

EDITORIAL

Reviews and Responses for **Prediction of Arrival Runway Occupancy Time and Exit Taxiway Using ADS-B Trajectories**

Authors: Kevin Hänggi, Jeremy Wilde, and Manuel Waltert

Reviewers: Gabriel Jarry and Yutong Chen

Editor: Ramon Dalmau

1. Original paper

The DOI for the original paper is <https://doi.org/10.59490/joas.2026.8465>

2. Review - round 1

2.1 Reviewer 1

This paper investigates the use of open-source ADS-B trajectory data (sourced from the Open-Sky Network) to predict Rapid Exit Taxiway (RET) selection and Arrival Runway Occupancy Time (AROT) at Zurich Airport. The authors propose a two-stage approach: a LightGBM model to predict RET probabilities, followed by a neural network that utilizes these probabilities along with time-variant trajectory snippets to predict AROT. The results demonstrate that models trained on public ADS-B data can achieve performance comparable to those relying on proprietary radar or A-SMGCS data, offering a scalable alternative for enhancing runway throughput. The paper is well-structured with a clear progression from problem definition to solution. The literature review is comprehensive, and the study clearly articulates its contribution by leveraging open data compared to existing studies relying on proprietary datasets. The description of the data pipeline, including collection, cleaning, and feature engineering, is exceptionally transparent and well-illustrated.

Minor comments:

1. Generalization of Methodology (Section 2.1.2) The authors manually defined spatial polygons using Google Maps. While effective for a single airport, this limits scalability. How might this methodology be generalized to multiple airports? Is there potential to automate the polygon generation process by integrating digital AIP data or OpenStreetMap features rather than relying on manual drawings?

2. Handling Outliers in Production (Section 2.1.4) The paper describes removing outliers, such as groundspeeds over 240 kt, during pre-processing. How would these thresholds apply in a production environment? If the live data stream contains a spurious value classified as an outlier, would the model fail to predict, or is there a fallback mechanism? Clarifying the robustness of the inference pipeline is recommended.

3. Feature Engineering (Section 2.1.5) The authors report high feature importance for categorical variables such as Aircraft Type and Airline. While this captures variance effectively, relying on high-cardinality identity features raises generalizability concerns. I suggest considering a future study fo-

cused on Feature Decomposition. Transitioning from categorical identities toward property-based feature engineering, such as aircraft-specific nominal approach speeds or airline-specific average stand distances, would evolve the model into a more physics-aware system. This transition would enhance robustness and utility for cold-start scenarios in operational deployment.

4. Data Splitting Strategy (Sections 2.1.5 & 2.2 & 2.3) The study employs a random split for training and testing. Ideally, a temporal split is recommended to rigorously prevent data leakage (e.g., correlated weather conditions across adjacent timeframes) and better simulate operational forecasting. While re-running the study is not necessary, this should be explicitly acknowledged as a limitation and a target for future enhancement, potentially requiring a larger dataset.

This study convincingly validates open-source ADS-B data as a scalable, high-performance alternative to proprietary systems for runway occupancy prediction. The methodology is robust and well-documented. Addressing the outlined points on spatial automation, feature engineering, and validation strategies would further strengthen the work's operational relevance. I recommend publication with minor revisions.

Recommendation: Revisions Required

2.2 Reviewer 2

The paper explores the use of publicly available ADS-B data to predict rapid exit taxiway usage and arrival runway occupancy time. The topic is practically relevant, and the direction aligns with current interest in leveraging open data for runway operations research. However, in its current form, the manuscript reads primarily as a feasibility study rather than a contribution that advances theoretical understanding or methodological rigour. The core issues lie in the lack of a clearly articulated theoretical contribution, the absence of meaningful comparative or ablation experiments, and the use of numerous parameters whose rationale is not adequately justified. Substantial revisions are required before the paper can be considered for publication.

1. The theoretical contribution of the paper is not clearly established. The work applies existing machine-learning techniques to a new data source but does not articulate what new conceptual insights, modelling principles, or generalizable mechanisms the study adds to the field. Without a clearer theoretical framing or methodological innovation, it is difficult to position the paper within the broader research landscape.

2. The manuscript lacks the necessary comparative experiments to validate the effectiveness of the proposed approach. The reported results only reflect the performance of the authors' chosen models, with no comparison against simpler baselines, alternative methods, or ablation studies that would isolate the contribution of individual components. As a result, the reader cannot assess whether the proposed modelling choices are justified or whether similar performance could be achieved with far simpler approaches.

3. The paper relies on a substantial number of parameter choices throughout data processing, feature construction, and experimental setup, but the rationale behind these choices is insufficiently explained. Without justification or sensitivity analysis, it is unclear whether these parameters are robust, data-driven, or merely heuristic. This weakens both the methodological soundness and the reproducibility of the study.

Recommendation: Revisions Required

3. Response - round 1

3.1 Response to reviewer 1

1. Generalization of Methodology (Section 2.1.2) The authors manually defined spatial polygons using Google Maps. While effective for a single airport, this limits scalability. How might this methodology be generalized to multiple airports? Is there potential to automate the polygon generation process by integrating digital AIP data or OpenStreetMap features rather than relying on manual drawings?

Response

We agree with the reviewer that the current manual polygon definition limits scalability. We have added a few sentences in Section 2.1.2 (Lines 144–147) clarifying that the approach can be generalised by automatically generating polygons from digital AIP data or OpenStreetMap features, and that the feasibility of such automation depends on the availability and quality of the underlying data.

2. Handling Outliers in Production (Section 2.1.4) The paper describes removing outliers, such as groundspeeds over 240 kt, during pre-processing. How would these thresholds apply in a production environment? If the live data stream contains a spurious value classified as an outlier, would the model fail to predict, or is there a fallback mechanism? Clarifying the robustness of the inference pipeline is recommended.

Response

We have added a clarifying statement in Section 4.2 (Lines 456–460) addressing the handling of filtered trajectory snippets in an operational context. The text now explains that the fixed-length input requirement of the model may limit direct applicability if snippets contain fewer than ten valid points, and that simple capping or imputation strategies could be applied, provided that the same preprocessing approach is used consistently during training.

3. Feature Engineering (Section 2.1.5) The authors report high feature importance for categorical variables such as Aircraft Type and Airline. While this captures variance effectively, relying on high-cardinality identity features raises generalizability concerns. I suggest considering a future study focused on Feature Decomposition. Transitioning from categorical identities toward property-based feature engineering, such as aircraft-specific nominal approach speeds or airline-specific average stand distances, would evolve the model into a more physics-aware system. This transition would enhance robustness and utility for cold-start scenarios in operational deployment.

Response

We have extended the discussion in Section 4.2 (Lines 473–476) to address the limitations of high-cardinality categorical features such as aircraft type and airline. The revised text highlights the potential of replacing identity-based features with numerical representations of underlying physical and operational characteristics to improve robustness and applicability in cold-start scenarios. In line with this, we now explicitly refer to Nguyen et al. [1], who replace categorical features with numerical equivalents. Furthermore, we have added this as potential future work in Section 5 (Lines 519–521).

4. Data Splitting Strategy (Sections 2.1.5 & 2.2 & 2.3) The study employs a random split for training and testing. Ideally, a temporal split is recommended to rigorously prevent data leakage (e.g., correlated weather conditions across adjacent timeframes) and better simulate operational forecasting. While re-running the study is not necessary, this should be explicitly acknowledged as a limitation and a target for future enhancement, potentially requiring a larger dataset.

Response

We strongly agree and have added the limitation of the random data split to Section 4.2 (Lines 430–435), where it is now integrated into the discussion of comparability with the literature and their employed splitting methods.

3.2 Response to reviewer 2

1. The theoretical contribution of the paper is not clearly established. The work applies existing machine-learning techniques to a new data source but does not articulate what new conceptual insights, modelling principles, or generalizable mechanisms the study adds to the field. Without a clearer theoretical framing or methodological innovation, it is difficult to position the paper within the broader research landscape.

Response

We have revised the introduction (Lines 94–95) by adding a sentence to clarify the contribution of the study. In particular, we now explicitly state that the aim of the paper is to apply state-of-the-art machine learning methods to a practically relevant operational problem.

This addition also clarifies that the work is positioned as application-oriented research, in line with the scope of the OpenSky Network Symposium.

2. The manuscript lacks the necessary comparative experiments to validate the effectiveness of the proposed approach. The reported results only reflect the performance of the authors' chosen models, with no comparison against simpler baselines, alternative methods, or ablation studies that would isolate the contribution of individual components. As a result, the reader cannot assess whether the proposed modelling choices are justified or whether similar performance could be achieved with far simpler approaches.

Response

We thank the reviewer for this comment. To address this point, we introduced additional comparative experiments for both the RET and AROT prediction models:

- For RET prediction, we added a naive baseline based on the most frequently used RET per aircraft type derived from the training set and evaluated it on the AROT dataset (Section 2.2: Lines 264–267; Section 3.1: Line 318; Section 4 (Discussion): Lines 349–354; Section 5: Lines 496–497).
- For AROT prediction, we trained an additional single-input model using only time-invariant features together with the corresponding values of groundspeed, gealtitude, and vertical rate at the prediction point (Section 2.3: Lines 298–304; Section 3.2: Lines 325–326; Section 4 (Discussion): Lines 440–445, 448–449; Section 5: Lines 501–503).

The corresponding results were added to Tables 1 and 2, and the discussion was extended accordingly. Additionally, the Jupyter notebooks used for the comparative models have been added to the GitHub repository.

3. The paper relies on a substantial number of parameter choices throughout data processing, feature construction, and experimental setup, but the rationale behind these choices is insufficiently explained. Without justification or sensitivity analysis, it is unclear whether these parameters are robust, data-driven, or merely heuristic. This weakens both the methodological soundness and the reproducibility of the study.

Response

We thank the reviewer for this comment and have carefully reviewed all parameter choices throughout the manuscript. We would like to address each aspect separately.

Regarding data processing, the parameters used in the pre-processing pipeline follow best practice for ADS-B data processing and are largely based on default parameters provided by the *traffic* library. The remaining thresholds, such as the altitude and groundspeed limits for trajectory snippets, were set based on inspection of the data distribution and reflect the observed extremes in the dataset. The minimum occurrence threshold for aircraft types and airlines was chosen to ensure statistical relevance of the categorical features used in the models.

Regarding the prediction point, we acknowledge that the choice of 4 NM was not explicitly justified in the original manuscript. We have added a sentence in Section 2.1.4 (Lines 184–188) referencing the prediction distances used in related work, specifically Martínez et al. [2], to motivate this choice.

Regarding the model architecture, the dual-input design was not arbitrary but resulted from stepwise testing, in which the dual-input architecture consistently outperformed a single-input baseline. We have added a brief statement to this effect in Section 2.3 (Lines 275–277).

Additionally, the GitHub repository containing the implementation is publicly available.

References

- [1] An-Dan Nguyen, Duc-Thanh Pham, Nimrod Lilith, and Sameer Alam. “Model generalization in arrival runway occupancy time prediction by feature equivalences”. In: *Proceedings of the 9th International Conference for Research in Air Transportation (ICRAT 2020)*. 2020. doi: [10.48550/arXiv.2201.11654](https://doi.org/10.48550/arXiv.2201.11654). URL: <https://arxiv.org/abs/2201.11654>.
- [2] Miguel García Martínez, Javier García Moreno, Rubén González Sendino, and Álvaro Rodríguez Sanz. “Resilient arrival runway occupancy time prediction for decision-making tool in Barcelona (LEBL) Airport”. In: *Proceedings of the 9th International Conference on Research in Air Transportation (ICRAT)*. Tampa, FL, USA, June 2020.