

Qualifying public flight data using onboard sensor records for MRO-targeted historical datasets

Emy Arts ^{*}, Martin Poppinga , Laura Keller , and Florian Raddatz 

German Aerospace Center, Hamburg, Germany

*Corresponding author: emy.arts@dlr.de

Abstract

Modern MRO strategies such as condition-based maintenance and predictive maintenance rely on data-driven methods. Public flight data is one of the main sources of big data in the aviation industry, however due to its crowd-sourced roots and open communication protocols as well as measurement, transmission and decoding errors, its accuracy can be uncertain. These uncertainties need to be taken into account when studying the aircraft's condition using this data. This work aims to investigate the quality of flight data from the OpenSky historical database by comparing them to data from the same flights sourced from our research aircraft sensor data. Our analysis shows that different trajectory parameters require different processing methods and have different accuracies. Commonly used filtering functions for OpenSky data can improve the accuracy but can also introduce artifacts, as such, their use should be considered individually per application. Position accuracy values provided in raw ADS-B messages can aid in providing confidence intervals for the final result of prediction algorithms.

Keywords: Maintenance Repair and Overhaul; Aircraft Trajectory; Flight Measurement Data

1. Introduction

Modern Maintenance Repair and Overhaul (MRO) strategies such as condition-based maintenance and predictive maintenance rely on data-driven methods. Open flight trajectory data is one of the main sources of big data in the aviation industry[1]; however, due to its crowdsourced roots, open communication protocols as well as measurement, transmission and decoding errors, its accuracy can be uncertain[2]. While data obtained directly from the aircraft's measuring system provide higher accuracy and sampling rates, this type of data is not often available. In the absence of flight measurement data, some applications of predictive maintenance or condition monitoring (such as usage-driven corrosion control[3], landing gear fatigue analysis[4] and engine performance analysis[5]) can use open trajectory data. However, when using the open trajectory data the uncertainties and inaccuracies that come with it need to be taken into account. This work aims to investigate the quality of flight data from the OpenSky historical database by comparing them to data from the same flights sourced from our research aircraft sensor data.

2. Methods

At the German Aerospace Center (DLR), flight data is stored in Digital Twin Storage and Application Hub (*twinstash*)[6] which was developed to be used in combination with our research flight operations. DLR research aircraft are equipped with flight test instrumentation that allows to record data from the Avionics Standard Communications Bus (ASCB) data bus as well as data from additionally equipped sensors such as for aeroelastics. Trajectory data in *twinstash* provide high-quality data at high sampling rates (10 Hz – 50 Hz) and often include redundant sensors for single parameters. However, the available data is limited to DLR research aircraft. The OpenSky Network (OSN) [7]

historical database on the other hand provides flight data of a large number of aircraft. These flights contain data from Automatic Dependent Surveillance–Broadcast (ADS-B) and Mode-S messages providing trajectory data at a 1-2 Hz sampling rate. To be able to use these flights in connection with other public MRO databases, such as Service Difficulty Reports as was done in [8], and develop better machine learning algorithms for predictive maintenance that can account for uncertainties, we aim to qualify them using data from the flight measurement system. This work presents the evaluation of the on board measured sensor data of our twin-engine mid-sized business jet for which 97 flights (mostly German domestic) were recorded both in twinstash as well as in OSN.

For the comparisons, we consider the latitude, longitude, geometric altitude from the Global Positioning System (GPS) as well as the barometric altitude, and ground speed from the Air Data System (ADS) both connected to the ASCB, though the systems allow a 20 Hz sampling rate, the real sampling rate is at 10 Hz. From OpenSky we use the trino historical database (state_vectors_data4 table) and the Trino raw data to identify the Navigation Accuracy Category for Position and Velocity (NACp & NACv) as well as the Vertical Estimated Position Uncertainty (VEPU) and Geometric Vertical Accuracy (GVA) values. As the timestamps are not always identical between the 2 sources a window of 30 s is used to find the accuracy values for a datapoint. We utilize the following data processing pipeline: For each flight we want to evaluate and have internal data stored, in the *twinstash* database, we look up the entry in the OpenSky Network database using the Trino API and download the corresponding data. We limit the comparison to the overlapping part of the data, as our internal data often starts before data is received by the OSN.

Comparison Metrics We first time-align the two sources: we take the *lastposupdate* timestamp of the OSN data and interpolate the position on the internal data, used as ground truth, to match the timestamp and convert units. (Preliminary tests showed no significant outliers or time-delayed entries in our internal data.) We only compare present data entries, data gaps are reported separately. Based on the aligned data we compute the position Root Mean Squared Error (RMSE) and the 95th percentile of the horizontal haversine distance. A constant time offset per flight, minimising the median position error, accounts for unsynchronised clocks and transmission latencies. To tolerate state-vector timing errors, and evaluate the track distance, we use the minimal distance of each entry to the flight track flown within ± 100 , ± 300 and ± 600 ms of its timestamp, as well as using a window of ± 2 s as track distance, which approximates the cross-track error. For the geometric altitude and the ground speed we report the RMSE, p95, Normalized Median Absolute Deviation (NMAD) and median bias. For the barometric altitude we report the NMAD value and the median bias. Altitudes and ground speed are evaluated at the report time only, without tolerance windows. Here we also report the fractions of reports fulfilling the provided VEPU and GVA for the vertical height.

We also repeated this process on filtered OSN data, namely a mean filter over a 5 s window, a median filter over 5 entries, and the *KalmanSmoother6D* filter of the traffic library[9], allowing to evaluate if those filters can improve the data quality in the presented cases. For each flight and each of the mentioned error scores we calculated how many position records are within the stated uncertainty (using reports NACp, VEPU and GVA entries), under the premise that our internal data is precise.

3. Results

Figure 1 shows comparisons of the original sensor data and the OpenSky data using different filters, and Figure 2 the error distribution of the different metrics, indicating improvements or decline of scores in presence of filters and tolerance levels. Table 1 shows a selection of scores.

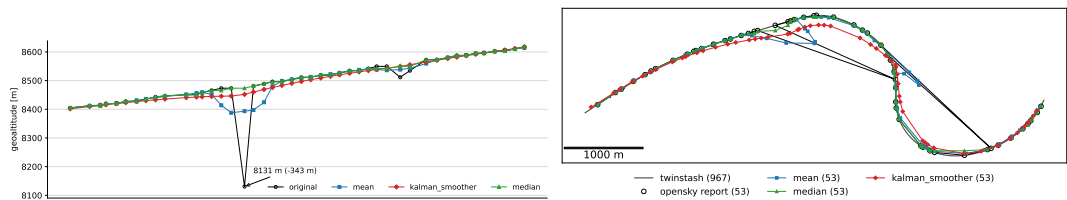


Figure 1. Altitude and horizontal data filtering in presence of outliers

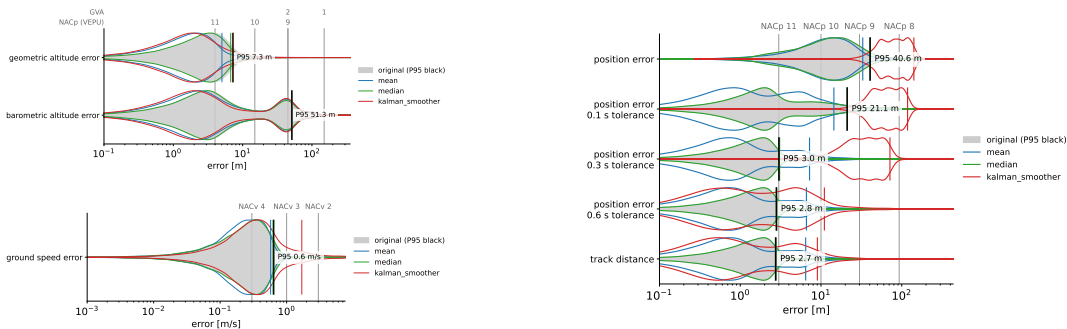


Figure 2. Error distribution in comparison to internal flight data, with varying tolerances and filters

4. Discussion

In this work we showed that based on a sample of 97 flights, 58 had positional accuracy data in the OpenSky historical database. The indicated accuracy in most cases reflects the error margin found comparing the data to the original sensor values on board the aircraft which on average is 20 m in horizontal position, 58 m altitude and 0.4 m/s for speed. For the horizontal position, a source of the error could be attributed to a time delay. Removing the temporal based error, by the error between single OSN data points to the track of the flight, reduced the error to less than less than 3 m. Using Kalman filtering removes altitude outliers (altitude RMSE 3.6 m instead of 57.7 m), but it is less suited if a precise horizontal position is required. Since it extrapolates from the estimated velocities, it diverges from the flown path in turns and across data gaps and runs about 0.5 s ahead of the report time, which raises the position p_{95} to 125 m and leaves only 19 of 58 flights within their claimed NACp along the path, although eliminating outliers, which corresponds to an improved RMSE. Though the filters shown in this work were effective at reducing outliers and overall error they provide a smoothing effect, removing small fluctuations and anomalies in the data including those that are not outliers. The use of OSN data for MRO application greatly varies depending on the specific use case. To fuse trajectory data with environmental data for environmentally driven degradation (e.g. corrosion) analysis the error margins as well as a smoothed curve is appropriate

Table 1. Key results over 97 flights: median over flights, and number of flights meeting the 95 % claim criterion.

Variant	Position [m]			Offset removed			Altitude [m]		GS [m/s]		Claim met (of 58 flights)		
	p_{95}	$tol_{0.3s}$	$track_{95}$	RMSE	p_{95}	$ \Delta t $ [s]	RMSE	p_{95}	RMSE	p_{95}	NACp $\pm 0.3s$	NACp track	VEPU GVA
Original	35.1	2.8	2.7	20.5	38.2	0.07	57.7	7.2	0.4	0.5	55	58	13 58
Mean	28.4	5.7	5.4	16.0	28.8	0.07	25.1	4.9	0.3	0.5	26	28	40 58
Median	35.1	2.8	2.8	20.6	38.0	0.07	3.8	6.6	0.3	0.5	55	58	13 58
Kalman	125.1	63.5	7.4	14.3	25.7	0.57	3.6	7.0	0.7	1.3	0	19	27 58

Offset removed: position errors after removing a constant time offset per flight that minimises the median position error. Claims: out of the 58 flights with NACp, VEPu and GVA claims; NACp time-aligned: 0. Insufficient NACv reports were found.

and reduces the amount of outliers in the data. On the other hand, other applications such as engine performance monitoring, gust detection and hard landing detection rely on small anomalies in the data. These use cases should use the trajectory data in its most unprocessed form and consider other forms of outlier detection (e.g. a plausibility check), take the given accuracy margins into account and use them to provide confidence intervals where applicable. Based on the analysis of this work, the accuracy ranges reported in the OSN historical database provide a good baseline.

While open aircraft trajectory databases such as OSN remains one of the main data sources for data-driven MRO its usage the inherent errors and uncertainties of the values into account.

Open data & Reproducibility statement

Refer to our repository to reproduce this research¹. As twinstash access cannot be granted outside of the DLR and flights had to be anonymised, it contains local files from twinstash as well as its OpenSky historical counterparts to allow users to access and reproduce a subset (ferry flights) of the relevant flights for this research. The times of these flights have artificially been set to January 2000.

Author contributions

EA: Conceptualisation, Methodology, Validation, Writing - Original Draft, Writing — review & editing. MP: Methodology, Software, Data Curation, Formal Analysis, Validation, Visualisation. LK: Software, Data Curation, Formal Analysis. FR: Writing — review & editing.

References

- [1] E. Adamopoulou and E. Daskalakis. “Applications and Technologies of Big Data in the Aerospace Domain”. In: *Electronics* 12.10 (2023). ISSN: 2079-9292. DOI: [10.3390/electronics12102225](https://doi.org/10.3390/electronics12102225).
- [2] J. Sun. *The 1090 Megahertz Riddle: A Guide to Decoding Mode S and ADS-B Signals*. 2nd ed. TU Delft OPEN Publishing, 2021. ISBN: 978-94-6366-402-8. DOI: [10.34641/mg.11](https://doi.org/10.34641/mg.11).
- [3] European Union Aviation Safety Agency. *EASA Artificial Intelligence Roadmap 2.0: A Human-Centric Approach to AI in Aviation*. Tech. rep. Accessed: 2026-08-05. EASA, 2023.
- [4] J. Hoole, S. Ramachandra, J. Booker, and J. Cooper. “Deriving Occurrence Variability in Fatigue Critical Turning Manoeuvres for Landing Gear Design from Air Traffic Data”. In: *Aerospace* 13.3 (2026). ISSN: 2226-4310. DOI: [10.3390/aerospace13030257](https://doi.org/10.3390/aerospace13030257).
- [5] M. Klein, S. Staudacher, K. Abu-Taa, and J. Schwengler. “Digital Service Twin to Assess Engine Deterioration and Remaining Part Life”. In: (2022). DOI: [10.1115/GT2022-82113](https://doi.org/10.1115/GT2022-82113).
- [6] S. Haufe, M. Bäßler, C. Pätzold, M. Tchorzewski, H. Meyer, E. Arts, and A. Kamtsiuris. “Digital Twins Storage and Application Service Hub (Twinstash)”. In: *Deutscher Luft- und Raumfahrt Kongress 2022*. Deutsche Gesellschaft für Luft- und Raumfahrt - Lilienthal-Oberth e.V., Bonn, 2023. DOI: [10.25967/570104](https://doi.org/10.25967/570104).
- [7] M. Schäfer, M. Strohmeier, V. Lenders, I. Martinovic, and M. Wilhelm. “Bringing up OpenSky: A large-scale ADS-B sensor network for research”. In: *IPSN-14 proceedings of the 13th international symposium on information processing in sensor networks*. IEEE, 2014, pp. 83–94.
- [8] E. Arts, H. Meyer, F. Raddatz, and G. Wende. “[Poster] Identifying Maintenance, Repair and Overhaul Events based on ADS-B Data”. In: *Journal of Open Aviation Science* 1.2 (2023). DOI: [10.59490/joas.2023.7220](https://doi.org/10.59490/joas.2023.7220).
- [9] X. Olive. “traffic, a toolbox for processing and analysing air traffic data”. In: *Journal of Open Source Software* 4 (2019), p. 1518. ISSN: 2475-9066. DOI: [10.21105/joss.01518](https://doi.org/10.21105/joss.01518).

¹<https://github.com/DLR-MO/qualify/tree/release-osn>