

EDITORIAL

## Reviews and Responses for Aerodrome Movement Monitoring Using ADS-B Data: A Case Study at Lommis Airfield

**Authors:** Alex Fustagueras and Manuel Waltert

**Reviewers:** Àlex Vallés Villanova, Junzi Sun, and Anastasia Lemetti

**Editor:** Tatiana Polishchuk

### 1. Original paper

The DOI for the original paper is <https://doi.org/10.59490/joas.2026.8455>

### 2. Review - round 1

#### 2.1 Reviewer 1

Abstract (p.1, final): “You may consider softening the final sentence to clarify that scalability refers to the potential of the pipeline, since full validation was conducted at a single airfield. Suggested wording: “The validated pipeline demonstrates the potential for a scalable path toward automated, data-driven movement reporting, with full end-to-end validation conducted at a single airfield and additional cross-airfield evidence shown at the model level.”

Figure 1 (p.2), typo: “In Figure 1, please correct the column header to ‘MOVEMENTS’ (currently misspelled).”

Acronyms (general): “After the first mention of ‘Lommis Airfield (LSZT)’, using ‘LSZT’ alone would improve conciseness (e.g., p.2 l.38, p.3 l.93).”

Literature paragraph (p.2 l.41): “The paragraph covering rule-based methods, ML approaches, DB-SCAN, XGBoost, and CNNs spans many concepts; splitting it into two paragraphs would improve readability without altering the technical content.”

Sampling length (p.7 l.221–226): “Briefly motivate the choice of 500 uniformly sampled points to clarify robustness across circuits of different durations.”

Duplicate paragraph (p.13 l.408–423 & l.424–439): “The explanation of LSTM underperformance appears duplicated; consider removing one instance to improve conciseness.”

Typo (p.6 l.187): “The string ‘peak/valley’ protrudes into the right margin (overfull line). A hyphenation or rephrasing may address this.”

Typo (p.15 l.520): “Replace ‘patterns.. Finally,’ with ‘patterns. Finally.’”

Typo (p.7 Figure 5): “Replace caption from a) ‘Right hand,’ with ‘Right-hand’”

UK/US consistency (p.8 l.240–243): “There is mixed UK/US spelling (e.g., ‘standardised’ vs. ‘standardization’). Consider unifying the paper to UK English for consistency.”

Acknowledgement (p.15 l.523–525): “Replace ‘two/three reviewers’ with the final number.”

Recommendation: Accept Submission

## 2.2 Reviewer 2

The paper is generally well written and well structured. It explains a data pipeline for detecting aerodrome traffic circuits for GA flights using ADS-B data. Below are my detailed comments:

1. It would be helpful to add an illustration of a traffic circuit in the introduction. It would also help to introduce the concept of the circuit earlier (rather than first mentioning it around page 3).
2. Consider removing: “To the best of our knowledge, no existing work offers a robust and generalisable automated approach for identifying aerodrome traffic circuits at small airfields on the basis of ADS-B data.” It reads a bit like a lot constrains for a first research, and it also makes the transition into “Therefore, this paper addresses...” feel slightly awkward.
3. The paragraph starting around line 78 and the paragraph starting around line 86 overlap in meaning (research questions vs. objectives). Consider using one overall research question, followed by three objectives, to reduce repetition.
4. “Trajectories were kept only if the aircraft descended to a minimum of 300 feet or less above the aerodrome field elevation” — please justify the choice of 300 ft (even if it is an informed/educated threshold).
5. Around line 135, the paragraph becomes a bit wordy. On rereading, it does not seem to add much beyond what is already stated before it.
6. Figure 3: the scale differs quite a lot from the first plot to the last. Is this intentional, and does it help communicate the point you want to make?
7. Line 190: there is a mixed use of citation styles. I suggest removing author names and keeping only “[28]” (same comment for line 237).
8. Figure 4: “Level” and “N/A” use the same colour. Could you change one so they are distinct? Also, around 1600, 2100, and 2200, peaks and valleys appear to overlap—could you explain why?
9. Related to the previous point: the detection step using SciPy for local maxima and minima is not very clear. Please provide more details on the method and parameters.
10. Line 214: what criteria were used for manual labelling of the ground truth? Did you use multiple labelers (e.g., to vote on edge cases)?
11. Page 8 lists the machine learning models, but it would be helpful to explain how/why these models were chosen (e.g., based on prior studies, suitability for the task, or simply to compare a range of approaches).
12. In Section 3, it could be useful to include a feature-importance analysis showing which features contribute most to the classification.
13. Consider adding to the Results section a few examples of failure cases (false positives/false negatives). It would also be informative to show cases where one model fails (e.g., LSTM) but another succeeds (e.g., 1D CNN). This could provide stronger insights than speculating about performance differences (line 413–415).

14. In the Discussion, it is unclear whether incomplete ADS-B data is due to coverage limitations or because some GA aircraft are not equipped with ADS-B. Do you have proposals to mitigate these limitations?

15. The Discussion appears to contain two paragraphs that are the same.

16. Line 520 has two full stops.

Recommendation: Revisions Required

### 2.3 Reviewer 3

The paper presents an end-to-end pipeline for automated aerodrome movement monitoring at small, non-towered airfields using ADS-B data. Its core contribution is the supervised detection of aerodrome traffic circuits, formulated as a binary classification task on segmented trajectory candidates. The topic is operationally relevant and addresses a genuine gap in GA aerodrome monitoring. The authors compare rule-based heuristics with several machine-learning models (Logistic Regression, Random Forest, LSTM, BLSTM, and 1D CNN), showing that ML, particularly a 1D CNN, substantially improves recall and overall performance. The methodology is clearly structured and reproducible, and the empirical evaluation is thorough, including segment- and flight-level analyses and end-to-end validation through a three-month case study at Lommiss Airfield. The results provide valuable comparative insights, notably the strong performance of simpler models and the advantages of CNNs over LSTMs for this task, while overall coverage is mainly constrained by ADS-B data availability.

Here are some minor comments:

Some abbreviations are introduced more than one time, e.g., Automatic Dependent Surveillance Broadcast (ADS-B), Collaborative Decision Making (A-CDM), Long Short-Term Memory (LSTM). Average Precision is introduced as AP and AP (PR) in different places. The abbreviation ML is used in the manuscript without being explicitly introduced, while the full term machine learning also appears. Using a consistent convention for introducing abbreviations at first occurrence throughout the manuscript would improve clarity.

Figure 8 is not explicitly referenced in the main text. The caption is also relatively long; part of the explanatory content could be moved into the main text when introducing the figure.

There appears to be a duplicated paragraph in the manuscript: one starting at line 408 and a repeated version starting at line 424. Please remove the duplication and retain a single instance.

In the sentence introducing Table 2, the text states that "the same evaluation metrics as in Table 1 are used," although the reported metrics differ slightly (e.g., ROC AUC and AP are not shown in Table 2). A small rephrasing would be good here.

In Section 4, it may be helpful to report how many aerodrome movement candidates were excluded during the manual labelling process, for example as a percentage of the total candidates. This would help quantify the extent of the filtering and clarify its potential impact on model bias and performance.

Recommendation: Revisions Required

## 3. Response - round 1

### 3.1 Response to reviewer 1

Abstract (p.1, final): "You may consider softening the final sentence to clarify that scalability refers to the potential of the pipeline, since full validation was conducted at a single airfield. Suggested

wording: “The validated pipeline demonstrates the potential for a scalable path toward automated, data-driven movement reporting, with full end-to-end validation conducted at a single airfield and additional cross-airfield evidence shown at the model level.”

Response

The abstract has been updated according to the suggestion.

Figure 1 (p.2), typo: “In Figure 1, please correct the column header to ‘MOVEMENTS’ (currently misspelled).”

Response

The figure has been corrected. It now states "MOVEMENTS".

Acronyms (general): “After the first mention of ‘Lommiss Airfield (LSZT)’, using ‘LSZT’ alone would improve conciseness (e.g., p.2 l.38, p.3 l.93).”

Response

We have updated the manuscript to use the abbreviation ‘LSZT’ for all subsequent mentions following the initial definition, ensuring a more concise narrative. As suggested, we have retained the full name at the start of major sections (e.g., Sections 3 and 4) to maintain clarity for the reader. The captions for Figures 2, 3, 5, and Table 3 have also been updated accordingly.

Literature paragraph (p.2 l.41): “The paragraph covering rule-based methods, ML approaches, DBSCAN, XGBoost, and CNNs spans many concepts; splitting it into two paragraphs would improve readability without altering the technical content.”

Response

The paragraph covering rule-based methods, ML approaches, DBSCAN, XGBoost, and CNNs has been split into two paragraphs.

Sampling length (p.7 l.221–226): “Briefly motivate the choice of 500 uniformly sampled points to clarify robustness across circuits of different durations.”

Response

The choice of 500 points was motivated by the need to standardise the input for the neural network while preserving the structural integrity of the flight path. Specifically:

- 500 points provide sufficient granularity to capture critical manoeuvres (e.g., base-to-final turns) even in longer circuits.
- Since the typical circuit length is contained within this range, 500 points represent a "sweet spot" that minimises heavy interpolation for short circuits and avoids excessive data loss during downsampling for longer ones.
- This fixed-length tensor ensures a uniform input for the model without incurring unnecessary computational overhead.

We have added a brief clarifying sentence to p. 7, lines 221–226 to make this motivation explicit.

Duplicate paragraph (p.13 l.408–423 & l.424–439): “The explanation of LSTM underperformance appears duplicated; consider removing one instance to improve conciseness.”

#### Response

We thank the reviewer for noting this oversight, the duplicated paragraph has been accordingly removed.

Typo (p.6 l.187): “The string ‘peak/valley’ protrudes into the right margin (overfull line). A hyphenation or rephrasing may address this.”

#### Response

We have amended the text to fix the overfull of p.6, l.187.

Typo (p.15 l.520): “Replace ‘patterns.. Finally,’ with ‘patterns. Finally.’”

#### Response

We have amended the text to fix the typographical error of p.15, l.520.

Typo (p.7 Figure 5): “Replace caption from a) ‘Right hand,’ with ‘Right-hand’”

#### Response

We have amended the text to fix the typographical error of p.7, Figure 5.

UK/US consistency (p.8 l.240–243): “There is mixed UK/US spelling (e.g., ‘standardised’ vs. ‘standardization’). Consider unifying the paper to UK English for consistency.”

#### Response

The manuscript has been thoroughly reviewed and edited to ensure consistent use of UK English spelling throughout (e.g., ‘standardisation’, ‘optimiser’, ‘recognising’). All instances identified by the reviewer, along with others found during a manual sweep, have been corrected

Acknowledgement (p.15 l.523–525): “Replace ‘two/three reviewers’ with the final number.”

#### Response

The final number of reviewers has been changed to ‘three’.

### 3.2 Response to reviewer 2

1. It would be helpful to add an illustration of a traffic circuit in the introduction. It would also help to introduce the concept of the circuit earlier (rather than first mentioning it around page 3).

#### Response

We appreciate the suggestion to bring this concept to the forefront. The definition and operational concept of the traffic circuit are introduced in the sixth paragraph of the Introduction (p.3, l.64–78). We believe this placement provides the necessary context immediately following the discussion of GA operational challenges.

2. Consider removing: “To the best of our knowledge, no existing work offers a robust and generalisable automated approach for identifying aerodrome traffic circuits at small airfields on the basis of ADS-B data.” It reads a bit like a lot constrains for a first research, and it also makes the transition into “Therefore, this paper addresses...” feel slightly awkward.

#### Response

The sentence has been removed as suggested, allowing for a more direct transition to the paper’s specific contributions.

3. The paragraph starting around line 78 and the paragraph starting around line 86 overlap in meaning (research questions vs. objectives). Consider using one overall research question, followed by three objectives, to reduce repetition.

#### Response

We agree with the reviewer that the research questions and objectives shared significant thematic overlap. We have consolidated these sections into a single, cohesive paragraph (p.3, lines 78–88). This revised structure presents one overarching research question followed by three specific technical objectives, improving readability and removing the identified repetitions.

4. “Trajectories were kept only if the aircraft descended to a minimum of 300 feet or less above the aerodrome field elevation” – please justify the choice of 300 ft (even if it is an informed/educated threshold).

#### Response

The 300 ft AGL threshold was selected as an informed middle ground between the initial approach detection (500 ft AGL) and the final landing determination (150 ft AGL) identified by **karboviak2018**. Our empirical observations during the preliminary phase showed that a 500 ft threshold introduced excessive noise from aircraft merely overflying or entering the vicinity, while 300 ft effectively isolated aircraft committed to a touch-and-go or landing sequence. We have updated Section 2 to explicitly state the criteria for AGL threshold selection.

5. Around line 135, the paragraph becomes a bit wordy. On rereading, it does not seem to add much beyond what is already stated before it.

#### Response

The paragraph of p.4 l.139-144 has been removed and its explanations integrated into the latter paragraph to improve conciseness.

6. Figure 3: the scale differs quite a lot from the first plot to the last. Is this intentional, and does it help communicate the point you want to make?

#### Response

The variation in scale is intentional and reflects the spatial extent of the *aerodrome circuit candidate* segments. The first subplot includes non-identified manoeuvres prior to the aircraft approaching within 1 NM of the runway. In contrast, the final two subplots focus on the immediate vicinity of the runway to distinguish between a valid traffic circuit and a rejected candidate. Maintaining a fixed scale for the first plot would result in a lack of visible data, while the current approach intuitively demonstrates why specific segments are discarded despite being initially flagged as candidates.

7. Line 190: there is a mixed use of citation styles. I suggest removing author names and keeping only “[28]” (same comment for line 237).

#### Response

The citation style of lines 190 and 237 has been standardised to numeric format as suggested.

8. Figure 4: “Level” and “N/A” use the same colour. Could you change one so they are distinct? Also, around 1600, 2100, and 2200, peaks and valleys appear to overlap—could you explain why?

#### Response

Figure 4 has been updated to use distinct colours for the “Level” and “N/A” states. The overlapping of peaks and valleys in the previous version was due to an erroneous configuration of the prominence parameter in SciPy’s `find_peaks()` function. By setting a prominence of 0.1, we now ensure that only significant vertical variations are identified, effectively resolving the overlap and improving the robustness of the peak/valley detection.

9. Related to the previous point: the detection step using SciPy for local maxima and minima is not very clear. Please provide more details on the method and parameters.

#### Response

The detection of local maxima and minima is implemented using the SciPy `find_peaks()` function, which identifies peaks based on a comparison of neighbouring values within a 1-D array. To ensure the method captures meaningful flight dynamics while ignoring signal noise, we applied a prominence threshold of 0.1. This parameter measures the vertical distance between a peak and its lowest contour line, ensuring only prominent features are extracted. Detailed documentation of this implementation can be found in **[api-scipy]**. We have updated Section 2 to explicitly state the criteria for peak selection.

10. Line 214: what criteria were used for manual labelling of the ground truth? Did you use multiple labellers (e.g., to vote on edge cases)?

#### Response

While the primary labelling was performed by a single researcher, edge cases were reviewed and validated through team discussions to ensure consistency. The classification criteria centred on identifying enclosed trajectory loops that lacked internal intersections and did not deviate excessively from the standard traffic circuit geometry. These criteria, along with illustrative examples of both accepted and rejected circuits, are detailed on page 7 (lines 205–212 and Figure 6).

11. Page 8 lists the machine learning models, but it would be helpful to explain how/why these models were chosen (e.g., based on prior studies, suitability for the task, or simply to compare a range of approaches).

#### Response

The selection of the five models was intended to evaluate a broad spectrum of architectural approaches, ranging from traditional classifiers (Logistic Regression, Random Forest) to deep learning models capable of capturing spatial and temporal dependencies (1D CNN, LSTM, GRU). This allowed us to benchmark whether circuit classification benefits more from local geometric features or long-range sequential memory. We have added a clarifying sentence to Section 3 to explicitly state the rationale for these choices.

12. In Section 3, it could be useful to include a feature-importance analysis showing which features contribute most to the classification.

#### Response

We agree that a feature-importance analysis provides valuable insight into the models' decision-making processes. Accordingly, we have added a new Figure 9 to Section 3.1.

This addition includes two paragraphs of text detailing the results. The analysis highlights that:

- The standard deviation of various parameters, particularly the runway centreline alignment (*Alignment\_std*), is the most significant predictor across both non-sequential models.
- Flight phase-related features, specifically those associated with *climb* and *descent*, show higher importance than other phases.
- There is a high level of consistency between the Random Forest (Gini importance) and Logistic Regression (signed coefficients) regarding the top-performing features.

A new sentence has also been added in the Discussion (Section 4) regarding these new findings.

13. Consider adding to the Results section a few examples of failure cases (false positives/false negatives). It would also be informative to show cases where one model fails (e.g., LSTM) but another succeeds (e.g., 1D CNN). This could provide stronger insights than speculating about performance differences (line 413–415).

#### Response

We appreciate the reviewer's suggestion to include a qualitative analysis of failure cases. To address this without significantly increasing the manuscript's length, we have added a brief discussion of characteristic failure modes in the Discussion (Section 4).

We examined misclassified cases and found that false negatives typically arose from severely incomplete ADS-B trajectories (where critical circuit legs were missing due to signal dropouts) or highly irregular circuit geometries that deviated substantially from standard patterns. False positives occurred primarily at ground-level taxi or landing roll-out segments where the LSTM misinterpreted the transition from descent to flat, low-altitude movement as the final phase of a circuit pattern. Specifically, we observed LSTM false positives on segments maintaining nearly constant altitude at field elevation (1550 ft at LSZT) with minimal runway distance variation (0–0.2 NM), where the sequential context of a preceding descent led the LSTM to hallucinate circuit completion despite the aircraft remaining on the ground. The 1D CNN correctly rejected these cases by focusing on local geometric features rather than sequential transitions. This analysis reinforces our interpretation that local geometric consistency, rather than long-range temporal coherence, is the primary discriminator for circuit classification.

By describing these specific instances, we believe the manuscript now provides the "stronger insights" requested without the need for additional extensive visualisation.

14. In the Discussion, it is unclear whether incomplete ADS-B data is due to coverage limitations or because some GA aircraft are not equipped with ADS-B. Do you have proposals to mitigate these limitations?

#### Response

We have clarified in the Discussion (p.13, lines 423–436) that data constraints arise from both a lack of on-board transponders in some GA aircraft and signal coverage gaps caused by terrain or a lack of ground receivers. To mitigate these issues, we have expanded the Conclusions and Outlook section to

suggest the integration of multi-source data and the strategic deployment of supplementary receivers in future research.

15. The Discussion appears to contain two paragraphs that are the same.

#### Response

We thank the reviewer for noting this oversight, the duplicated paragraph has been accordingly removed.

16. Line 520 has two full stops.

#### Response

We have amended the text to fix the typographical error of p.15 l.520.

### 3.3 Response to reviewer 3

Some abbreviations are introduced more than one time, e.g., Automatic Dependent Surveillance Broadcast (ADS-B), Collaborative Decision Making (A-CDM), Long Short-Term Memory (LSTM). Average Precision is introduced as AP and AP (PR) in different places. The abbreviation ML is used in the manuscript without being explicitly introduced, while the full term machine learning also appears. Using a consistent convention for introducing abbreviations at first occurrence throughout the manuscript would improve clarity.

#### Response

We thank the reviewer for identifying these inconsistencies. We have performed a comprehensive audit of the manuscript to ensure a consistent convention for abbreviations. Specifically:

- All technical terms, including ADS-B, A-CDM, LSTM, are now defined at their first occurrence only, with the acronym used consistently thereafter.
- The notation for Average Precision has been standardised to AP throughout the text and figures to avoid confusion.
- We have ensured that Machine Learning (ML) is explicitly introduced in the Introduction and that the abbreviated form is used in all subsequent sections.

With these changes we hope to have improved the clarity and flow of the manuscript.

Figure 8 is not explicitly referenced in the main text. The caption is also relatively long; part of the explanatory content could be moved into the main text when introducing the figure.

#### Response

A new paragraph has been added in Section 3 introducing Figure 8, and the caption has been significantly reduced.

There appears to be a duplicated paragraph in the manuscript: one starting at line 408 and a repeated version starting at line 424. Please remove the duplication and retain a single instance.

Response

We thank the reviewer for noting this oversight, the duplicated paragraph has been accordingly removed.

In the sentence introducing Table 2, the text states that "the same evaluation metrics as in Table 1 are used," although the reported metrics differ slightly (e.g., ROC AUC and AP are not shown in Table 2). A small rephrasing would be good here.

Response

We have rephrased the introductory sentence for Table 2 to accurately reflect the metrics presented, removing the incorrect reference to Table 1 to ensure clarity.

In Section 4, it may be helpful to report how many aerodrome movement candidates were excluded during the manual labelling process, for example as a percentage of the total candidates. This would help quantify the extent of the filtering and clarify its potential impact on model bias and performance.

Response

We have added the requested breakdown to Section 4 to quantify the filtering process. Out of the 3,297 labelled segments, 695 (21.1%) were identified as *traffic circuits* and 2,422 (73.5%) as *non-circuits*. The remaining 180 segments (5.46%) were categorized as *omitted* due to ambiguity or data quality issues. This explicit quantification helps clarify the dataset's composition and the extent of the manual filtering phase.

## 4. Review - round 2

### 4.1 Reviewer 1

Thank you to the authors for their careful and thorough response to the reviewer comments.

Comments have been adequately addressed, and the requested corrections have been implemented in the revised manuscript. I appreciate the authors' efforts in improving the clarity, consistency, and presentation of the paper.

At this stage, I do not have further substantive comments, and I do not consider an additional round of revision necessary.

Recommendation: Accept Submission

### 4.2 Reviewer 2

The authors have carefully addressed all comments raised in the previous review round. The revisions have improved the clarity, and overall quality of the manuscript. I have no further comments. I recommend the manuscript for publication in its current form.

Recommendation: Accept Submission

### 4.3 Reviewer 3

The authors have addressed all my comments.

Recommendation: Accept Submission