Predictive modeling of inbound demand at major European airports with Poisson and Pre-Scheduled Random Arrivals

Carlo Lancia[a,*], Guglielmo Lulli[b]

[a]Leiden University Mathematical Institute, Niels Bohrweg 1, 2333 CA, Leiden, NL
[b]Lancaster University Management School, Bailrigg, Lancaster, LA1 4YX, UK

Abstract

This paper presents an exhaustive study on the arrivals process at eight major European airports. Using inbound traffic data, we define, compare, and contrast a data-driven Poisson and Pre-Scheduled Random Arrivals (PSRA) point process on their ability to predict future demand. As part of this analysis, we show the weaknesses and somehow the difficulties of using a Poisson process to model with good approximation the arrivals stream. On the other hand, our innovative and simple data-driven PSRA model, captures and predicts with higher accuracy the main properties of the typical arrivals stream. These results have important implication on modeling and simulation-based analyses of the inbound traffic aiming to improve the use of available capacity thus reducing air traffic delays. In a nutshell, the results lead to the conclusion that, whenever possible, PSRA should be preferred over any process in the family of Poisson, as it provides more accurate air traffic predictions.

Keywords: Transportation, Air traffic, Stochastic optimization, Demand prediction, Data-driven modeling
2010 MSC: 90B06, 62P30

1. Introduction

Air congestion is a regular and persistent phenomenon in the air traffic system in both the US and Europe. Over the years, air traffic demand increased at a much faster pace compared to the increment of air-traffic-system capacity. Although in the last decade we have witnessed a mitigation of congestion phenomena –air traffic demand has not recovered yet since the 2008 economic crisis– the latest air traffic statistics published by EUROCONTROL show a significant

*Corresponding author.
Email addresses: c.lancia@math.leidenuniv.nl (Carlo Lancia), g.lulli@lancaster.ac.uk (Guglielmo Lulli)
URL: https://sites.google.com/view/clancia (Carlo Lancia)

deterioration of on-time performance in the European Civil Aviation Conference area. The average delay per flight is at its highest in the last 10 years (EUROCONTROL, 2016a). To get a good grasp of the level of congestion, 7,167 flights were canceled and 107,426 delayed in Europe between November 28 and December 27, 2016. The situation was even worse in the US, as the statistics of canceled and delayed flights showed double digits (FlightStats). However, for the sake of completeness, the number of controlled flights in the US is much larger: 15.3 million in the US versus 9.9 million flights in Europe in 2015 (EUROCONTROL & FAA, 2015).

Airports are the most relevant bottlenecks of the air traffic system. The Arrival Sequencing and Metering Area (ASMA) additional time –which is a proxy for the average arrival runway queuing time of the inbound traffic flow– during times when the airport is experiencing high demand, is an indicator of airport congestion (Cappelleras). In 2015, the average ASMA additional time at the top 30 European airports amounted to 2.27 minutes per arrival, increasing by about 18% with respect to the previous year. The ASMA performance deterioration in 2015 was largely driven by an increase in average additional ASMA time at London Gatwick, Stockholm Arlanda, Dublin, and Brussels. London Heathrow has by far the highest level of average additional ASMA time in Europe, which is almost 9 minutes per arrival, followed by London Gatwick with more than 4 minutes per flight (EUROCONTROL, 2016b). Similar situations occur in the US, although with less contrast in additional time reported across airports (EUROCONTROL & FAA, 2015). This situation occurs despite the fact that the principal airports in Western and Central Europe are treated as *fully coordinated*, meaning essentially that the number of flights that can be scheduled per hour (or other unit of time) is not allowed to exceed airport *declared capacity* (de Neufville & Odoni, 2003). In the U.S., scheduling limits are applied only to airports of the New York region, Washington Reagan, and Chicago O'Hare airport, under the High Density Rule.

Starting from the pioneering work of Blumstein (1959), airport operations has attracted the interest of the scientific community in the attempt to alleviate congestion. Many quantitative methods have been developed to understand the various causes of congestion. These methods try to ameliorate the level of congestion by detecting possible actions for improving the use of capacity and reducing delays. In particular, a great amount of work has been devoted to study the arrivals process at airports and the corresponding queues. Given the nature of the phenomenon, most of these studies rely on either queuing theory or simulation models. To estimate congestion with accuracy, models should include both *i*) fluctuations of the arrival demand over time due to hub-and-spoke operations carried out at major airports and *ii*) randomness affecting the arrivals. Koopman assumed that the statistics of arrivals follow a Poisson law, but with an arrival rate that is a strongly-varying function of time according to quantities actually observed at airports. According to Hengsbach & Odoni (1975), the assumption of Poisson arrivals for airport demand has two very appealing properties: *i*) it is mathematically tractable and is consistent with observations at major airports, and *ii*) it has been extensively used in the transportation literature.

Poisson arrivals have been assumed to study the arrival stream at several airports: J.F. Kennedy and La Guardia (Koopman, 1972), Toronto Person (Bookbinder, 1986), and Boston Logan (Hengsbach & Odoni, 1975) among others. Yet, this assumption has been corroborated only in more recent times (Willemain et al., 2004). In that paper, the authors examined data on arrivals to nine major US airports during December 2003 for evidence of exponentiality in the distribution of the Expected Time of Arrival (ETA) as estimated when the aircraft were 100 miles from their destinations under the assumption that the arrival rates changes are relatively slow over time. The analysis confirmed the nearly-exponential character of the intervals between projected ETAs. However, there are some inherent issues with Poisson arrivals. First, the Poisson process –by definition– is not capable of capturing any correlation between arrivals at consecutive time periods. This leads to an overestimation of the queue length presumably because the uncaptured correlation is negative (Caccavale et al., 2014). The overestimation of the queue length has a strong impact on the determination of control actions (decisions) to use efficiently the available capacity: models adopting Poisson processes may overestimate congestion and yield too *conservative* decisions. Second, if we model the arrival stream as a non-homogeneous Poisson process, a possibly large number of parameters has to be estimated. Third, the use of projected arrivals (Willemain et al., 2004) might not be adequate in a European context, where the arrival rates at several major European airports tend to change *rather fast*, as highlighted by the demand histograms of Figure 4 in Section 3.2.

To overcome these issues, Guadagni et al. (2011) have recently proposed to model the arrival stream at airports with a Pre-Scheduled Random Arrivals (PSRA) process, which is obtained from a deterministic schedule by superimposing Independent Identically Distributed (IID) random delays. The list of actual arrival time is the result of *mixing-up* the fixed schedule by the addition of random perturbations. The resulting process –which is known since the 60's (Kendall, 1964)– was able to provide very good fit with real data from London Heathrow airport (Caccavale et al., 2014). Further, the PSRA process is easy to study numerically, and some significant analytical results have been recently achieved by Lancia et al. (2018). Nikoleris & Hansen (2012) used PSRA to develop a single-server queuing model for trajectory-based aircraft operations that accounts explicitly for varying levels of imprecision in meeting prescribed times of arrival at airspace fixes. With the purpose of gaining insight into the generation of the observed delays and balancing congestion delays more efficiently between ground and en-route, Gwiggner & Nagaoka (2014) compared two single-server queuing models ($\cdot/D/1$ and $\cdot/G/1$ in Kendall's notation) using both Poisson and PSRA as arrival processes. From the analysis on the east-bound arrivals at Tokyo International Airport, they concluded that PSRA and a Poisson stream behave equivalently during moderate congestion but differ substantially during very high congestion. However, this comparison is indirect because based on a queue. Analysis of radar data gave in fact both arguments in favor and against the hypothesis of Poisson arrivals.

In this paper, we present data-driven models for both PSRA and Poisson and compare their performances in predicting future demand. Shifting the focus on demand-prediction accuracy offers more appropriate metrics for comparing those models in a stochastic optimization framework. An important element of novelty introduced by this work is the use of a regression model for PSRA delays ξ_i (see (3) below) instead of a parametric distribution (Ball et al., 2001; Guadagni et al., 2011; Nikoleris & Hansen, 2012). The use of a regression model allows to model flight delays and yields precise prediction of the demand. The paper introduces elements of novelty also in the definition of the Poisson process, which is learned from the data using an original combination of online change-point detection and clustering.

We study the arrival process in the period from June 15 to September 15, 2016, at some of the busiest and most congested airports in Europe, namely, London Heathrow (International Civil Aviation Organization (ICAO) code: EGLL), London Gatwick (EGKK), Frankfurt am Main (EDDF), Amsterdam Schiphol (EHAM), and Paris C. De Gaulle (LFPG). As we are also interested in the modeling of medium-intensity traffic, we include in the dataset arrivals at three other important airports: Madrid Barajas (LEMD), Rome Fiumicino (LIRF), and Athens International (LGAV).

Inter-arrival data seemingly suggest that the underlying arrival stream is Poisson. Nevertheless, using such a process to model the arrival stream with good approximation presents some weaknesses and difficulties, which we describe in detail in the following sections. On the other hand, PSRA combine a simple formulation with good predictive qualities of the inbound arrival stream. The results presented below are relevant to analyses and simulation-based studies of the air traffic system. Indeed, (fast-time) simulation is one of the most common tools used by practitioners and experts of Air Navigation Service Providers and Network Manager to detect fine-tuned control actions to improve the performances of the air traffic system and to alleviate congestion especially at airports. The approach described herein will allow more accurate analysis, and therefore better decision-making. By product, it will also contribute to the improvement of the ATM-system efficiency: even a small reduction in terms of average ASMA time can have a huge impact in terms of fuel costs, greenhouse emissions, air traffic controllers' workload and safety.

Summarizing, the contribution of this paper is three-fold: *i*) verify the results of Willemain et al. (2004) in a European context, showing at the same time that this finding does not directly support the assumption of Poisson arrivals due to serial correlations in the arrival stream; *ii*) propose an innovative data-driven approach to the modeling of the inbound stream, showing all procedural details to define a non-homogeneous Poisson process and PSRA; *iii*) compare the processes obtained this way on the task of predicting future demand.

The remainder of this paper is organized as follows. In Section 2, we describe the dataset and the data-analysis methodology used for this study. Section 3 presents the finding of the paper: in Sections 3.1 and 3.2 we study the exponentiality of the inter-arrival times and its modeling consequences; in Section 3.3, we show how to construct a time-dependent Poisson and a PSRA process in a

Table 1: Size of inbound sample for each airport considered. Observation period: from June 15 to September 15, 2016.

airport name	ICAO code	sample size
Frankfurt am Main International Airport	EDDF	58167
London Gatwick Airport	EGKK	39746
London Heathrow Airport	EGLL	56716
Amsterdam Airport Schiphol	EHAM	63279
Madrid Barajas International Airport	LEMD	48162
Charles de Gaulle International Airport	LFPG	60122
Athens International Airport	LGAV	29503
Rome Fiumicino International Airport	LIRF	43333

data-driven manner; these processes are compared in Section 3.4 on their capabilities of predicting future demand. Finally, in Sections 4 and 5 we discuss the results and provide closing comments.

2. Data and Methods

2.1. Data

Inbound flight data were extracted from EUROCONTROL's Demand Data Repository (DDR) between June 15 and September 15, 2016. The choice of the summer period aims at reducing the variability of traffic patterns, which might be caused by adverse weather conditions. Table 1 displays the total count of inbound flights in the study sample for each airport.

We queried the DDR to extract both *M1 and M3 flight plans*, respectively the last flight plan agreed with NETWORK MANAGER (EUROCONTROL) and the flight trajectory actually flown. For M1 we use the adjective *regulated* while for M3 we use *observed*. From M1 and M3 flight plans, we obtain t^{M1} and t^{M3}, i.e. the time at which the aircraft enter a cylinder of 40 NM (Nautical Mile) around the airport according to the corresponding flight plan. This procedure is in agreement with the computation of the ASMA times (Cappelleras). Indeed, the passage time at 40 NM is a proxy for the time when the flight starts the approach phase and is handed over to the Terminal Control. This time could have been measured with more accuracy by considering the instantaneous latitude and longitude of each aircraft as reported by the DDR, yet the general data analysis methodology and results illustrated hereafter remain valid. Unless explicitly stated, times are local.

Timestamps of the passage at 40 NM form a Time Series (TS) for each airport. Once the TS has been created, *inter-arrival times* are defined as the time lapse in seconds between two successive events. As the arrival rate is not constant, this TS has no fixed frequency. Since timestamps are measured as precisely as the nearest second, the inter-arrival TS generally contains ties, i.e. a set of two (or more) equal values.

2.2. Exponentiality of the inter-arrival times

We investigate evidence of exponentiality in the inter-arrivals through a QQ-plot using theoretical quantiles from the Weibull distribution

$$f_W(x; \lambda, \beta) = \frac{\beta}{\lambda} \left(\frac{x}{\lambda}\right)^{\beta-1} e^{-(\frac{x}{\lambda})^{\beta}} . \tag{1}$$

The use of a Weibull is in line with Willemain et al. (2004) and is preferable over an exponential law because the *shape parameter* β can appreciably modify the probability of observing small inter-arrivals, i.e. a large number of arrivals in a fixed interval, while the chance of observing large inter-arrivals still decays exponentially fast. The presence of ties in the sample is overcome by using the discrete version of (1) (Nakagawa & Osaki, 1975; Barbiero, 2013)

$$P_W(X = x; q, \beta) = q^{x^{\beta}} - q^{(x+1)^{\beta}} , \tag{2}$$

where q plays now the role of the *scale parameter* λ in (1). When $\beta = 1$ then (1) and (2) become respectively an exponential distribution and a geometric Probability Mass Function (PMF). QQ-plots are drawn for three different time intervals, namely, 08:00–09:30, 12:00–13:30, and 18:00–19:30, local time; these are meant to capture different operational phases of the airports considered in this study, especially for those hosting hub-and-spoke operations. We use the Kolmogorov-Smirnov test (Arnold & Emerson, 2011) to evaluate goodness of fit.

2.3. Arrival process: average demand and serial correlations

Typical characteristics of the inbound stream are assessed with an exploratory analysis of both average properties and serial correlations of the demand; the latter is a key point that can motivate the use the PSRA process. Those characteristics of the demand are investigated by aggregating the arrivals TS by intervals of ten minutes. The reason for choosing an interval length of ten minutes is two-fold. On the one hand, it is not too large to overlook changes in the regime of the underlying (stochastic) arrival process. On the other, it is not too small to capture noisy variations of the demand that would challenge the interpretability of the results.

Average demand is estimated per interval of ten minutes, yielding a daily average profile of the demand; typical fluctuations in the daily average are described by 95% point-wise confidence intervals. We look for evidence of serial correlations in the arrival stream by computing the Autocorrelation function (ACF) on the premise that the capacity of both en-route sectors (airspace) and airports impose constraints (dependencies) on the number of arrivals in consecutive time intervals. Stationarity of the arrivals TS is achieved by taking first-order differences and then checked in a 24-hour window by the augmented Dickey-Fuller test (Fuller, 2009; Seabold & Perktold, 2010). Further, we explore periodicity of the demand using a continuous wavelet transform based on the Ricker wavelet (Ryan, 1994). We conclude the descriptive analysis of the arrival stream by computing demand correlations over consecutive time intervals.

2.4. Data-driven modeling of the arrival processes

Two models for the inbound stream at airports are presented: a non-homogeneous Poisson process and PSRA. Instead of doing inference on the Poisson intensity $\lambda(t)$ or the distribution of the PSRA delays, we adopt data-driven modeling procedures described in the following subsections. The two models will be then compared and contrasted on their prediction capabilities in Section 3.4.

2.4.1. Construction of Poisson process

We approximate the intensity of the time-dependent Poisson process with a step-function. The intensities and the corresponding time intervals are computed using first the PELT algorithm (Killick et al., 2012) to detect change-points[1] in the arrival stream, and then DBSCAN (Ester et al., 1996; Pedregosa et al., 2011) to cluster change-points and estimated intensities in the (t, λ) plane; the intensity of the learned Poisson process is obtained by the centroid of those clusters. The use of PELT is appealing for its low computational cost compared to other change-point detection methods[2] (Killick et al., 2012). DBSCAN is appropriate because it works with areas of high/low density points, which is exactly the situation depicted in Figure A.2. Both behavior and performance of PELT strongly depend on the settings of the penalty function. Thus, Appendix A in the supplementary material offers a sensitivity analysis using penalties based on Akaike Information Criterion (AIC) (Akaike, 1998) (our choice) and Modified Bayesian Information Criterion (MBIC) (Zhang & Siegmund, 2007) (the R package default). The details of the procedure are described by Algorithm 1 below.

2.4.2. Construction of PSRA process

For PSRA, we simulate the process

$$t_i = t_i^{M1} + \xi_i \,, \tag{3}$$

where t_i^{M1} is the expected (aka regulated) arrival time at 40 NM, ξ_i is a delay that is generated from a regression model for $\{t_i^{M3} - t_i^{M1}\}_i$, and t_i^{M3} is the observed arrival time at 40 NM. The model for the delays ξ_i's is obtained by training an ε-support vector regression model Cristianini & Shawe-Taylor (2000) on the following features: flight origin (national, continental, or intercontinental), arrival time according to the M1 flight plan, and day of the week. The last two features are encoded as two-dimensional cyclic variables using sine/cosine transformations. An inner cross-validation with consecutive monthly split is used to search the optimal value of the parameters C and ε on a logarithmic grid: ε defines a margin of tolerance within which no penalty (C) is associated

[1] Time points where either the mean or the variance of the arrival TS undergoes a *structural* change.

[2] As a matter of fact, Functional Pruning Optimal Partitioning (Maidstone et al., 2017) might perform even better since the intensity of the process is the only parameter that is going to change under the null. However, PELT usage is established and its code well-documented.

Algorithm 1 Identification of data-driven *Poisson* process

$B \leftarrow$ time series of binned arrivals
$H_0 \leftarrow$ 'Poisson'
$T, L \leftarrow \text{PELT}(B, H_0)$
$n, C \leftarrow \text{DBSCAN}(T, L)$
$\{C$ is a list of length n that contains 2-dimensional lists of clustered t, λ-points$\}$
$\hat{T} \leftarrow [\ ]$
$\hat{L} \leftarrow [\ ]$
for $i = 1$ **to** n **do**
  $\hat{t}_k, \hat{l}_k \leftarrow$ `centroid(C[i])`
  Append $\hat{t}_k$ to $\hat{T}$
  Append $\hat{l}_k$ to $\hat{L}$
end for
$\lambda^*(t) \leftarrow \text{step}(\hat{T}, \hat{L})$ {A step function taking on value $\hat{\lambda}_i$ for $\hat{t}_i \leq t < \hat{t}_{i+1}$}
return $\lambda^*(t)$

in the training loss function. The limited number of features used and its impact on this research are discussed in Section 4. The simulation procedure is outlined in Algorithm 2.

Algorithm 2 Simulation of data-driven PSRA process

$T^{M1} \leftarrow$ sequence of *regulated* arrival times at 40 NM
$T \leftarrow$ empty list
$X \leftarrow$ matrix of features
$\Xi \leftarrow$ epsilon-support vector regression
$C, \varepsilon \leftarrow$ optimal hyper-parameters via inner cross validation
for each t in T^{M1} and each row x in X **do**
  $\xi \leftarrow \Xi(x)$
  Append $t + \xi$ to T
end for
return T

2.5. Prediction capabilities of Poisson and PSRA

PSRA and Poisson stream are compared on the capability of predicting the average weekly demand in the last week of the data set and the daily demand in the last day. An outer cross-validation with consecutive weekly splits is used to compare true demand with that predicted by model 3. Cross-validation of the Poisson model is not possible, because our formulation assumes that the intensity $\lambda(t)$ is 24-hour periodic. Evidence supporting this assumption is given in Section 3.2. Further, this assumption has the desirable property of simplifying the model formulation.

All figures and statistical analyses were produced using `Python v.3.6` and `R v.3.3` (via `rpy2`). The code used for generating the analyses is freely available at the address `https://github.com/clancia/air-traffic-data-driven-modelling/blob/master/air_traffic_analysis.ipynb`.

3. Results

In this section, we present the results of our analysis using the dataset and the methods described in Section 2.

3.1. Exponentiality of the inter-arrival times

Figures 1-3 show, for each of the eight airports and three time intervals considered, the QQ-plot of the inter-arrivals against the corresponding fitted Weibull (2). Regardless of the time interval, there is quite a good accordance between empirical and theoretical quantiles in the bulk of the distribution. This can be observed as a general flat adherence of the QQ-plot onto the 45-degrees dotted line and should be interpreted as the capability of Weibull inter-arrivals to describe small-to-moderate inter-arrival times, i.e. situations of high demand. However, the goodness of the fit deteriorates on the tails and it typically shows over-dispersion, which can be severe at EDDF, EHAM, LEMD, and LFPG. A remarkable exception is EGLL, for which the demand fluctuates around the value of 40 aircraft/hour (corresponding to an average inter-arrival of 90 seconds). Accordingly, EGLL exhibits the smallest degree of over-dispersion on the tails among the airports considered in this paper.

Table 2 reports the parameters q and β, the mean of the fitted distribution, the Kolmogorov-Smirnov D-statistic, and the p-value of the corresponding goodness-of-fit test for each time interval and airport considered. The fitted shape-parameter β is always fairly close to 1, meaning that the fitting Weibull looks like an exponential/geometrical. Very often we can reject the null hypothesis of Weibull inter-arrivals at the 1% significance level. This is due to the large size of the sample considered, which makes the test very powerful even against small deviations from the expected behavior.

3.2. Arrival process: average demand and serial correlations

Figure 4 shows the daily mean profile of the demand along its 95% pointwise confidence band. Most of the airports analyzed in this paper show the characteristic *wavy pattern* for arrivals, typical of airports hosting hub-an-spoke operations. A notable exception is EGLL, which has a fairly constant mean arrival rate and a corresponding arrival process that is characterized by a single regime: the demand oscillates around 7 arrivals every 10 minutes –the declared arrival capacity is 45 aircraft/hour. This also explains why Heathrow shows the best fit with a nearly-exponential distribution in Figures 1–3.

The correlograms in Figure 5 highlight the presence of two kinds of serial correlation in the demand. First, all airports present a statistically significant strong negative lag-1 correlation. These correlations cannot be appreciated in

Table 2: Parameters and goodness of fit of the fitted discrete Weibull (2). The parameter β is measured in seconds. The interpretation of the p-values of the goodness-of-fit test is not straightforward due to the large sample size.

time	airport	q	β	mean	D-stat.	p-value
08:00–09:30	EDDF	0.992	1.046	93.131	0.020	0.03
	EGKK	0.998	1.127	214.506	0.032	0.02
	EGLL	0.995	1.135	103.273	0.024	<0.01
	EHAM	0.993	1.118	82.209	0.033	<0.01
	LEMD	0.996	1.064	166.540	0.019	0.21
	LFPG	0.993	1.143	71.336	0.033	<0.01
	LGAV	0.996	1.000	261.533	0.021	0.40
	LIRF	0.994	1.060	129.793	0.023	0.03
12:00–13:30	EDDF	0.994	1.064	122.668	0.018	0.13
	EGKK	0.997	1.166	137.588	0.033	<0.01
	EGLL	0.995	1.158	90.262	0.026	<0.01
	EHAM	0.994	1.157	74.990	0.035	<0.01
	LEMD	0.994	1.041	133.073	0.016	0.29
	LFPG	0.996	1.119	127.154	0.027	<0.01
	LGAV	0.994	1.000	175.291	0.013	0.74
	LIRF	0.992	1.067	88.120	0.021	0.01
18:00–19:30	EDDF	0.990	1.040	82.466	0.015	0.12
	EGKK	0.997	1.167	135.505	0.034	<0.01
	EGLL	0.996	1.199	85.679	0.031	<0.01
	EHAM	0.993	1.195	61.895	0.033	<0.01
	LEMD	0.995	1.095	116.381	0.013	0.43
	LFPG	0.995	1.130	93.981	0.027	<0.01
	LGAV	0.996	1.064	189.090	0.013	0.77
	LIRF	0.994	1.072	117.987	0.016	0.21

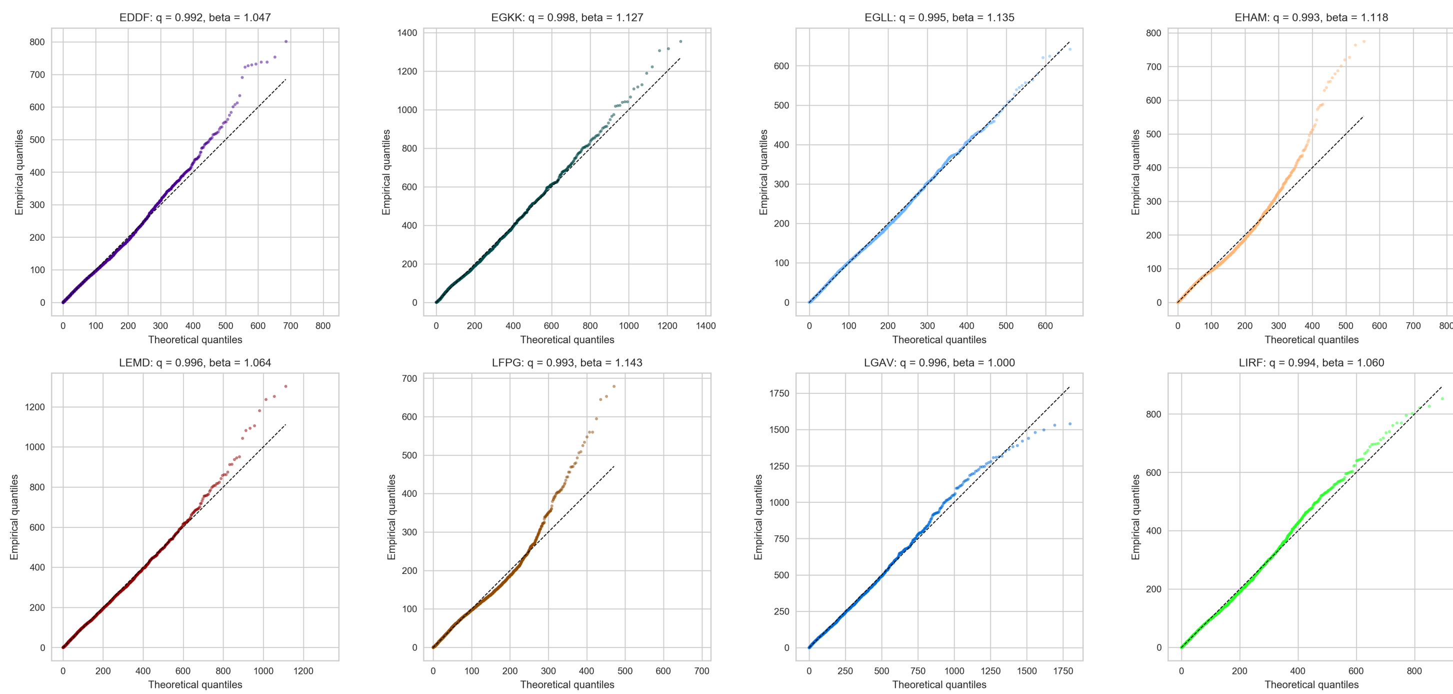

Figure 1: QQ-plot of inter-arrivals at 40 NM in the period 08:00 – 09:30. Theoretical quantiles obtained from (2); parameters listed in Table 2; times are local.

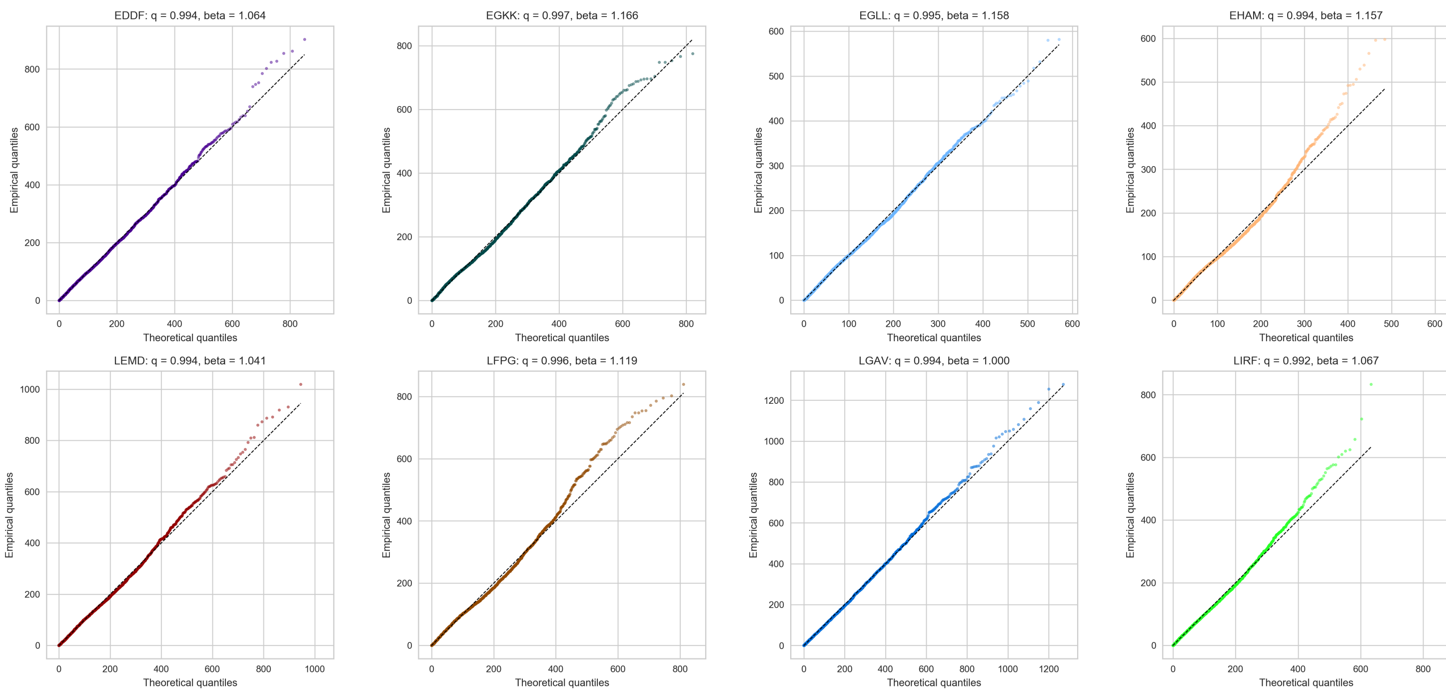

Figure 2: QQ-plot of inter-arrivals at 40 NM in the period 12:00 – 13:30 local time. Theoretical quantiles obtained from (2); parameters listed in Table 2; times are local.

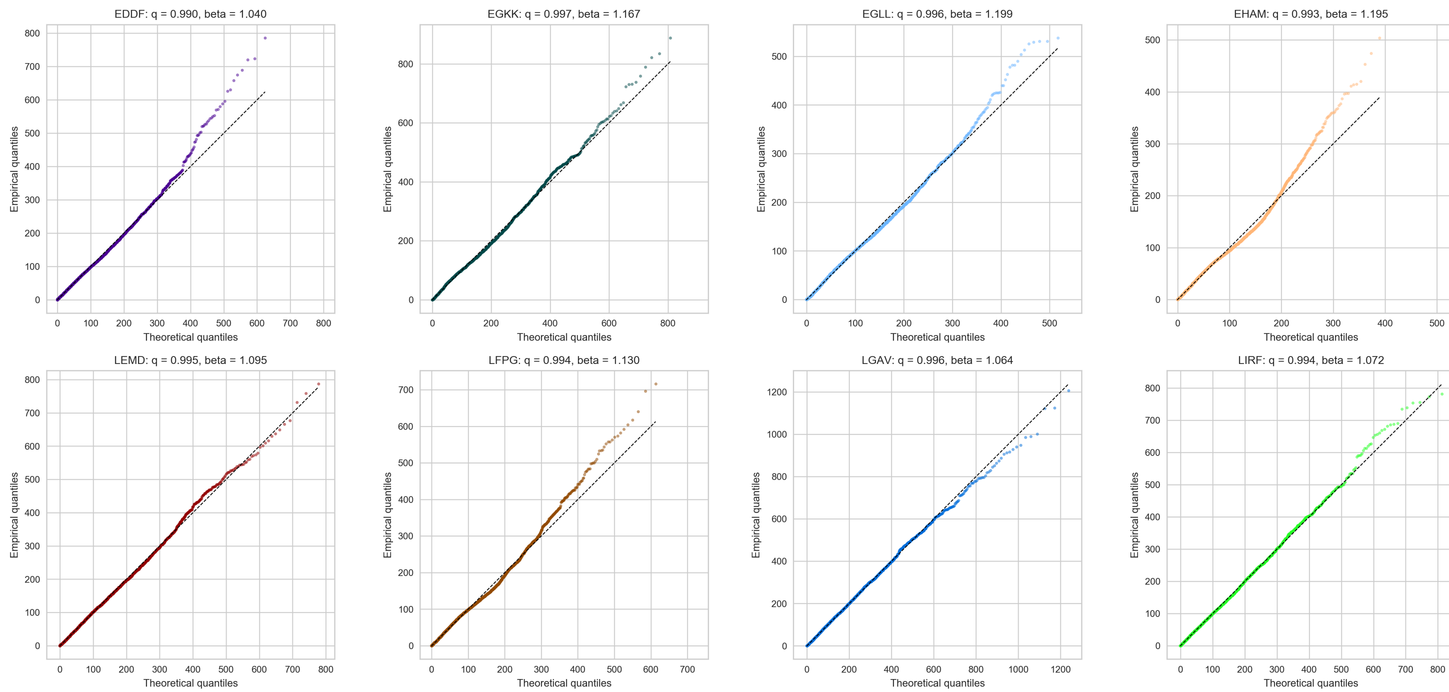

Figure 3: QQ-plot of inter-arrivals at 40 NM in the period 18:00 – 19:30 local time. Theoretical quantiles obtained from (2); parameters listed in Table 2; times are local.

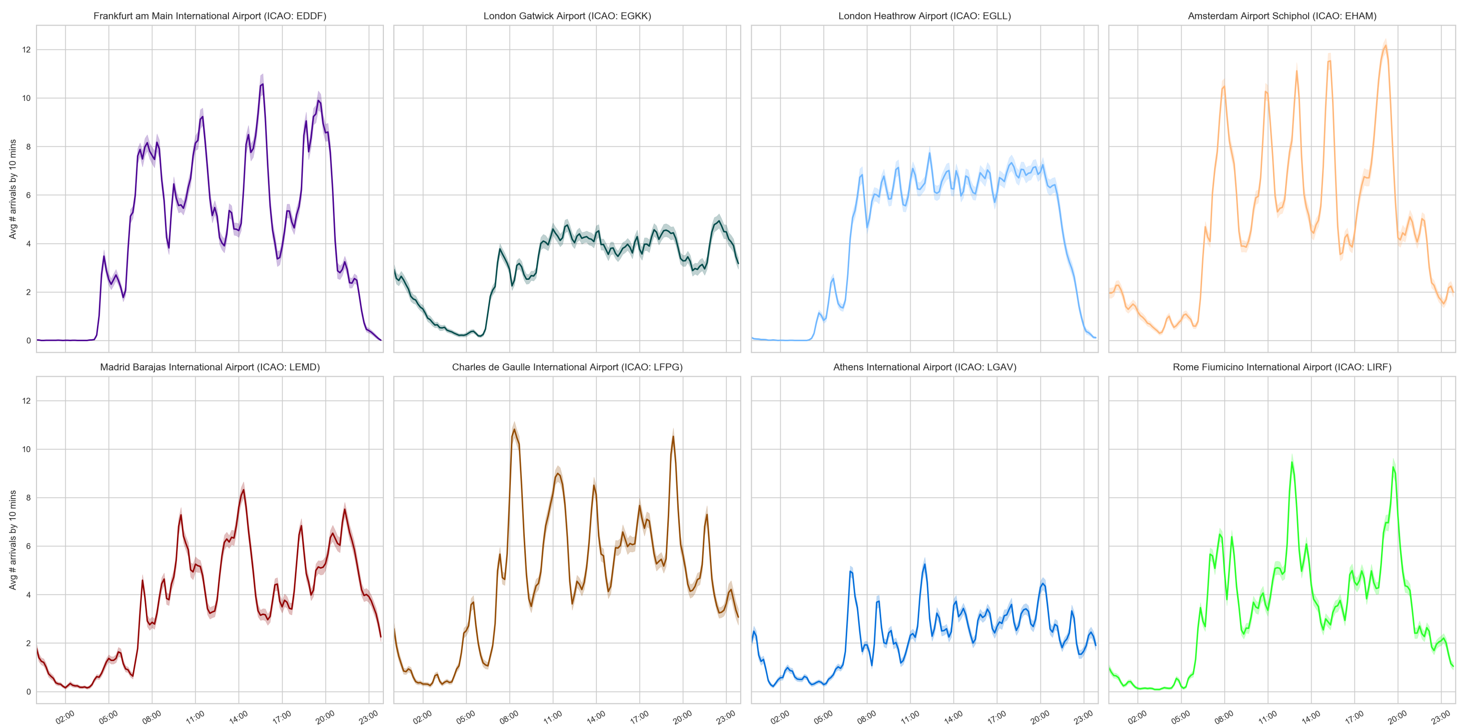

Figure 4: Average demand aggregated by 10 min. Shaded areas display 95% point-wise confidence bands.

Table 3: Values of lag-1 autocorrelations.

ICAO code	lag-1 autocorr.
EDDF	-0.447
EGKK	-0.526
EGLL	-0.440
EHAM	-0.359
LEMD	-0.466
LFPG	-0.415
LGAV	-0.479
LIRF	-0.535

full from Figure 5 because the y-axis is clipped at ± 0.2 for making other correlations more readable. For this reason, their exact numerical value is reported in Table 3. The presence of these correlations suggests that the net variation of the demand[3] over an interval of 10 minutes is negatively correlated with the demand variation in the following 10 minutes. In other words, intervals where the demand increases (resp. decreases) are more likely to be followed by intervals where the demand decreases (resp. increases). This property, which can be also observed in Figure 4, might have an interesting connection with capacity constraints. In fact, if an interval of increased demand were likely followed by another interval of increased demand, then the capacity of the airport could be temporarily exceeded.

Second, many airports show the presence of statistically significant correlations at lags of 1, 2, and 3 days. These correlations are not strong in absolute terms, yet they are the strongest shown by the correlograms, and their relatively low magnitude can be explained by the presence of natural daily variation of the demand evolution in a very large sample. Appendix B in the supplementary material offers a more-in-depth analysis of these serial correlations through a continuous wavelet transform of the demand. This analysis shows that correlations at lag of one or more days are of practical significance. Thus, we have the following:

Key Point 1. Significant correlations at lags multiple of one day motivate the idea of learning a daily-periodic Poisson process in the next section.

Figure 6 shows the Pearson's correlation $\rho_{t_i,t_{i+1}}$ between the *demand variation* in the intervals $[t_i, t_{i+1})$ and $[t_{i+1}, t_{i+2})$. The demand variation is computed as the difference between the number of arrivals observed from t_i^{M3} data and the number of arrivals that were expected according to t_i^{M1} data. The value of $\rho_{t_i,t_{i+1}}$ is mostly negative, especially during normal operations hours. The majority of these correlations are different or borderline-different from zero at a 5% significance level. This finding is somehow expected in view of the lag-1

[3]The demand TS is made stationary by taking first-order differences, see Section 2.3.

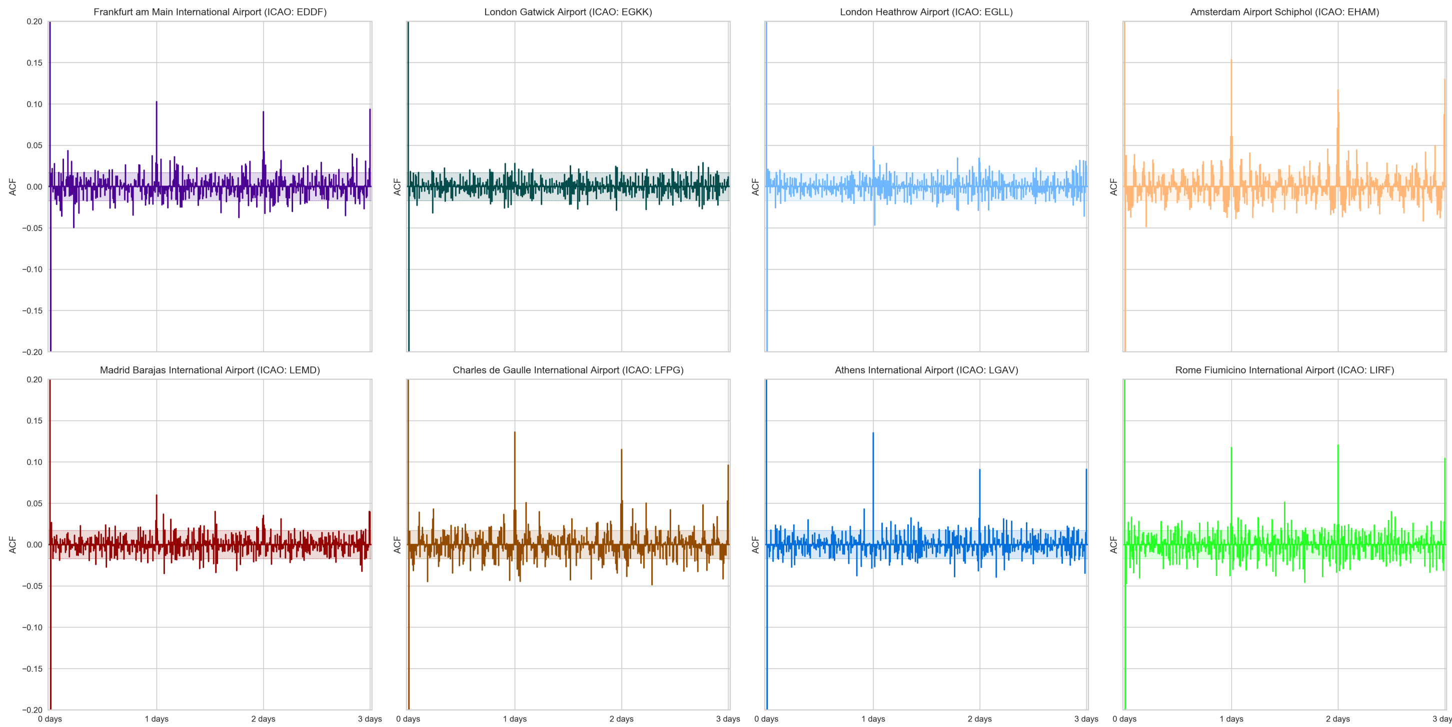

Figure 5: Autocorrelation of arrivals TS aggregated by intervals of 10 minutes. Shared areas mark the limits of statistical significance at the 99% level by Bartlett's Formula. For the sake of readability, the y-axis of the autocorrelation (1st column) is capped at ± 0.2. The values of the lag-0 and lag-1 correlations are off scale and reported by Table 3.

autocorrelations reported by Table 3.

Key Point 2. The sign of the correlations is in line with the general result that PSRA generate negatively autocorrelated streams (Guadagni et al., 2011). Please note that Guadagni et al. (2011) compute correlations from the observed inbound stream, but have a equally-spaced-in-time arrival schedule. In our formulation, pre-scheduled arrivals are not equally spaced, but come from the M1 flight plan. Accordingly, the correct quantity to consider is the demand variation.

The negative sign of correlations is a very appealing feature of the model as it captures bounds on the available capacity. If demand variations were mostly uncorrelated or even positively correlated, random fluctuations in the inbound flow might temporarily exceed the capacity of airports and/or terminal sectors. On the contrary, the M1 schedule imposes a structure in the sequence of arrivals (3). For the sake of completeness, correlations computed on simulations of 3 are shown in Appendix C of the supplementary material.

3.3. Data-driven modeling of the arrival processes

Using the data-driven methodology described in Section 2.4, we now detail the parameters characterizing the inbound-stream models, i.e. Poisson and PSRA, at the considered airport. We begin with the construction of the Poisson process, which is daily-periodic in our formulation. This modeling choice is motivated by Key Point 1 above.

3.3.1. Construction of Poisson process

Figure 7 shows the DBSCAN clustering of change-points identified by PELT and the daily average rate of arrivals at 40 NM per 10-minute intervals. For each cluster, thin dashed lines mark the average time of the day $\hat{t}_i$ and the corresponding average Poisson intensity $\hat{\lambda}_i$, where i is the index of the cluster. In view of the 24-hour periodicity of the demand highlighted in Section 3.2, we define our data-driven Poisson model by a periodic step-function, which takes on value $\hat{\lambda}_i$ for $\hat{t}_i \leq t < \hat{t}_{i+1}$. The values of $\hat{t}_i$ and $\hat{\lambda}_i$ are reported by Table D.1 in Appendix D of the supplementary material.

Remark 1. Note that the values of $\hat{\lambda}$ are substantially in line with the fitted values from Table 2, since $\lambda \approx 60 \times \text{mean}^{-1}$ in the approximation of exponential inter-arrivals.

3.3.2. Construction of PSRA process

Tables 4 and 5 show the results of a nested cross validation of PSRA process (3), where the delays ξ_i are obtained via a regression model. The model is validated on its capability of predicting the aggregated demand and not the delays ξ_i. Based on 5 consecutive two-weekly splits, the inner cross validation estimates the best parameters C and ε of the support vector machine on a logarithmic grid $C = 10^{-2}, \ldots, 10^{5}; \varepsilon = 10^{-5}, \ldots, 10^{2}$. The outer cross validation

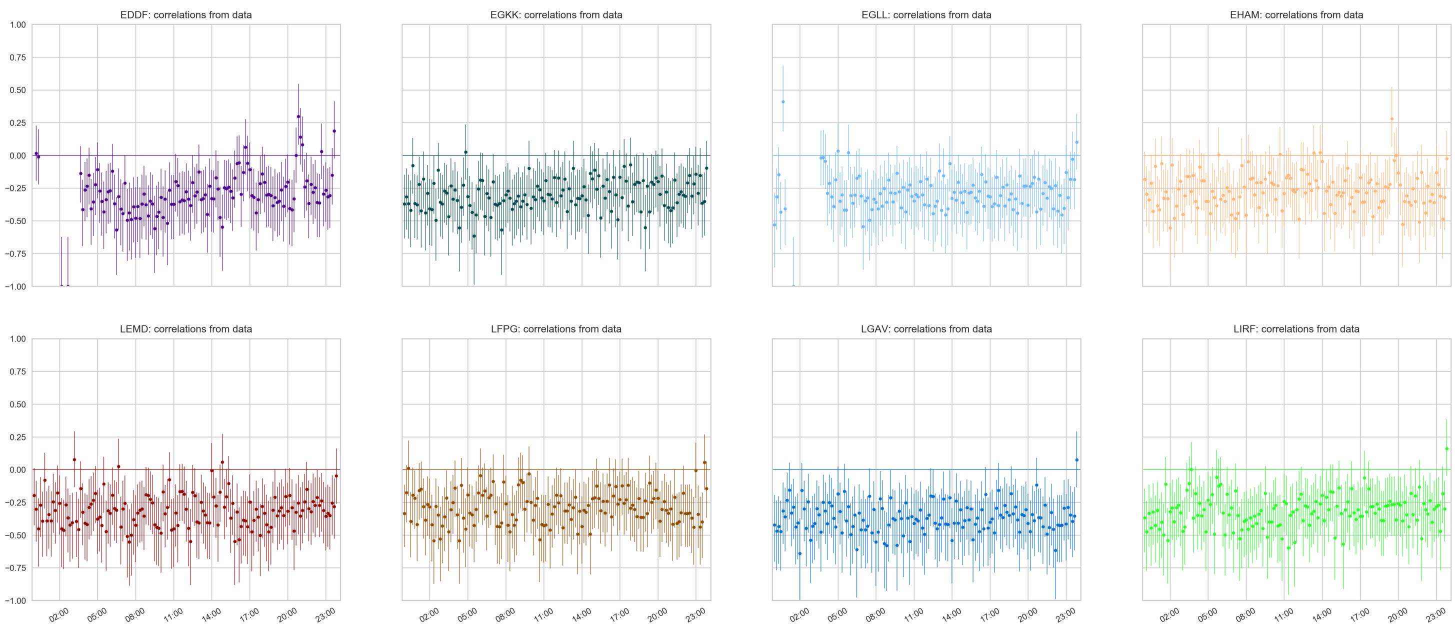

Figure 6: Correlations from t_i^{M3} data. Error bars show 95% confidence interval for Pearson's ρ.

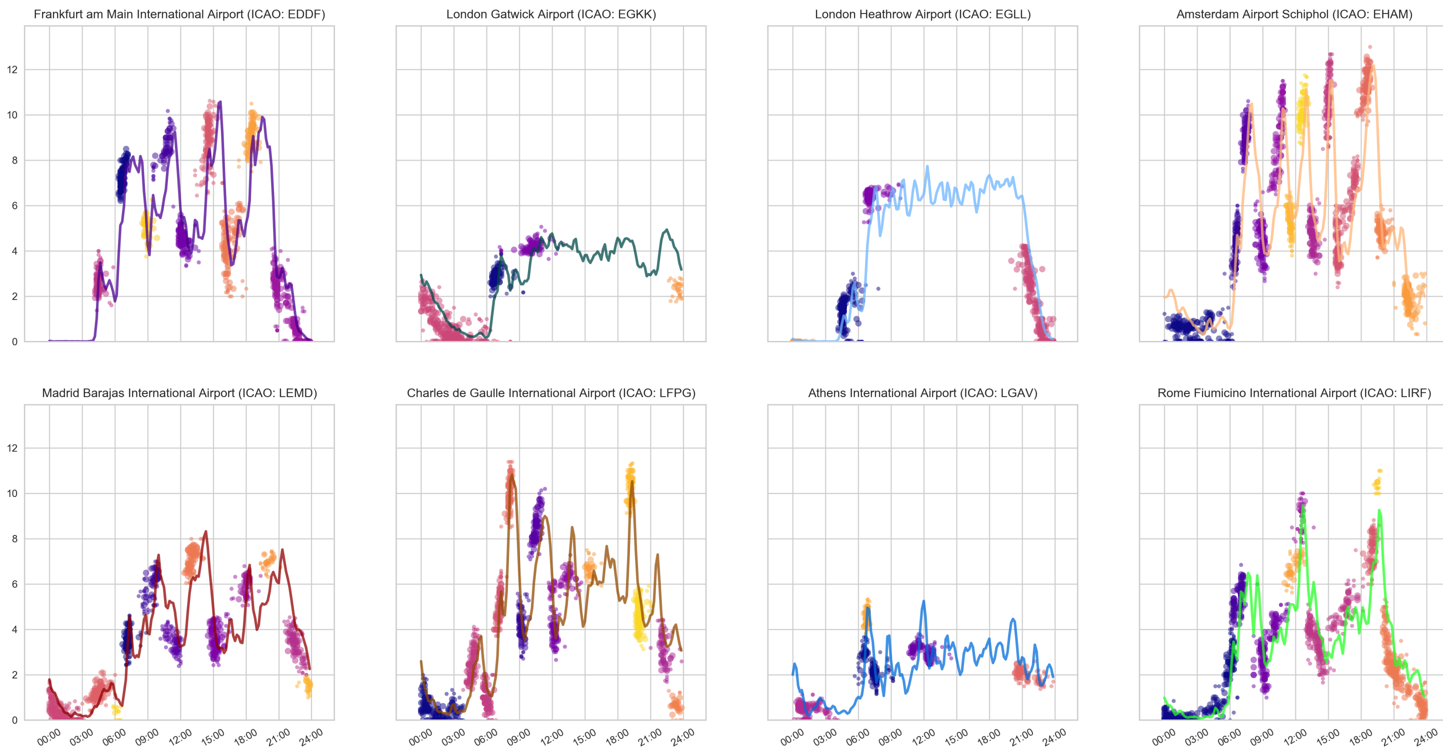

Figure 7: Data-driven modeling of non-homogeneous Poisson process. The PELT algorithms returns a sequence of couples (t_j, λ_{t_j}), where λ_{t_j} is the intensity of the Poisson process in the interval $[t_j, t_{j+1})$. The couples are successively clustered via DBSCAN (outliers are not displayed). The clusters in the plane (t, λ) are shown superimposed to the daily average arrivals at each airport.

Table 4: Inner cross-validation of PSRA model (3). The table shows the best values for the parameters of the support vector machine regression model using a sequential two-weekly split, and the corresponding average values of Mean Absolute Error (MAE), Mean Squared Error (MSE), and r^2 using 11-fold consecutive weekly splits.

airport	C	ε	MAE	MSE	r^2
EDDF	100.0	100.000	0.055680	0.330237	0.034639
EGKK	1000.0	0.001	0.024724	0.088728	0.027090
EGLL	100.0	0.100	0.042192	0.257391	0.026009
EHAM	10.0	100.000	0.037729	0.296439	0.023622
LEMD	100.0	100.000	0.042015	0.205384	0.024280
LFPG	100.0	0.010	0.025096	0.147226	0.018576
LGAV	1000.0	1.000	0.034228	0.119433	0.042989
LIRF	1000.0	1.000	0.024727	0.146581	0.026880

Table 5: Outer cross-validation of PSRA model (3). The table shows average value (μ) and standard deviation (σ) of Mean Absolute Error (MAE), Mean Squared Error (MSE), and r^2 using 11-fold consecutive weekly splits.

airport	MAE (μ)	MAE (σ)	MSE (μ)	MSE (σ)	r^2 (μ)	r^2 (σ)
EDDF	1.591	0.071	5.531	0.366	0.594	0.036
EGKK	1.436	0.043	3.911	0.204	0.126	0.069
EGLL	1.582	0.061	5.414	0.370	0.509	0.033
EHAM	1.831	0.078	6.490	0.530	0.530	0.041
LEMD	1.563	0.044	4.765	0.281	0.378	0.035
LFPG	1.700	0.059	5.477	0.323	0.500	0.033
LGAV	1.199	0.040	2.752	0.156	0.141	0.052
LIRF	1.356	0.037	3.844	0.143	0.514	0.026

evaluates demand prediction on a 11-fold cross validation based on consecutive weekly splits. The r^2 metric is often around 0.5, meaning that the model is capturing a good part of the variance of the demand, while the mean absolute error is about 1.5 aircraft/10 minutes.

Key Point 3. It is clear that even a simple regression model like the one used for the delays is performing quite well.

3.4. Prediction capabilities of Poisson and PSRA

The prediction capabilities of both Poisson and PSRA are compared by Figures 8 and 9. The first figure shows the average-demand prediction obtained for the last week included in the dataset (September 5–11, 2016), while the second figure shows the prediction of the demand for the last day (September 14, 2016). Results are plotted as difference between true and predicted demand (solid: PSRA, dotted: Poisson). Figure 8 clearly shows that PSRA are much more accurate than Poisson in predicting the average future demand. Table 7

Table 6: Scores of prediction tasks for PSRA and Poisson model. True and predicted aggregated demand, averaged in the week September 5-11, 2016, are compared using the following scores: Mean Absolute Error (MAE), Mean Squared Error (MSE), and r^2.

airport	prediction	model	MAE	MSE	r^2
EDDF	Sep 5-11	PSRA	0.451	0.398	0.964
		Poisson	1.624	5.199	0.531
EGKK	Sep 5-11	PSRA	0.320	0.187	0.920
		Poisson	0.748	0.928	0.602
EGLL	Sep 5-11	PSRA	0.395	0.355	0.957
		Poisson	0.959	1.661	0.798
EHAM	Sep 5-11	PSRA	0.629	0.875	0.921
		Poisson	1.262	3.635	0.672
LEMD	Sep 5-11	PSRA	0.555	0.593	0.894
		Poisson	0.876	1.491	0.733
LFPG	Sep 5-11	PSRA	0.520	0.516	0.932
		Poisson	1.199	2.764	0.635
LGAV	Sep 5-11	PSRA	0.289	0.143	0.906
		Poisson	0.704	0.869	0.427
LIRF	Sep 5-11	PSRA	0.354	0.252	0.956
		Poisson	1.668	6.104	-0.064

compares models by presenting Mean Absolute Error (MAE), Mean Squared Error (MSE), and r^2 score for this prediction task at each airport. A close look at Figure 9 shows that, for the prediction on a single day, the demand predicted by the Poisson model fluctuates more than PSRA around the true value. Table 7 shows indeed that PSRA still have a higher predictive accuracy also in this task, due to a smaller error and a higher r^2 score.

4. Analysis

Analysis of inter-arrival times at 40 NM showed a fair accordance between data and a Weibull distribution; this was already known in the literature for *projected* inter-arrivals at some major US airports (Willemain et al., 2004). The shape parameter of the fitted Weibull was sufficiently close to unity to seemingly justify the *classical* assumption of Poisson arrivals[4]. The null hypothesis of Weibull-distributed IID inter-arrivals could often be rejected based on the Kolmogorov-Smirnov goodness-of-fit test, yet this finding requires caution: due to the large sample size, the test is extremely powerful (especially at the most congested airports) and likely to flag even small deviations of the empirical distribution from the theoretical one. The average demand plotted in Figure 4

[4]Necessary and sufficient condition for a random process to be Poisson is that inter-arrival times are *independent* and exponentially distributed.

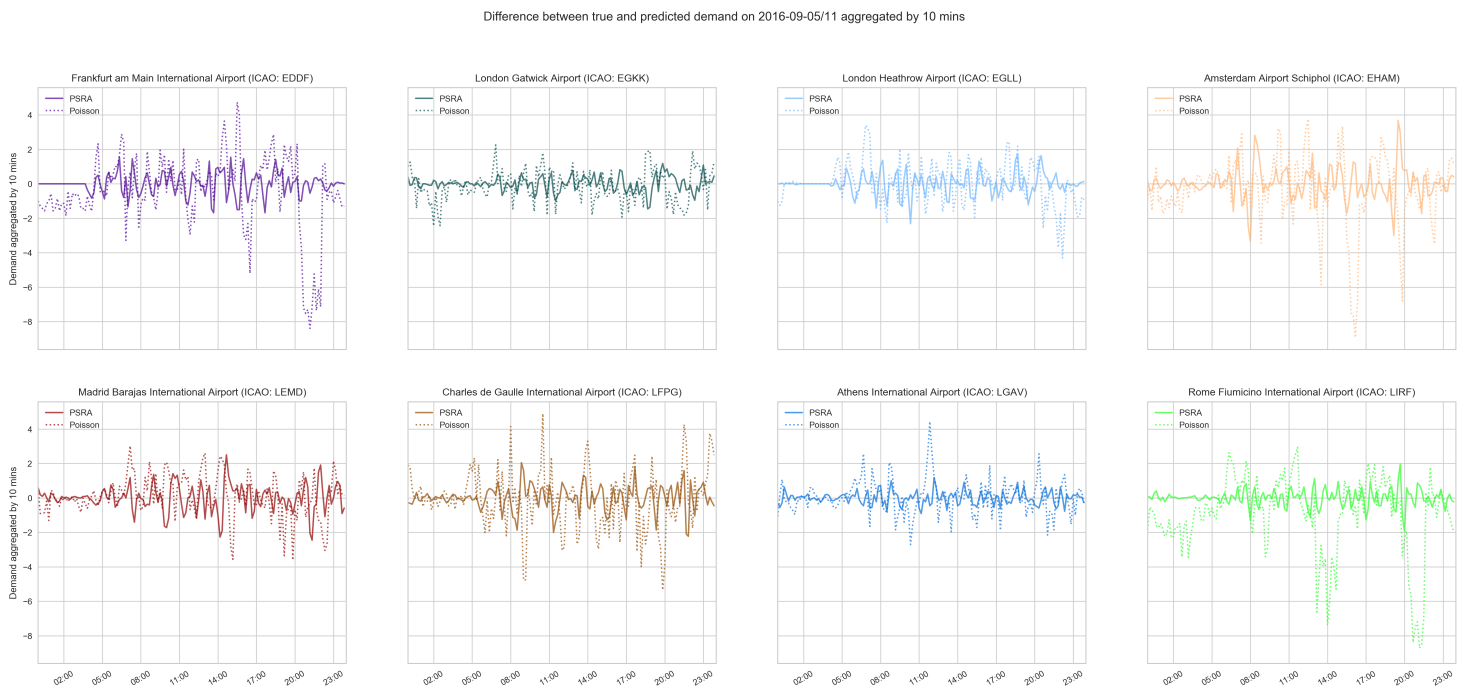

Figure 8: Difference between true and predicted average demand between September 5 and 11, 2016. A solid line shows the prediction of the PSRA while the dotted one that of Poisson process defined by Table D.1.

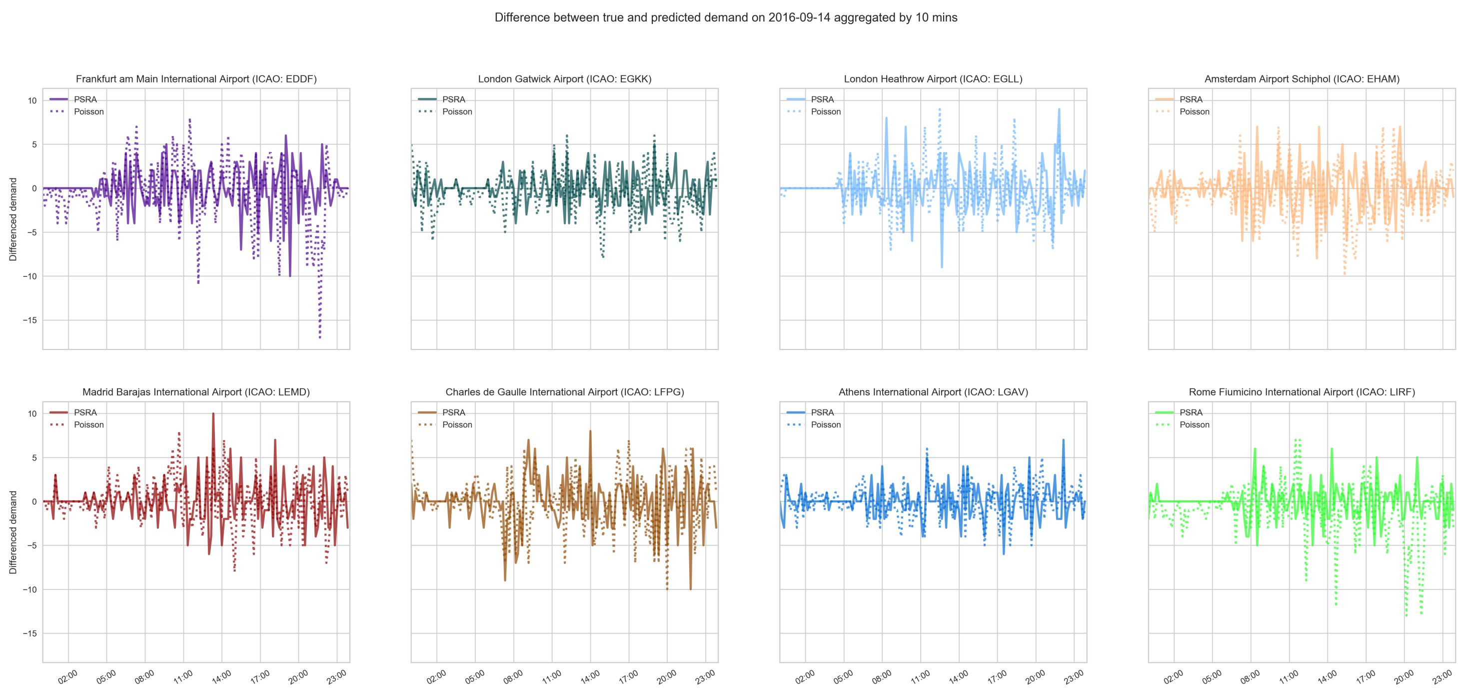

Figure 9: Difference between true and predicted demand on September 14, 2016. A solid line shows the prediction of the PSRA model while the dotted one that of Poisson process defined by Table D.1.

Table 7: Scores of prediction tasks for PSRA and Poisson model. True and predicted aggregated demand on September 14, 2016 are compared using the following scores: Mean Absolute Error (MAE), Mean Squared Error (MSE), and r^2.

airport	prediction	model	MAE	MSE	r^2
EDDF	Sep 14	PSRA	1.611	5.472	0.597
		Poisson	2.597	11.972	0.119
EGKK	Sep 14	PSRA	1.292	3.069	0.286
		Poisson	1.660	5.174	-0.203
EGLL	Sep 14	PSRA	1.646	5.979	0.427
		Poisson	1.944	8.236	0.211
EHAM	Sep 14	PSRA	1.826	6.493	0.541
		Poisson	2.375	10.347	0.268
LEMD	Sep 14	PSRA	1.618	5.688	0.332
		Poisson	2.042	8.611	-0.012
LFPG	Sep 14	PSRA	1.958	7.778	0.170
		Poisson	2.785	13.813	-0.475
LGAV	Sep 14	PSRA	1.243	3.396	-0.072
		Poisson	1.549	4.771	-0.506
LIRF	Sep 14	PSRA	1.222	3.444	0.561
		Poisson	2.222	9.097	-0.159

suggests that the goodness-of-fit might be related to the wavy variation of the demand. On the one hand, QQ-plots in Figures 1-3 show that the best fit is achieved by airports where the demand is stable over time, e.g. EGKK or EGLL. On the other hand, the selected time windows often span moments of lower and higher demand, so that a better fit might be probably obtained by letting the parameter β vary over time.

The analysis of the demand correlations in Figure 6 suggests that a process with independent increments (e.g. a time-dependent Poisson) might be not realistic anyway. Nevertheless, Poisson processes are a popular choice for modeling inbound air traffic Gwiggner & Nagaoka (2014). Thus, we proposed an innovative, data-driven approach to the modeling of a non-homogeneous Poisson process by use of PELT (a change-point detection algorithm) and DBSCAN (clustering). Poisson arrivals defined this way were contrasted with a PSRA point process, where the observed arrival time arises as the sum of the last-agreed arrival time and a random fluctuations. A parametric distribution is the typical modeling solution for the the delays ξ_i in (3), but we introduced an element of novelty by adopting also for PSRA a data-driven approach. We used a support vector regression model but other choices, like ordinary/generalized least squares or other penalized regressions, are viable options worth to be explored.

The PSRA model presented here could be hastily criticized on the limited number of features used. While it is true that a larger number of covariates could be obtained from the DDR, recent constrains imposed to the use of such repository limited the amount of data that we could use. Further, the use of meteorological conditions as a predictor of delay would arguably increase the training performance of the model, but worsen its testing metrics due to the high uncertainty associated with weather forecasts. Yet Tables 6 and 7 shows that, PSRA often score a r^2 much larger than 0 even with a small number of features, indicating that a substantial part of demand variance is already captured by the model. *On the premise that the predictive power of such formulation of PSRA could only increase if a larger number of features were used, this work validated the idea that PSRA are preferable over a Poisson model for modeling inbound air traffic.*

The predictive accuracy of PSRA over the Poisson process is demonstrated by Figures 8 and 9, where Poisson arrivals show a larger fluctuation of the predicted demand around the true value. As a consequence, using such arrival process in a queue model would lead to overestimating the congestion. As reported by Caccavale et al. (2014), such overestimation can be gross.

One could argue that the accuracy of the Poisson process could be increased by modeling the intensity of the process on a finer time scale, i.e. by estimating it in small prescribed intervals –it is well known that the Maximum Likelihood Estimator (MLE) of the parameter λ is the sample mean. Thus, a Poisson process with $\lambda(t)$ *forced* to vary every 10 minutes would be a model that exactly reproduces the daily average aggregated demand. While it is not obvious that such highly-parametric model would score much better than the one presented here, it would still fail to capture the correlation structure of the arrival data because a non-homogeneous Poisson process has independent increments regardless of the functional form of the intensity $\lambda(t)$. Conversely, our PSRA model preserves the right correlation structure of the demand by inheriting it from the regulated flight plan, see Appendix C in the supplementary material.

5. Conclusions

In this paper, we offered a thorough analysis of inbound air traffic at eight European airports. We developed two data-driven models, one in the family of Poisson arrivals and the other in the family of PSRA. We compared their capabilities of predicting future demand. In all the considered airports, the PSRA process provides better predictions on the arrival stream and, *in view of these results, we recommend to adopt, whenever possible, PSRA instead of any process in the family of Poisson arrivals.*

The proposed approach should be envisioned in simulation-based analysis of air traffic management initiatives at either strategic or planning phase. More specifically, it can be used as an engine to support and assess flight schedule development and strategic slot allocation schemes. The analysis can target either the single airport and the implications on its operations or the ATM network and its overall performances. In the latter case, the proposed model will be one of the atomic components of a larger embedded simulation model. On a longer time horizon, the proposed model can be used to evaluate long term growth initiatives such as airport expansions. Several airports around the world, like London Heathrow and Rome Fiumicino in Europe, are planning to increase their capacity. Thus, it would be advisable to have accurate studies on the possible gains in terms of available capacity and performances of the system.

The proposed model for inbound air traffic demand is also recommendable to fine-tune Traffic Management Initiatives on a shorter time scale such as Ground Delay Programs. However, in this case it might be advisable to extend the predicting models to include weather features, similarly to Liu et al. (2017) and Gopalakrishnan & Balakrishnan (2017).

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike* (pp. 199–213). Springer.

Arnold, T. B., & Emerson, J. W. (2011). Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions. *The R Journal*, *3*, 34–39. URL: http://journal.r-project.org/archive/2011-2/RJournal_2011-2_Arnold+Emerson.pdf.

Ball, M., Vossen, T., & Hoffman, R. (2001). Analysis of demand uncertainty effects in ground delay programs. In *4th USA/Europe Air Traffic Management R&D Seminar* (pp. 51–60).

Barbiero, A. (2013). *DiscreteWeibull: Discrete Weibull distribution*. URL: http://CRAN.R-project.org/package=DiscreteWeibull r package version 1.0.

Blumstein, A. (1959). The landing capacity of a runway. *Operations Research*, *7*, 752–763.

Bookbinder, J. H. (1986). Multiple Queues of Aircraft Under Time-Dependent Conditions. *INFOR*, *24*, 280–288.

Caccavale, M. V., Iovanella, A., Lancia, C., Lulli, G., & Scoppola, B. (2014). A model of inbound air traffic: The application to Heathrow airport. *Journal of Air Transport Management*, *34*, 116–122.

Cappelleras, L. (). Performance Indicator – Additional ASMA Time. Available at http://ansperformance.eu/references/methodology/additional_asma_time_pi.html.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X. et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (pp. 226–231). volume 96.

EUROCONTROL (2016a). *CODA DIGEST Q2 2016*. Technical Report Eurocontrol. Available at http://www.eurocontrol.int/sites/default/files/content/documents/official-documents/facts-and-figures/coda-reports/digest-q2-2016-final.pdf.

EUROCONTROL (2016b). *Performance Review Report on European Air Traffic Management Performance in 2015*. Technical Report Eurocontrol. Available at www.eurocontrol.int/documents/performance-review-report-european-air-traffic-management-performance-2015.

EUROCONTROL, & FAA (2015). *Comparison of Air Traffic Management-Related Operational Performance: U.S./Europe*. Technical Report Federal Aviation Administration Available at `https://www.faa.gov/air_traffic/publications/media/us_eu_comparison_2015.pdf`.

FlightStats (2017). `www.flightstats.com` [Accessed: June, 2017].

Fuller, W. A. (2009). *Introduction to statistical time series* volume 428. John Wiley & Sons.

Gopalakrishnan, K., & Balakrishnan, H. (2017). A comparative analysis of models for predicting delays in air traffic networks. In *12th USA/Europe Air Traffic Management R&D Seminar* (pp. 1–10).

Guadagni, G., Ndreca, S., & Scoppola, B. (2011). Queueing systems with pre-scheduled random arrivals. *Mathematical Methods of Operations Research*, *73*, 1–18.

Gwiggner, C., & Nagaoka, S. (2014). Data and queueing analysis of a japanese air-traffic flow. *European Journal of Operational Research*, *235*, 265–275.

Hengsbach, G., & Odoni, A. R. (1975). *Time dependent estimates of delays and delay costs at major airports*. Technical Report Cambridge, Mass.: MIT, Dept. of Aeronautics & Astronautics, Flight Transportation Laboratory, 1975.

Kendall, D. G. (1964). Some recent work and further problems in the theory of queues. *Theory of Probability & Its Applications*, *9*, 1–13.

Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, *107*, 1590–1598.

Koopman, B. O. (1972). Air-Terminal Queues under Time-Dependent Conditions. *Operations Research*, (pp. 1089–1114).

Lancia, C., Guadagni, G., Ndreca, S., & Scoppola, B. (2018). Asymptotics for the late arrivals problem. *Mathematical Methods of Operations Research*, . URL: `https://doi.org/10.1007/s00186-018-0643-3`. doi:`10.1007/s00186-018-0643-3`.

Liu, Y., M. Hansen, D. Z., Liu, Y., & Pozdnukhov, A. (2017). Modeling ground delay program incidence using convective and local weather information. In *12th USA/Europe Air Traffic Management R&D Seminar* (pp. 1–10).

Maidstone, R., Hocking, T., Rigaill, G., & Fearnhead, P. (2017). On optimal multiple changepoint algorithms for large data. *Statistics and Computing*, *27*, 519–533. URL: `https://doi.org/10.1007/s11222-016-9636-3`. doi:`10.1007/s11222-016-9636-3`.

Nakagawa, T., & Osaki, S. (1975). The discrete Weibull distribution. *IEEE Transactions on Reliability*, *24*, 300–301.

de Neufville, R., & Odoni, A. (2003). *Airport Systems: Planning, Design and Management*. McGraw-Hill.

Nikoleris, T., & Hansen, M. (2012). Queueing models for trajectory-based aircraft operations. *Transportation Science*, *46*, 501–511.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Ryan, H. (1994). Ricker, Ormsby, Klander, Butterworth – A Choice of Wavelets. Available at `http://74.3.176.63/publications/recorder/1994/09sep/sep94-choice-of-wavelets.pdf`.

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference* (p. 61). volume 57.

Willemain, T. R., Fan, H., & Ma, H. (2004). Statistical analysis of intervals between projected airport arrivals. *Rensselaer Polytechnic Inst., DSES Tech. Rept*, (pp. 38–04).

Zhang, N. R., & Siegmund, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, *63*, 22–32.