engineering proceedings

MDPI

Proceeding Paper

Reference Trajectories: The Dataset Enabling Gate-to-Gate Flight Analysis †

John Fitzgerald [1,*], Enrico Spinielli [2], Allan Tart [1] and Rainer Koelle [2]

1 OpenSky Network, Eyzälg 23, 3400 Burgdorf, Switzerland; allantart@gmail.com
2 Eurocontrol, Rue de la Fusée 96, 1130 Brussels, Belgium; enrico.spinielli@eurocontrol.int (E.S.); rainer.koelle@eurocontrol.int (R.K.)
* Correspondence: fitzgerald@opensky-network.org

† Presented at the 9th OpenSky Symposium, Brussels, Belgium, 18–19 November 2021.

Abstract: Without a doubt, a publicly verifiable data is required to ensure a strong, transparent and independent air traffic management performance review system. Community sourced data (such as ADS-B/Mode S provided by OpenSky Network and others alike) has been used to analyse different aspects of air traffic management. The main drawback of such ADS-B data is the lack of crucial pieces of information that need to be inferred. On the other hand, Eurocontrol has used correlated position reports (CPRs) gathered from European Air Navigation Service Providers (ANSP) to conduct some of its actual/flown trajectory oriented performance analysis. The availability and the granularity of the CPRs vary between Eurocontrol Member States, making it difficult to perform accurate wide-scale studies. Using the strengths of both data sources would obviously result in great benefits. This paper describes the first step in creating a pan-European Flight Table (FT) and its supporting reference trajectories (RT). It is expected that the resulting dataset will be made available for the general public and that the work will continue to improve in scope and accuracy.

Keywords: open data; data fusion; crowdsourced data

Citation: Fitzgerald, J.; Spinielli, E.; Tart, A.; Koelle, R. Reference Trajectories: The Dataset Enabling Gate-to-Gate Flight Analysis. *Eng. Proc.* **2021**, *13*, 14. https://doi.org/10.3390/engproc2021013014

Academic Editor: Junzi Sun

Published: 28 January 2022

1. Introduction

Societal challenges and political priorities revolving around the operational efficiency of air transport operations and the associated environmental footprint will require higher levels of transparency and publicly verifiable data. In order to ensure a strong, transparent and independent air traffic management performance review system, it is important to integrate and validate data from different sources in order to create a value added set of analytic data.

The Operational Performance Review Service of the Aviation Intelligence Unit (AIU), a unit within Eurocontrol Agency's Directorate European Civil-Military Aviation, has used correlated position reports (CPRs) gathered from European air navigation service providers (ANSP) to conduct some of its actual/flown trajectory oriented performance analysis. Essentially the CPRs are air traffic control (ATC) radar position reports correlated with flight information. So next to the 4D surveillance position (location and timestamp), CPRs provide data fields for the aerodrome of departure (ADEP), aerodrome of destination (ADES), and aircraft identification. The availability of CPR data varies between different Eurocontrol member states. In some areas, Eurocontrol has no or partial CPR coverage due to the fact that either member states are not delivering this data feed at all or are reducing the frequency of the reports.

Community sourced surveillance position data, like ADS-B/Mode S, has been used to analyse different aspects of air traffic management [1–4]. However, the move from surveillance sensor information to flight-oriented analytic data is a challenge. The main drawback of such ADS-B data is the lack of crucial pieces of information that need to be inferred (e.g., ADEP/ADES, on/off-block times, take-off/landing times, and airspace intersection). Developing and implementing algorithms for inferring the aforementioned information is not a trivial task.

Under the sponsorship of Eurocontrol's Performance Review Commission, AIU together with OpenSky Network has established a joint project to create a comprehensive dataset that researchers from different fields can use without the need to replicate data preparatory steps and spend too much effort to "fill in the gaps" within the CPRs and crowdsourced surveillance data themselves.

The aim of the project is to establish a pan-European Flight Table (FT) and its supporting reference trajectories (RT).

The FT hold some key data and useful milestones about the flight (e.g., entry/exit points together with respective 4D positions, ADEP/ADES, take-off runway, top of climb, etc.).

The RT, on the other hand, is essentially a collection of 4D positions. The link between the tables is established using ICAO 24 bit aircraft address, call-signs or any other data that help uniquely identify a unique flight and its flown path.

The scope of FT and RT is defined as the point of entry and exit within the European air navigation system. For inner-European flights this represents the parking position at departure aerodrome and parking position at destination. Thus, entry and exit will reflect the respective entry- or exit-4D position for flights arriving to or departing from European aerodromes. However, it may also include flights departing outside Europe with a destination in Europe. In this case, the entry position corresponds to the intersection of the flight path with the European airspace volume or vice versa for European departures. Further enriched flight table information can be constructed from the trajectory information. The respective movement milestones expand the pure surveillance data and establish an analytic dataset for operational performance analysis and other air transport monitoring. For example the actual take-off time of a flight can be determined based on the open crowdsourced data.

This paper provides a comprehensive overview of the dataset. In addition to describing the data and its key features, algorithms used to build the dataset are described. This includes the merging procedure of the two datasets, ADEP/ADES estimation/confirmation procedure, take-off/landing time estimation, off-/on-block time estimation.

Inevitably, there are data gaps within the RT due to the particularities of underlying datasets. We will highlight and quantify those gaps in dedicated sections in the paper. By doing so, we will perpetuate the way for further discussions about how to improve the RT quality.

2. Input Data

The first iteration of FT and RT is based on two datasets:

1. Correlated position reports (CPRs) provided by Eurocontrol and
2. `flights_data4` table provided by OpenSky Network.

Due to the nature of the information contained in the CPRs, this dataset is not publicly accessible, whereas `flights_data4` is available trough Impala Shell maintained by OpenSky Network. Access to latter is granted according to the OpenSky Network Terms of Use (https://opensky-network.org/about/terms-of-use, accessed on 26 January 2022). Before taking a closer look at RTs let us give an overview about the the dataset that the RTs are based on.

2.1. Correlated Position Reports

Air traffic flow management in Europe is performed by the Network Manager (NM) in collaboration with stakeholders shown on Figure 1. NM's overarching objective is to provides the means to measure and balance demand (Aircraft Operators willingness to fly) and capacity (availability of Air Traffic Controllers) so as to maintain the highest safety levels. To provide real-time updates to flight plan based and pre-calculated trajectories, European air navigation service providers (ANSPs) send correlated position reports (CPRs) to the NM. These updates support a network level air traffic flow situation awareness, and as such improve the tactical traffic counts for the European air traffic control sectors. Potential bottlenecks (more flights than a sector can handle) can be identified and flow management initiatives taken. In essence CPRs are aircraft location reports (radar reports) correlated with flight information (i.e., flight identification, assigned Mode 3/A squawk).

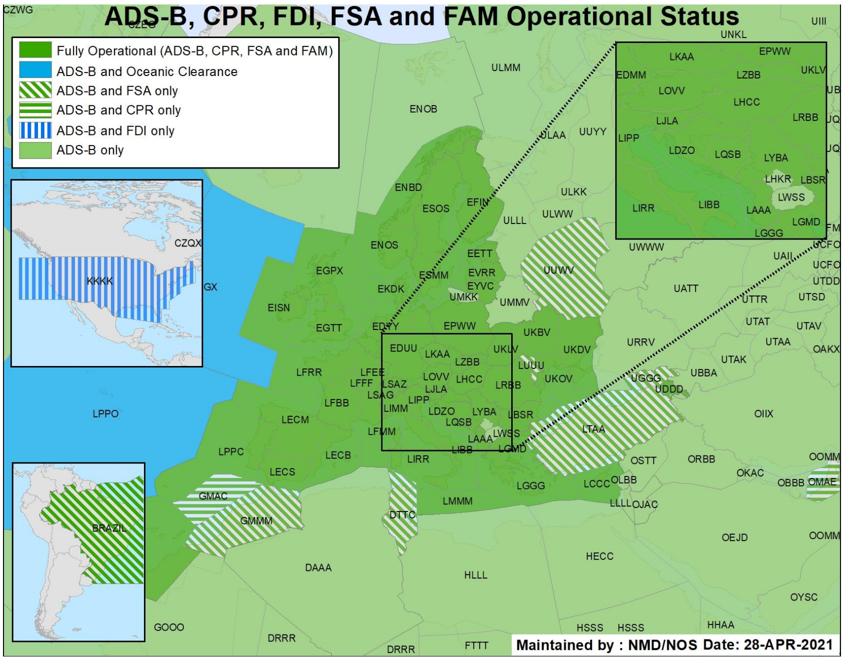

Figure 1. ADS-B, CPR, FDI FSA and FAM operational status at Eurocontrol (https://www.eurocontrol.int/service/data-collection-service, accessed on 26 January 2022).

The CPRs that correspond to a given flight are given a unique identifier, TACT ID, and are stored daily in a file of around 200 MB. Due to the particularities of the data archiving processes, the CPRs corresponding to a given flight for a given day might be stored in multiple files.

The time period of the current study is September 2019.The corresponding 258,050,101 correlated position reports originate from 34 countries.

In essence, the CPRs should give a comprehensive overview about any given flight taking place in European airspace. Post-operational analyses can build on the historic CPRs. However, limitations for performance monitoring exist. For example, the scope of the NM role focuses on the airborne part of a flight. Thus, CPRs do not cover the airport surface movements and start/break off in the proximity of an aerodrome. Varying update rates of the CPRs (ranging from 30 s to over 1 min) and differing surveillance data processing/tracking systems by the member states result in a low-fidelity trajectory that may miss flight events (e.g., start/end of turns, start/end of climb/descent) that impact the utility for performance analyses. It is also worth noting and is also shown on Figure 1 that there are areas where Eurocontrol is missing CPRs from partners and hence, the FT and RT constructions need to do without this information.

2.2. OpenSky Network Flights Table

Crowdsourced open surveillance data initiatives such as OpenSky Network offer access to high-fidelity air traffic trajectory data. The second dataset used for creating the Flight Table (FT) and its corresponding reference trajectories (RT) comes from the `flights_data4` table.

These 'flights' in OpenSky Network are aggregated representations of the path flown by the aircraft. It means that not all the position reports (those are stored in `state_vectors` _data4 table) are given, instead significant way-points are stored in the table. The rules about how the way-points are set and stored are:

- The first point is set immediately after the the aircraft's expected departure, or after the network received the first position when the aircraft entered its reception range.
- The last point is set right before the aircraft's expected arrival, or the aircraft left the networks reception range.
- There is a way-point at least every 15 min when the aircraft is in-flight.
- A way-point is added if the aircraft changes its track more than 2.5°.
- A way-point is added if the aircraft changes altitude by more than 100 m (330ft).
- A way-point is added if the on-ground state changes.

In general, a flight is uniquely identifiable by its ICAO 24 bit aircraft address, its call-sign and the date and time the flight took place.

In addition to the aggregated flown path, the table also includes estimated departure and arrival airports for the given flight.

As a flights_data4 table contains much less data about a aircraft trajectory than state_vectors4 and therefore requires less effort to handle, it is well suited for analysis where high position report granularity is not needed.

The main weak point of the flights representation of flown trajectory in OpenSky Network is that it heavily depends on network coverage. If an airport is located in poorly covered region, the flights from/to that airport are well under represented.

There were more than 2.9 million flights in the flights_data4 table that took place in September 2019. Only a portion of them departed and/or arrived to Europe and potentially would have a match in CPRs.

3. Flight Table and Associated Reference Trajectories

The goal of the project is to create a Flight Table (FT) and its associated reference trajectories (RTs) that contains all the details necessary to perform in-depth analysis about air navigation efficiency in European airspace while keeping the amount and complexity of the data relatively low. The project builds on the currently utilised performance indicators for assessing operational air navigation system efficiency. The set of indicators could be replicated with the proposed Flight Table.

It is well understood that the Flight Table will not be the go-to solution for all the use cases. For example, analysis where high position report update frequency is expected (e.g., safety incident investigations) will benefit from the underlying trajectory.

The aim is to logically provide the following information:

- Flight identifier
- Aircraft identifier i.e., ICAO 24 bit address and registration;
- ADEP/ADES;
- Off-/on-block time, i.e., time of aircraft leaving gate and arriving to gate;
- Runway entry/exit time includes take-off and landing times;
- 40 nmi 4D position, i.e., (lon/lat/altitude/timestap) with respect to departure (destination) airport reference point (ARP);
- 100 nmi 4D position;
- Top of climb, i.e., 4D position of aircraft reaching its cruising altitude;
- Flight trajectory, i.e., sequence of 4D positions (from state_vectors4)

During the first iteration we plan to create a FT table that is based on the flights_data4 table in OpenSky Network. In this table, (a) the ADES/ADEP data fields are updated with the data received from the CPRs and (b) the flights_data4 table is augmented with the take-off/landing times. In order to do the things listed above, the CPRs and flights in flights_data4 table need to be linked together. In the following subsection, a short overview about joining the datasets is given.

3.1. The Joining Procedure

In an ideal world, there would be a single unique key to join two related datasets. However, the underlying paradigm of both sources is also reflected in the data fields. The CPR exchange format dates back to times where aircraft where uniquely identified based on their control region and squawk (i.e., Mode 3/A). The continual growth of air traffic and the increasing limitations of "traditional" secondary surveillance systems then lead to the introduction of Mode S and the 24 bit aircraft address (sometimes referred to as 'ICAO 24 bit address').

Given that the ICAO 24 bit address (`arcaddress` in CPR and `icao24` in `flights_data4` table) was a not always present in the CPR data it has been decided to use the callsign (`aircraftId` in CPR and `callsign` in `flights_data4` table) to join on.

Due to the fact that there can be up to 10% of flights a day that reuse a callsign, there is a need to differentiate between these potential clashes in matching CPR position reports with state vector ones.

The matching is then performed adopting a temporal heuristics whereby a comparison of `firstseen` and `lastseen` data points in OpenSky Network `flights_data4` table is performed with `etfmsTimestamp` data field in CPR table. In the future iterations, in order to improve the matching robustness, it might be beneficial to include some spatial heuristics into the join algorithm as well. As such longitude/latitude values and/or altitude, heading and speed might also be considered.

On average, more than 77% of the CPRs were linked to a flight in `flights_data4` table.

For approximately 87% of those CPRs that did not have positive match, the callsign was not present in the OpenSky Network table. The flight might have taken place outside OpenSky Network coverage area or belong to a VFR flight for which there is generally no records in Eurocontrol because a relevant flight plan has not been submitted.

For the remaining set of CPRs, that did have matching callsign in OpenSky Network flights table, but did not end-up with a positive match, further investigation is needed to find a root cause of the issue.

3.2. Improving ADEP/ADES Estimates

The first step in enhancing the `flights_data4` was to improve the departure/arrival airports using the CPRs. As the CPR make use of actual flight information from filed flight plans, the ADEP/ADES fields in CPRs are considered as ground truth. So, for all those flights present in `flights_data4` table, which have a matching CPRs associated with them, a check is conducted between `estDepartureAirport(estArrivalAirport)` in `flights_data4` and `ades` (`adep`) in `cpr` respectively. If a mismatch is found, data from CPR is taken and substituted into the OpenSky Network flights table.

DEPARTURE AIRFIELD:

For approximately 75% of the cases the CPRs and the flights in OpenSky Network considered table have a matching ADEP fields. From the 25% of mismatching departure airports, approximately 20% of cases, OpenSky Network table had `NULL` - meaning, the OpenSky Network departure airfield estimation was not able to provide a viable estimate. For the remaining 5% of the cases, the algorithm produced a wrong estimate. This might be caused by the fact that if two or more airfields are in close proximity and OpenSky Network does not have a good coverage in the area, the *first seen* position of an aircraft is closer to the "wrong" airport. An example of such case is shown on Figure 2.

The statistic about the ADEP predictions in OSN database is shown on Figure 3. It reveals that approximately 75% of time, the OSN `flights_data4` table contains right departure airport, in approximately 20% of cases, the respective field in the table is empty, and in approximately 5% of cases, the prediction proved to be wrong.

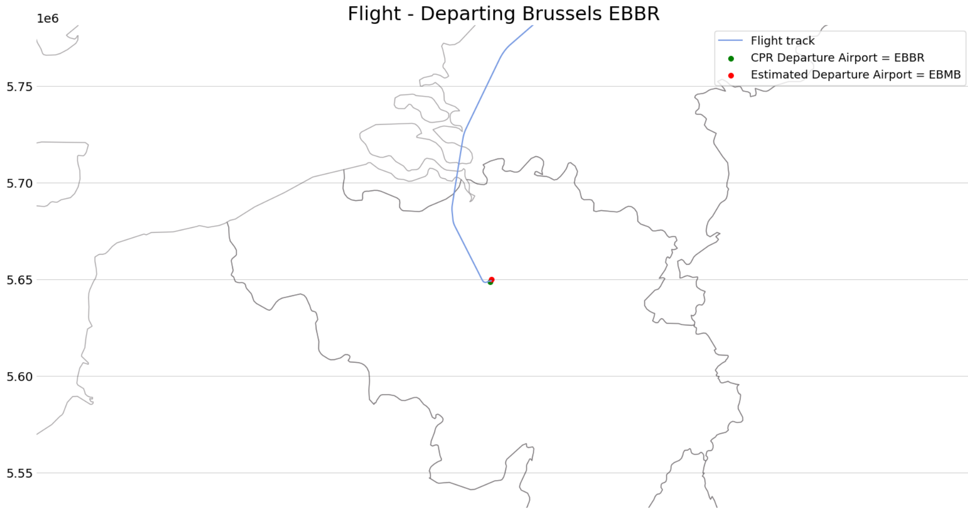

Figure 2. An example of wrong ADEP: departing flight from Brussels.

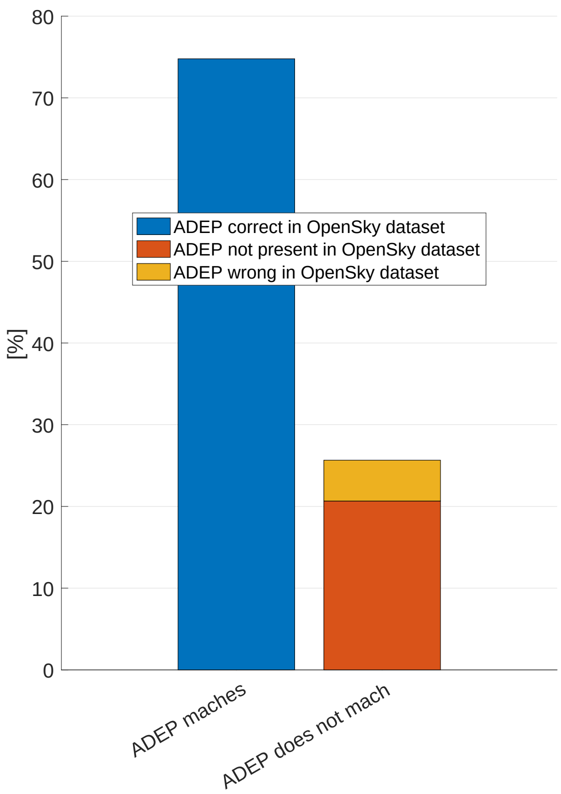

Figure 3. Some statistics about the accuracy of ADEP predictions in OSN database.

ARRIVAL AIRFIELD:

For approximately 57% of the cases, the CPRs and the flights in OpenSky Network considered table have a matching ADEP fields. From the 43% of mismatching departure airports, approximately 33% of cases OpenSky Network table had `NULL` - meaning, the OpenSky Network departure airfield estimation was not able to provide viable estimate. For the remaining 10% of the cases, the algorithm produced a wrong estimate. As for the departing case, this might be caused by the fact that two or more airfields are in close proximity and OpenSky Network does not have a good coverage in the area, the *last seen* position of an aircraft is closer to "wrong" airport.

The statistic about the ADEP predictions in OSN database is shown on Figure 4. It reveals that less than 60% of time, the OSN `flights_data4` table contains right departure airport, in approximately 35% of cases, the respective field in the table is empty and in approximately 9% of cases, the prediction proved to be wrong.

The cause for rather large ADEP and ADEP prediction accuracy differences need to be investigated further.

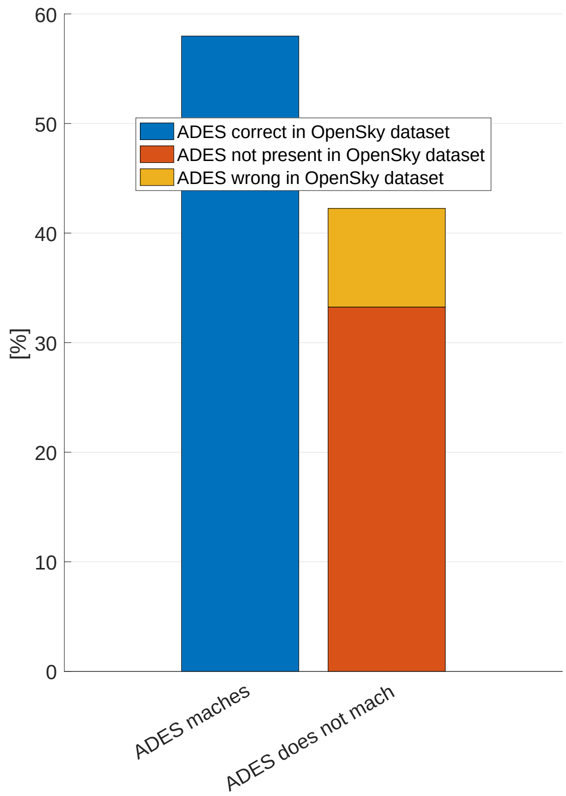

Figure 4. Some statistics about the accuracy of ADES predictions in OSN database.

3.3. Take-off Time Estimation

Although, the CPRs contain a field called 'Estimate Take-Off Date and Time' (EOBT) it is considered unreliable as it corresponds to so called scheduled take off time and this is usually quite different from the actual take off time.

Position report contained in the `state_vectors_data4` are used for estimating the actual take-off time.

In short the procedure works as follows: A sequence of trajectory way-points are selected from the beginning of the flight so to satisfy the following criteria:

- Distance to airport reference point (ARP) is less than 7500 m (A search area for Frankfurt airport is shown on Figure 5 as an example);
- Altitude is less than 1524 m above the aerodrome;
- Velocity is greater than 20 m per second.

If this sequence of trajectory way-points are rising in altitude, a first way-point with velocity greater than 20 m per second and altitude less than 200 m minus the altitude of the airport is chosen. This timestamp is then blessed as the flight take-off time.

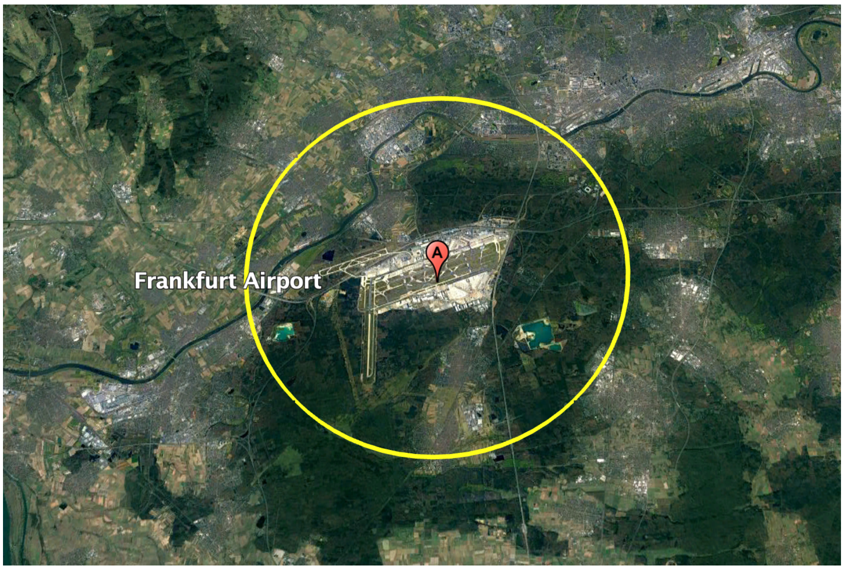

Figure 5. Circle with 7500 m radius around Frankfurt airport.

The results are shown on Figure 6. On the figure, the absolute difference between estimated take-off time and the flight take-off time available in Airport Operator Data Flow dataset (APDS) [5] is shown. In current analysis, the APDS is considered as ground truth. It shows that in more than 70% of cases, the difference between the estimated take-off time and take-off time in APDS is less than 1 min.

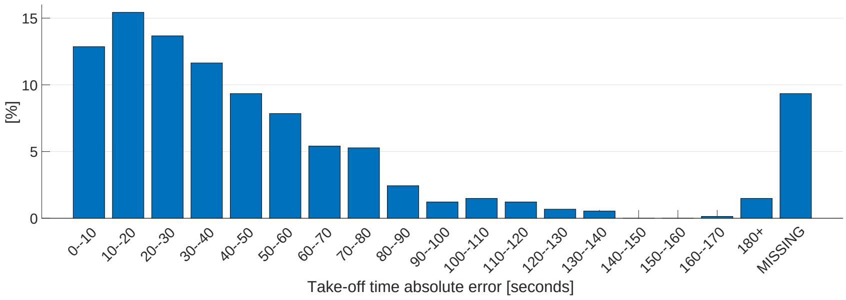

Figure 6. Take-off time estimation accuracy.

4. Conclusions

This paper demonstrates the general feasibility of the initial implementation of a pan-European flight table and associated reference trajectories. This was achieved by enriching an open and crowdsourced high-fidelity dataset with correlated position reports. This supports the creation of a pan-European Flight Table with a set of flight event milestones (e.g., actual take-off/landing time). The initial Flight Table serves as a basis for replicating the operational performance indicators. The dataset containing FTs will be made available for public use and provides an excellent basis for analysis in the scope of air traffic management.

It is expected that the data accuracy will improve over time as more sophisticated algorithms will be employed and the both CPR and crowdsourced data coverage will improve. This may also serve as a future model of cooperation between day-to-day operations and open community and research efforts.

Developments can focus on added value applications removing the need to invest effort and resources on the cumbersome data preparatory steps.

This project is a key stepping stone in establishing an open data based Flight Table for the European region. The latter will support political decision-makers and the public to validate claims made by the air transport and air navigation community.

Author Contributions: Conceptualisation, J.F., E.S., A.T. and R.K.; methodology, J.F. and E.S.; software, J.F.; validation, A.T. and R.K.; resources, R.K.; data curation, J.F.; writing—original draft preparation, A.T.; writing—review and editing, J.F., E.S. and R.K.; visualization, A.T.; supervision, R.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting reporter results can be found from https://opensky-network.org/datasets/flights_data5_develop_sample/, accessed on 26 January 2022.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Filippone, A.; Parkes, B. Evaluation of commuter airplane emissions: A European case study. *Transp. Res. Part D Transp. Environ.* **2021**, *98*, 102979. [CrossRef]
2. Koelle, R.; Peeters, S.; Spinielli, E. Building Back Better—Democratization of Performance Monitoring with Open Data. In Proceedings of the 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 3–7 October 2021.
3. Schumann, U.; Poll, I.; Teoh, R.; Koelle, R.; Spinielli, E.; Molloy, J.; Koudis, G.S.; Baumann, R.; Bugliaro, L.; Stettler, M.; et al. Air traffic and contrail changes during COVID-19 over Europe: A model study. *Atmos. Chem. Phys. Discuss.* **2021**, *21*, 7429–7450. [CrossRef]
4. Spinielli, E. Investigating ADS-B MOPS Compliance using Open Data. In Proceedings of the 7th OpenSky Workshop 2019, Zurich, Switzerland, 21–22 November 2019; pp. 135–143. [CrossRef]
5. EUROCONTROL Specification for Operational ANS Performance Monitoring-Airport Operator Data Flow. 2019. Available online: https://www.eurocontrol.int/sites/default/files/content/documents/single-sky/specifications/EUROCONTROL%20Specification%20APDF%20Edition%201.0%20final%20web.pdf (accessed on 26 January 2022).