

1 **State-of-the-Art Analysis of United States Flight Event and Surveillance Data Coverage**  
2 **and Future Research Directions**

3  
4 **Susan Hotle**

5 Department of Civil and Environmental Engineering  
6 Virginia Polytechnic Institute and State University, Blacksburg, VA 24061  
7 Email: shotle3@vt.edu

8  
9 **Kyle Titlow**

10 Spatial Statistician/Geographer  
11 US Department of Transportation  
12 Office of Spatial Analysis and Visualization  
13 Washington, DC 20590  
14 Email: kyle.titlow@dot.gov

15  
16 **Mehdi Hashemipour**

17 Senior Data Scientist  
18 US Department of Transportation  
19 Office of Spatial Analysis and Visualization  
20 Washington, DC 20590  
21 Email: m.hashemipour@dot.gov

22  
23 **Ed Strocko**

24 Director  
25 US Department of Transportation  
26 Office of Spatial Analysis and Visualization  
27 Washington, DC 20590  
28 Email: ed.strocko@dot.gov

29  
30  
31  
32  
33 **Keywords:** SWIM, OpenSky Network, Flight Event, Flight Surveillance, Coverage

1 **ABSTRACT**

2 The Automatic Dependent Surveillance-Broadcast (ADS-B) mandate along with advances in  
3 government and third-party flight tracking systems have allowed researchers to analyze aspects of  
4 aviation operations that were not possible beforehand. The purpose of this study is to provide a 1)  
5 summary of flight event data available to the public, 2) comparison of API surveillance data  
6 available in the United States (US) and 3) evaluation of flight coverage across the data sources.  
7 This analysis considers the Federal Aviation Administration's (FAA) Aviation System  
8 Performance Metrics (ASPM), Bureau of Transportation Statistic's (BTS) Airline Service Quality  
9 Performance System (ASQP), Official Aviation Guide's (OAG) On-Time Performance Flight  
10 dataset, FAA System Wide Information Management (SWIM) and OpenSky Network (OSN). It  
11 reports the spatial and market segment (i.e. commercial, general aviation, air cargo) coverage,  
12 helping researchers identify the appropriate dataset for their studies. This study found that OSN  
13 has worldwide coverage of enroute flights, with limited airport surface movement. FAA SWIM  
14 contains several surveillance datasets, two enroute (only one includes Alaska and Hawaii) and one  
15 airport surface datasets. Combining sources allows one source to overcome a spatial limitation of  
16 another to generate a complete trajectory. The flight counts by source are reported with a  
17 discussion of their discrepancies, where surveillance sources do not capture all flights from the  
18 flight event datasets and vice versa. Therefore, surveillance should be considered supplemental  
19 and not a replacement for measuring flight counts. This paper then discusses future research  
20 directions given surveillance availability.

## 1 INTRODUCTION

2 The operational efficiency and on-time performance of the National Airspace System (NAS)  
3 are important research areas given their implications on safety, the environment and the economy.  
4 This research field is supported by flight on-time performance datasets collected by government  
5 agencies and third-party vendors. Examples include the Federal Aviation Administration's (FAA)  
6 Aviation System Performance Metrics (ASPM), Bureau of Transportation Statistic's (BTS)  
7 Airline Service Quality Performance (ASQP) and the Official Airline Guide's (OAG) On-Time  
8 Performance reports. These datasets, which report flight gate-Out, wheels-Off, wheels-On, gate-  
9 In (OOOI) times, vary in terms of aggregation, flight coverage and user cost.

10 Flight surveillance Application Programming Interfaces (APIs) have become available to the  
11 public at no cost, providing individual flight messages in near real-time. Each message includes a  
12 flight's latitude, longitude and timestamp. API examples include the FAA's System Wide  
13 Information Management (SWIM) and third-parties such as OpenSky Network (OSN). Spatial  
14 coverage of these surveillance sources is continuing to improve with time, where one of the SWIM  
15 API feeds started releasing position data from the Automatic Dependent Surveillance-Broadcast  
16 (ADS-B) system to users in 2022 (1).

17 Given the number of aviation operational datasets available to the public and the changing  
18 coverage of these datasets, many researchers in aviation are unaware of the flight event and  
19 surveillance available and the associated limitations with each. Therefore, this paper serves to fill  
20 the literature gap by providing a 1) summary of flight event performance data available to the  
21 public, 2) comparison of API surveillance data available in the United States (US) and 3)  
22 evaluation of flight coverage and counts across the data sources. The more refined analysis of  
23 comparing the OOOI times for individual flights across the sources is outside the scope of this  
24 paper and is recommended for future research.

25 The APIs considered in this study are FAA SWIM and OSN. This analysis records the airlines  
26 in each dataset and the coverage of different market segments (i.e. air cargo flights, general  
27 aviation flights and airport ground vehicles). This paper ends with a flight counts comparison  
28 across the sources, determination if surveillance sources can be a substitute for flight event sources  
29 when it comes to flight counts and a discussion on possible future research directions given the  
30 advances in surveillance data availability. Overall, the motivation for this paper is to inventory  
31 the current state of flight event and surveillance flight performance data available to the public,  
32 removing unnecessary data exploration for researchers.

## 33 LITERATURE REVIEW

34 Studies on the US aviation industry's on-time performance are supported by datasets collected  
35 by government agencies and third-party vendors, with example datasets outlined in **Table 1**. The  
36 datasets in **Table 1** all contain scheduled and actual OOOI times and have supported research  
37 studies given their detailed information in an easy-to-access format. These datasets have varying  
38 levels of aggregation, flight coverage and user cost, where the limiting factors to researchers are  
39 italicized in **Table 1**.  
40

1

**TABLE 1 Flight event operational data sources**

<b>Agency</b>	<b>FAA</b>	<b>BTS</b>	<b>OAG</b>
<b>Dataset</b>	ASPM (11)	ASQP (27, 28)	On-Time Performance Flight Data (29)
<b>Flights/Airlines</b>	Flights to/from 77 US ASPM airports and 27 ASPM carriers (30)	<i>Flights from US air carriers that receive <math>\geq 0.5\%</math> of domestic scheduled passenger revenues. Currently 15 airlines</i>	Flights from over 500 global airlines and 1200 airports
<b>Cost</b>	Free	Free	<i>Subscription cost</i>
<b>Availability</b>	<i>Airport-hour-level publicly available. Flight-level available to FAA employees and contractors (31)</i>	Flight-level publicly available	Flight-level to customer

2

3 ASPM is collected for flights to/from 77 US airports and flights from 27 carriers. Some of  
4 these carriers have ceased operation. The publicly-available version is aggregated to the airport-  
5 hour level, allowing researchers to quantify delay for a specific hour at an airport, but not for a  
6 specific flight. Internal flight-level ASPM information has individual flight OOOI times. If an  
7 actual gate-out or gate-in time is unknown, it is estimated based on the unimpeded taxi time with  
8 relation to the wheels-off or wheels-on times (2). This data has supported studies on reducing fuel  
9 burn (3), departure demand prediction (4), airport runway configuration prediction (5), as well as  
10 delay in the terminal airspace (6), during the taxiing phase (7, 8), at the departure runway (9) and  
11 delay by cause (10). The Air Traffic Control (ATC) causal factors of delay are then reported in  
12 the Operations Network (OPSNET) dataset on the ASPM website (11, 12). OPSNET includes the  
13 causal factor (i.e. weather, volume, equipment, runway and other) and the ATC delay program  
14 applied (i.e. Expected Departure Clearance Time, Ground Stop, and Traffic Management  
15 Initiative).

16 ASQP data is collected through the Part 234 of the U.S. Department of Transportation's  
17 (DOT's) Regulations (13), where reporting airlines are mandated by congress to self-report their  
18 flight on-time performance data on a monthly basis. A reporting airline is defined as an air carrier  
19 that "accounted for at least 0.5 percent of domestic scheduled-passenger revenues in the most  
20 recently reported 12-month period" (13). Prior to 2018, the regulation was 1 percent (14). The 12  
21 carriers in ASPM not in ASQP include Air Canada (ACA), Mesa Airlines (ASH), ExpressJet  
22 Airlines (ASQ), Air Wisconsin (AWI), Compass Airlines (CPZ), FedEx (FDX), GoJet Airlines  
23 (GJS), Trans States Airline (LOF), Piedmont Airlines (PDT), Horizon Air (QXE), CommuteAir  
24 (UCA) and United Parcel Service (UPS). During the COVID-19 pandemic, ASQ, UCA, and LOF  
25 ceased operations and are no longer operating carriers. ASQP is used to enforce the maximum  
26 tarmac rules in the passenger Fly Rights (15) and includes the airline-reported causal factors for  
27 delay (i.e. carrier, weather, NAS, security and late aircraft). It has supported studies on optimizing  
28 aircraft routing (16), clustering airlines based on on-time performance trends (17), the impact of  
29 weather on flight cancellations (18) and measuring the effects of delay propagation (19).

30 OAG is a third-party product with worldwide coverage. It is used by airports for tracking  
31 arrivals and departures. Airport usage of OAG is typically advertised on the airport's website (for

1 example Willard Airport (20)). While OAG has a greater coverage of airlines and airports than  
2 the other two flight event datasets, it is only available through a subscription cost. OAG has  
3 supported studies not only in operations research, such as predicting air service loss to small  
4 communities (21) and overall airline network planning strategies during a shock event (22, 23),  
5 but also airline revenue management (24) to provide both the universal and feasible itinerary  
6 choice sets available to the airline customer. In these studies, it is important to have all carriers  
7 and itineraries considered in the models, meaning the other two sources with limited reporting  
8 carriers are not sufficient.

9 While there are additional aviation operations datasets publicly available, this analysis solely  
10 focuses on flight event datasets that include scheduled and actual OOOI times. Therefore, BTS'  
11 T-100 (25), Airline Origin and Destination Survey (DB1B) (26) and the previously stated FAA  
12 OPSNET are outside the scope of this analysis.

13 Recently, near real-time flight tracking APIs have become available. They report at an  
14 individual flight-level unlike ASPM, contain more airlines than ASQP and are free unlike OAG.  
15 Together, these characteristics of the APIs provide new opportunities for research. The APIs  
16 considered for this paper are FAA's SWIM and OSN. Each emits flight messages (i.e. pings) that  
17 contain the time, latitude and longitude of the flight along with other flight characteristics, such as  
18 aircraft ID and aircraft type. These API datasets are free, publicly-available and support studies  
19 such as weather impact on flight trajectories (32) and estimating emissions (33).

20 SWIM started in 2007 (34). SWIM contains multiple aviation-related API feeds with  
21 connection information hosted on the SWIM Industry-FAA Team (SWIFT) portal. API's hosted  
22 on the SWIM Industry-FAA Team (SWIFT) portal include Aeronautical Information Management  
23 Federal Notice to Air Missions System (AIM FNS), Integrated Terminal Weather Service (ITWS),  
24 SWIM Flight Data Publication Service (SFDPS), SWIM Terminal Data Distribution System  
25 (STDDS), Time Based Flow Management (TBFM) and Traffic Flow Management System  
26 (TFMS). The data feeds are still evolving. For example, the SWIM Flight Data Publication  
27 (SFDPS) feed started emitting ADS-B position data in 2022 (35), to put into perspective the  
28 recency of system advancements.

29 Alternatively, the OSN API was the result of a research project that began in 2012 (36, 37) and  
30 is managed by a non-profit organization headquartered in Switzerland. It is based on  
31 crowdsourcing, where volunteers set up line-of-sight receivers to track ADS-B equipped flights  
32 (38), meaning spatial coverage continually improves. Given the improvements in the APIs, this  
33 analysis records their current coverage in comparison to the flight event datasets.

## 34 35 **DATA**

36 This section covers what the OSN and FAA SWIM data contains and its format, followed by  
37 the study time period collected for this comparison analysis.

## 38 39 **API Connections**

40 Each API in this analysis is near real-time with a constant data channel to consumers. The  
41 FAA hosts the SWIFT portal with a collection of SWIM APIs. The SFDPS, SWIM Terminal Data  
42 Distribution System (STDDS) and Traffic Flow Management System (TFMS) are all available via  
43 the SWIM portal (39). **Figure 1** shows the XML format of a sample flight. This format can be  
44 imported into a database format using a read XML function in a programming language.

```

<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<ns5:MessageCollection xmlns:ns5="http://www.faa.aero/nas/3.0"
xmlns:ns2="http://www.fixm.aero/base/3.0" xmlns:ns3="http://www.fixm.aero/flight/3.0"
xmlns:ns4="http://www.fixm.aero/foundation/3.0">
  <message xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="ns5:FlightMessageType">
    <flight xsi:type="ns5:NasFlightType" centre="ZTL" source="TH" system="ATL" timestamp="2024-02-26T22:48:38.136Z">
      <arrival xsi:type="ns5:NasArrivalType" arrivalPoint="K15J"/>
      <controllingUnit xsi:type="ns2:IdentifiedUnitReferenceType" unitIdentifier="ZTL"
sectorIdentifier="05"/>
      <departure xsi:type="ns5:NasDepartureType" departurePoint="SVI144034"/>
      <enRoute xsi:type="ns5:NasEnRouteType">
        <position xsi:type="ns5:NasAircraftPositionType" targetPositionTime="2024-02-26T22:48:33Z"
positionTime="2024-02-26T22:48:33Z" reportSource="SURVEILLANCE">
          <actualSpeed>
            <surveillance uom="KNOTS">104.0</surveillance>
          </actualSpeed>
          <altitude uom="FEET">5600.0</altitude>
          <position xsi:type="ns2:LocationPointType">
            <location srsName="urn:ogc:def:crs:EPSG::4326">
              <pos>34.434722 -85.563056</pos>
            </location>
          </position>
          <targetAltitude uom="FEET">5600.0</targetAltitude>
          <targetPosition srsName="urn:ogc:def:crs:EPSG::4326">
            <pos>34.434722 -85.563056</pos>
          </targetPosition>
        </position>
      </enRoute>
    </flight>
  </message>
</ns5:MessageCollection>

```

Figure 1 Format from SWIM feed

OSN is a third-party API. The format of the OSN Live API data is describe on (40) and shown in Table 2. The OSN website provides API connection codes to retrieve the flight messages (i.e. individual flight pings) from the API and append each flight message to the overall collection of flight messages in a database. The API connection codes are available in REST, Java and Python and the user can choose which they prefer.

TABLE 2 Format from OSN

Index	callsign	icao24	origin_country	time_position	true_track	velocity	vertical_rate	on_ground	altitude_baro	altitude_geo	geometry
0	MSR986	01012b	Egypt	2024-02-28 21:58:45	64.82	256.37	0	no	10058.4	9989.82	POINT (-69.6133 42.703)
1	MSR986	01012b	Egypt	2024-02-28 21:59:39	64.82	256.37	0	no	10058.4	9989.82	POINT (-69.6133 42.703)
2	MSR986	01012b	Egypt	2024-02-28 21:59:39	64.82	256.37	0	no	10058.4	9989.82	POINT (-69.6133 42.703)
3	MSR986	01012b	Egypt	2024-02-28 22:00:25	64.98	256.62	0	no	10058.4	9997.44	POINT (-69.4623 42.7552)
4	MSR986	01012b	Egypt	2024-02-28 22:00:52	65.02	257.08	0	no	10058.4	9997.44	POINT (-69.4117 42.7725)

API Data Collection

This analysis uses a data collection across the SWIM and OSN API's. The data collection was on 01-20-2024 for 2 hours (01:30-03:30 UTC) across all four API sources (OSN, SFDPS, STDDS, TFMS), simultaneously. The data collection focused on the US, including Alaska and Hawaii and occurred at a time where all time zones within the US would have commercial flights. It is then compared with the flight counts reported in the flight event datasets. The data was collected for a short duration for two reasons. First, the study results would be the same with a longer duration, where the spatial coverage and flight message frequency do not change over time. Second, mapping the flight messages spanning several days or weeks would be misleading, where the maps would show the sources as having better coverage than in reality. Mapping a week of

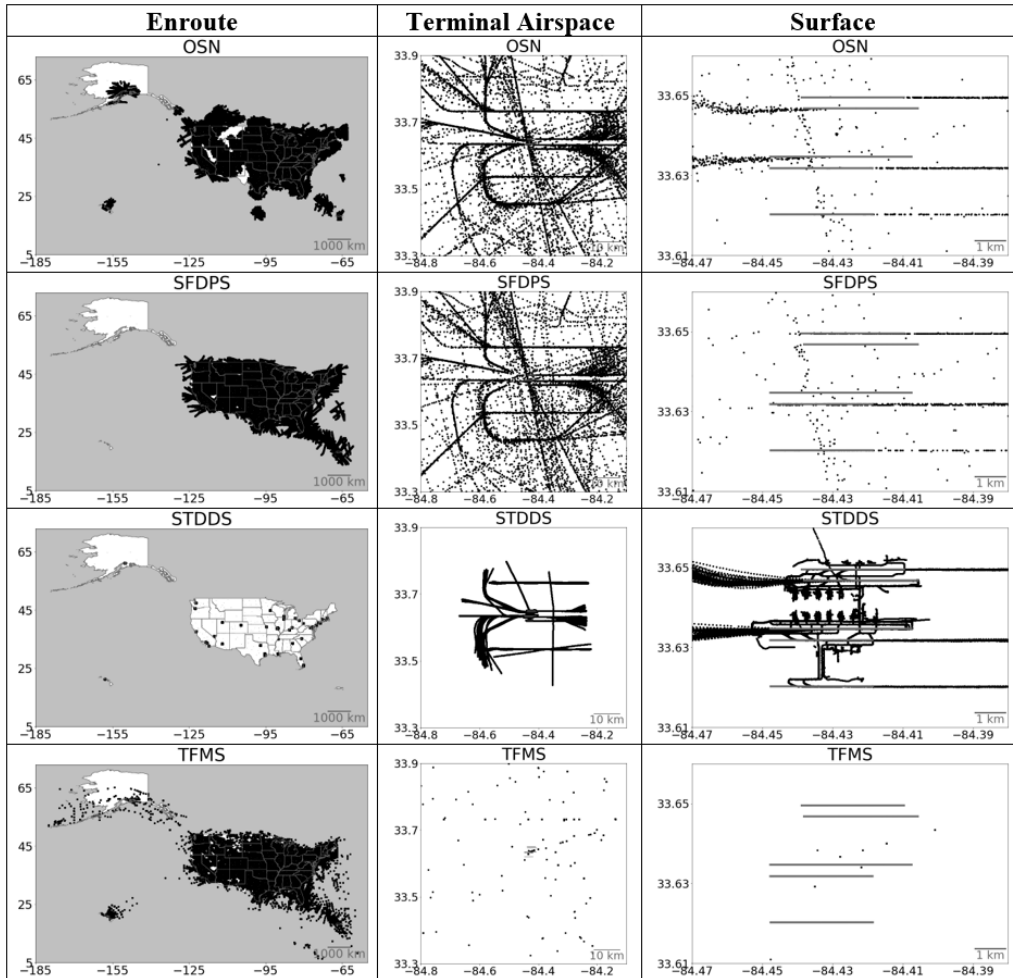
1 flight messages around an airport makes it impossible to infer the daily coverage at the airport and  
 2 spatial distancing between flight messages since the density of flight messages around the airport  
 3 would be high. It is easier to infer the flight coverage of one week if only a few hours of flight  
 4 messages are mapped. Also, the data for a longer period of time makes it harder to draw attention  
 5 to the nuanced daily characteristics intrinsic to each feed.

6

7 **API Data Summaries**

8 Using the collected surveillance data, **Figure 2** shows the coverage of the US, the terminal  
 9 airspace surrounding airports and the airport surface. Each source has different strengths and  
 10 weaknesses, where there is an indirect relationship between a source’s enroute coverage of the US  
 11 and its coverage of the airport surface. STDDS covers the airport surface, sometimes even to the  
 12 gate, but the signal dissipates at approximately 20 km from the airport, giving it a poor enroute  
 13 coverage. Another observation is the surface coverage of enroute sources. Both OSN and SFDPS  
 14 provide flight information near or on the runway and a few messages on the airport surface. At  
 15 the time of the data collection, ATL was operating in west-flow, meaning OSN has sufficient  
 16 runway coverage for both arrivals and departures, but SFDPS only has coverage of arrival runway  
 17 information. Departures start transmitting data through SFDPS after exiting the departure runway.  
 18 This limited runway coverage for departures in SFDPS was observed at other airports.

19



20  
 21

**Figure 2** API data source coverage with Atlanta Hartsfield-Jackson Airport (ATL)

1 Comparing the three enroute APIs shows the differing coverage of the US. As stated before,  
2 OSN is worldwide, but only its US coverage is shown in **Figure 2** for this analysis. OSN and  
3 TFMS have Hawaii and Alaska, but SFDPS does not. SFDPS and TFMS have fewer spatial gaps  
4 over the continental US, where OSN has a few dead zones throughout the US. OSN and SFDPS  
5 get close to the airport, but TFMS does not. In the 2 hours collected, there were only 8 flight  
6 messages on or near the ATL surface from the TFMS source. The TFMS source also has  
7 significant spatial distance between the flight messages due to the message frequency. This means  
8 that although SFDPS and TFMS have similar continental US coverage, SFDPS provides a more  
9 detailed flight trajectory due to its higher message frequency.

10 The spatial coverage results would remain the same if the data collection had spanned a  
11 larger time frame. The 2 hours were chosen based on when all time zones in the US simultaneously  
12 have significant commercial flight traffic, including Alaska and Hawaii. This ensured that the  
13 coverage maps in **Figure 2** are accurate. If several days or weeks had been collected, STDDS  
14 would still only cover 43 airports, SFDPS would still exclude Hawaii and Alaska and only STDDS  
15 would have quality surface movement information.

16 Limiting the data collection to 2 hours actually provided more insight into **Figure 2**  
17 findings. If several weeks of flight messages had been consolidated, the surface-level maps would  
18 have shown both east and west flow coverage simultaneously at ATL, masking the changing  
19 coverage of arrivals and departures (i.e. SFDPS has better coverage for arrivals than departures  
20 near the runway). Also, the 2 hours show only 8 flight messages for TFMS at the surface-level in  
21 **Figure 2**. A longer time period would have shown a high density of flight messages reporting on  
22 the airport surface, misleading the readers to conclude from our paper that TFMS has better surface  
23 coverage than it actually has. A smaller time frame emphasizes how much distance is between  
24 flight messages for TFMS in the terminal airspace. Having too much data can bleed out the small-  
25 and local-level elements of each feed that are key to understanding their utility at airports.

26 An overall summary of the characteristics for each API is in **Table 3**, where all are free to  
27 the public. Three of the four cover the enroute phase of flight and contain limited flight coverage  
28 on the airport surface during the taxiing phase. The opposite is true of the STDDS feed, which  
29 provides detailed surface information at 43 US airports since these airports are equipped with  
30 Airport Surface Detection System- Model X (ASDE-X) (41) or Airport Surface Surveillance  
31 Capability (ASSC) (42). The remaining US airports have limited coverage on the airport surface,  
32 where enroute sources may have a few surface data points. Some unequipped airports receive  
33 STDDS ground coverage if it is near a STDDS airport. For example, Willow Run Airport (YIP)  
34 receives limited coverage from the Detroit Metropolitan Wayne County Airport's (DTW)  
35 instrumentation.

36 In the SWIM feeds, the message frequency is dictated by the API. The measured message  
37 frequency from the different API sources are shown in **Table 3**. Messages with greater than 120  
38 seconds gap were removed from percentile statistics. A flight might enter a dead zone leading to  
39 a large time gap between successive messages. It was found, for example, that the SFDPS feed  
40 provides a median of 1 message per aircraft every 6 seconds. Only the OSN API allows users to  
41 customize the data collection frequency.

42  
43  
44  
45  
46

1

**TABLE 3 Surveillance flight API data sources**

<b>Agency</b>	<b>FAA SWIM</b>			<b>OSN</b>
<b>Dataset</b>	<b>SFDPS</b>	<b>STDDS</b>	<b>TFMS</b>	<b>OSN</b>
<b>Flight Phase</b>	Enroute	Surface	Enroute	Enroute
<b>Coverage</b>	US, excluding Hawaii and Alaska	43 US airports	US, including Hawaii and Alaska	Worldwide
<b>Average Message Frequency</b>	25 <sup>th</sup> =1 second 50 <sup>th</sup> = 6 seconds 75 <sup>th</sup> =12 seconds	25 <sup>th</sup> = 1 second 50 <sup>th</sup> = 1 second 75 <sup>th</sup> =1 second	25 <sup>th</sup> =37 seconds 50 <sup>th</sup> =60 seconds 75 <sup>th</sup> =61 seconds	User-defined, For this analysis: 25 <sup>th</sup> =3 seconds 50 <sup>th</sup> =6 seconds 75 <sup>th</sup> =9 seconds
<b>Cost</b>	Free	Free	Free	Free
<b>Availability</b>	Flight message level publicly available	Flight message level publicly available	Flight message level publicly available	Flight message level publicly available

2

3

**METHODOLOGY**

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

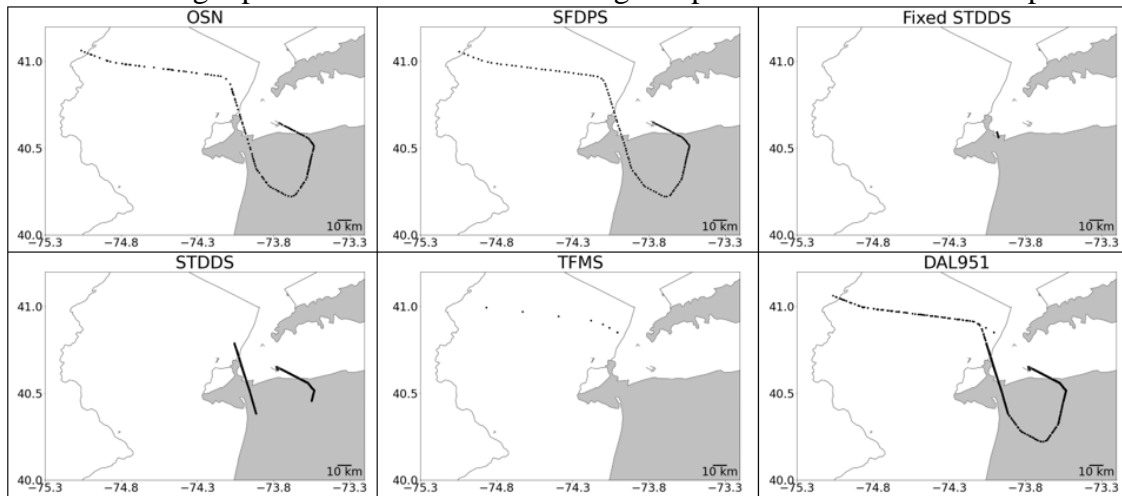
Merging the sources together creates a flight trajectory as shown in **Figure 3**. This merge is based on the flight ID field (e.g. DAL951). However, considerations are needed to salvage data missing a flight ID. Sometimes STDDS receives location data from a plane, but the flight ID is unknown. In this case, STDDS reports the timestamp and location data with an all-number flight ID (e.g. 197572) to the STDDS user. Once STDDS receives the flight ID, the API starts including the flight ID in the API messages, but cannot fix the prior all-number flight ID messages. Therefore, it is possible to recover a message's flight ID if it is close in terms of time and distance to a message from any of the API sources with a known flight ID. Once a single message in an all-number flight ID message series is recovered, that known flight ID number is applied to the entire trajectory.

For the flight ID recovery process, the all-number flight-ID messages were linked to the closest messages with a known flight ID based on three components: latitude, longitude and time. Only messages where a flight was traveling between 200-400 mph were considered for the merge. Flights traveling less than 200 mph are typically at the gate, taxiing or on the runway, meaning there are numerous other flights located within a short distance. Greater than 400 mph takes out flights flying near an airport that is not its origin or destination airport. We use miles per hour (mph) as the unit because this analysis contains both enroute and surface flights.

We used the KDTree capability in Python to find a message's nearest neighbor using the three components (43). This tree-based algorithm to defining the nearest neighbor is described in detail in Maneewong and Mount (44). Once the nearest neighbor was identified, it had to meet maximum latitude, longitude and time differences thresholds as the nearest neighbor may be located too far away to be the same flight. Therefore, an all-number flight ID was merged with a known flight ID if it met one of the following criteria:

- 1) Low-speed: if both flights are traveling 200-300 mph and are less than 0.005 miles apart and less than 0.002 hours apart.
- 2) High-speed: if both flights are traveling 300- 400 mph and are less than 0.015 miles apart and less than 0.004 hours apart.

1 An example of an arrival flight into John F. Kennedy International Airport (JFK) is shown in  
 2 **Figure 3**. It had an all-number (unknown) aircraft ID in STDDS that was found to be part of the  
 3 DAL951 flight using the KDTree algorithm. This portion of the trajectory is shown as the “Fixed  
 4 STDDS.” It looks like TFMS was slightly off the trajectory that the other sources showed because  
 5 we extracted the forecasted TFMS trajectory instead of actual to see the accuracy of the forecast  
 6 (i.e. target position). The actual position was in the XML data’s nested child (shown in **Figure 1**),  
 7 making it difficult to process into a database format. The TFMS forecast is accurate when the  
 8 flight travels a straight path but is not accurate during sharp turns in the terminal airspace.



9  
10 **Figure 3 Merging trajectory sources for a single arrival flight**

11 **RESULTS**

12 This section focuses on the API surface coverage of flights, which helps extract flight OOOI  
 13 times, allowing for an estimation of flight counts at airports. It shows coverage of individual  
 14 market segments and is organized as follows: commercial, general aviation, air cargo and airport  
 15 ground vehicle movements. This section ends with a flight count comparison across the flight  
 16 event and surveillance data sources during the study time period.

17  
18 **Market Segment**

19 The commercial flight coverage and the airlines included are based on a sensitivity analysis,  
 20 as defining an airline in the APIs is somewhat subjective. Some may suggest a company has to  
 21 own a certain number of planes and/or operate a certain number of flights minimum to be  
 22 considered an airline. **Table 4** outlines the sensitivity analysis of the number of airlines included  
 23 in each dataset by changing the minimum number of messages required to be considered an  
 24 included airline. Airlines with at least one continuous plane of messages during the study period  
 25 are in **Table 4**. The airlines with at least 10 are bolded. For example, the 10 planes scenario shows  
 26 airlines with at least 10 planes worth of continuous messages from a source. Using this logic,  
 27 TFMS has 1 message per minute per flight, so at least 1,200 messages from TFMS were needed  
 28 during the 2 hours to be considered. STDDS has the least number of airlines as the number of  
 29 messages needed does not consider the spatial coverage of each source.

30 The results show the APIs capture more airlines than the ASPM reporting carriers and the  
 31 ASQP reporting carriers. Additionally, the APIs receive information of all flights from the airlines.  
 32 This contrasts with ASQP where reporting carriers do not report non-revenue, deadheading flights.

1 Therefore, ASQP would better quantify passenger delay, while the APIs would provide a more-  
 2 robust estimate of plane delay.  
 3

4

**TABLE 4 List of airlines with at least 1 continuous plane of messages**

	SFDPS				STDDS		TFMS			OSN			
1	<b>AAL</b>	<b>DAL</b>	JSX	SKQ	<b>AAL</b>	JZA	<b>AAL</b>	<b>FDX</b>	ROU	<b>AAL</b>	DLH	LYM	THY
2	<b>AAV</b>	DAT	JTZ	<b>SKW</b>	AAV	LYM	<b>AAV</b>	FDY	<b>RPA</b>	<b>AAV</b>	<b>EDV</b>	MXV	TRF
3	ABX	DLH	<b>JZA</b>	SUB	ACA	<b>NKS</b>	ABX	<b>FFT</b>	SCX	ABX	<b>EJA</b>	NDU	TSC
4	<b>ACA</b>	DWI	KLM	<b>SWA</b>	AIP	PDT	<b>ACA</b>	FLE	SIA	<b>ACA</b>	<b>ENY</b>	<b>NKS</b>	<b>UAL</b>
5	AFR	<b>EDV</b>	LBQ	SWG	AMF	QXE	AFR	GJS	SKQ	AFR	EPI	OST	UCA
6	AIP	<b>EJA</b>	LXJ	SWQ	ASA	<b>RPA</b>	AIP	GTI	<b>SKW</b>	AIP	<b>ERU</b>	OXF	<b>UPS</b>
7	AJT	<b>ENY</b>	LYM	TAI	ASH	<b>SKW</b>	AMF	GXA	SUB	AMF	EVA	PCM	VAR
8	AMF	EPI	MRA	TAP	ASI	<b>SWA</b>	AMX	HAL	<b>SWA</b>	AMX	<b>FDX</b>	PDT	VIV
9	AMX	ERU	MXV	THY	ATN	<b>UAL</b>	ASA	IBE	SWG	ANA	FDY	POE	VJA
10	<b>ASA</b>	EVA	<b>NKS</b>	TSC	AWI	UCA	<b>ASH</b>	<b>JBU</b>	SWQ	<b>ASA</b>	<b>FFT</b>	PTR	VOI
11	<b>ASH</b>	<b>FDX</b>	PCM	<b>UAL</b>	BWR	UPS	ATN	<b>JIA</b>	TAP	<b>ASH</b>	FLE	QXE	VTE
12	ASI	FDY	<b>PDT</b>	UCA	<b>DAL</b>	WJA	AWI	JLG	THY	ASI	GJS	ROU	VTM
13	ATN	<b>FFT</b>	POE	<b>UPS</b>	<b>EDV</b>		AZU	JSX	TSC	ATN	GTI	<b>RPA</b>	VXP
14	AWI	FLE	PTR	VAR	EJA		BAW	JZA	<b>UAL</b>	AWI	GXA	SCA	WCP
15	AZU	FRG	PXT	VJA	<b>ENY</b>		BMJ	KLM	UCA	BAW	HAL	SCU	WEN
16	BAW	GJS	QFA	VOI	EVA		BVN	LBQ	<b>UPS</b>	BMJ	IBE	SCX	<b>WJA</b>
17	BMJ	GLO	QXE	VOS	FDX		CMP	LXJ	VAR	CAL	<b>JBU</b>	SIA	WSN
18	BVN	GTI	RAX	VTE	FDY		CXK	LYM	VJA	CJT	<b>JIA</b>	SKQ	
19	BWA	GXA	RJA	VTM	<b>FFT</b>		<b>DAL</b>	MRA	VOI	CKS	JLG	<b>SKW</b>	
20	CKS	HAL	ROU	VXP	GJS		DLH	MXV	VTE	CMP	JSX	SLI	
21	CMP	IBE	<b>RPA</b>	WCP	GTI		<b>EDV</b>	<b>NKS</b>	VXP	CSJ	<b>JZA</b>	<b>SWA</b>	
22	CPT	IRO	SCU	<b>WJA</b>	HAL		<b>EJA</b>	PCM	<b>WJA</b>	CSN	KAL	SWG	
23	CSJ	<b>JBU</b>	SCX	WSN	<b>JBU</b>		<b>ENY</b>	PDT	WSN	CXK	KLM	SWQ	
24	CTL	<b>JIA</b>	SIA		<b>JIA</b>		ERU	POE		<b>DAL</b>	LBQ	TAI	
25	CXK	JLG	SIL		JSX		EVA	QXE		DAT	LXJ	TAP	

5 *Note:* Bolded are airlines with at least 10 continuous planes of messages  
 6

7

8 General aviation is included in both FAA SWIM and OSN. General aviation is not  
 9 included in ASQP as it is not operated by the reporting carriers. Some general aviation flights  
 10 to/from the ASPM 77 airports are included in ASPM. **Figure 4** highlights the general aviation  
 11 STDDS capabilities at Salt Lake City Airport (SLC). For the purposes of this study, general  
 12 aviation was defined as flights with aircraft IDs beginning with N (i.e. N#). To get an accurate  
 13 count of all general aviation, the aircraft IDs would need to be compared with the FAA aircraft  
 14 registry (45). Using Google Maps (46), the general aviation facility location identified by the  
 flight tracks is validated in **Figure 4**.



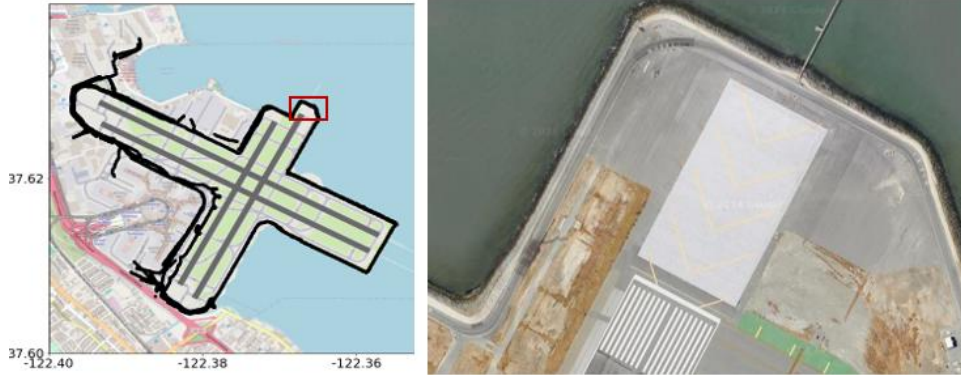
**Figure 4 Select general aviation at SLC (a) trajectories and (b) aerial view**

Similar to general aviation facilities, surveillance captures air cargo facility location and airline carrier usage using the flight trajectories (when the airport is instrumented). **Figure 5** shows the air cargo flights operated by Fedex (FDX) and United Parcel Service (UPS) based on the flightIDs provided by the API. The air cargo facilities located by the FDX and UPS flights are validated using Google Maps. Once the air cargo flight terminal locations are identified from the trajectories, the flight type can potentially be assumed for mixed use carriers based on which terminal the flight uses.



**Figure 5 Select FDX and UPS air cargo at SLC (a) trajectories and (b) aerial view**

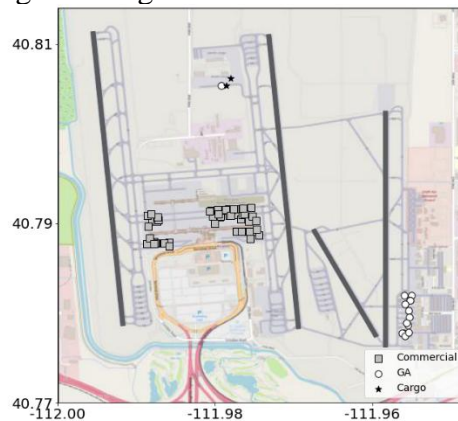
Another interesting operation included in FAA SWIM is airport ground vehicle movements, as shown in **Figure 6**. These operations only show up in STDDS and have an aircraft ID that start with “OPS” in the dataset. All vehicle movements are contained in the airport boundaries and sometimes located on airport roadways that would not fit an airplane. This is validated in **Figure 6** when Google Maps shows the vehicle locations with respect to access roads. These OPS airport ground vehicles are not included in any of the flight level datasets (ASPM, ASQP and OAG) or in OSN.



**Figure 6 Airport ground vehicles at SLC (a) trajectories and (b) aerial view**

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11

Combining the commercial, cargo and general aviation trajectories and parsing out the gate-out and gate-in locations shows the terminal locations for each market segment, as shown in **Figure 7**. Air cargo flights were defined based on the carriers that carried freight but not passengers in 2023 according to the T-100 segment (47). Commercial flights included in **Figure 7** are all flights from airlines that transported at least 100,000 passengers in 2023 according to the T-100 segment (47). This logic allows for easy facility identification surveillance-equipped airports. Not all flights have defined gate-out/gate-in events as some start emitting position information after already having left the gate or end before reaching the gate.



**Figure 7 Airport facility by market segment**

12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27

### Flight Counts by Data Source

**Table 5** shows the number of operations during the study period by airport and data source. The first and last parts of the study period were used to set-up the fused surveillance event extraction, therefore the counts in **Table 5** are for 01-20-2024 between 02:00-03:00 UTC. The fused sources detected wheels-off event when 1) a flight increases in speed by at least 55 mph in the runway polygons generated from the FAA (48) with a final speed of at least 100 mph or 2) the flight trajectory starts within 4 miles of an airport’s centroid, with the centroid coordinates defined by the FAA (48). The distance of 4 miles was chosen to capture the distance between runway ends and the centroid for large area airports, such as Denver (DEN) airport. The second criterion allows for flights taking off from an airport with poor surface coverage to still be counted. Wheels-on events had the opposite criteria as the wheels-off events. Miles per hour was used instead of knots due to being a surface speed. The speed thresholds were primarily defined based on 1) general aviation aircraft standards as it contains aircraft types with much lower approach speeds than

1 commercial aircraft types and 2) the accuracy of the flight location information. While 55 mph  
2 seems high for a taxiing speed, this constraint allows position imprecision between successive 1-  
3 second messages.

4 **Table 5** compares the wheels-off/wheels-on flight counts from the flight event datasets and  
5 extracted from the surveillance sources fused together. Flights missing both an aircraft ID and tail  
6 number in a data source were not included in the counts. There are several reasons for the  
7 discrepancies between the different counts. In addition to the differing reporting airlines and  
8 airports across the sources, general aviation flights have a higher coverage in the fused surveillance  
9 sources. However, if a flight's trajectory starts/ends more than 4 miles from an airport, it would  
10 not show up in the fused surveillance sources, but it would in OAG and possibly ASPM and ASQP  
11 if in the reporting airport or airline lists. Also, deadheading flights would be reported in all sources  
12 except ASQP. To add to the complexity, the timestamps in ASPM, ASQP and OAG are reported  
13 to the minute, but the surveillance data is to the second. This means the flights included in the first  
14 or last minute of the study period will differ. Even with a short study duration, **Table 5** already  
15 shows discrepancies in the count of flight operations across the different sources. This is due to  
16 several factors, most notably the differing spatial coverage and commercial, general aviation, cargo  
17 flight coverage, as discussed earlier. The discrepancies between sources would also be present  
18 with longer study timeframes.

19 Overall, it is difficult to answer the question of exactly how many total flights departed or  
20 arrived at an airport during a specific timeframe given the differing coverages of the sources. Each  
21 source contains a different subset of the flights, airports, airlines and market segments. In the case  
22 of SFB in **Table 5**, there are 4 wheels-on events that were reported by surveillance that were not  
23 included in any of the Flight Event sources. These wheels-on events were from an airline "LFA"  
24 as reported by OSN. It is possible that these flights occurred and were not reported by the other  
25 sources as SFB is not an ASPM airport, LFA is not a required airline to report to BTS's ASQP  
26 data and it may not be schedule service to be included in OAG. Alternatively, it could be a data  
27 error of the flight or the OSN system. There is no way to confidently conclude the true flight  
28 counts of SFB based on the sources in **Table 5** alone.

29 At this point, it is evident the raw surveillance sources cannot replace the flight event  
30 databases as sometimes the surveillance counts are lower compared to other datasets. Therefore,  
31 surveillance should be considered supplemental to the flight event datasets instead of a  
32 replacement. However, manipulating the raw surveillance data could improve the accuracy, where  
33 it is possible to fill in an estimated wheels-on timestamp when the flight's destination airport and  
34 speed is known along with the distance from the airport the flight lost signal. Also, surveillance  
35 does not include scheduled gate-out/gate-in times, so the flight event datasets are needed for that  
36 information.

37 These datasets can be used for quality checks on each other for major carriers and airports,  
38 but it is difficult to identify which source is correct when there are discrepancies in flight counts.  
39 This is also an issue when the OOOI times do not agree across sources. Further, most researcher  
40 do not have access to all flight event and surveillance sources to make these quality checks. Future  
41 research could consider how to improve flight event data coverage available to researchers, while  
42 ensuring the quality of the data and listing its limitations clearly. For example, some Visual Flight  
43 Rules (VFR) flights and military flights do not show up in any of the datasets.

1

**TABLE 5 Wheels-off/wheels-on flight counts by source**

	Flight Event			Surveillance Sources
	ASPM	ASQP	OAG	Fused OSN, SFDPS, STDDS
<b>STDDS AIRPORT</b>				
<b>Large</b>				
<b>ATL</b>	68/75	69/72	67/72	75/82
<b>LAX</b>	30/43	21/30	26/42	37/32
<b>MSP</b>	33/29	36/17	33/25	32/20
<b>ORD</b>	69/39	62/25	71/40	93/49
<b>Medium</b>				
<b>SNA</b>	6/19	4/13	9/14	8/18
<b>STL</b>	7/12	8/13	7/11	7/10
<b>Small</b>				
<b>PVD</b>	1/1	0/0	1/1	1/1
<b>SDF</b>	1/9	2/5	1/8	1/8
<b>NOT STDDS AIRPORT</b>				
<b>Medium</b>				
<b>RDU</b>	6/16	4/16	5/14	4/16
<b>SMF</b>	12/15	8/11	9/11	0/0
<b>Small</b>				
<b>BUF</b>	2/2	3/2	2/3	3/3
<b>SFB</b>	1/0	1/0	1/1	1/5

2

3

**LIMITATIONS OF RESEARCH**

4

5

6

7

8

9

10

11

12

13

**DISCUSSION OF FUTURE RESEARCH**

14

15

16

17

18

19

20

21

22

23

24

Flight tracking APIs provide flight location and timestamp information, making possible aviation operations research studies that could not have been conducted before. Now with the spatial and market segment coverage of each API known, this section discusses future research directions. This discussion is not exhaustive as there are numerous ways the API data could be used.

In addition to the data quality control research area previously discussed, the first research area that could be analyzed is trends in the general aviation sector. Prior to the APIs, there was not sufficient data on this available to researchers. General aviation demand could be quantified during shock events, similar to the Coronavirus Disease of 2019 (COVID-19) or a recession. Also, the impact of new Advanced Air Mobility (AAM) technologies on general aviation could be measured when implemented.

1 The second research area made possible by surveillance is airport facility usage. Researchers  
2 can analyze how terminal facilities (e.g. commercial, general aviation, cargo) are used today and  
3 determine which airport surface Geographic Information Systems (GIS) layouts (i.e. aprons,  
4 taxiways, runways) promote a more efficient taxiing operation, in terms of distance, time, logistics  
5 and equity. On a more refined scale, an airport layout analysis with the APIs could quantify the  
6 capacity of a double lane apron area instead of a single lane apron area and other similar capacity  
7 comparisons of airport elements.

8 The third is enroute efficiency, where the throughput and capacity of airways could be  
9 quantified in near-real time. This would also provide insight into the XTM system needs and  
10 potential challenges of AAM implementation. Specifically, AAM would likely be restricted from  
11 airspace used by non-AAM flights in order to not have to contact ATC. These APIs could show  
12 where in the US non-AAM flights operate at AAM flight levels that would cause airspace conflicts.  
13 This would occur mostly near airports.

14 While flight tracking APIs have the ability to expand operations research with the previously  
15 discussed topics, the field of aviation revenue management will not be able to take advantage of  
16 the APIs and need to still rely on airline schedules data from OAG. Revenue management models  
17 require information about all feasible alternatives (i.e. commercial flight alternatives). The APIs  
18 do not capture all commercial flights as sometimes the flight event counts in ASQP were higher  
19 than those from surveillance.

## 20 21 **CONCLUSIONS**

22 ADS-B mandates along with spatial coverage improvements in both government and third-  
23 party flight tracking APIs have made it possible for researchers to analyze aspects of aviation  
24 operations that were not possible beforehand. This analysis reported the spatial and market  
25 segment coverage of each source for researchers to identify the best data for their studies and  
26 limitations associated with each. It was found that OSN had a custom user-defined flight message  
27 frequency. It has worldwide coverage and, therefore, had flights across the US, including Alaska  
28 and Hawaii. Its focus is in the enroute flight phase, with limited to no surface information. In  
29 FAA SWIM, the message rate is set by the FAA, ranging from every second by STDDS and every  
30 minute by TFMS. There are two enroute datasets (SFDPS and TFMS), with only TFMS containing  
31 Alaska and Hawaii, and one airport surface movements dataset STDDS. Combining sources is  
32 possible, where one source can overcome a limitation of another. Also, the APIs cover more  
33 airlines than the mandated reporting airlines in ASQP and ASPM. Overall, these API technologies  
34 can support research studies at a highly disaggregate level for studies in several areas of aviation,  
35 but cannot be considered a replacement for the information provided in the flight event datasets.

36

1 **DATA AVAILABILITY**

2 The datasets generated and/or analyzed during the current study have different data availabilities.  
3 ASQP is publicly-available on the BTS website. All other datasets in this study are not publicly-  
4 available. The flight-level ASPM, SWIM and OSN data require data user agreements with the  
5 data providers. OAG is accessible via a paid data subscription.  
6

7 **AUTHOR CONTRIBUTIONS**

8 The authors confirm contribution to the paper as follows: study conception and design: S. Hotle,  
9 K. Titlow, M. Hashemipour, E. Strocko; data collection: S. Hotle, K. Titlow, M. Hashemipour;  
10 analysis and interpretation of results: S. Hotle and K. Titlow; draft manuscript preparation: S.  
11 Hotle, K. Titlow. All authors reviewed the results and approved the final version of the manuscript.  
12

13 **DECLARATION OF CONFLICTING INTERESTS**

14 The authors declared no potential conflicts of interest with respect to the research, authorship,  
15 and/or publication of this article.  
16

17 **FUNDING**

18 The authors disclosed receipt of the following financial support for the research, authorship, and/or  
19 publication of this article: This research was supported by the U.S. Department of Transportation  
20 (DOT) through the Oak Ridge Institute for Science and Education (ORISE) Program.  
21

22 **REFERENCES**

- 23 1. Federal Aviation Administration (n.d.). SWIM News.  
24 [https://www.faa.gov/air\\_traffic/technology/swim/news](https://www.faa.gov/air_traffic/technology/swim/news)  
25
- 26 2. Federal Aviation Administration (n.d.) ASPM: Individual Flights: Definitions of Variables.  
27 [https://aspm.faa.gov/aspmhelp/index/ASPM\\_\\_Individual\\_Flights\\_\\_Definitions\\_of\\_Variables.ht](https://aspm.faa.gov/aspmhelp/index/ASPM__Individual_Flights__Definitions_of_Variables.htm)  
28 [ml](https://aspm.faa.gov/aspmhelp/index/ASPM__Individual_Flights__Definitions_of_Variables.htm)  
29
- 30 3. Kang, L., & Hansen, M. (2018). Improving airline fuel efficiency via fuel burn prediction  
31 and uncertainty estimation. *Transportation Research Part C: Emerging Technologies*, 97, 128-  
32 146.  
33
- 34 4. Wang, L., Mykityshyn, A., Johnson, C., & Marple, B. D. (2020). Deep learning for flight  
35 demand forecasting. *arXiv preprint arXiv:2011.04476*.  
36
- 37 5. Puranik, T. G., Memarzadeh, M., & Kalyanam, K. M. (2023, October). Predicting airport  
38 runway configurations for decision-support using supervised learning. In *2023 IEEE/AIAA 42nd*  
39 *Digital Avionics Systems Conference (DASC)* (pp. 1-9). IEEE.  
40
- 41 6. Alsalous, O., & Hotle, S. (2021). Modeling Arrival Flight Times within the Terminal  
42 Airspace. *Transportation Research Record*, 2675(10), 1-13.  
43
- 44 7. Hotle, S., Baszczewski, B., & Gulding, J. (2016). Using ASDE-X Surveillance for Taxi-Out  
45 Time Benchmarking and Delay Estimation. In *16th AIAA Aviation Technology, Integration, and*  
46 *Operations Conference* (p. 3749).

- 1 8. Balakrishna, P., Ganesan, R., & Sherry, L. (2010). Accuracy of reinforcement learning  
2 algorithms for predicting aircraft taxi-out times: A case-analysis of Tampa Bay  
3 departures. *Transportation Research Part C: Emerging Technologies*, 18(6), 950-962.  
4
- 5 9. Esmaeilzadeh, E., & Mokhtarimousavi, S. (2020). Machine learning approach for flight  
6 departure delay prediction and analysis. *Transportation Research Record*, 2674(8), 145-159.  
7
- 8 10. Ohi, S. J., & Kim, A. M. (2020). Count models to represent the impacts of weather and  
9 infrastructure on flight disruptions. *Transportation Research Record*, 2674(5), 510-521.  
10
- 11 11. Federal Aviation Administration (n.d.). FAA Operations & Performance Data  
12 <https://aspm.faa.gov/>  
13
- 14 12. Federal Aviation Administration (n.d.) The Operations Network (OPSNET)  
15 <https://aspm.faa.gov/opsnet/sys/main.asp>  
16
- 17 13. National Archives and Records Administration (1987). Part 234- Airline Service Quality  
18 Performance Reports. <https://www.ecfr.gov/current/title-14/chapter-II/subchapter-A/part-234>.  
19
- 20 14. Federal Aviation Administration (n.d.) Airline Service Quality Performance (ASQP).  
21 [https://aspm.faa.gov/aspmhelp/index/Airline\\_Service\\_Quality\\_Performance\\_\(ASQP\).html#:~:te  
22 xt=Starting%20in%20CY%202018%2C%20air,domestic%20scheduled%20passenger%20carrier  
23 %20revenues](https://aspm.faa.gov/aspmhelp/index/Airline_Service_Quality_Performance_(ASQP).html#:~:text=Starting%20in%20CY%202018%2C%20air,domestic%20scheduled%20passenger%20carrier%20revenues).  
24
- 25 15. U.S. Department of Transportation (2023) Fly Rights  
26 <https://www.transportation.gov/airconsumer/fly-rights>  
27
- 28 16. Yan, C., & Kung, J. (2018). Robust aircraft routing. *Transportation Science*, 52(1), 118-133.  
29
- 30 17. Truong, D. (2016). Developing Airline Segmentation Based on The On-time  
31 Performance. *Int J Aeronautics Aerospace Res*, 3(5), 131-140.  
32
- 33 18. Goodman, C. J., & Griswold, J. D. S. (2019). Meteorological impacts on commercial aviation  
34 delays and cancellations in the continental United States. *Journal of applied meteorology and  
35 climatology*, 58(3), 479-494.  
36
- 37 19. Atallah, S., & Hotle, S. (2022). Evaluation of Airport Size and Delay Causal Factor Effects  
38 on Delay Propagation Dissipation. *Transportation Research Record*, 2676(3), 608-620.  
39
- 40 20. IFLYCU (2024). Flight Tracker. <https://iflycu.com/info/flightrackingmenu/>  
41
- 42 21. Atallah, S., & Hotle, S. L. (2019). Assessment of contributing factors to air service loss in  
43 small communities. *Transportation Research Record*, 2673(6), 598-606.  
44
- 45 22. Atallah, S., Hotle, S. L., & Mumbower, S. (2018). The evolution of low-cost Carrier  
46 operational strategies pre-and post-recession. *Journal of Air Transport Management*, 73, 87-94.

- 1 23. Mumbower, S. (2022). Airline market exit after a shock event: Insights from the COVID-19  
2 pandemic. *Transportation Research Interdisciplinary Perspectives*, 14, 100621.  
3
- 4 24. Bockelie, A., & Belobaba, P. P. (2018). Total revenue optimization with the ancillary  
5 marginal demand and ancillary marginal revenue transformation heuristics. In *AGIFORS 58th*  
6 *Annual Symposium (Anna Valicek Paper Competition)*, Tokyo, Japan.  
7
- 8 25. Bureau of Transportation Statistics (n.d.) Air Carrier Statistics (Form 41 Traffic)- All  
9 Carriers <https://www.transtats.bts.gov>  
10
- 11 26. Bureau of Transportation Statistics (n.d.) Airline Origin and Destination Survey (DB1B)  
12 <https://www.transtats.bts.gov>  
13
- 14 27. Bureau of Transportation Statistics (n.d.) Origin and Destination Survey  
15 [https://www.transtats.bts.gov/Fields.asp?gnoyr\\_VQ=FGJ](https://www.transtats.bts.gov/Fields.asp?gnoyr_VQ=FGJ)  
16
- 17 28. Federal Aviation Administration (n.d.) Airline Service Quality Performance (ASQP)  
18 <https://aspm.faa.gov/aspmhelp>  
19
- 20 29. OAG (n.d.) On-Time Performance Flight Data. [https://www.oag.com/on-time-performance-](https://www.oag.com/on-time-performance-data)  
21 [data](https://www.oag.com/on-time-performance-data)  
22
- 23 30. Federal Aviation Administration (n.d.) Aviation System Performance Metrics (ASPM).  
24 [https://aspm.faa.gov/aspmhelp/index/Aviation\\_Performance\\_Metrics\\_%28APM%29.html](https://aspm.faa.gov/aspmhelp/index/Aviation_Performance_Metrics_%28APM%29.html)  
25
- 26 31. Federal Aviation Administration (n.d.) FAA Operations & Performance Data.  
27 <https://aspm.faa.gov/Control/Users/sysMailTo.asp>  
28
- 29 32. Deshannavar, U. (2022). High Dimensional Weather Data Used in a Deep Generative Model  
30 to Predict Trajectories of Aircraft. *Journal of Airline Operations and Aviation*  
31 *Management*, 1(1), 80-88.  
32
- 33 33. Sun, J., Basora, L., Olive, X., Strohmeier, M., Schäfer, M., Martinovic, I., & Lenders, V.  
34 (2022, September). OpenSky Report 2022: Evaluating Aviation Emissions Using Crowdsourced  
35 Open Flight Data. In *2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC)* (pp. 1-  
36 8). IEEE.  
37
- 38 34. Federal Aviation Administration (2023) SWIM Questions & Answers  
39 [https://www.faa.gov/air\\_traffic/technology/swim/questions\\_answers](https://www.faa.gov/air_traffic/technology/swim/questions_answers)  
40
- 41 35. Federal Aviation Administration (2024) FAA’s SWIM Flight Data Publication Service  
42 (SFDPS) Features Enhanced Location Data  
43 [https://www.faa.gov/air\\_traffic/technology/swim/news](https://www.faa.gov/air_traffic/technology/swim/news)  
44
- 45 36. OpenSky Network (n.d.) The OpenSky Network. <https://opensky-network.org/about/about-us>  
46

- 1 37. Matthias Schäfer, Martin Strohmeier, Vincent Lenders, Ivan Martinovic and Matthias  
2 Wilhelm. "Bringing Up OpenSky: A Large-scale ADS-B Sensor Network for Research".  
3 In *Proceedings of the 13th IEEE/ACM International Symposium on Information Processing in*  
4 *Sensor Networks (IPSN)*, pages 83-94, April 2014.
- 5
- 6 38. OpenSky Network (n.d.) Get a Receiver. [https://opensky-network.org/contribute/get-a-](https://opensky-network.org/contribute/get-a-receiver)  
7 receiver
- 8
- 9 39. Federal Aviation Administration (n.d.) SWIFT Portal. <https://portal.swim.faa.gov/>
- 10
- 11 40. OpenSky Network (2021) The OpenSky Network API documentation <https://OpenSky>  
12 [Network.github.io/opensky-api/](https://OpenSkyNetwork.github.io/opensky-api/)
- 13
- 14 41. Federal Aviation Administration (2023) Airport Surface Detection Equipment, Model X  
15 (ASDE-X) [https://www.faa.gov/air\\_traffic/technology/asde-x](https://www.faa.gov/air_traffic/technology/asde-x)
- 16
- 17 42. Federal Aviation Administration (2023) ADS-B Airport Surface Surveillance Capability  
18 (ASSC) [https://www.faa.gov/air\\_traffic/technology/adsb/atc/assc](https://www.faa.gov/air_traffic/technology/adsb/atc/assc)
- 19
- 20 43. SciPy (2024) `scipy.spatial.KDTree`  
21 <https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.KDTree.html>
- 22
- 23 44. Maneewongvatana, S., & Mount, D. M. (1999). Analysis of approximate nearest neighbor  
24 searching with clustered point sets. arXiv preprint [cs/9901013](https://arxiv.org/abs/cs/9901013).
- 25
- 26 45. Federal Aviation Administration (n.d.) Aircraft Inquiry.  
27 <https://registry.faa.gov/aircraftinquiry>
- 28
- 29 46. GoogleMaps (2024). GoogleMaps. <https://www.google.com/maps>
- 30
- 31 47. Bureau of Transportation Statistics (n.d.) Air Carriers: T-100 Segment (All Carriers).  
32 [https://www.transtats.bts.gov/Fields.asp?gnoyr\\_VQ=FMG](https://www.transtats.bts.gov/Fields.asp?gnoyr_VQ=FMG)
- 33
- 34 48. Federal Aviation Administration (2024) 28 Day NASR Subscription.  
35 [https://www.faa.gov/air\\_traffic/flight\\_info/aeronav/Aero\\_Data/NASR\\_Subscription/](https://www.faa.gov/air_traffic/flight_info/aeronav/Aero_Data/NASR_Subscription/)