The Open Performance Data Initiative: A Foundation Supporting the European Open Science Alliance for ATM Research

Quinten Goens, Rainer Koelle, Enrico Spinielli
Aviation Intelligence Unit
EUROCONTROL
Brussels, Belgium

Tatjana Bolić, Andrew Cook
Centre for ATM Research
University of Westminster
London, United Kingdom

Martin Strohmeier
OpenSky Network
Burgdorf, Switzerland

***Abstract*—There exists an abundance of challenges for the current and future air transport. A highly efficient air navigation system will be a key enabler to meet the political and - especially - the environmental targets. In order to demonstrate progress and validate claims and performance benefits it will be essential to independently reproduce and verify results. Previous research identified 'open data' as a key ingredient and postulated a roadmap towards an Open Science Alliance. One of the first stepping stones is the Open Performance Data Initiative. As a first step of the initiative, a conceptual trajectory reduction approach was applied. With a focus on the European context, this resulted in comprehensive datasets containing flight events, associated measurements, and flight lists, covering 33.9 million flights based on publicly available crowd-sourced ADS-B data, from January 2022 to July 2024. An initial quantitative comparison between airport traffic and operations, as monitored by EUROCONTROL and the established open datasets, reveals a high degree of consistency in trends and its suitability for supporting research and day-to-day performance monitoring. The research also showed that vertical flight inefficiency for arrivals is currently underestimated based on data granularity. The work presented in this paper builds a foundation for the proposed Open Science Alliance roadmap. Further continued development and expansion of the flight list, available flight events and measurements, will provide a more comprehensive access for reproducible and open research.**

I. Introduction

The year 2024 saw air traffic movements rebound to pre-pandemic levels and observed levels of air traffic management (ATM) constraints in terms of air traffic flow management (ATFM) delay [1] exceeding pre-pandemic levels clearly showing massive capacity shortcomings. There is a strong political push to curb the climate impact of air transport. In 2022 ICAO adopted a long-term aspirational goal to reach net-zero carbon emission for international aviation by 2050. The agenda of the European Union targets a 55% emission reduction by 2030 that also includes air transportation. The 'Flightpath 2050' establishes wider general ambitions for European aviation including competitiveness next to sustainability / climate change adaptation and operational efficiency [2]. Accordingly, there is an abundance of challenges facing European air transportation and air traffic management.

With technological advances requiring a certain lead time, increasing operational efficiency can be an immediate measure. In this context, it will be essential to verify the results of research or claimed performance benefits of deployed solutions or changes. Applying scientific principles, most notably opening the relevant data to wider research and interested audiences, offers the prospect of better informed political decision-making, strategic planning, and addressing public expectations by providing *open data* as a *common baseline*. Combined with value added and targeted validation activities these baseline data may serve as an open-source ground truth.

Barriers to open data for air transport related research and analysis exists. Traditionally air transport data is processed by government organisations or dedicated service providers. Constraints range across the full spectrum of legal, technical, and organisational concerns. The emergence of novel data collection and processing techniques as well as the adoption of data-driven approaches in research and operations provide a basis for a paradigm change. Previous research proposed the setup of an Open Science Alliance for ATM Research and laid out a series of steps [3]. The Open Performance Data Initiative (OPDI) represents a first step in lowering the barriers by establishing a set of datasets that take away the first-mile problem (i.e. access to and preparation of the air transport trajectory data).

This paper describes the success of the Open Science Alliance proponents in setting up the OPDI. The goal was to conceptualise and implement one of the first building blocks of the roadmap. This included tackling the challenges of large-scale data ingestion, processing, and exposure to research and practitioners. The associated rollout establishes a platform for further development and engagement with the community to foster the development of novel and open methods, including the open review and validation of published results [4].

The contribution of this paper comprises the

- *conceptual approach* to provide essential flight events (i.e. a trajectory reduction) for a majority of air transport research and reducing the data preparatory step by providing a common baseline;

SESAR Innovation Days 2024
12 - 15 November 2024. Rome

- *integration and pre-processing* of the supporting open air transport data for wider use by research and interested practitioners; and
- an *initial set of milestone events for analysis* to demonstrate the implementation, approach, and accessibility of air traffic movement data based on e.g., geospatial tessellation.

II. The Quest for Open Data on Air Transport

Despite the general acknowledgement of the utility of 'open data' for research and reproducibility, the progress to date is limited across all disciplines [5]. Interestingly, the move towards open data saw a boost under the umbrella of 'good government' practices. The past decade witnessed an increasing number of 'open data' policies aiming at making data and information widely and proactively available to the public [6]. Associated guidelines build on the benefits for public policy, science, and society in general by stimulating research, inform evidence-based policy and decision-making, and provide transparency for the interested public. A constant theme is the lowering of existing barriers in terms of access and use of the data.

Within the ATM domain, the topic of reproducibility is not fundamentally new. Bourgois and Sfyroeras analysed the role and availability of open data in air transport research examining over 300 research articles and identifying the most used data types, sources and access policies. The analysis showed that 70% of research in air transport is heavily reliant on data curated by governmental bodies and of limited access to research or the wider public. Given the freedom of information acts in place, data is more accessible in the US. The paper sees Europe lagging considerably behind in terms of open research and reproducibility. Thus, Europe is missing out on entrepreneurship, innovation and scientific discovery [7].

Throughout the recent years several authors provided showcases for utilising open data for air transport and ATM related research. This research spans from applications for transparent performance monitoring (e.g. [8], [9], [10], [11]), analyses of dedicated operational problems (e.g. [12], [13], [14], [15]), to the research and design of novel data-driven methods (e.g. [16], [17], [18]). Bolic et al. proposed the setup of an Open Science Alliance and supporting high-level roadmap [3]. Within SESAR, requests for open data access have become increasingly articulated over the years, especially with the emergence of data-driven methods and artificial intelligence/machine learning. Still, limitations exist in sharing underlying datasets across different SESAR projects or making these datasets available to other researchers and practitioners. The work proposes a framework of an open science based alliance with the goal to setup (1) open access to scientific methods and data utilised; (2) open access to (analytical) code and methods; and (3) open review of reported analyses/research to advance the state-of-the-art in Europe. Fig. 1 shows a potential roadmap to implement the proposed framework.

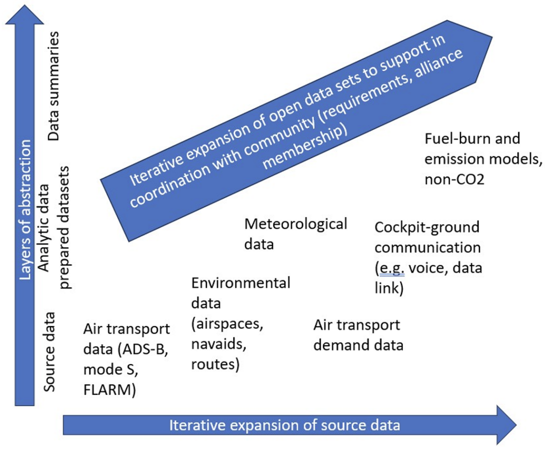

Figure 1: Potential roadmap for expanding ATM related open data in Europe [3].

A key enabler for addressing the goal of the Open Science Alliance is the emergence of open air transport data. Crowd-sourcing of open air transport data has become a significant trend in recent years, providing valuable insights for various research applications and performance oriented analyses. The OpenSky Network (OSN) is a prominent example of a platform that collects and shares openly air traffic control data from aircraft globally [19]. Strohmeier [20] examined the backgrounds and typical usage patterns of OpenSky's users, both academic and non-academic.
OSN data was instrumental in numerous studies (c.f. Fig. 2). For example, OSN-derived flight connectivity data was a key enabler to assess the scale and impact of air traffic constraints during the COVID-19 pandemic [12].

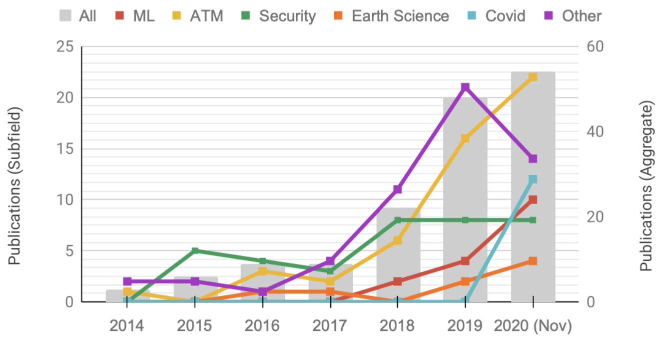

Figure 2: Publications with OpenSky data from 2014 to November 2020. Overall numbers as bars on the right vertical axis. Subfield numbers as lines on the left vertical axis. Publications may have multiple fields [20].

The Engage 2 Knowledge Transfer Network (KTN) inquired about challenges to current research [21]. According to the report, data availability and openness of data access was identified as a key blocking point in SESAR's Exploratory Research as witnessed by the difficulty to acquire required data and restrictions of its use (e.g. non-disclosure agreements, limited to a specific project/task). The authors concluded that the lack of open data presents a "barrier to improving experimental comparability across projects" [3] and precludes reproducibility. A series of barriers are reflected in Table I.

TABLE I: BARRIERS TO OPEN DATA IN THE AIR NAVIGATION CONTEXT.

Barrier	Type	Importance
Data ownership and control	Legal/regulatory	Issues around who owns and controls data usage, especially when data is collected from multiple stakeholders, even if at public transportation sectors
Reputational repercussions	Reputational	Risk of negative perception or damage to reputation from data analyses
Privacy concerns	Privacy	Concerns about individual or organisational privacy being compromised
Business considerations	Business	Issues related to confidentiality, legal liabilities, and business-critical data
National security and defence issues	National security	Concerns over the disclosure of sensitive information that could affect national security or defence operations, needing to define what is legitimate usage of data
Costs	Financial	Expenses associated with data collection, processing, storage, and maintenance. Prohibitive costs and complexities in installing necessary technologies for data collection and processing
Public interest and societal well-being	Ethical/public policy	Balancing individual privacy rights with the broader societal benefits, especially in the context of climate change and mobility

Throughout the past years, the idea of a 'Open Science Alliance for ATM Research' formed in Europe. It aims to increase transparency and access to underlying data in the field. The Alliance advocates for open data in a framework that promotes independent verification and validation of reported impacts and performance levels. Benefits of adopting an open science approach include reduced innovation cycles in ATM, increased transparency, and trust between research institutions and citizens. Open science practices involve open access to publications, open data, open source/code, open methodology, open peer review, and open educational resources. When implemented properly, these practices foster greater efficiency in research, increased collaboration, higher levels of verification/validation, and reduced duplication.

III. Setting up the Open Data Foundation

A. *Overview*

This paper focusses on a fundamental first step in establishing an open data environment in Europe; the Open Performance Data Initiative (OPDI). As indicated in Fig. 1 future expansion of the data sets will include complementary operational (e.g. ATFM measures, flight plans) and environmental (e.g. airspace and sector capacities/sectorisation) data. The integration of other air transportation communities can also support to augment the open data sets with - inter alia - passenger, connectivity, or multi-modal related data.

Currently the OPDI data sets are based on OSN crowd-sourced ADS-B data for the period January 2022 to June 2024 and comprise: (1) a flight list; (2) flight events for each flight; (3) a list of measurements associated to a flight event; and (4) a 5-second state vector[1] representation of the flight trajectory as recorded by OSN.

In particular:

- The **flight list** contains records of all flights that were operated in the European airspace, indicating the flight ID, aerodrome of departure/destination (ADEP/ADES) and (UTC) day of flight. Based on the underlying data source, the flight list provides information on all types of flights (e.g. commercial and general aviation) and covers 33.9 million flights[2].
- **Flight events** reduce trajectory state vectors of a flight to a set of 'significant' events encoded as a 4D position (trajectory ID, 3D + timestamp), flight event type and contextual info. Current examples of flight events are: top of climb and descent; start and end of level segments; landing, take-off, runway entry and exit, parking position entry and exit and taxiway entry and exit.
- A supporting **list of measurements** provides additional quantitative information for each flight event. The currently available types of measurements are distance travelled and time passed since the first observation of the flight trajectory. Future releases of the OPDI will include additional measurements such as e.g., cumulative fuel burn or gaseous emissions. Such measurements will be based on commonly agreed and validated models.

Fig. 3 depicts the overall approach as a series of data ingestions, processing steps and data set generations. For the flight event data generation three distinct algorithms were developed to account for different types of flight events: (1) Flight Phase Based Events, (2) Crossing Type Events and (3) Airport Based Events.

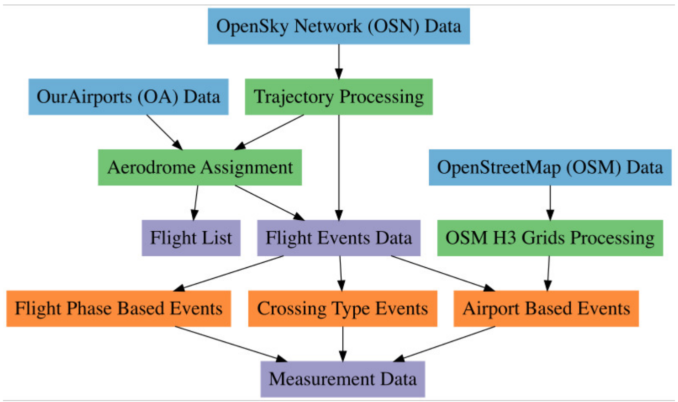

Figure 3: An overview of the data processing flow of the OPDI (blue: input data, green: processes, purple: OPDI datasets, orange: event algorithms).

[1] A *state vector* is a 1-sec snapshot of the state of an aircraft (timestamp, position, altitude, . . .) derived from ADS-B/Mode S messages as received by OSN.

[2] For this paper, the database comprises 33.9 million flights for the period 2022 through June 2024 and exceeds the flights accounted for by EUROCONTROL (in the EUROCONTROL area) by about 9.4 million flights. This is due to the underlying ADS-B data including non-IFR movements and covering a slightly wider area than the EUROCONTROL area to account for future entries (e.g., Iceland) to the EUROCONTROL area and the interface to the Middle East and North Africa.

B. Data Ingestion

The OPDI builds on open air transport data from OSN state vectors representing position reports of aircraft every 5 seconds within a defined area[3] covering EUROCONTROL Member States airspace. The state vectors are delivered via a cloud object storage service (MinIO) and then ingested into a EUROCONTROL-managed Apache Hive [4] database table for further processing using Apache Spark [5]. The large volumes involved require a bulk download to the local computing infrastructure rather than the use of various state vector querying mechanisms[6] freely available via OSN.

Following the declared aim to maintain the project fully open-source, additional sources of publicly available data were integrated for the aeronautical reference information. These sources comprise OurAirports [22] for aerodrome related data (e.g. airport location, size category) and OpenStreetMap (OSM) [23] for geographical data on airport infrastructure elements (e.g. runways, taxiways, parking positions).

While the use of open-source data provides open access and thus enhances reproducibility by third parties, limitations on open data are applicable. As per example, the OpenSky Network coverage is dependent on the location and number of ADS-B receivers of the crowd sourced network, in addition to this, data quality might be influenced by external factors such as GPS jamming. Other limitations, to the OpenStreet Map and OurAirports data, might be potential missing data. The Open Science Alliance aims to enhance the publicly available data where possible by placing ADS-B receivers at places with bad coverage (to enhance the crowd sourced ADS-B network coverage) and by enhancing open source data repositories such as OpenStreetMap and OurAirports with additional industry data.

C. Trajectory Processing

After receiving the OSN state vectors, trajectory identifiers (or `track_ids`) are assigned in monthly batches. The assignment process starts by grouping the state vectors based on the aircraft's Mode S transponder code (field `icao24`) and the flight's ID (`callsign`). Each group is further segmented if there is a gap greater than 30 minutes between subsequent state vectors, with each segment assigned a `splitnumber`. Each resulting trajectory is assigned a unique identifier, which is the SHA256 hash[7] of the concatenated `icao24`, `callsign`, and `splitnumber` values, along with the year and month of the `event_time`. This method ensures the uniqueness of trajectory identifiers throughout the dataset.

[3]The area is geographically contained by the following coordinates: from 25.87°N, 26.75°E in the southwest to 49.66°N, 70.26°E in the northeast.

[4]Apache Hive is a distributed, fault-tolerant data warehouse system that enables analytics at a massive scale.

[5]Apache Spark is a multi-language engine for executing data engineering, data science, and machine learning on single-node machines or clusters.

[6]See https://opensky-network.org/data/data-tools.

[7]SHA 256 is a part of the cryptographic hash algorithms function that transforms an input text into a fixed-length string of 256 bits for unique identification purposes (aka 'fingerprint').

D. Aerodrome Assignment and Flight List

For the construction of a flight list, the aerodrome of departure (ADEP) or destination (ADES) needs to be determined for each trajectory. To enhance computational efficiency, the implementation utilises Uber's H3 hexagonal geospatial indexing system [24]. Circular detection areas of H3 hexagons are constructed around each medium or large airport with a radius of 30 NM using a resolution level of 7 (hexagon's average edge length of ~1.4 km), balancing spatial precision and processing performance. The flight trajectories are filtered for those state vectors with an altitude below 40,000ft in order to focus on take-off and landing.

Each latitude/longitude position point is then assigned an H3 index. These H3-augmented state vectors are superimposed onto the pre-calculated airport cells. This allows to map each state vector to the set of potential aerodromes in the proximity. In case of multiple candidate aerodromes, the minimal initial (final) great circle distance between the aircraft's state vector and the Aerodrome Reference Point (ARP) is computed to determine the most probable ADES (ADES). Alternative (less) plausible candidates are noted as potential airports of departure (`ADEP_P`) or destination (`ADES_P`). These candidates are stored in the dataset to account for edge cases such as ambiguous departure/arrival locations near dense clusters of airports (e.g. London or Paris), or flight paths that exhibit irregular behaviours, such as diversions, touch-and-go operations, or aircraft repositioning.

E. Flight Event & Measurement Data

A combination of three methods is used to determine different flight event types: (1) flight phase based events, (2) airport based events and (3) crossing type events. For the measurement data, the state vectors of each trajectory are augmented with the (cumulative) great circle distance travelled and time passed (since it was first seen). Accordingly, when a state vector can be associated to a flight event the associated measurements are available.

1) Flight Phase Based Events: The state vectors of each trajectory are classified into various flight phases: Ground phase (GR), Level segment phase (LVL), Cruise phase (CR), Descent phase (DE), and Climb phase (CL). Based on such a classification, the following flight events can be identified for each trajectory:

- **`level-start/level-end`**: the first/last state vector in each *level segment* phase (LVL);
- **`top-of-climb/top-of-descent`**: the first/last state vector of the first/last *cruise* phase; and
- **`take-off/landing`**: the first state vector of the *climb* phase (CL) following a *ground* phase (GR) / the first state vector of the ground phase (GR) following a descent phase (DE).

The methodology was implemented leveraging the phase labelling algorithms detailed in OpenAP [8] [8] using fuzzy

[8]See https://openap.dev/ - an open model for aircraft performance and emission estimations.

matching techniques to handle uncertainties in the flight dynamics, such as noisy or missing data. This fuzzy approach enables to assign state vectors to predefined phase patterns, even when the exact boundaries between phases are somewhat unclear. This accounts for real-world operational data, where sensor noise or varying update rates can create ambiguities. The algorithms were adapted to a native PySpark implementation, improving processing speed, particularly for large-scale datasets.

2) Airport Based Events: Events on the runway, taxiways and apron are paramount when assessing airport operational performance and associated entry and exit times of aircraft entering these structures are thus desired:

- **`entry-runway`** / **`exit-runway`**: the actual entry/exit event/time of a flight operating on a runway;
- **`entry-taxiway`** / **`exit-taxiway`**: the actual entry/exit event/time of a flight operating on a taxiway; and
- **`entry-parking_position`** / **`exit-parking_position`**: the entry event/time of an aircraft at its designated stand or event/time of an aircraft vacating the stand.

For this purpose, the layout elements of large and medium airports in Europe were retrieved (where available) from OSM [23]. The polygons of these elements were then associated to H3 hexagons. Fig. 4 depicts the runway, taxiway, and apron/stand system of (a part of) London Heathrow Airport (EGLL).

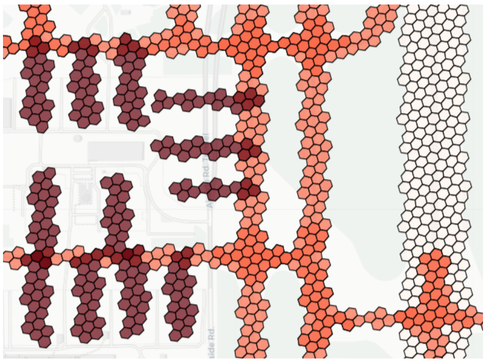

Figure 4: Zoomed-in H3 hexagonal grid displaying parts of London Heathrow airport, EGLL, (orange: taxiway, bordeaux: parking position, white: runway)

To detect ground events for a flight, trajectories are filtered to retain only state vectors with an altitude lower than 500 feet with respect to elevation of the aerodrome. The H3 resolution for this task is 12 (hexagon edge length 10.8 m).

3) Crossing Type Events: Crossing type events mark the crossing of a certain geographical or altitude boundary. For this first implementation predefined flight level (FL50/70/100/245) crossings were determined in a two-stage approach: (1) a smoothed average FL is calculated for each state vector and (2) the flight values are compared between each subsequent state vector. The first time a flight crosses an FL of interest, the crossing state vector is recorded as a **`first-xing`** event. The last time it crosses such an FL, the crossing state vector is recorded as a **`last-xing`** event. Future work will include the determination of geospatial crossings (e.g. state borders and airspace boundaries).

F. Resulting Datasets

The three datasets account for:

- a flight list with 33.9 million flights;
- a flight events dataset of 738.9 million events; and
- a measurement dataset of 1.5 billion measurements.

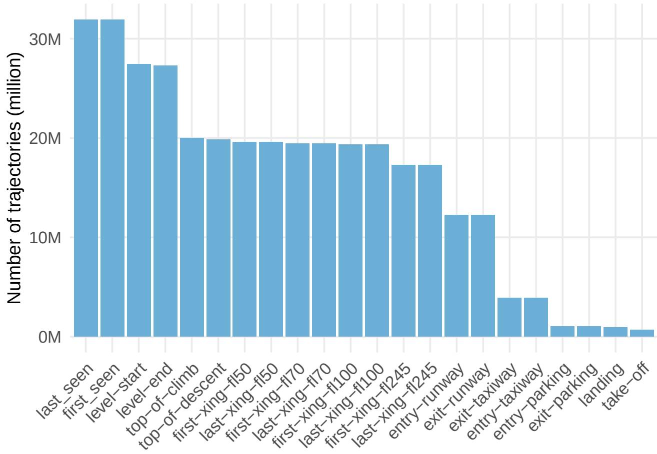

Figure 5: Number of trajectories which contain a certain flight event at least once.

The detection of flight events is dependent on and constrained by various factors. For this work the predominant drivers were (1) the ADS-B/Mode S receiver coverage and (2) the availability of geographical data in OSM. Fig. 5 shows the number of detected flight events for the processed flights within the European region. It suggests that the airborne phase of flights is well covered. A lower level of event detection is evident for the ground segment, with more landings detected than take-offs. This discrepancy results from the take-off detection algorithm, which compares current state vectors with previous ones. Due to limited airport coverage, previous state vectors are often sparse at take-off, impacting detection accuracy.

IV. Results

A. Principal Approach

To ensure the robustness of the processed data, including the event identification and its supporting geospatial associations, several verifications were conducted by comparing the flight list and flight event data against a verified source. The combination of open data and associated network processes strengthens the validity of the OPDI data. For example, members of the Open Science Alliance or external parties can contribute, through supporting data quality assurance processes, by validating that the open data can serve as a reliable common baseline.

Within the European context, the monitoring of operational performance for air navigation services (ANS) at and around airports relies on data collected under the Airport Operator

Data Flow (APDF)[9] [25]. The APDF data includes information on flights arriving and departing at airports, as well as details on the stands and runways used.

A comparison of the observed airport traffic will reveal whether the OPDI captures similar flight movements as those monitored by EUROCONTROL. Furthermore, verifying the frequencies of detected runway usage against a reliable data source will confirm the robustness of the airport event data. To account for some severe service disruptions at the OSN data center affecting the data extracted from late 2023 and 2024, data was excluded from the results due to a lack of observed trajectories in some months. Work is on-going to reprocess the underlying source data and the study datasets. The updated datasets will be made available in future releases. Accordingly, the results presented may differ within margins over time.

B. Flight Movements At Major Airports

In terms of completeness, airport flight movements for the top-10 airports were compared to the APDF EUROCONTROL data[10]. Fig. 6 shows the percentage of flights movements identified compared to the movements reported.

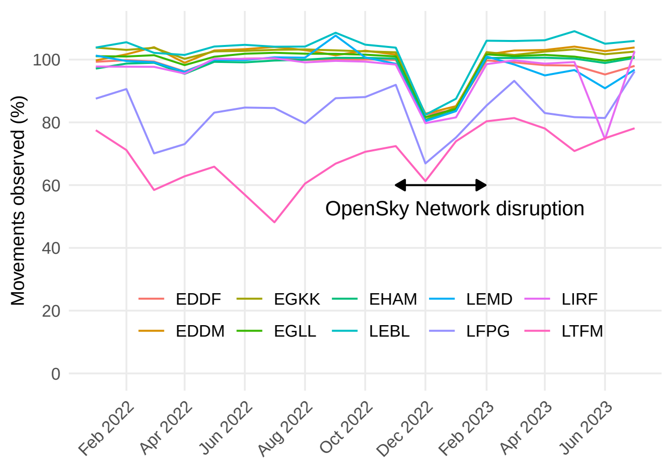

Figure 6: Percentage of flight movements observed for 10 major airports by the OPDI flight list compared to EUROCONTROL APDF reporting.

For most airports, the observed flight movements percentage by the OPDI ranges around or above a hundred percent. EUROCONTROL mainly monitors commercial flights and does not monitor non-commercial aircraft as has been detailed in the EUROCONTROL Data Snapshot #35[11]. Due to this, the OPDI is able to provide insights on how other - or non-commercial - traffic influences operational performance at European airports.

The two major exceptions are Istanbul (LTFM) and Paris Charles de Gaulle (LFPG). Coverage at these airports is limited[12] in 2022 and 2023. In addition, due to the Russian war in Ukraine, ADS-B signal disruptions from GPS jamming activities[13] around the Bosporus impacted traffic data collection for flights operating at and around Istanbul airport.

C. Airport Event Coverage

For each flight landing at an airport, it is expected that at least one runway, taxiway, and parking entry event will be observed. To measure this, we count the number of flights at each airport with at least one of these events for all three event types. These counts are normalized against the total number of flights observed at the airport, as listed in the OPDI flight list. The OPDI flight list has been shown to be robust when compared with EUROCONTROL's data in the previous section. Using this method, we calculate the (minimal expected) detection rate for runway, taxiway, and parking entry events at each airport.

The top ten airports (c.f. previous section) were examined for airport event coverage. The detection rates are presented in Fig. 7 in decreasing order of the overall detection rate.

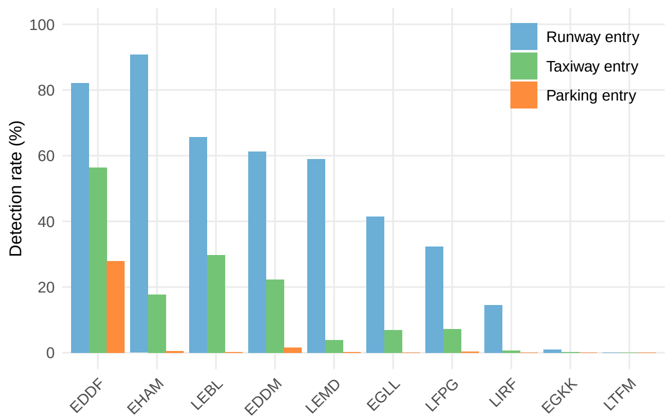

Figure 7: Airport detection rates of airport events for 10 major airports.

This confirms the pattern identified in Fig. 5. Runway entries are the most commonly observed type suggesting a good level of coverage. Lower detection of taxiway entries and lastly by parking positions suggest the need to increase sensor coverage for the taxiway system and aprons. The previously detected outliers for Istanbul (LTFM) and Paris Charles de Gaulle (LFPG) are amongst the lowest detection rates.

ADS-B signals may further be obstructed by airport buildings. Accordingly, entry events occurring closer to the terminal observe generally a lower detection rate. For the detection of parking and taxiway entries additional work is required to improve the availability of geospatial data on taxiways and parking positions in OSM. This problematic is amplified for smaller airports which show a lower level of tagging in OSM.

One solution to improve coverage at airports is to place OSN ADS-B receivers at airports. For example, as part of the OPDI, a receiver was placed at Vilnius airport by the end of

[9]The APDF collects data from more than 80 airports in Europe under the EUROCONTROL Performance Review System and the Single European Sky Performance and Charging Regulation.

[10]https://www.eurocontrol.int/Economics/DailyTrafficVariation-Airports.html

[11]See https://www.eurocontrol.int/publication/eurocontrol-data-snapshot-35-moving-towards-higher-levels-transparency-measuring.

[12]https://opensky-network.org/network/facts

[13]https://gpsjam.org/

2022. The effects on the observed event detection rate can be seen in Fig. 8.

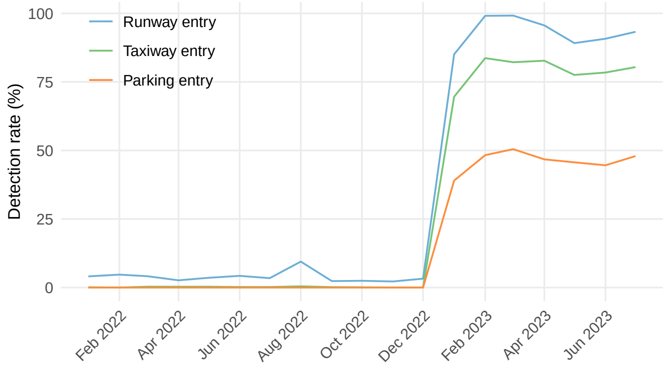

Figure 8: Monthly detection rates of airport events for Vilnius airport (EYVI).

D. Airport Event Accuracy

To assess the airport event accuracy of the runway entry, we compared the monthly runway entry counts versus the runway usage as reported by the APDF. As an example, the results for Amsterdam airport (EHAM) in Fig. 9 and for Frankfurt airport (EDDF) in Fig. 10.

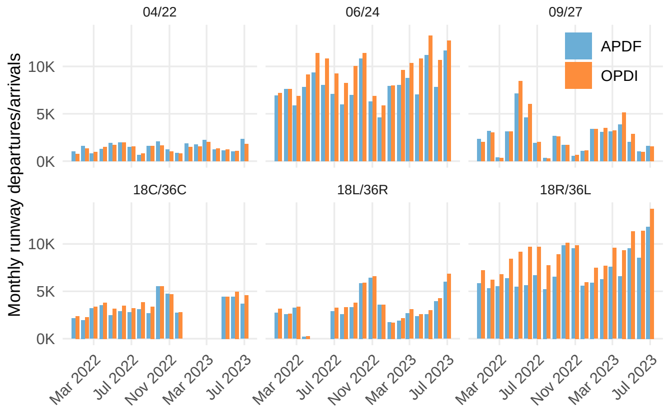

Figure 9: Monthly number of runway departures/arrivals per runway for Amsterdam airport (EHAM) as detected by the OPDI runway entry flight events and the reported EUROCONTROL APDF data.

At Amsterdam airport (EHAM) the trends are mostly consistent across the majority of runways with the exception of runways 06/24 and 18R/36L. The observed deviation for the utilisation of runway 06/24 could be its central location and the potential assignment of a runway event of aircraft taxiing to other runways. Runway 18R/36L might be flown over at low altitude by departing and arriving flights resulting in an additional association to this runway. This anomalous incongruence warrants further investigation.

Due to major renovations runway 18C/36C (Zwanenburgbaan) was closed between January 2, 2023, and mid-April 2023[14] which is reflected in Fig. 9. A similar observation can be made for runway 18L/36R (Aalsmeerbaan), which was closed for maintenance from April 4 to mid-July 2023 [15]. Knowledge about such closures could be mapped to other performance observations as notable events affecting operational patterns. Such data is currently not openly available and confirms a major driver for the Open Science Alliance goal.

The four runways departure and arrival counts at Frankfurt (EDDF) show a high agreement with the APDF data (c.f. Fig. 10). There's one exception, runway 07R/25L from March to April 2023, where further investigation is needed.

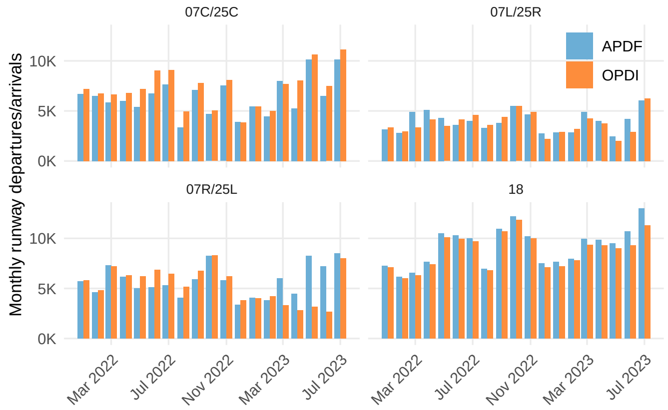

Figure 10: Monthly number of runway departures/arrivals per runway for Frankfurth airport (EDDF) as detected by the OPDI runway entry flight events and the reported EUROCONTROL APDF data.

E. Vertical Flight Efficiency

In this study, flight events and associated measurements were utilized to assess the vertical flight efficiency (VFE) performance for arriving flights at Amsterdam (EHAM) during 2022. This initial demonstration for VFE monitoring emulates the Performance Review Unit (PRU) algorithm[16]. Level segments within 200NM from EHAM and past the top-of-descent were considered to calculate the average time-in-level-flight per arrival. This method is an approximate emulation, as it employs different phase labelling methodologies. The PRU monitoring data uses trajectories with an average update rate of 37 seconds. This is in contrast to this study using flight events based on an update rate of 5 seconds. Fig. 11 compares the average time-in-level-flight for arrivals into EHAM with the VFE indicator monitored by the PRU.

While similar, OPDI determines a consistently higher time spent in level flight. This is unsurprising and demonstrates the benefit of utilising a higher-frequency reference as the actual change in level is detected on a more fine-grained basis. This implies that the current level of VFE in Europe underestimates the actual average time flown in level. This work identified a

[14]https://www.schiphol.nl/en/you-and-schiphol/news/provisional-runway-maintenance-schedule-2023/

[15]https://nieuws.schiphol.nl/aalsmeerbaan-drie-maanden-buiten-gebruik-voor-groot-onderhoud/

[16]https://ansperformance.eu/efficiency/vfe/

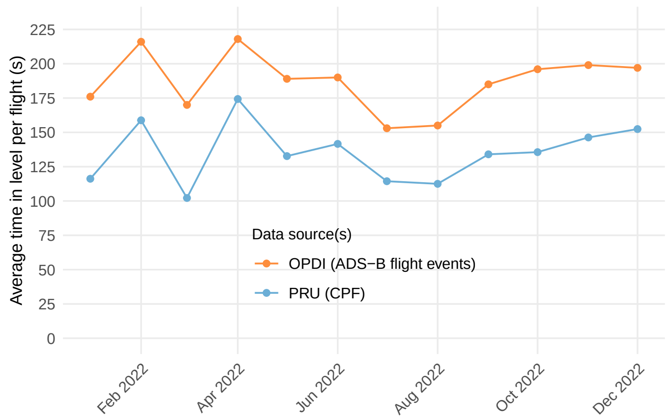

Figure 11: The time in level flight (within 200 NM of airport of destination) after top of descent for flights arriving at Amsterdam airport as calculated by both the PRU using CPF trajectories and by using OPDI flight event data.

significant portion of about ~16% of level segments that are shorter than 37 seconds which may merit a revision of the VFE algorithm. Future research may further quantify the differences in flight phase labelling and its impact on the overall measure.

V. Conclusion

The roadmap for an Open Science Alliance and establishing the means for reproducibility of air traffic management related research and realised performance benefits is well documented [3]. This paper presented the initial step towards establishing such an environment: the OPDI. It evidences the combination of resources across different organisations to establish a data environment for research, practitioners and strategic planners, and the interested public in general for the European context. The conceptual framework and initial rollout of the supporting datasets addresses the higher-level goal of establishing a ‘ground truth’ reference for the European context.

This paper documents the ongoing research and development and setup of the OPDI, building on openly available air transport data [19], supporting open aeronautical information data, the processing pipeline researched and developed in this paper, and an initial rollout of reference datasets and their high-level validation. The OPDI focusses currently on the European context, which is a limitation, as air transport and ATM problems ultimately require a gate-to-gate coverage. This paper implemented and showed the utility of a geospatial tesselation approach to balance the processing costs/resources and associated identification of flight milestone events. The results presented confirm operational performance measures at the top-10 airports in Europe and provided insights on the potential underestimation of the current VFE monitoring.

The developed approach allowed to map the existing airport surface coverage and may serve as a basis for the deployment and integration of sensors for open data collection. Experience from recent sensor deployments (e.g. Lithuania, Vilnius airport [EYVI]) demonstrated the utility of integrating additional sensors, both in terms of general coverage and ensuring surface monitoring of the maneuvering area. Working level contacts exist with other regions, e.g. United States and Brazil, to augment the coverage as a network of initiatives and complementing datasets. This can help to reduce the processing overhead on a single unit/contributor while supporting the overall goal of establishing a global open air transport data repository. An essential idea of the Open Science Alliance is to integrate, complement, and coordinate the capabilities of different organisations to establish value added datasets for ATM research, operational planing and monitoring, or informing political/strategic decision-making.

The conceptual approach and roll-out presented in this paper can form the basis for a series of added-value actions:

- *research* - availability of a series of pre-processed event milestone data sets reducing the ‘first mile’ hurdle for research in air transport and widening the user base.
- *operational performance monitoring* - higher level of transparency as a unified and open data base is available within Europe to validate operational benefits and achieved/reported performance levels for a considerable share of the current performance indicators within the European context.
- *political decision-making & strategic planning* - based on the openly available and reproducible monitoring, political decision-making and strategic planning can tap into ‘ground truth data’ as a pan-European harmonised data repository emerges supporting local, national, and pan-regional considerations.

The uptake of the data environment and future interaction within the Open Science Alliance will help to reshape the conceptual approach and identify additional event milestones. As a proponent of the Alliance, the Engage 2 will continue to publicise the OPDI in its activities – to the funded PhD students, funded catalyst projects, at the workshops, summer schools and other events. Engage 2 will facilitate the communications and eventual requests for new ‘events’ to be added to OPDI. Furthermore, another proponent of the Alliance, the SESAR JU’s Scientific Committee will continue publicising the use of OPDI and open data in general within the SESAR. The future roadmap may extend the open data to wider transport mobility initiatives, not only aviation. Based on this and the political discussion revolving around the climate impact of air transport, a series of additional data sources, e.g. open meteorological data, will help to improve the utility. Next to the classical aerodrome based weather products (e.g. METAR) further enhanced pan-regional weather products (e.g. re-analysis [26]) including forecasts are becoming available on an open data basis. Accordingly, this paper presented a first step of the envisaged roadmap.

REFERENCES

[1] EUROCONTROL, "Summer 2024 – Overview of Network performance — eurocontrol.int." [Online]. Available: https://www.eurocontrol.int/press-release/summer-2024-overview-network-performance

[2] A. Krein and G. Williams, "Flightpath 2050: Europe's vision for aeronautics," in *Innovation for sustainable aviation in a global environment*, IOS Press, 2012, pp. 63–71.

[3] T. Bolic, A. Cook, R. Koelle, E. Spinielli, Q. Goens, and M. Strohmeier, "Roadmap for a European open science alliance for ATM research," *13th SESAR Innovation Days*, 2023.

[4] "OPDI - Open Performance Data Initiative (OPDI)." [Online]. Available: https://www.opdi.aero/. [Accessed: 15-Sep-2024]

[5] N. Chakravorty, C. S. Sharma, K. A. Molla, and J. K. Pattanaik, "Open science: Challenges, possible solutions and the way forward," *Proceedings of the Indian National Science Academy*, vol. 88, no. 3, pp. 456–471, 2022.

[6] E. U. P. Office, "Open Data in Europe 2023 | data.europa.eu — data.europa.eu." [Online]. Available: https://data.europa.eu/en/publications/open-data-maturity/2023

[7] M. Bourgois and M. Sfyroeras, "Open Data for Air Transport Research: Dream or Reality?" in *Proceedings of The International Symposium on Open Collaboration - OpenSym '14*, 2014, pp. 1–7 [Online]. Available: http://dl.acm.org/citation.cfm?doid=2641580.2641602. [Accessed: 04-Sep-2018]

[8] J. Sun, J. M. Hoekstra, and J. Ellerbroek, "OpenAP: An open-source aircraft performance model for air transportation studies and simulations," *Aerospace*, vol. 7, no. 8, p. 104, 2020.

[9] R. Koelle, "Open source software and crowd sourced data for operational performance analysis," in *Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017)*, 2017.

[10] S. Reitmann and M. Schultz, "An adaptive framework for optimization and prediction of air traffic management (sub-) systems with machine learning," *Aerospace*, vol. 9, no. 2, p. 77, 2022.

[11] E. Spinielli, R. Koelle, K. Barker, and N. Korbey, "Open flight trajectories for reproducible ANS performance review," *Proceedings of the SIDs*, 2018.

[12] M. Strohmeier, X. Olive, J. Lübbe, M. Schäfer, and V. Lenders, "Crowdsourced air traffic data from the OpenSky Network 2019–2020," *Earth System Science Data*, vol. 13, no. 2, pp. 357–366, 2021 [Online]. Available: https://essd.copernicus.org/articles/13/357/2021/

[13] H. Hardell, A. Lemetti, T. Polishchuk, and L. Smetanová, "Evaluation of the sequencing and merging procedures at three European airports using Opensky data," *Engineering Proceedings*, vol. 13, no. 1, p. 13, 2022.

[14] J. Sun, A. Tassanbi, P. Obojski, and P. Plantholt, "Evaluating transatlantic flight emissions and inefficiencies using space-based ADS-b data," in *Proceedings of the 13th SESAR innovation days, sevilla, spain*, 2023.

[15] J. Sun, X. Olive, E. Roosenbrand, C. Parzani, and M. Strohmeier, "OpenSky report 2024: Analysis of global flight contrail formation and mitigation potential," in *Proceedings of the 42th digital avionics systems conference*, 2024.

[16] H. Ali, D.-T. Pham, S. Alam, and M. Schultz, "A deep reinforcement learning approach for airport departure metering under spatial–temporal airside interactions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 23933–23950, 2022.

[17] X. Olive and L. Basora, "A python toolbox for processing air traffic data: A use case with trajectory clustering," in *7th OpenSky workshop 2019*, 2019, pp. 73–60.

[18] X. Olive, L. Basora, B. Viry, and R. Alligier, "Deep trajectory clustering with autoencoders," in *ICRAT 2020, 9th international conference for research in air transportation*, 2020.

[19] M. Schäfer, M. Strohmeier, V. Linders, I. Martinovic, and M. Wilhelm, "Bringing Up OpenSky: A Large-Scale ADS-B Sensor Network for Research," in *13th IEEE/ACM International Symposium on Information Processing in Sensor Networks (IPSN)*, 2014, pp. 83–94.

[20] M. Strohmeier, "Research Usage and Social Impact of Crowdsourced Air Traffic Data," in *8th OpenSky Symposium 2020*, 2020, p. 1 [Online]. Available: https://www.mdpi.com/2504-3900/59/1/1. [Accessed: 14-Sep-2024]

[21] A. Cook, D. Schaefer, M. Bourgois, P. Hernandez, G. Tanner, and T. Bolic, *Technical Report 622*, vol. Technical Report 622. Engage KTN, 2022.

[22] OurAirports, "Open data at OurAirports," 2022. [Online]. Available: https://ourairports.com/data/. [Accessed: 11-Oct-2022]

[23] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org." 2017 [Online]. Available: https://www.openstreetmap.org

[24] "H3 geospatial indexing system." [Online]. Available: https://h3geo.org/. [Accessed: 15-Sep-2024]

[25] Performance Review Unit, "EUROCONTROL Specification for Operational ANS Performance Monitoring – Airport Operator Data Flow," EUROCONTROL, Brussels, Belgium, EUROCONTROL-SPEC-175, Jan. 2019 [Online]. Available: https://www.eurocontrol.int/publication/eurocontrol-specification-operational-ans-performance-monitoring

[26] European Centre for Medium-Range Weather Forecasts, "ECMWF Reanalysis v5 — ecmwf.int." [Online]. Available: https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5