 35 (5), 531–551. https://doi.org/10.1111/juaf.12012.

FAA, 2008. Review and Approval of Aviation Forecasts. Retrieved from https://www.faa.gov/airports/planning_capacity/media/approval_local_forecasts_2008.pdf.

FAA, 2015. Airport Master Plans. Advisory Circular 150/5070-6B.

FAA, 2016. Passenger Boarding (Enplanement) and All-Cargo Data for U.S. Airports - Previous Years. Retrieved August 22, 2016, from https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/.

FAA, 2017. Terminal Area Forecast (TAF). Retrieved May 12, 2017, from https://www.faa.gov/data_research/aviation/taf/.

Fischer, E., 2011. Airport Communications: The Future is Wireless [Internet]. Airport Technology. [cited 2019 June 18]. Available from: http://www.airport-technology.com/features/feature118643/.

Flyvbjerg, B., 2008. Curbing optimism bias and strategic misrepresentation in planning: reference class forecasting in practice. Eur. Plann. Stud 16 (1), 3–21. https://doi.org/10.1080/09654310701747936.

Flyvbjerg, B., Skamris Holm, M.K., Buhl, S.L., 2005. How (In)accurate are demand forecasts in public works projects?: the case of transportation. J. Am. Plann. Assoc. 71 (2), 131–146. https://doi.org/10.1080/01944360508976688.

Frechtling, D., 2012. Forecasting Tourism Demand, 1st ed. Routledge, London. https://doi.org/10.4324/9780080494968.

Fuellhart, K., Ooms, K., Derudder, B., O'Connor, K., 2016. Patterns of US air transport across the economic unevenness of 2003–2013. J. Maps 12 (5), 1253–1257. https://doi.org/10.1080/17445647.2016.1152917.

Goetz, A.R., Szyliowicz, J.S., 1997. Revisiting transportation planning and decision making theory: The case of Denver International Airport. Transport. Res. Part A: Policy Practice 31 (4), 263–280. https://doi.org/10.1016/S0965-8564(96)00033-X.

Goetz, A.R., Vowles, T.M., 2009. The good, the bad, and the ugly: 30 years of US airline deregulation. J. Transp. Geogr. 17 (4), 251–263.

Green, R.K., 2007. Airports and economic development. Real Estate Econ. 35 (1), 91–112. https://doi.org/10.1111/j.1540-6229.2007.00183.x.

Hyndman, R.J., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22 (4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001.

Jiang, B., Liu, C., 2019. Managerial optimism in a competitive market. Prod. Operat. Manage. 28 (4), 833–846. https://doi.org/10.1111/poms.12952.

Johnson, D., Hess, S., Matthews, B., 2014. Understanding air travellers' trade-offs between connecting flights and surface access characteristics. J. Air Transport Manage. 34, 70–77. https://doi.org/10.1016/j.jairtraman.2013.08.001.

Kahneman, D., Tversky, A., 1977. Intuitive Prediction: Biases and Corrective Procedures. Defense Technical Information Center. Retrieved from http://www.dtic.mil/docs/citations/ADA047747. https://doi.org/10.1017/CBO9780511809477.031.

Kain, J.F., 1990. Deception in dallas: strategic misrepresentation in rail transit promotion and evaluation. J. Am. Plann. Assoc. 56 (2), 184–196. https://doi.org/10.1080/01944369008975758.

Ke, J., Zheng, H., Yang, H., Chen, X.M., 2017. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. Transport. Res. Part C: Emerg. Technol. 85, 591–608. https://doi.org/10.1016/j.trc.2017.10.016.

Kim, A., Ryerson, M.S., 2018. A long drive: interregional airport passenger leakage in the U.S. Tourism Manage. 65, 237–244. https://doi.org/10.1016/j.tourman.2017.10.012.

Kim, S., Shin, D.H., 2016. Forecasting short-term air passenger demand using big data from search engine queries. Autom. Constr. 70, 98–108. https://doi.org/10.1016/j.autcon.2016.06.009.

Lovallo, D., Kahneman, D., 2003. Delusions of success. Harvard Bus. Rev. 81 (7), 56–63.

Lee, D., Luengo-Prado, M.J., 2005. The impact of passenger mix on reported "Hub Premiums" in the U.S. Airline Industry. South. Econ. J. 72 (2), 372. https://doi.org/10.2307/20062116.

Li, Max Z., Suh, Daniel Y., Ryerson, Megan S., 2018. Visualizing aviation impacts: Modeling current and future flight trajectories with publicly available flight data. Transport Res. D: Tr. E. 63, 769–785.

Maldonado, J., 1990. Strategic planning: An approach to improving airport planning under uncertainty, Unpublished master's thesis Cambridge. Massachusetts Institute of Technology, MA.

Makridakis, S., Bakas, N., 2016. Forecasting and uncertainty: A survey. Risk Decis. Anal. 6 (1), 37–64. https://doi.org/10.3233/RDA-150114.

May, M., Hill, S.B., 2006. Questioning airport expansion—A case study of Canberra International Airport. J. Transp. Geogr. 14 (6), 437–450. https://doi.org/10.1016/j.jtrangeo.2005.10.004.

Moore, T.G., 1986. US airline deregulation: Its effects on passengers, capital, and labor. J. Law Econo. 29 (1), 1–28.

Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., Damas, L., 2013. On predicting the taxi-passenger demand: A real-time approach. In: Portuguese Conference on Artificial Intelligence. Springer, Berlin, Heidelberg, pp. 54–65.

Mosbah, S., Ryerson, M.S., 2016. Can US metropolitan areas use large commercial airports as tools to bolster regional economic growth? J. Plann. Lit. 31 (3), 317–333.

Parker, J., Cropper, P., Shao, L., 2012. A calibrated whole building simulation approach to assessing retrofit options for Birmingham airport. Proc. IBPSA—England.

Pickrell, D.H., 1992. A desire named streetcar fantasy and fact in rail transit planning. J. Am. Plann. Assoc. 58 (2), 158–176. https://doi.org/10.1080/01944369208975791.

Redondi, R., Malighetti, P., Paleari, S., 2012. De-hubbing of airports and their recovery patterns. J. Air Transport Manage. 18 (1), 1–4. https://doi.org/10.1016/j.jairtraman.2011.04.002.

Ryerson, M., Hansen, M., 2013. Capturing the impact of fuel price on jet aircraft operating costs with engineering and econometric models. Transport. Res. Part C: Emerg. Technol. 33, 282–296. https://doi.org/10.1016/j.trc.2011.05.015.

Ryerson, M., Kim, H., 2013. Integrating airline operational practices into passenger airline hub definition. J. Transp. Geogr. 31, 84–93. https://doi.org/10.1016/j.jtrangeo.2013.05.013.

Ryerson, M.S., 2016. Building air service sustainability: analytical approach to documenting air carrier incentive programs in airport sustainability plans. Transport. Res. Rec.: J. Transport. Res. Board 2569, 1–15. https://doi.org/10.3141/2569-01.

Ryerson, M.S., Woodburn, A., 2014. Build airport capacity or manage flight demand? How regional planners can lead American aviation into a new frontier of demand management. J. Am. Plann. Assoc. 80 (2), 138–152. https://doi.org/10.1080/01944363.2014.961949.

Suh, D., Ryerson, M.S., 2017. Frameworks for adaptive airport planning and techniques for a new era of planning. Transport. Res. Rec.: J. Transport. Res. Board 2603, 65–77. https://doi.org/10.3141/2603-07.

U.S. Department of Transportation, 2009. The Importance of Transportation Forecasting. Retrieved August 22, 2016, from https://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/subject_areas/trending_and_forecasting/workshop_for_transportation_forecasters/importance_of_transportation_forecasting/html/index.html.

Wijnen, R.A.A., Walker, W.E., Kwakkel, J.H., 2008. Decision support for airport strategic planning. Transport. Plann. Technol. 31 (1), 11–34. https://doi.org/10.1080/03081060701835670.

Wachs, M., 1989. When planners lie with numbers. J. Am. Plann. Assoc. 55 (4), 476–479.

Wadud, Z., 2011. Modeling and forecasting passenger demand for a new domestic airport with limited data. Transp. Res. Rec. 2214 (1), 59–68. https://doi.org/10.3141/2214-08.

Wei, Y., Chen, M.C., 2012. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transport. Res. Part C: Emerg. Technol. 21 (1), 148–162. https://doi.org/10.1016/j.trc.2011.06.009.

Wei, W., Hansen, M., 2005. Impact of aircraft size and seat availability on airlines' demand and market share in duopoly markets. Transport. Res. Part E: Logist. Transport. Rev. 41 (4), 315–327. https://doi.org/10.1016/j.tre.2004.06.002.

Xie, G., Wang, S., Lai, K.K., 2014. Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches. J. Air Transport Manage. 37, 20–26. https://doi.org/10.1016/j.jairtraman.2014.01.009.

Xu, S., Chan, H.K., Zhang, T., 2019. Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach. Transport. Res. Part E: Logist. Transport. Rev. 122, 169–180. https://doi.org/10.1016/j.tre.2018.12.005.