

EDITORIAL

Reviews and Responses for Generation of Vertical Profiles with Neural Ordinary Differential Equations Trained on Open Trajectory Data

Authors: Gabriel Jarry and Xavier Olive

Reviewers: Raúl Sáez, Manuel Waltert, and Yutong Chen

Editor: Martin Strohmeier

1. Original paper

The DOI for the original paper is <https://doi.org/10.59490/joas.2026.8457>

2. Review - round 1

2.1 Reviewer 1

General assessment:

In this paper, the authors use neural ordinary differential equations to learn vertical dynamics from ADS-B data. The paper is well structured and well-written, and the contribution is clear. Many details about the methodology are given, so the study is highly reproducible. I only have some minor comments/doubts that could be briefly clarified in the paper.

2.2 Reviewer 2

General assessment:

The authors train a model based on Neural Ordinary Equations (Neural ODEs) to generate vertical profiles of transport aircraft. Their approach uses both ADS-B and Mode S Enhanced Surveillance data obtained through the OpenSky Network. In doing so, the authors contribute to the literature by providing a reproducible and open-source Neural ODE framework built entirely on open-access data. I would like to thank the authors for this excellently written paper, which addresses an important gap in the existing literature. In the following, I will highlight a few minor observations intended to further strengthen the manuscript.

2.3 Reviewer 3

General assessment:

The paper addresses an interesting problem and demonstrates that Neural ODEs can be trained on open ADS-B/Mode S data for vertical dynamics reconstruction. The direction is promising, but several aspects require strengthening.

3. Response - round 1

3.1 Response to reviewer 1

Comment 1: Minimizing local deviations in stable control regimes

In line 269, the authors wrote “These local deviations, while minor in magnitude, highlight the inherent difficulty for data-driven models to reproduce stable control regimes governed by subtle autopilot dynamics.” Would there be any way to minimize these deviations when working with data-driven models?

Response

We thank the reviewer for this insightful point. Minimizing these drifts is indeed a key challenge in moving data-driven models toward operational use. Incorporating physical constraints into the training process (e.g., penalizing energy changes during level-flight labels) would force the model to respect the steady-state dynamics that BADA inherently assumes.

Change in manuscript: We have added a clarification in Section 6 regarding potential strategies to minimize these deviations. Specifically, we discuss the use of physics-based regularization, such as penalizing energy changes during level-flight segments, to enforce steady-state equilibrium conditions and reduce cruise drifts.

Comment 2: Challenges of learning lateral dynamics

Do the authors think that learning lateral dynamics from open data is a bigger challenge than the vertical part? I am just curious about why the authors left that part out of their study. I am not asking the authors to perform such a study now, as it will be part of future work; I am rather interested in which might be the challenges compared to the vertical dynamics estimation when applying such a methodology.

Response

While lateral dynamics are intuitively less complex from a purely kinematic standpoint, our decision to focus on the vertical-longitudinal profile was motivated by the fact that it represents the most challenging modelling task. Specifically, accurately capturing the energy balance governing climb and acceleration requires learning unobservable physical dependencies, such as the aircraft’s mass and the energy share factor (how available power is partitioned between potential and kinetic energy). Lateral dynamics, although involving challenges such as turn detection and the distinction between orthodromic (great circle) and loxodromic paths, remain largely decoupled from these fundamental propulsion constraints. We therefore prioritized the vertical axis as the core complexity of trajectory reconstruction.

Change in manuscript: We have added a clarification in Section 6 explaining the deliberate focus on vertical-longitudinal dynamics due to the underlying physics of energy partitioning, which represents a higher modelling challenge than lateral kinematics.

Comment 3: Comparison with QAR data results (Reference 5)

I understand the authors applied the same methodology to QAR data in Reference 5. Could the authors include a brief discussion/comparison of how the results obtained in this paper compare to those obtained in reference 5? I think this will further strengthen the contribution of this paper.

Response

We thank the reviewer for this suggestion. The methodology applied to QAR data in Reference 5 indeed shares the same Neural ODE architecture, but operates on fundamentally different input data. QAR (Quick Access Recorder) data provides high-frequency measurements (typically 4–25 Hz) with direct

access to internal flight parameters such as fuel flow, engine thrust settings (N1), aircraft mass, and precise autopilot targets. In contrast, ADS-B and Mode S EHS data are lower-frequency (around 1 Hz or less) and require the reconstruction of control intent from noisy and irregular signals, as discussed in Section 6. This introduces significant uncertainty in variables such as selected altitude, selected speed, and vertical rate targets, which propagates into both training and evaluation.

As a result, we expect the QAR-based model to benefit from cleaner inputs and richer physical observability (mass, thrust, fuel flow), leading to higher accuracy and better interpretability. The ADS-B approach, while more scalable and globally accessible, operates under noisier conditions and lacks direct propulsion information. Despite these limitations, the Neural ODE framework demonstrates promising vertical profile reconstruction quality, suggesting that the continuous-time formulation provides sufficient regularization to absorb inconsistencies in open surveillance data. The key trade-off is therefore between data accessibility and physical interpretability: ADS-B enables large-scale trajectory modelling without proprietary data, but cannot support absolute performance metrics such as fuel consumption.

Change in manuscript: We have added a comparative discussion in Section 6 highlighting the differences between ADS-B and QAR-based approaches. We emphasize that QAR data benefits from direct access to propulsion parameters (N1, fuel flow, mass) and cleaner control targets, whereas ADS-B requires noisy intent reconstruction. Despite these differences, the Neural ODE framework achieves promising trajectory reconstruction performance, demonstrating robustness to input uncertainty.

3.2 Response to reviewer 2

Minor Findings

Comment 1:

Lines 34–35: I suggest removing the “vertical,” part of the sentence, as “longitudinal dynamics of transport aircraft” should be sufficient.

Response

We agree and have removed “vertical” from the sentence.

Change in manuscript: Removed “vertical” from “longitudinal dynamics” in the Discussion section.

Comment 2:

Line 40: I would even argue that control inputs are not “limited in observability” but rather hidden.

Response

We agree that “hidden” is more accurate and have made this change.

Change in manuscript: Changed “limited observability of control inputs” to “hidden control inputs” in the Introduction.

Comment 3:

Line 57: It’s just cosmetics, but please cite first source [1] and then source [2].

Response

Corrected the citation order.

Change in manuscript: Swapped citation order in Literature Review section.

Comment 4:

Lines 78–79: I suggest not using the term “physical guarantees” but rather “flyability” of the generated trajectories.

Response

We agree that “flyability” is more specific to the aviation domain and have made this change.

Change in manuscript: Replaced “physical guarantees” with “flyability” in the Literature Review.

Comment 5:

Line 115: I would use the term “ICAO aircraft type designators” instead of “ICAO type designators”.

Response

Corrected to use the full official ICAO terminology.

Change in manuscript: Changed to “ICAO aircraft type designators” in Section 3 (Data Sources).

Comment 6:

Methods section: I suggest reviewing tense coordination in the Data Sources and Methods sections. Generally, present tense should be used for “standard methods”, figures, tables, and the paper itself. For your procedures, however, it is recommended to use the past tense (e.g. see my comment on Line 118).

Response

We have reviewed the entire Data Sources and Methods sections to ensure consistent tense usage: present tense for general methods and past tense for our specific procedures. Three corrections were made.

Change in manuscript: Corrected tense coordination throughout Sections 3 and 4.

Comment 7:

Line 118: I would write “For each selected aircraft type, we retrieved...”.

Response

We have added “type” as suggested.

Change in manuscript: Changed “For each selected aircraft” to “For each selected aircraft type” in Section 3.

Comment 8:

Lines 114–120: What is the geographic area of observation? (Worldwide?)

Response

The geographic coverage is indeed worldwide, as ADS-B data from the OpenSky Network provides global coverage. We have clarified this in the manuscript.

Change in manuscript: Added “worldwide” to clarify geographic coverage in Section 3.

Comment 9:

Lines 133–134: Please consider adding a proper citation (or at least a URL) to the ECMWF data.

Response

We have added the proper citation to the ERA5 reanalysis dataset.

Change in manuscript: Added citation for ERA5 in Section 3.

Comment 10:

Line 134: Please consider a proper citation of the fastmeteo library (J. Sun and E. Roosenbrand, “Fast contrail estimation with Open-Sky data,” Journal of Open Aviation Science, vol. 1, 2023. doi: 10.59490/joas.2023.7264).

Response

We have added the citation as suggested.

Change in manuscript: Added citation for fastmeteo library in Section 3.

Comment 11:

Table 4, description of selected altitude: I suggest writing “...extracted from EHS messages”.

Response

Corrected for consistency with the terminology used throughout the paper.

Change in manuscript: Changed “EH messages” to “EHS messages” in Table 4.

Comment 12:

I see no reference to Figure 1 in the text. I think this would be crucial for the understanding of the methodology.

Response

We agree and have added a reference to Figure 1 in the methodology introduction.

Change in manuscript: Added Figure 1 reference in Section 4 introduction.

Comment 13:

Figure 1: Is there a reason why the number 1 in $e_1(t)$ is not subscripted (similar to $e_0(t)$)?

Response

This was an oversight. We have corrected the notation to use subscript.

Change in manuscript: Corrected $e_1(t)$ notation in Figure 1.

Comment 14:

Paragraph on Lines 181–193: In the current manuscript, it is hard to understand what the exact inputs and outputs of the model are. Does the description given here refer to the architecture of the derivative layer? In addition, it could be helpful to include a brief reminder of the structural layer.

Response

We agree that this section needed clarification. We have restructured Section 4 to follow the logical data flow. First, we added an itemized overview immediately after the state vector definition to explicitly describe the model's inputs, the sequential processing through the Trajectory and Derivative layers, and the final ODE integration. Second, we reordered the detailed descriptions so that the analytical Trajectory Layer formulas appear before the neural Derivative Layer implementation, ensuring a cohesive narrative.

Change in manuscript: Restructured Section 4 to include an itemized architecture overview after Equation 1 and reordered subsections to match the logical flow (Trajectory Layer formulas before Derivative Layer details).

Comment 15:

To ensure the model is correctly understood: The input is a state $x(t)$, from which $e_1(t)$ is derived. This information is then passed to the derivative layer to obtain the gradient for each parameter at $x(t)$. Put differently, the model outputs the direction in which the state parameters should evolve from the observed state at time t to the next state at time $t+1$. I would recommend stating this explicitly earlier in the paper.

Response

We fully agree with this description of the model flow. As detailed in our response to Comment 14, we have added an explicit itemized description at the beginning of Section 4 stating exactly this: inputs $x(t)$ are processed by the trajectory layer to derive $e_1(t)$, which is then passed to the derivative layer to obtain the gradients. This flow is now stated explicitly earlier in the paper, as recommended.

Change in manuscript: See changes for Comment 14 (added itemized architecture overview).

Comment 16:

Do we agree on the fact that $u(t)$ and $e_0(t)$ do not need to be determined autoregressively and are available for all t ?

Response

We fully agree. $u(t)$ and $e_0(t)$ are indeed exogenous inputs defined for the entire flight duration. We have added an explicit note in Section 4 (after Equation 5) to clarify that these variables are available for all t and are not generated autoregressively.

Change in manuscript: Added explicit clarification in Section 4 that $u(t)$ and $e_0(t)$ are exogenous inputs available for all t .

Comment 17:

Line 194: I find it a bit misleading to call it the “trajectory layer” as it is not trained and purely analytical.

Response

We appreciate the reviewer's perspective on this terminology. We chose the term “layer” to reflect its structural role as a functional block within a **differentiable computation graph**. In modern deep learning frameworks, a “layer” defines a transformation of the data, regardless of whether it contains learnable parameters. This is consistent with standard components like pooling, normalization, or fixed

positional encoding layers.

In our PINN architecture, this module is implemented directly in PyTorch to ensure it is a first-class component that propagates gradients. To avoid any ambiguity, we have consistently explicitly labeled it as **analytical** in the manuscript (e.g., at Line 183). This nomenclature emphasizes its role as a modular architectural component while being transparent about its deterministic nature.

Change in manuscript: The terminology was maintained for architectural consistency. We have double-checked that its “analytical” and non-trainable nature is clearly specified in the model description to prevent any misunderstanding.

Comment 18:

Line 198: The derivative layer is actually the identity for V_z and V_{GS} because they were computed in the trajectory layer. It could be nice to show this in Figure 1.

Response

We have added clarification in the text that V_z and V_{GS} are computed analytically in the trajectory layer and passed directly without transformation.

Change in manuscript: Added clarification in Section 4 about V_z and V_{GS} identity mapping.

Comment 19:

Line 200: I would mention that Figure 1 describes only the estimation to go from $x(t)$ to $x(t+1)$, and that the full trajectory is built autoregressively.

Response

We agree that this distinction is crucial. We have updated the caption of Figure 1 to explicitly state that it represents the single-step derivative estimation, while the full trajectory is generated autoregressively via ODE integration.

Change in manuscript: Updated Figure 1 caption to clarify single-step estimation vs autoregressive generation.

Comment 20:

Figure 1 & Lines 195, 202: In the text, you mention a “trajectory layer”; in the figure, however, this layer is referred to as “trajectory analytics”. I suggest using the same term in both the text and the figure.

Response

We have corrected Figure 1 to use “Trajectory Layer” to match the text.

Change in manuscript: Changed “Trajectory Analytics” to “Trajectory Layer” in Figure 1.

Comment 21:

Line 181: I suggest writing “In our implementation, we used a ...” (this ties into my comment regarding tense coordination).

Response

Corrected as part of the tense coordination review.

Change in manuscript: Changed “we use” to “we used” in Section 4.

Comment 22:

I would put Equation 5 earlier in the section to better explain the implemented model. Moreover, I think the reference to Table 4 comes too late.

Response

We agree with the reviewer that formalizing the ODE model early improves readability. We have moved Equation 5 (and its description) to adhere to the structure proposed in the introduction of Section 4, placing it immediately before the detailed breakdown of the layers. Regarding Table 4, we have added a reference to it (“Summary of features”) immediately where the state and context variables are introduced, ensuring the reader has access to the feature definitions early in the section.

Change in manuscript: Moved Equation 5 to the beginning of the methodology section and added early reference to Table 4.

Comment 23:

Line 218: Is the loss computed on the state vector at a specific time T , or on a full aircraft trajectory (from $t = 0$ to T)?

Response

This is an important clarification. The loss is computed on partial trajectories generated autoregressively. Specifically, we use sequences of $N = 60$ time steps. The ODE is integrated forward from t_0 , and the loss is accumulated over all N predicted states $x(t_1) \dots x(t_N)$ comparing them against the ground truth. We have updated Equation 10 and the surrounding text to make this explicit.

Change in manuscript: Updated loss function description and Equation 10 to explicitly specify the summation over autoregressive sequences of $N = 60$ time steps.

Comment 24:

Line 223: I would say “the weights of the derivative layer” to stay consistent with Figure 1.

Response

We agree that this phrasing is precise and consistent with the figure. We have replaced “weights of the neural ODE” with “weights of the derivative layer” in the text.

Change in manuscript: Changed “weights of the neural ODE” to “weights of the derivative layer” in Section 4.

Comment 25:

Line 209: To match the formatting of Line 202, I suggest typesetting “derivative layer” in bold on Line 209 as well.

Response

We have followed this recommendation. To ensure consistency throughout the text, we have typeset “**derivative layer**” in bold at every occurrence in Section 4.

Change in manuscript: Applied bold formatting to all instances of “derivative layer” in Section 4.

Comment 26:

Line 214: Since RK4 is only used once, I suggest removing the abbreviation.

Response

We have added the full definition. The text now reads “4th order Runge–Kutta (RK4)”.

Change in manuscript: Defined RK4 as “4th order Runge–Kutta”.

Comment 27:

Line 255: What exactly is meant by “...at first point of each flight”? Do you mean an aircraft on the runway ready for take-off?

Response

The reviewer is correct to point out this ambiguity. This refers to the first data point available in the ADS-B recording for a given flight, which serves as the initialization point for the integration. We have clarified the text to read: “applied at the first available point of the recorded trajectory (depending on ADS-B coverage)”.

Change in manuscript: Clarified that the starting point corresponds to the first available point of the recorded trajectory.

Comment 28:

Why is Figure 3 referred to before Figure 2 in the text?

Response

We have reordered the figures to ensure correct sequential referencing.

Change in manuscript: Reordered figures in Section 5.

Comment 29:

Line 260: Are you comparing the errors for a whole aircraft trajectory estimated autoregressively, or predicted point by predicted point (where the input of the model is always an observed point, and not an estimated point)? Did you look at the propagation of error when trajectories are built autoregressively to predict several time steps in advance? Do you consider mitigation measures?

Response

This is a crucial point indeed. The evaluation is performed **autoregressively** on the full trajectory (from $t = 0$ to T). The errors reported are the accumulated errors over the entire flight, which is a much more rigorous test than point-by-point prediction. We have added a sentence in Section 5 (Results) to make this explicit.

Change in manuscript: Clarified in the Results section that evaluation metrics are computed on full autoregressive trajectories ($t = 0$ to T).

Comment 30:

Figure 2: Does the data shown in orange correspond to y_{true} (see Equation 6)? In my view, the “orange data” contains quite a few outliers and fluctuations (especially in the CAS and vertical speed signals) that may not necessarily reflect the actual underlying flight dynamics. It might therefore be worth considering showing the filtered/smoothed data instead of the raw data in Figure 2.

Response

We deliberately show raw, unfiltered ADS-B data to demonstrate the model’s robustness to measurement noise and quantization artifacts inherent in open surveillance data. This transparency is important for reproducibility and shows that the model can handle real-world data quality.

Change in manuscript: No change (justified choice to show raw data).

Comment 31:

Figure 3: The selected value in the vertical speed signal is difficult to read due to the current color styling. Perhaps choosing a different color or visual encoding would improve clarity.

Response

Indeed there is a lot of information to bring together and we tried many combinations, to no avail. We think though that this is not critical on this plot.

Change in manuscript: No change.

Comment 32:

Figure 2: Since “altitude (in ft)”, “TAS (in kts)”, and “vertical speed (in ft/min)” are titles, I recommend capitalizing at least the first letter of each phrase.

Response

Since this change is purely stylistic and does not affect clarity or interpretation of the figure, we have chosen to keep the current formatting, consistent with other figures.

Change in manuscript: No change.

Comment 33:

Lines 256–264: I suggest combining the two paragraphs into one, as they describe the same aspect.

Response

We agree that these two paragraphs discuss the same results. We have merged them into a single paragraph to improve the flow of the text.

Change in manuscript: Merged the paragraph starting with “Overall...” with the preceding one.

Comment 34:

Figure 3: I appreciate Table 3, as it provides a clear and informative overview. However, it is somewhat difficult to extract the absolute numerical values (and therefore the resulting errors) from this visualization. Is there a way to summarize the numerical errors (e.g. those discussed in Lines 265 to 272) in a more accessible form?

Response

We appreciate the reviewer's observation. Figure 3 is intentionally designed as a comparative visualization to emphasize regions where the predictive model performs worse than BADA, rather than to allow direct reading of absolute numerical errors. The numerical values referenced in Lines 265–272 are provided in Table 3 for clarity.

Change in manuscript: No change.

Comment 35:

Lines 273–292: I recommend that the authors more clearly separate the presentation of the results (Section 5) from the discussion and interpretation (Section 6). This distinction would help readers differentiate between the study's empirical findings and the authors' interpretation of those findings.

Response

We agree with the reviewer that these paragraphs interpret the results and discuss potential causes of discrepancy (BADA methodology, selected altitude noise, mass assumptions) rather than presenting raw findings. We have moved this discussion to Section 6, grouping it under a new subsection "Interpretation of Benchmarking Results". We also merged the discussion on selected altitude noise with the corresponding limitation paragraph to avoid redundancy.

Change in manuscript: Moved the interpretative discussion from the Results section to the Discussion section.

3.3 Response to reviewer 3

Comment 1:

The manuscript should clarify more explicitly why existing physics-based and sequence-based methods cannot address the identified gap, and why Neural ODEs are necessary in this context.

Response

This is a fundamental point. We chose Neural ODEs to bridge the gap between rigid physics-based models (BADA) and black-box sequence models (LSTMs). Unlike BADA, our model learns latent operational factors (mass, airline policy) from data. Unlike LSTMs, the continuous-time formulation naturally handles the irregular sampling inherent to ADS-B data and allows for the seamless integration of physical constraints (kinematic equations). We have added an explicit justification at the end of Section 2.

Change in manuscript: Added a paragraph in the Related Work section explaining the necessity of Neural ODEs for handling irregular sampling and integrating physical constraints compared to sequence-based and purely physics-based methods.

Comment 2:

The theoretical contribution of the work needs to be articulated more clearly, beyond the engineering implementation, indicating what conceptual or methodological novelty the paper introduces.

Response

The primary contribution is methodological: we demonstrate that enforcing analytical kinematic constraints within a Neural ODE framework acts as a sufficient regularizer to reconstruct physically consistent dynamics even when key variables (mass, thrust) are unobservable. This goes beyond engineering by establishing “hybrid modelling” as a valid theoretical approach to recover high-fidelity dynamics from low-fidelity, incomplete open data. We have explicitly stated this contribution in the Introduction.

Change in manuscript: Added a statement in the Introduction clarifying the methodological contribution regarding hybrid modelling as a compensation for unobservable states.

Comment 3:

The fairness of the comparison with BADA needs to be justified, and potential sources of bias, such as noise, mass assumptions, and differences in control input handling, should be acknowledged.

Response

We fully agree. To address this, we have added a dedicated subsection “Interpretation of Benchmarking Results” in Section 6 (Discussion). This section explicitly acknowledges the sources of bias, including the noise in Mode S inputs (selected altitude), the difference in mass assumptions (fixed for BADA vs learned for Neural ODE), and the fact that BADA is a generic model not designed for individual trajectory reproduction with noisy controls.

Change in manuscript: See changes for Reviewer 2 (creation of “Interpretation of Benchmarking Results” subsection).

Comment 4:

The rationale and sources of key experimental settings and parameters should be documented, including thresholds, network configuration, ODE solver choice, and loss weighting, to improve transparency and reproducibility.

Response

We appreciate the concern for reproducibility. Most of these details were already present in Section 4:

- **Network Configuration:** 3-layer backbone (48 neurons) + 2-layer heads (48 neurons), ReLU activations (Section 4).
- **ODE Solver:** Explicit Euler scheme, as RK4 was deemed unnecessary for this sampling rate (Section 4).
- **Loss Weighting:** Inverse standard deviation of each feature (Eq. 10 and text below).
- **Training:** AdamW optimizer ($lr = 10^{-4}$), Batch size 512, 40k steps (Section 4).

Furthermore, to ensure absolute transparency and reproducibility, we have released the full source code, pre-trained weights, and preprocessing scripts on GitHub.

Change in manuscript: No definition added as the text already contained all specified parameters and GitHub repo available.