



**HAL**  
open science

## A Year Into the Pandemic: a Passenger Perspective on its Impact at Paris-Charles de Gaulle Airport

Clara Buire, Geoffrey Scozzaro, Aude Marzuoli, Eric Feron, Daniel Delahaye

► **To cite this version:**

Clara Buire, Geoffrey Scozzaro, Aude Marzuoli, Eric Feron, Daniel Delahaye. A Year Into the Pandemic: a Passenger Perspective on its Impact at Paris-Charles de Gaulle Airport. 2021 IEEE International Conference on Big Data' au sein du workshop 'Applications of Big Data in the Transport Industr, Dec 2021, Virtual, France. hal-03474780

**HAL Id: hal-03474780**

**<https://enac.hal.science/hal-03474780>**

Submitted on 10 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Year Into the Pandemic: a Passenger Perspective on its Impact at Paris-Charles de Gaulle Airport

Clara Buire \*  
*Ecole Nationale de l'Aviation Civile*  
Toulouse, France  
clara.buire@enac.fr

Geoffrey Scozzaro \*  
*Ecole Nationale de l'Aviation Civile*  
Toulouse, France  
geoffrey.scozzaro@enac.fr

Aude Marzuoli  
*Georgia Tech*  
Atlanta, United States  
amarzuoli3@gatech.edu

Eric Feron  
*King Abdullah University of Science and Technology*  
Thuwal, Saudi Arabia  
eric.feron@kaust.edu.sa

Daniel Delahaye  
*Ecole Nationale de l'Aviation Civile*  
Toulouse, France  
daniel.delahaye@enac.fr

**Abstract**—The COVID-19 pandemic has profoundly affected the air transportation system, its structure, its reliability, and its dynamics. Passengers have in turn significantly adapted their behavior. Through a case study at Paris-Charles de Gaulle airport, the present paper examines the new traffic network, the fact that delays remain high despite a drop in flight volume, the significant decrease in aircraft load factors and the change in passenger behavior at the airport.

**Keywords**—Air Transportation, Big Data, COVID-19, Air passengers

## I. INTRODUCTION

The COVID-19 pandemic has jeopardized the air transportation system with a dramatic reduction in passenger volume, attributable to both officially imposed travel limitations and public fear of the virus spread. According to the International Civil Aviation Organisation (ICAO) [1], the air transportation system suffers from a 60% decline in worldwide passenger volume in 2020. The International Air Transport Association (IATA) [2] expects a long-term loss of two years growth. Airlines have had to drastically adapt their offers to survive the economic crisis. Many industry predictions do not foresee the recovery of 2019 passenger volume before 2024 at the earliest [1]. The preliminary estimates announced an overall reduction of 39% to 40% of seats offered by airlines in 2021 compared to 2019 levels [1]. While the pandemic is still going on, this paper provides an account of air transportation operations, especially from a passenger perspective, since January 2020. In 2021, a moderate recovery of domestic travel has begun. International travel remains stagnant. Throughout the dynamic changes in travel recommendations, airports have overhauled their security processes and operations to accommodate new public health measures. The massive drop in passenger volume has led many airports to adopt a new configuration. For instance, at Paris-Charles de Gaulle airport (CDG), only one terminal (terminal 2E and 2F) has been operating since the

beginning of the pandemic compared with three terminals during normal operations.

The availability of massive amounts of data is bringing multiple ways to examine the air transportation system during the COVID-19 pandemic. Recent literature in the past year has initially focused on the structural modifications of the air traffic network. Sun et al. [3] provided an account of the early impact of the pandemic on network structure until May 2020, showing that each airport lost half of its connections on average. European airports appear to have been hit the hardest when compared with US and Chinese airports. Albers et al. [4] studied the responses of airlines to the COVID-19 outbreak until May 2020, highlighting their heterogeneity. Adrienne et al. [5] measured the consequences of massive aircraft grounding and the ensuing operational challenges at airports. Schultz et al. [6] estimated the increase of aircraft turn around time due to post-pandemic requirements. Mueller [7] quantified the change in spatial connectivity in Europe until June 2021.

Perhaps all operators now wonder how passengers can be convinced to fly again, hence the need to investigate COVID-19 repercussions on air passenger behavior. Iacus et al. [8] from the European Commission were amongst the first to provide projections of air passenger traffic in May 2020, stressing the impact of aviation contraction on the growth domestic product and jobs worldwide. Kim et al. [9] obtained passenger data in South Korea and they observed that domestic and international demand were responding differently to the pandemic. Lamb et al. [10] surveyed both business and leisure travelers to determine the factors that influenced their propensity to fly during the pandemic. Budd et al. [11] leveraged traditional surveys to categorize both passengers in terms of their propensity to fly and the safety measures in terms of their impact on restoring public confidence in air travel. Monmousseau et al. [12] analyzed the impact of travel restrictions on relations between air passengers and airlines in the US through the use of Twitter and Bureau of Transportation

\* : these authors equally contributed to this paper

and Statistics (BTS) data.

The present paper aims at providing passenger-centric views of pandemic impact on the air transportation system, focusing on passengers departing from, arriving at, or transiting through CDG. These views include the changes of easily reachable origin and destination airports, the evolution of flight delays, the evolution of load factors and passenger perception of airport departure processing time. Two large data sets collected before and during the pandemic have been used to synthesize the observation presented below. The first data set consist of all flight traffic, including actual passenger volume per flight. The second data set consists of time stamps of all passenger scanning operations within CDG. Among the main results, the number of airports on average daily connected to CDG was reduced by 40%. Over a time period when traffic volume was reduced by 75%, the average flight delay increased. Aircraft load factors dropped by 20% during the pandemic. Finally, the share of connecting passengers at CDG for European flights increased by 8% during the pandemic, which might reflect a consolidation of the hub and spoke model used by airlines.

This paper is organized as follows: Section II details the contents of the data sets supporting the analysis. Section III evaluates the network and flight delays changes during COVID-19 pandemic. Section IV quantifies the decrease of aircraft load factors. Section V characterizes passenger arrivals at security screenings and connecting passenger volumes before and during the pandemic. Section VI concludes the paper and provides threads for further investigations.

## II. DATA DESCRIPTION

Two data sets supported the analysis reported in this paper. The first data set contains flight data in and out of CDG airport collected from January 2017 to August 2021, totaling over 1.7 million records. An example of a record is presented in Table I.

Flight characteristics	Values
Estimated in/off block time	2019/06/12 06:10
Actual in/off block time	2019/06/12 05:47
Flight number	AF748
Terminal	T2E
Boarding area	S4
Airline IATA code	AF
Airport IATA code	CKY
Movement type	Arrival
Aircraft registration	FGLZ0
Passenger volume carried	242

TABLE I: Example of one row of the air traffic data set

Each record characterizes a flight operated at CDG: scheduled departure or arrival time at the gate, actual departure or arrival time at the gate, flight number, terminal, boarding area, IATA airline code, destination or origin airport, movement type, aircraft registration and passenger volume carried.

The second data set is composed of passenger time stamps at the airport collected during June 2019 and June 2020, totaling over to 1.6 million records. An example of one record of passenger scans is displayed in Table II.

Row characteristics	Values
Date	2019/06/15
10 min interval	04:10 - 04:20
Flight number	AF1066
Terminal	Terminal 2E
Boarding area	C2F-F2
Position type	PIF
Passenger volume tracked	3

TABLE II: Example of one row of the passenger scan data set. For the flight AF1066 on June 15th, three passengers exited the T2E security screening between 04:10 and 04:20.

Each record provides the date, a ten minutes interval, a flight number, terminal, boarding area, position type of the operator performing the scan, and the volume of passengers tracked. The different position types include check-in, security screening entrance or exit and boarding area.

Position type	Description	Passenger volume
ACC-PAF	Customs entrance	285868
ACC-PIF	Security screening entrance	481927
CAB	Check-in	355471
SHOP	Shop	2595
PAF	Customs exit	9318
PIF	Security screening exit	1707071
PTE	Boarding gate	272173
PTE-AUTO	Automatic boarding gate	24317

TABLE III: Passenger volume tracked at each position aggregated on June 2019 and 2020

Upon closer examination of the recorded volume at each position, see Table III, the most represented position is the one denoted by “PIF” corresponding to the security screening exit. At that position, the outbound passenger volume is at least three times higher than at any others. According to CDG operators, it provides the most reliable estimate for outbound passengers.

## III. NETWORK EVOLUTION

### A. Evolution of the network structure

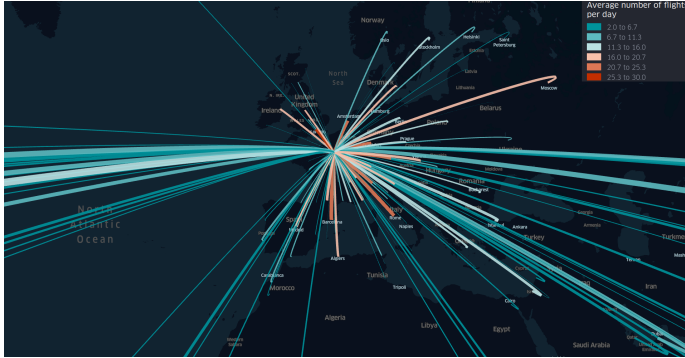
In this section, the evolution of CDG’s integration within the air transportation network is investigated. The COVID-19 crisis has led to a decrease in traffic volume and a shrinkage in the number of destinations reached, due to border closures and travel restrictions. For the rest of the study, the set of airports connected to CDG is denoted by  $\mathcal{A}$  and the set of days of the year  $y$  is denoted by  $\mathcal{D}_y$ . The average number of flights per day over a year  $y$  between CDG and an airport  $a_i \in \mathcal{A}$  is defined by

$$\forall a_i \in \mathcal{A}, n_y^{a_i} = \frac{\sum_{d \in \mathcal{D}_y} n_d^{a_i}}{\text{card}(\mathcal{D}_y)},$$

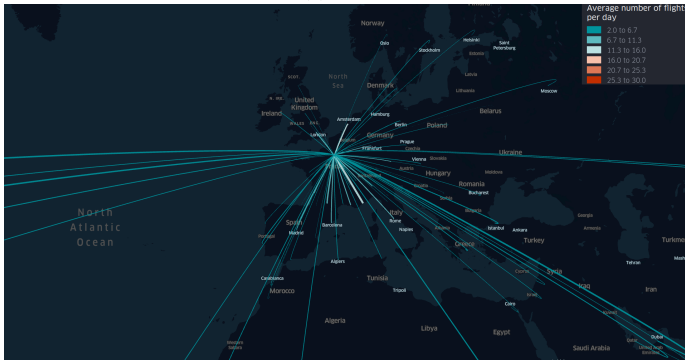
where  $n_d^{a_i}$  is the number of flights between CDG and airport  $a_i$  on day  $d$ . Similarly, the average number of passengers carried per day  $d$  during a given year  $y$  between CDG and an airport  $a_i \in \mathcal{A}$  is computed as

$$\forall a_i \in \mathcal{A}, q_y^{a_i} = \frac{\sum_{d \in \mathcal{D}_y} q_d^{a_i}}{\text{card}(\mathcal{D}_y)},$$

where  $q_d^{a_i}$  the number of passengers carried between CDG and airport  $a_i$  on day  $d$ . For each airport connected to CDG, these values are computed for 2019 and 2020 are displayed on Figures 1a and 1b respectively.



(a) 2019



(b) 2020

Fig. 1: Network overview for the years 2019 (a) and 2020 (b). Each arc corresponds to a connection between CDG and an airport  $a_i$ . The color of arcs is based on the average number of flights (a red arc indicates a high flight volume while a blue one corresponds to a lower volume). The stroke denotes the passenger volume carried (the wider the arc, the higher the number of passengers carried). A filter has been applied to remove arcs with less than one round-trip per day on average.

First, one can notice that CDG is highly connected to European airports in 2019 as shown in Figure 1a. Also, since CDG is a hub located in France, many feeder flights are operated between CDG and other French airports. Transatlantic arcs show a relatively low average number of flights per day. However, the number of passengers carried is amongst the highest, because long-haul flights tend to be operated by larger aircraft. On the contrary, some destinations such as Moscow have a higher frequency, but the number of passengers carried is lower. Thus, these flights are very likely to be operated by smaller aircraft. Regarding the post-COVID network starting March 2020, both flight and passenger volumes decrease drastically. The disappearance of arcs shown in Figure 1b indicates destinations for which, on average, less than a round-trip per day is operated. This is the case, for instance, with Boston, Oslo or Rio de Janeiro airports. Table IV presents

the evolution of the number of airports directly connected to CDG.

	2019	2020	2021
Number of airports with on average at least a daily connection with CDG	218	83	138

TABLE IV: Number of airports directly connected with CDG from March 15th to August 30th on 2019, 2020 and 2021. This represents the number of airports with at least one connection with CDG on average per day.

The number of direct links with at least one flight per day on average is reduced by three from 2019 to 2020. However, one can notice a reversal for the year 2021 which indicates that the air transportation system has cautiously started its recovery. Flights frequency at CDG has been increasing and stands above half its 2019 level during 2021.

Figure 2 displays the number of departing flights operated at CDG and the daily number of deaths in France due to COVID-19 on 2020.

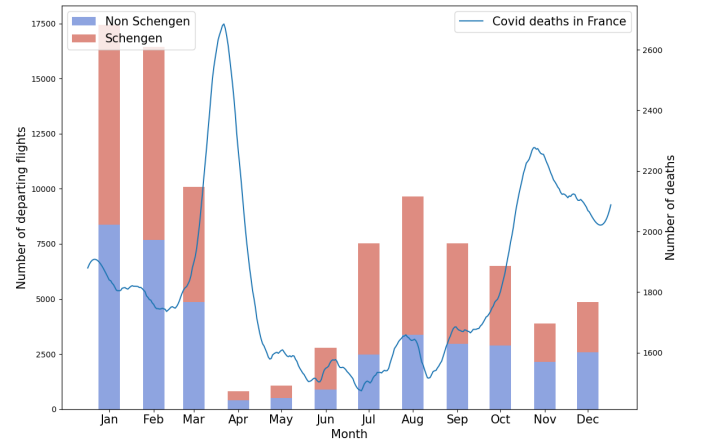


Fig. 2: Barplot of departing flights from CDG airport for 2020, stacked by destination : Schengen space/non-Schengen space. The amount of deaths by day due to COVID-19 in France is also displayed.

First, lower traffic volumes appear to occur when higher deaths are reported. This reflects travel restrictions enforced by governments in response to the spreading of the virus. A distinction is made between flights within Schengen area (named Schengen flights) and flights with an origin or a destination out of Schengen area (named non-Schengen flights). The Schengen area is a group of 26 European countries that abolished border controls at their mutual borders. Before the COVID-19 crisis on March 2020, the number of flights operated within and out of Schengen space are approximately the same. However, due to travel restrictions, these numbers changed during summer 2020. Indeed, the European union closed its borders to other countries. Hence, the share of Schengen flights at CDG increased. This trend slowly decreases by the end of the year to return to an equilibrium.

For a closer look at impacted destinations, the connections

at CDG between a set of 9 French airports and 25 foreign airports, which were the most connected with CDG pre-pandemic, are studied. For each, the daily average number of potential connections is computed. A connection is considered as potential if the departing time of the connected flight is between 60 minutes and 300 minutes later than the arrival time of the first flight. Indeed, according to [13], the minimum connection time advised at CDG is 60 minutes. Regarding the upper bound, one can consider that passengers are not interested in having a connection time higher than five hours. An origin-destination (OD) connection matrix can be generated, and its evolution between two consecutive years is computed. The time period considered is between March 15th to August 30th for 2019, 2020 and 2021. Figure 3a and 3b display the result of the evolution between 2019 and 2020 and the evolution between 2020 and 2021 respectively.

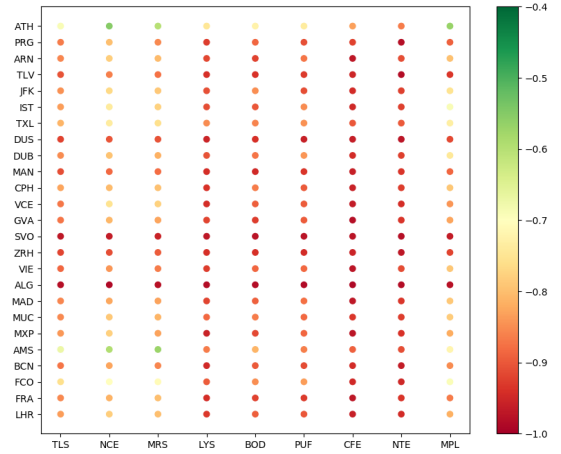
In 2020, Alger (ALG) and Moscow (SVO) were the destinations for which the connectivity had been the most reduced. Indeed in 2020, the European Union closed its border to Russia for a longer time than other countries such as the United States for instance. Algeria closed its borders from March 2020 until June 2021. Meanwhile, connections to Amsterdam (AMS) airport, another major European hub, were least impacted. Regarding the evolution from 2020 to 2021, the air traffic recovery has partially started. For instance, Nantes (NTE), Bordeaux (BOD) and Lyon (LYS) airports have significantly increased their connections to CDG. Also, New-York (JFK), AMS and ALG airports are starting to reconnect with CDG. However, Berlin (TXL) airport closed in 2020, making connections infeasible.

### B. Study of the network asymmetry

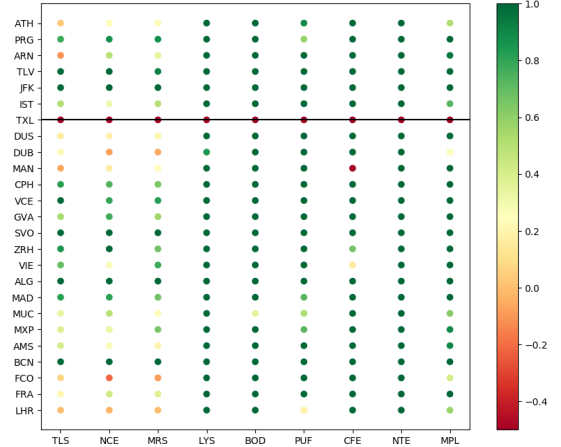
Because of travel restrictions, imbalances in volumes forced airlines to operate flights under new conditions, changing their entire network structure. Furthermore, depending on the country, the inflow and outflow of passengers could be quite different, with countries evacuating their citizens.

Therefore the asymmetry between the flights arriving and departing at CDG airport, as well as the passengers are analyzed. Connections are aggregated for Europe, United-States and Asia respectively. Let  $\mathcal{A}_c$  denote the set of airports on the continent  $c$ ,  $n_{a_i,d}^{\text{arr}}$  the number of flights arriving from airport  $a_i$  at CDG during a day  $d$ , and  $n_{a_i,d}^{\text{dep}}$  the number of flights departing from CDG to airport  $a_i$  during the same day  $d$ .  $q_{a_i,d}^{\text{arr}}$  refers to the number of passengers arriving from airport  $a_i$  at CDG during the day  $d$  and  $q_{a_i,d}^{\text{dep}}$  the number of passengers leaving CDG to airport  $a_i$ . Thus, for each continent and each day, the flights asymmetry coefficient  $ac^{\text{F}}$  is defined as

$$ac_d^{\text{F}} = \frac{\sum_{a_i \in \mathcal{A}_c} n_{a_i,d}^{\text{arr}} - n_{a_i,d}^{\text{dep}}}{\sum_{a_i \in \mathcal{A}_c} n_{a_i,d}^{\text{arr}} + n_{a_i,d}^{\text{dep}}} \in [-1, 1].$$



(a) 2019-2020



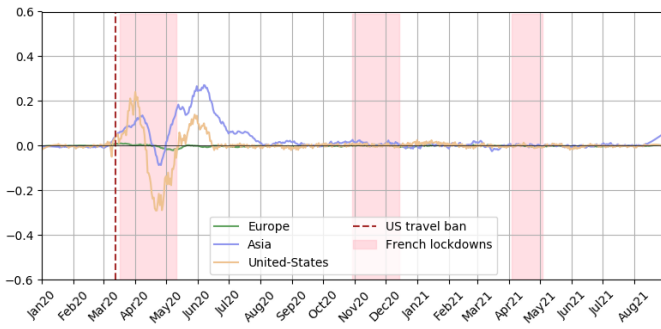
(b) 2020-2021

Fig. 3: Evolution in the average number of potential connections per day between two consecutive years for the same period of March 15th to August 30th. (a) displays the evolution in 2020 compared to 2019. Note that the scale is between -40% and -100% since no connection was improved in 2020 due to the pandemic. (b) displays the evolution in 2021 compared to 2020. The scale is between -50% and 100% as the air traffic recovery has already begun in 2021.

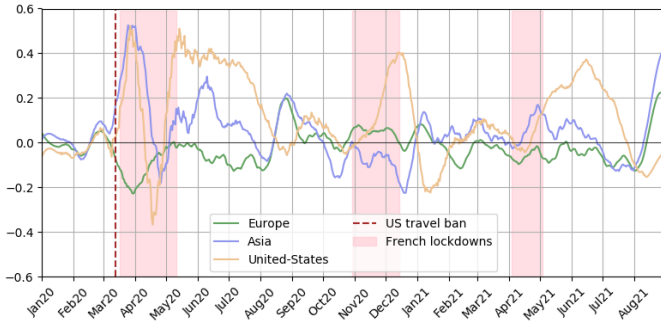
Similarly, the passenger asymmetry coefficient  $ac^{\text{P}}$  for a day  $d$  is defined as

$$ac_d^{\text{P}} = \frac{\sum_{a_i \in \mathcal{A}_c} q_{a_i,d}^{\text{arr}} - q_{a_i,d}^{\text{dep}}}{\sum_{a_i \in \mathcal{A}_c} q_{a_i,d}^{\text{arr}} + q_{a_i,d}^{\text{dep}}} \in [-1, 1].$$

For each day of the study, the flight and the passenger asymmetry coefficients are computed. In order to capture the global tendency and smooth the curve, a Savitsky-Golay filter is applied [14]. The sliding window width is fixed to a month and the polynomial degree to 3. The results for flights and passengers are displayed in Figure 4a and Figure 4b respectively.



(a) Flight asymmetry coefficient



(b) Passenger asymmetry coefficient

Fig. 4: Asymmetry coefficient for the year 2020-2021. A positive asymmetry coefficient for a destination means that on average, the arrival volume at CDG is higher than the departure volume from CDG. On the contrary, a negative asymmetry coefficient corresponds to a higher departure volume to this destination. The dates of the US travel ban announcement and of the French nationwide lockdowns are plotted.

Regarding the flight volumes (Figure 4a), an asymmetry was significant between March and August 2020. During this period, there were up to 20% more flights arriving from Asia than going there. The stringent restrictions enforced in Asia to contain the virus propagation might have discourage passengers to go to this continent. Hence, airlines most likely adapted their offer by reducing direct flights to Asia. The asymmetry coefficient for the United State is positive for the beginning of the 1st French lockdown. This coefficient then steadily decreases from 15% at the end of March until -25% at the end of April. It returns to 0 at the end of the lockdown. After August 2020 up to August 2021, the flight volume arriving and the flight volume departing are balanced for each continent. Nevertheless, this is not the case for passengers (Figure 4b). For instance, a significant imbalance is noted from May to August 2021 for passengers between CDG and US. Therefore, flights going to the US during this period were presumably less filled than returning flights. Furthermore, during lockdowns in France, French citizens were not authorized to travel out of the country. Thus, the number of arriving passengers from the US and Asia was higher. In terms of absolute volumes, Table V highlights the fact that both aircraft and passengers have

relocated during the pandemic, with a net inflow of passengers from the United States and Asia in particular.

	Net flights 2020	Net flights 2021	Net passengers 2020	Net passengers 2021
EU	-0.1%	-0.1%	-1.0%	-2.0%
ASIA	3.1%	0.5%	4.6%	4.3%
US	0.8%	0.0%	18.7%	5.6%

TABLE V: Traffic flow balance at CDG from March 15th to August 30th on 2020 and 2021 for flights and passengers. Each percentage corresponds to the difference between the volume arriving at CDG and the volume departing from CDG divided by the total volume.

### C. Delay analysis

The delays at CDG are analyzed for the years 2017 to 2021. Let  $\mathcal{F}_d$  be the set of flights operated at CDG on day  $d$ . For each flight  $f \in \mathcal{F}_d$  that took off or landed at CDG on that day,  $\mathbb{1}_f$  is defined as:

$$\mathbb{1}_f = \begin{cases} 1 & \text{if the flight } f \text{ was delayed} \\ 0 & \text{otherwise.} \end{cases}$$

The amount of delay of each flight is denoted by  $d_f$ . For an on-time flight,  $d_f=0$ . Thus, the average delay for one day is computed as

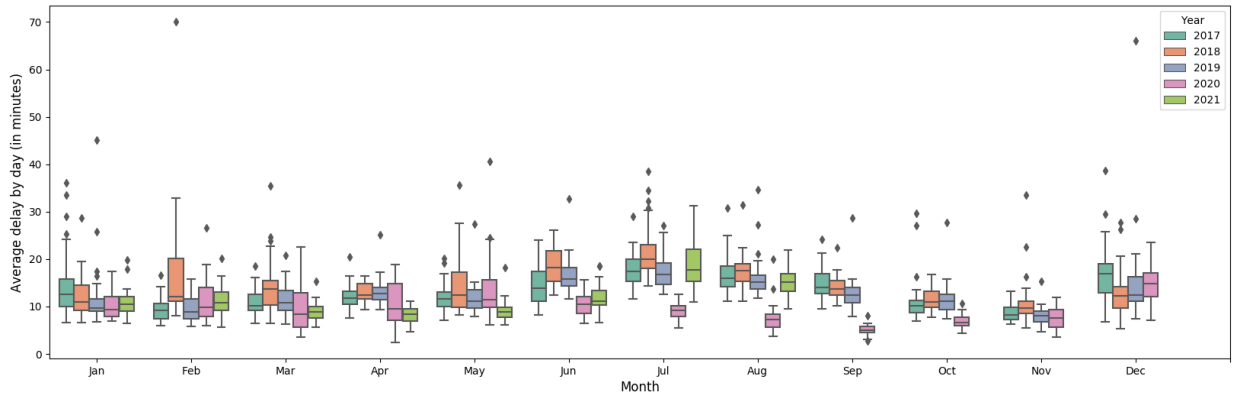
$$\delta_d = \frac{\sum_{f \in \mathcal{F}} d_f}{\sum_{f \in \mathcal{F}} \mathbb{1}_f}.$$

The average delay is computed for each day of the study. Then, the distribution of these delays for each month is analyzed, as displayed in Figure 5a. For each month, the total number of flights operated at CDG is represented in Figure 5b.

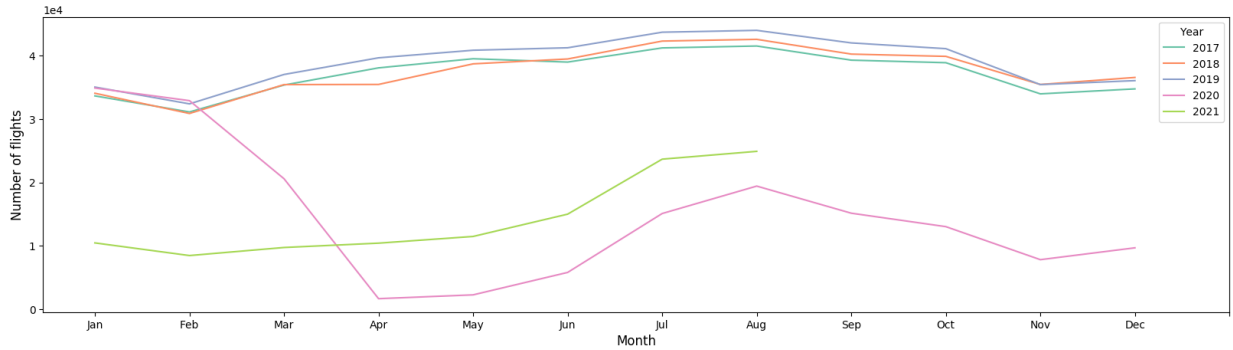
As one can observe on Figure 5b, before the COVID-19 crisis, July and August were the months with the highest number of flights : around 43,000 flights scheduled compared to 31,000 in February. This seasonal increase in air traffic volume leads to a congested airport which results in higher flight delays. Significant delays can also be observed in winter due to weather conditions. For instance, in February 2018, several flights were severely impacted by a snowstorm at CDG airport.

During the COVID-19 pandemic, one could have expected lower delays since the volume of flights operated dropped, and hence flight congestion should have been alleviated. This is the case for summer 2020. However, during December 2020, the delayed flights were on average more impacted than the previous year even if the flight volume was four times lower. These high delays can be due to an unexpected increase in air traffic volume between November 2020 and December 2020 as observed in Figure 5b. Indeed, the late communication by the French government of the end of the second lockdown that occurred in December 2020 might have led to an underestimated air traffic recovery. Versatile French government's decisions and their late announcements during the pandemic led to high uncertainties. This lack of visibility





(a) Boxplot of average delay at CDG from all months between January 2017 and August 2021. Only delayed flight are considered.



(b) Number of flights by month from all months between January 2017 and August 2021

Fig. 5: Average delay and flight volume per month from January 2017 to August 2021.

deteriorates air transportation stakeholders' decisions since anticipation is crucial for them. For instance, CDG airport establishes the staff volume 45 days in advance. Thus, in December 2020, the staffing contract was established before the end of the lockdown was announced by authorities. The same conclusion can be made for May 2020 at the end of the first lockdown in France. The high delays observed are probably due to an underestimated recovery demand. Following this undervaluation, the low delays in summer 2020 might reflect an overreaction of the air traffic system and thus an excess capacity.

This subsection brings several observations. First, the paradigm of higher congestion leading to higher delays ([15], [16]) has been weakened during the pandemic. Secondly, the reliability of operations seems to be a key lever to improve airport welfare.

#### IV. EVOLUTION OF PASSENGER VOLUME

In this section, a clustering of days based on flight and passenger volumes is performed. Aircraft load factors before and during the pandemic are inferred and analyzed.

##### A. Characterization of days based on traffic and passenger volumes

For each day of the time period studied, the following feature vector is computed:

$$v_d = ((n_d^{a_1} \quad q_d^{a_1}) \dots (n_d^{a_N} \quad q_d^{a_N})),$$

where  $n_d^{a_i}$  is the number of flights between CDG and airport  $a_i \in \mathcal{A}$ , and  $q_d^{a_i}$  is the number of passengers carried between CDG and airport  $a_i \in \mathcal{A}$  on that day.

A K-means cluster technique [17] is then applied to these feature vectors. The number of clusters  $k$  is set to 2 since it is the value that maximizes the silhouette score. As explained in [18], the silhouette coefficient is calculated by taking into account the mean intra-cluster distance  $a$  and the mean nearest-cluster distance  $b$  for each data point. The silhouette coefficient  $s$  for a sample is

$$s = \frac{b - a}{\max(a, b)}.$$

Results are displayed in Figure 6. In addition, Figure 7 shows the distribution of days colored by year.

Figure 6 indicates a clear distinction between pre- and post-COVID19 days of airport operations. Up to March 15th, 2020, all days belong to cluster 0 (blue) while after this date, days are gathered into cluster 1 (orange). The following day the French government announced the first lockdown. In addition to the clusterization, the curve drawn by the scatter plot can be analyzed. The slope represents the average number of passengers per aircraft. A higher slope means that for the same number of flights, more passengers are carried. Thus, a higher slope translates either higher aircraft load factors or

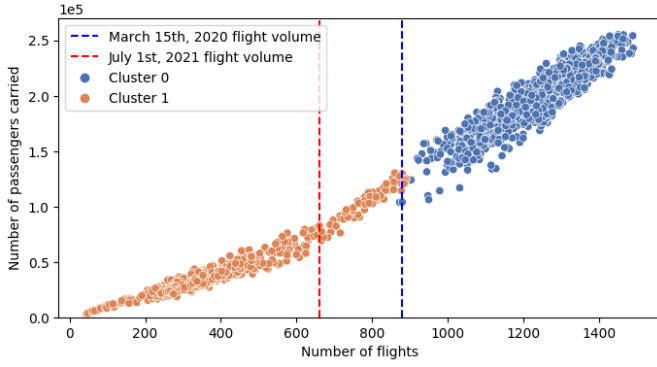


Fig. 6: Passenger volume carried versus daily flight volume. A K-means clusterization has been applied. The color of each point refers to its cluster. The number of flights on March 15th, 2020 and July 1st, 2021 are represented by the blue line and red line respectively.

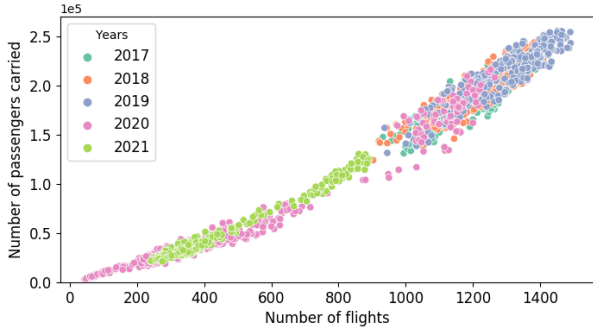


Fig. 7: Passenger volume carried in function of number of flights for each day at CDG colored by year.

a higher share of long-haul aircraft. A change in the slope can be observed in Figure 6 on July 1st 2021. Days after this date and before the first lockdown in France are on the right side of this border. Days during the COVID-19 pandemic until June 30th are located on the left side of the border. As explained in Figure 2, the number of long distance travel flights decreased at the beginning of COVID-19 pandemic. This is observed through a reduced slope on the left side of the slope border in Figure 6. The load factors might also have been reduced and this is studied in the next subsection. One can notice in Figure 7 that days of 2021 fill the gap between 2020 days and pre-COVID-19 days. The summer 2021 days are located on the right side of the slope border. This trend may indicate the beginning of the recovery of the air transportation system.

### B. Load factors evolution

The load factors of aircraft arriving to and departing from CDG between 2017 and 2021 have been computed. Aircraft type and seat configuration are not directly accessible in the data set at hand. An estimate of the number of available seats for each aircraft was inferred by considering the maximum number of passengers carried for the last five years, using the aircraft tail number. Denote this value  $q_{ac}^{\max}$ . Let  $F_{ac}^y$  denote

the set of flights operated by the aircraft  $a_c$  during the year  $y$ . The average load factor for each aircraft  $a_c$  for a year  $y$  is computed as

$$\bar{l}f_{ac} = \frac{\sum_{f \in F_{ac}^y} \frac{q_{ac}^f}{q_{ac}^{\max}}}{\text{card}(F_{ac}^y)}.$$

Figure 8 displays the distribution of aircraft load factors for the last five years. During nominal years (2017,2018,2019),

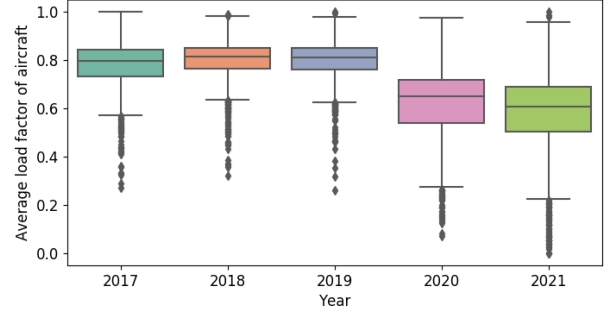


Fig. 8: Boxplots of average aircraft load factors for years 2017 to 2021.

the average load factor of these aircraft was close to 80% with a small standard deviation. Moreover, the lowest load factor was always greater than 20%: it was not worth flying with lower load factor. After the COVID-19 outbreak, load factors dropped and plummeted to values never reached before. On average, aircraft flew with 20% fewer passengers. 25% of aircraft flew at least half-empty in 2021. Finally, several flights were operated with a load factor lower than 10%. Such low load factors reflect hard airline constraints. For instance, pilot who have to be current with regulations must be flying a minimum number of hours every month [19]. Rotation and maintenance constraints may also have incited airlines to maintain unprofitable flights to bring back aircraft and crews to desired locations. The passenger asymmetry combined with the flight symmetry between CDG and the US in 2021 represented in Figure 4 corroborate these possible explanations.

Aircraft load factors are also subject to higher variability during the COVID-19 crisis. Indeed, the boxplots' interquartile ranges during the pandemic are twice wider than the ones of pre-COVID19 years as one can observe in Figure 8. Several hypothesis which could explain this observation are listed below:

- 1) Airlines' constraints may have induced flights with load factors well below the minima during normal years contributing to this higher variability.
- 2) Late information from governments about border closures and travel restrictions may have exacerbated the difficulty to forecast travel demand. Since airlines rely on these models to build their fleet and crew assignment, uncertainties may lead to poor quality decisions and additional costs. For instance, assigning a large aircraft



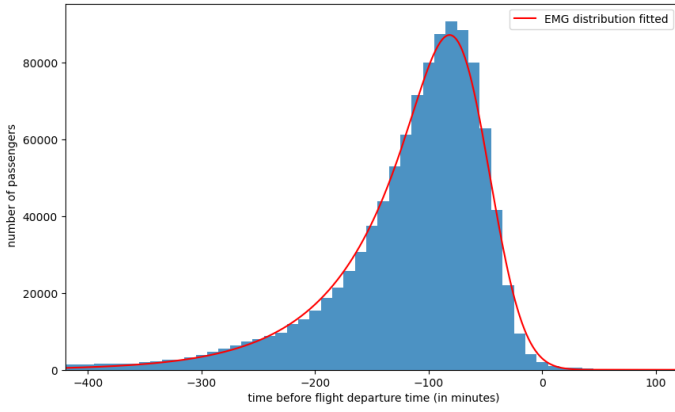


Fig. 9: Passenger arrival time distribution at security screening system on June 2019. In this plot, the time before flight departure time is the difference between the scheduled departure time and the time stamped at the security screening position. The data is fitted with an Exponentially Modified Gaussian Distribution (EMG) displayed in red.

to a flight with low demand is economically not suitable for airlines.

A flight-centric analysis has been led through this section. A complementary analysis through passenger data within airport terminals is presented in the next section. This analysis highlights how air passengers and airlines react to the COVID-19 crisis.

## V. ANALYSIS OF PASSENGER SCANS

This section focuses on analyzing passenger scans at security screening points. Passenger scans at the exit of security system were collected during June 2019 and June 2020. For each flight, the buffer time between its passenger arrival times at security screening locations and the scheduled departure time is estimated. This information is valuable to understand how early passengers arrive at the airport before their flights. Comparing these distributions on June 2019 and June 2020 provides insights on how passengers change their behavior during COVID-19 pandemic. These scans do not take into account connecting passengers, unless their connection requires exiting the secured transit zones and re-entering them. However, the actual number of passengers who embarked on each flight is accessible through traffic data records. Therefore the share of connecting passengers at CDG was inferred and compared in June 2019 and June 2020. Such analysis provides insights on how airlines adapted their offer during the COVID-19 crisis.

### A. Inferring passenger arrival times at security screening before their flights

The passenger arrival times distribution at the security checkpoint on June 2019 is presented in Figure 9. The distribution is asymmetric with a tail on the left side of the distribution. More than 40% of passengers scan between 100min and 60min before their flights while 6% of passengers which are more conservative arrive from four up to seven hours

before the departure. A probability density distribution can be used to model this passenger arrival process. A normal distribution captures the most part of passengers arriving around 90 min before their flights while a 'reversed' exponential distribution seems better fitted to model the distribution's left tail. Hence, an Exponentially Modified Gaussian distribution (EMG) seems suitable to model passenger arrival process. The EMG's probability density function is a convolution between a Normal distribution and an Exponential distribution [20]. It is characterized by three parameters  $(\mu, \sigma, \lambda)$  (two parameters from the Normal distribution and one from the Exponential distribution). EMG is used for example in biomedical science [21] or even to model the visual reaction time [22]. The probability density function (PDF) of such distribution is

$$f_{\text{EMG}}(x; \mu, \sigma, \lambda) = \frac{\lambda}{2} e^{\frac{\lambda}{2}(2\mu + \lambda\sigma^2 - 2x)} \operatorname{erfc}\left(\frac{\mu + \lambda\sigma^2 - x}{\sqrt{2}\sigma}\right),$$

where

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt.$$

The EMG distribution parameters can be estimated to fit the data through the Maximum Likelihood Method. Noting the set of observations  $S = \{s_1, s_2, \dots, s_n\}$  (i.e passenger arrivals), this method consists in finding the set of parameters  $(\mu, \sigma, \lambda)$  which maximizes the following expression:

$$L(S; \mu, \sigma, \lambda) = \prod_{i=1}^n f_{\text{EMG}}(s_i; \mu, \sigma, \lambda).$$

Different probability distributions are compared to model the passenger arrival process. The data is fitted with all of them according to the Maximum Likelihood Method. For each distribution the Mean Absolute Error (MAE) is computed. Denote  $Z = \{z_1, z_2, \dots, z_n\}$  and  $Y = \{y_1, y_2, \dots, y_n\}$  respectively the predicted and the observed values. The mean absolute error is computed as

$$\text{MAE}(Z, Y) = \frac{\sum_{i=1}^n |z_i - y_i|}{n}.$$

Results are displayed in Table VI.

Probability density distribution	MAE
EMG distribution	1218
Log Normal distribution	4640
Gamma distribution	5705
Beta distribution	5930
Normal distribution	8334

TABLE VI: Comparison of different MAE obtained by fitting the passenger arrival distribution with several probability density distributions.

One can observe that the EMG distribution outperforms the others with a MAE at least four times lower than the other distributions tested. It captures the high peak of arrivals around 90 minutes before flight departures while also catching the earlier arrivals. The relative mean absolute error (RMAE) is computed by dividing the MAE by the average number of passengers per time step. The EMG distribution has a RMAE

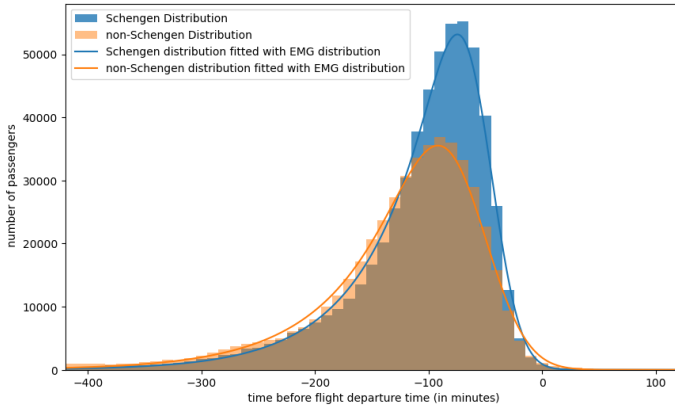


Fig. 10: Passenger arrival time distribution at security screening system on June 2019. The blue and orange distributions refer to the Schengen and non-Schengen passengers respectively. Each one of them is fitted with an EMG distribution.

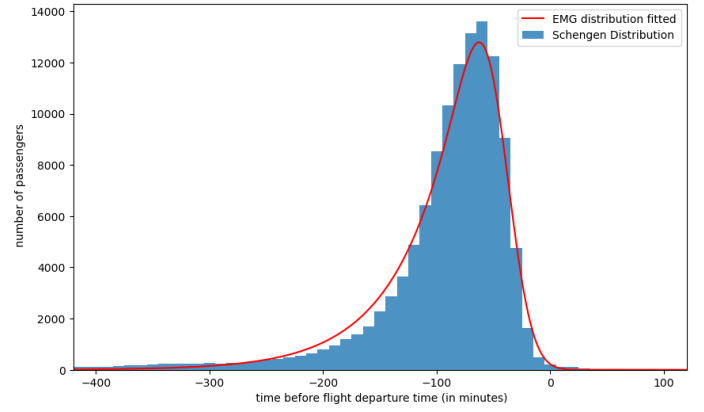


Fig. 11: Passenger arrival time distribution for Schengen flights at security screening system on June 2020. Passenger arrival times for non-Schengen flights could not be inferred for this period.

lower than 6.5%. This distribution is then selected to model passengers arrival process at security screening.

Using such parametric distribution is helpful to characterize and compare passenger arrival process on June 2019 and 2020. Each parameter's variation of the function  $f_{EMG}$  can be interpreted in the following way:

- 1) if  $\mu$  increases, passengers tend to arrive later at the airport (note that  $\mu$  is negative)
- 2) if  $\sigma$  increases, passenger arrivals are more scattered across time
- 3) if  $\lambda$  increases, the exponential decay which characterizes the left tail of the distribution is strengthened. This translates into fewer volume of passengers arriving way earlier before their departure flight time.

Parameters	June 2019 Schengen	June 2019 non-Schengen	June 2020 Schengen
$\mu$	-49.776	-59.358	-41.391
$\sigma$	21.542	31.698	18.046
$\lambda$	0.0174	0.0155	0.0197

TABLE VII: Parameters obtained by fitting passenger distribution with EMG distribution. A distinction is made between June 2019 and 2020 and between Schengen and non-Schengen passengers.

Figures 10 and 11 depict passenger arrivals at security screening in June 2019 and June 2020 respectively. Passenger distribution are fitted with an EMG distribution and their relative parameters are summarized in Table VII. In the following analysis, a distinction is made between flights within and out of Schengen area. Passenger scans for non-Schengen flights on 2020 are removed from the analysis because they were inconsistent due to a reconfiguration of an airport position. For this position, outbound passengers but also connecting passengers were tracked without distinction. Thus, the inference of passenger arrivals from the ground was not possible.

According to Table VII, the parameter  $\mu$  is ten minutes inferior for non-Schengen passengers. This means that passengers exit from security screening on average ten minutes before Schengen passengers. Also the exponential parameter is higher for Schengen arrivals than for non-Schengen arrivals. Therefore the number of conservative passengers on Schengen flights is lower than the one on non-Schengen flights. Finally, the higher standard deviation parameter  $\sigma$  for non-Schengen passengers indicates a broader arrival time distribution. According to these observations, one can conclude that passengers are more conservative when border controls are required. Some passengers are even more careful by arriving up to seven hours before the schedule time of their departing flights.

The opposite conclusion can be obtained by comparing Schengen passenger arrivals in 2019 and 2020. During COVID-19 pandemic, passengers tend to arrive height minutes later ( $\mu_{2020} = \mu_{2019} + 8$ ). Moreover, fewer conservative passengers are observed ( $\lambda_{2020} > \lambda_{2019}$ ). Passenger arrival times distribution is also narrower on June 2020 ( $\sigma_{2020} < \sigma_{2019}$ ). One possible explanation is that the queuing time reduction within the terminals due to airport decongestion was more important than the increase in processing time due to sanitary measures (PCR test verification, social distancing, etc) in passengers' perception. Thus passengers tend to be less conservative and reduce their buffer time at the departure airport. Another hypothesis is that passengers who arrived really early for their flights before the pandemic correspond to infrequent flyers. Hence, the burdensome travel planning due to COVID-19 might have discouraged these passengers to travel and consequently reduce the share of conservative passengers in 2020.

### B. Inferring connecting passengers

The number of connecting passengers for each flight is inferred by subtracting the number of passenger scans at security screening from the actual number of passengers carried. Connecting passengers for non-Schengen flights on 2020 could

not be inferred for the same reason explained in the previous subsection. Denote  $\pi_f$  the connecting passenger share of flight  $f$  and defined as

$$\pi_f = 100 \times \frac{\text{pax}_f^c}{\text{pax}_f^o + \text{pax}_f^c} \%,$$

where  $\text{pax}^c$  and  $\text{pax}^o$  represent the number of connecting passengers and the number of outbound passengers of flight  $f$  respectively.  $\pi_f = 0\%$  means that all passengers arrived from the ground for this flight  $f$ . On the contrary,  $\pi_f = 100\%$  means that all passengers of  $f$  had a previous connected flight at CDG.  $\pi_f$  is computed for each flight of June 2019 and 2020 departing from CDG. The connecting passenger share distributions for these two months are displayed in Figures 12 and 13.

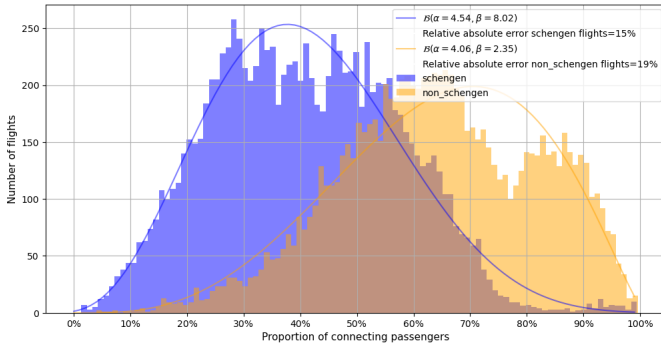


Fig. 12: Connecting passenger share distribution for departing flights at CDG on June 2019. A distinction is made between Schengen (blue) and non-Schengen (orange) flights.

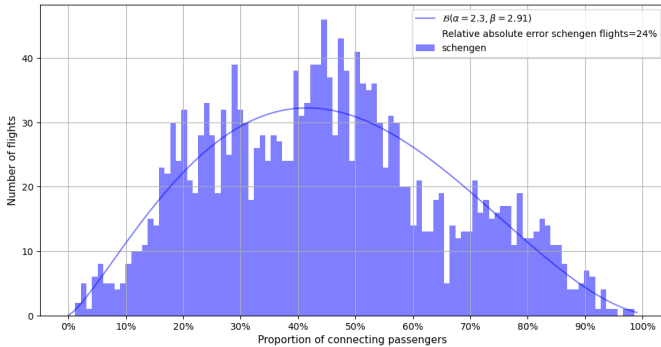


Fig. 13: Connecting passenger share distribution for departing flights at CDG to the Schengen Area on June 2020. Non-Schengen connecting passengers could not be inferred for this period.

The data is fitted with a Beta distribution by using the maximum likelihood method. This distribution is defined into the interval  $[0,1]$ . Hence it seems suitable since  $\pi_f$  is a proportion. The Beta distribution is characterized by the following probability density function:

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du}.$$

The 2019 data displayed in Figure 12 seems well fitted with a Beta distribution, especially for Schengen flights with a relative MAE of 15%. The fitting with a Beta distribution performs worse for 2020 data (Figure 13) with a relative MAE of 24%. This performance deterioration is partially explained by the drop in flight volume. The expected value and the variance of each Beta distribution are computed and compared. The expected value of Beta distribution is equal to

$$\mathbf{E}(X) = \frac{\alpha}{\alpha + \beta}.$$

The variance can be calculated as

$$\mathbf{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Table VIII presents the different expected values and variance obtained by fitting connecting passenger share distribution with a Beta distribution.

	Schengen 2019	non-Schengen 2019	Schengen 2020
$\mathbf{E}()$	36%	63%	44%
$\mathbf{Var}()$	1.7%	3.13%	3.97%

TABLE VIII: Moments obtained by fitting connecting passenger share distributions with a Beta distribution.

The expected share of connecting passengers on June 2019 is 36% for Schengen flights and 63% for non-Schengen flights. This seems relevant since CDG airport is one of the largest European hubs. Thus a lot of feeder flights arrive at CDG to ensure connections with long distance flights. Regarding 2020, the Schengen distribution is wider than in 2019 and the expected number of connecting passengers increases by 8%. One explanation could be that airlines have reinforced their hub and spoke model. Indeed, COVID-19 pandemic has deeply impacted air traffic and the number of passengers has plummeted on 2020. One solution for airlines to fill their aircraft is to remove direct flights between small cities and increase feeder flights to gather passengers at their hub. This explanation is consistent with the higher number of connecting passengers observed in June 2020.

This section has provided insights on consequences of COVID-19 crisis at CDG through the analysis of passenger scans. The magnitude of change in passenger behavior at the airport and in airlines' strategies during the pandemic have been measured. However, this study does not into account passengers' perception of the service quality provided by air transportation stakeholders. Incidents between airlines' agents and passengers refusing to wear a mask [23] or the reduction in the number of opened facilities at the airport have been frequent these last months. Such events may have deteriorated the customer satisfaction.

## VI. CONCLUSION AND FUTURE WORK

This paper sheds light on how COVID-19 has impacted passengers at Paris-Charles de Gaulle airport. The main contributions of this paper are summarized below:

- 1) Despite the sudden plunge of air traffic volume, the average flight delay has not significantly decreased. It

even increased in December 2020 and February 2021 with a traffic volume three to four times lower than the previous years. Hence, other factors than airport congestion (airspace capacity reduction, weather or airport staff reduction) should be a focus of attention to explain such delays.

- 2) In the aftermath of the pandemic, aircraft load factors have plummeted. One in four flights have been operated with less than half of the seats occupied in 2021, i.e. a drop of 25% compared to pre-pandemic load factors. These poorly filled flights are most likely not profitable for airlines. Assuming that airline decisions are taken to maximize their profit (or minimize the economic loss in this context), such observation underlines the economic weight of airline constraints (pilot qualifications, crew scheduling, aircraft maintenance). Airlines might have compensate the economic loss of load factor reduction by air freight but this information was not available. Besides, individuals have clearly relocated across continents as shown by the imbalance in passenger net flow.
- 3) Exponentially Modified Gaussian distribution is suitable to model passenger arrival process at the airport.
- 4) Due to airport decongestion, passengers on Schengen flights tend to arrive later at the departure airport during the pandemic.
- 5) The share of connecting passengers at CDG has increased during COVID-19 on European flights. This might reflect a renewed focus on the hub and spoke model by airlines. The drop in demand incited suppliers to replace point-to-point flights by connecting flights through their hub airport.

Future work aims at designing recovery strategies taking into account the large uncertainty in demand, in terms of network structure, schedule design and aircraft assignment to routes. A feedback from airlines would be valuable to go into analysis in depth and reinforce proposed insights highlighted through this study. A qualitative analysis on passenger satisfaction such as in [12] would complement the presented study. Such analysis would help to understand the impact of COVID-19 pandemic on the air transportation system through the passengers' eyes.

#### ACKNOWLEDGMENT

The authors would like to thank Paris-Charles De Gaulle airport for providing the data and for their willingness to share their knowledge and expertise. The authors are thankful to the TRANSIT project for its financial support.

#### REFERENCES

- [1] IATA. Effects of novel coronavirus (COVID-19) on civil aviation: Economic impact analysis. [https://www.icao.int/sustainability/Documents/COVID-19/ICAO\\_Coronavirus\\_Econ\\_Impact.pdf](https://www.icao.int/sustainability/Documents/COVID-19/ICAO_Coronavirus_Econ_Impact.pdf), October 2021.
- [2] IATA. COVID-19 : An almost full recovery of air travel in prospect. <https://www.iata.org/en/iata-repository/publications/economic-reports/an-almost-full-recovery-of-air-travel-in-prospect/>, 2021.
- [3] Xiaoqian Sun, Sebastian Wandelt, and Anming Zhang. How did COVID-19 impact air transportation? a first peek through the lens of complex networks. *Journal of Air Transport Management*, 89:101928, 2020.
- [4] Sascha Albers and Volker Rundshagen. European airlines strategic responses to the COVID-19 pandemic (january-may, 2020). *Journal of air transport management*, 87:101863, 2020.
- [5] Nena Adrienne, Lucy Budd, and Stephen Ison. Grounded aircraft: An airfield operations perspective of the challenges of resuming flights post COVID. *Journal of Air Transport Management*, 89:101921, 2020.
- [6] Michael Schultz, Jan Evler, Ehsan Asadi, Henning Preis, Hartmut Fricke, and Cheng-Lung Wu. Future aircraft turnaround operations considering post-pandemic requirements. *Journal of Air Transport Management*, 89:101886, 2020.
- [7] Falko Mueller. Examining COVID-19-triggered changes in spatial connectivity patterns in the european air transport network up to june 2021. *Research in Transportation Economics*, page 101127, 2021.
- [8] Stefano Maria Iacus, Fabrizio Natale, Carlos Santamaria, Spyridon Spyrtatos, and Michele Vespe. Estimating and projecting air passenger traffic during the COVID-19 coronavirus outbreak and its socio-economic impact. *Safety Science*, 129:104791, 2020.
- [9] Myeonghyeon Kim and Jeongwoong Sohn. Passenger, airline, and policy responses to the COVID-19 crisis: The case of South Korea. *Journal of air transport management*, page 102144, 2021.
- [10] Tracy L Lamb, Scott R Winter, Stephen Rice, Keith J Ruskin, and Austin Vaughn. Factors that predict passengers willingness to fly during and after the COVID-19 pandemic. *Journal of air transport management*, 89:101897, 2020.
- [11] Thomas Budd, Pere Suau-Sanchez, Nigel Halpern, Deodat Mwesiumo, and Svein Bråthen. An assessment of air passenger confidence a year into the COVID-19 crisis: A segmentation analysis of passengers in norway. *Journal of Transport Geography*, page 103204, 2021.
- [12] Philippe Monmousseau, Aude Marzuoli, Eric Feron, and Daniel Delahaye. Impact of covid-19 on passengers and airlines from passenger measurements: Managing customer satisfaction while putting the us air transportation system to sleep. *Transportation Research Interdisciplinary Perspectives*, 7:100179, 2020.
- [13] CDGFacile. Informations passagers en correspondance cdg. <https://cdgfacile.com/informations-passagers-correspondance-cdg/>, 2021.
- [14] Ronald W Schafer. What is a Savitzky-Golay filter?[lecture notes]. *IEEE Signal processing magazine*, 28(4):111–117, 2011.
- [15] Gregory D. Glockner. Effects of air traffic congestion delays under several flow-management policies. *Transportation Research Record*, 1517(1):29–36, 1996.
- [16] Alexandre Jacquillat and Amedeo R Odoni. An integrated scheduling and operations approach to airport congestion mitigation. *Operations Research*, 63(6):1390–1410, 2015.
- [17] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. The global k-means clustering algorithm. *Pattern recognition*, 36(2):451–461, 2003.
- [18] Ketan Rajshekhhar Shahapure and Charles Nicholas. Cluster quality analysis using silhouette score. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 747–748. IEEE, 2020.
- [19] Helen Massy-Beresford, Thierry Dubois, Adrian Schofield, Ben Goldstein, Lori Ranson, and Sean Broderick. Airlines choose different paths for keeping pilots ready. <https://aviationweek.com/air-transport/maintenance-training/airlines-choose-different-paths-keeping-pilots-ready>, 2021.
- [20] Scott Haney. Practical applications and properties of the exponentially modified gaussian (emg) distribution. 2011.
- [21] A Golubev. Exponentially modified peak functions in biomedical sciences and related disciplines. *Computational and mathematical methods in medicine*, 2017, 2017.
- [22] Evan M Palmer, Todd S Horowitz, Antonio Torralba, and Jeremy M Wolfe. What are the shapes of response time distributions in visual search? *Journal of experimental psychology: human perception and performance*, 37(1):58, 2011.
- [23] Hugo Martin. Ruckus in the skies: What happens when airline passengers refuse to wear masks. <https://www.latimes.com/business/story/2021-05-05/airline-passengers-mask-rules-faa-fine-zero-tolerance>, 2021.