

Inferred Tactical Command Sampling for Stochastic Vertical Profile Generation Using ADS-B Data

Alex Fustagueras ^{*,1} Manuel Waltert ¹ Xavier Olive ² and Gabriel Jarry ³

¹Centre for Aviation, Zurich University of Applied Sciences, Winterthur, Switzerland

²ONERA DTIS, Université de Toulouse, France

³Aviation Sustainability Unit, EUROCONTROL Innovation Hub, Brétigny-sur-Orge, France

*Corresponding author: alex.fustagueras@zhaw.ch

Abstract

Simulation-based air traffic studies require synthetic flight profiles that reproduce the variability of real operations while remaining physically feasible. Historical observations may provide insufficient samples, while conventional performance models can under-represent operational variability. This paper investigates the generation of synthetic vertical flight profiles by introducing variability through inferred aircraft commands rather than directly through the resulting profiles. From historical ADS-B and Mode S observations, we infer sequences of altitude targets, speed schedules, and total specific energy plans. These form an empirical library from which new command sequences are sampled and propagated through a physics-guided Neural Ordinary Differential Equation flight dynamics model. The approach is evaluated using 4237 Airbus A320-family flights on five European routes. Independent comparison with Mode S selected altitudes shows that 90.8 % of inferred altitude targets agree within 100 ft. Replaying the inferred commands reconstructs the observed altitude profiles with a median mean absolute error of 143 ft, although only 74.7 % of inferred altitude plateaux are captured within 250 ft at their midpoint, compared with 99.97 % for the observations. The generated vertical profiles reproduce the broad altitude structure and timing of observed operations, with the closest distributional agreement during level flight and descent. Larger discrepancies remain during climb, particularly in altitude, and 48 % to 82 % of generated profiles fall within the observed operational spread, depending on the route. The approach thus enables the generation of varied vertical flight profiles through explicit, inspectable command sequences.

Keywords: Trajectory Generation; Tactical commands; Vertical profile Modelling; Neural Ordinary Differential Equations; ADS-B

Abbreviations: ADS-B: Automatic Dependent Surveillance–Broadcast, BDS: Comm-B Data Selector, FL: Flight Level, MAE: Mean Absolute Error, RDP: Ramer–Douglas–Peucker, RMSE: Root Mean Square Error

1. Introduction

A wide range of air traffic management applications, including conflict detection [1], capacity assessment [2], and environmental analysis [3], rely on aircraft trajectory data. While some applications operate directly on observed trajectories, simulation-based studies and Monte Carlo assessments often require large sets of trajectories representative of specific routes, aircraft types, weather conditions, or traffic configurations. Historical observations alone cannot always provide sufficiently large and representative samples, especially when evaluating rare operating conditions, future procedures, or route-specific uncertainty. This limitation can be addressed by generating synthetic trajectories that reproduce the statistical characteristics of real operations while capturing the observed operational variability and remaining physically feasible [4].

Existing approaches to synthetic aircraft trajectory generation can broadly be divided into two categories: model-based methods, which generate trajectories from aircraft dynamics and performance

models, and data-driven methods, which learn trajectory characteristics directly from observed flight data. Classical model-based approaches remain the reference for physically consistent trajectory computation. Point-mass dynamics and aircraft performance models such as BADA [5] provide interpretable descriptions of aircraft motion, while open frameworks such as OpenAP [6] combine physics-based equations with parameters calibrated from open flight data, making performance-based simulation more accessible for research. However, these approaches still require the specification or estimation of quantities such as aircraft state, configuration, mass, weather, and intent. As a result, they may misrepresent the diversity of real operations, over-representing nominal behaviour and under-representing the variability introduced when airline procedures, controller interventions, or tactical speed and altitude management deviate from it.

At the same time, the increasing availability of open surveillance data through initiatives such as the OpenSky Network [7] has enabled data-driven methods to learn trajectory variability directly from historical observations. In terminal manoeuvring areas (TMA), probabilistic models have been used to predict and simulate operational flows from historical trajectories [8]. Similarly, synthetic trajectory-generation methods have demonstrated that realistic aircraft trajectories can be learned from observed flows using multivariate density models and copulas [9], while more recent work has extended this idea through deep generative trajectory models [10] and procedure-aware traffic models [11]. While these methods can represent patterns and variability present in the data, they generally introduce stochasticity at the level of the resulting trajectories.

Trajectory variability has both a lateral and a vertical component. This paper addresses the vertical profile only, that is, altitude, speed, and vertical rate along the flight. Modelling the trajectory as a stochastic object does not by itself expose the operational inputs that shape this profile. Here, the observed altitude, speed, and vertical rate profiles are not merely output patterns to be reproduced, but the visible consequences of underlying aircraft state, atmospheric conditions, operational intent, and tactical interventions by pilots and controllers [12]. Hybrid approaches have begun to introduce learned variability within physics-based trajectory models. Pepper and Thomas [13], for example, infer a stochastic correction to the nominal thrust profile from radar data and propagate sampled corrections through BADA to generate climb trajectories. Their formulation shows that variability can be learned in a latent model input rather than imposed directly on the resulting trajectory.

A more explicit input-based formulation is provided by physics-guided Neural Ordinary Differential Equations (Neural ODEs), which combine known analytical relations with control variables inferred from data [14]. Jarry *et al.* [15] applied this principle to aircraft vertical dynamics by coupling analytical kinematic relations with inferred altitude, speed, and vertical rate targets within a physics-informed Neural ODE trained on either Quick Access Recorder (QAR) data or open ADS-B and Mode S data [16]. Their work shows that these partially observed control targets can be inferred from open surveillance data and propagated through the learned dynamics to reconstruct physically consistent vertical profiles, while leaving open the question of how to generate multiple plausible trajectories under the same operational conditioning.

This paper addresses this gap by modelling the variability of inferred selected altitude targets, speed schedules, and total specific energy plans. The command events, speed schedules, and level-time allocations inferred from the flights of a route form an empirical library, from which new command sequences are sampled and propagated through a physics-guided Neural ODE to generate multiple plausible vertical profiles. More specifically, we investigate (i) whether the command representation inferred from historical ADS-B and Mode S observations retains sufficient information to reconstruct the corresponding flight, and (ii) whether sampled command sequences, when propagated through a physics-guided Neural ODE, generate trajectory pools that reproduce the variability observed under the same operational conditioning, here the route and the aircraft family.

2. Methods

The methodology comprises three main stages: inferring command sequences from observed aircraft trajectories, evaluating whether these sequences reproduce the observed flight profiles when applied to a flight dynamics model, and generating synthetic trajectories from newly sampled command sequences for comparison with observed operations. Section 2.1 introduces the dataset and its construction, while Section 2.2 develops the inferred command representation. Section 2.3 assesses the fidelity of these commands through replay using the flight dynamics model. Section 2.4 describes how new command sequences are sampled, and Section 2.5 how they are propagated through the flight dynamics model and compared with observed flights.

2.1 Dataset Construction

We retrieved historical flight data from the OpenSky Network [7] for aircraft operating on five European routes: London Heathrow–Barcelona (EGLL–LEBL), London Heathrow–Lisbon (EGLL–LPPT), Barcelona–Zurich (LEBL–LSZH), Zurich–Barcelona (LSZH–LEBL), and Zurich–Lisbon (LSZH–LPPT). For each route, we retrieved 1,000 flights from April to August 2024 for the two routes from London and from January to October 2024 for the other three, keeping the longest-tracked flights listed by OpenSky. These routes are mostly flown by the Airbus A320 family (A319, A320, A321, A20N, and A21N), on which the flight dynamics model was trained.

For each flight retrieved from OpenSky, two types of surveillance data are used. ADS-B provides the timestamped reference trajectory, including barometric altitude, ground speed, track, and vertical rate. All altitudes are barometric: flight levels (FL) refer to the standard setting of 1013.25 hPa, and feet to a local QNH. Mode S Enhanced Surveillance, decoded following Olive et al. [17] and implemented in *rs1090*, provides the altitude selected by the pilots on the autopilot control panel (MCP/FCU) through BDS register 4,0, which is retained as an independent reference. Its state registers, BDS 5,0 (true airspeed and roll angle) and BDS 6,0 (Mach and indicated airspeed), supply the observed speeds for the speed schedule and energy computation. In addition, ERA5 reanalysis [18] supplies temperature and wind, interpolated along each trajectory using *fastmeteo* [19].

The raw messages are cleaned and each flight passes two quality checks, described in the supplementary material together with the number of flights retained at each stage. Since the retrieval does not filter by aircraft type, flights by other types are then removed, using the OpenSky aircraft database, which leaves 4237 A320 family flights in the final dataset. For each route, these flights are randomly divided into two subsets: the *library set*, comprising 2,120 flights in total, from which new command sequences are sampled, and the *held-out set*, comprising 2,117 flights, whose observed trajectories serve as the reference for evaluating the generated trajectories.

2.2 Inferred Command Representation

Replaying or generating the vertical profile of a flight requires a sequence of commands specifying the aircraft's altitude and speed targets as well as its commanded vertical motion. These commands cannot be read reliably from the surveillance data. The selected altitude is reported in BDS 4,0 Comm-B replies, but the BDS code is transmitted only in the ground interrogation and not in the reply, so a third party observing the replies must infer it from the message content [20]. The decoded selected altitude is therefore affected by decoding artefacts and spurious step changes [16]. Moreover, the speed targets selected in the aircraft's flight management system are not reported at all. We therefore infer the required commands from the observed trajectory and store them as an inferred command representation, a compact, time-dependent specification of the inputs required by the flight dynamics model. Throughout this study, the term command is used in this operational sense, as a quantity that drives the flight dynamics model, rather than as a recovered cockpit setting or pilot input.

The command representation consists of three inferred quantities. First, the inferred selected altitude target $h_{sel}^*(t)$, where t denotes the elapsed time since the beginning of the flight, identifies the levels at which the aircraft was held. Second, the inferred speed schedule specifies a true airspeed (TAS) target $V^*(t)$ at every second. Finally, the inferred total specific energy plan $p^*(t)$ specifies the rate of change of the aircraft's total specific energy. The flight path angle $\gamma^*(t)$ is then derived from the energy plan and the speed schedule through the aircraft's energy balance. The resulting command sequence is

$$u(t) = (h_{sel}^*(t), V^*(t), \gamma^*(t)), \quad (1)$$

where the asterisk distinguishes quantities belonging to the command representation from observed aircraft states.

2.2.1 Selected Altitude Target

The inferred selected altitude target is a surrogate for the levels at which the aircraft was held, rather than a reconstruction of the altitude selected in the cockpit. A level-off may result from an altitude constraint associated with a standard instrument departure (SID) or standard terminal arrival route (STAR), a vertical profile managed by the aircraft's flight management system, or a temporary clearance issued by an air traffic controller. A level is detected wherever the aircraft's vertical rate, smoothed by two passes of a bilateral filter [21], remains within ± 250 ft/min for at least 10 s. It is retained only if the observed altitude also remains within ± 250 ft, or ± 50 ft at and above FL100, over that interval. The mean altitude of each retained level is rounded to the nearest 500 ft, or 100 ft below FL100 when BDS 4,0 reports a local QNH¹ setting, so that targets from different flights can be compared and recombined. This rounded altitude becomes the target from the end of the preceding level. The resulting h_{sel}^* is a step-held sequence of altitude targets, with the runway altitude at departure and the observed altitude at the end of the flight serving as its initial and final values, respectively. Neither represents an inferred selected level.

Two durations are derived for each altitude target. The first, τ_{target} , is the time between the introduction of a new target h_{sel}^* and the first entry of the observed altitude into the ± 250 ft band centred on it. The second, $\tau_{plateau}$, is the remaining time at that target until the next change, referred to as the target plateau. Each transition between successive altitude targets, together with its τ_{target} and $\tau_{plateau}$, defines a command event. Figure 1 illustrates these quantities for a *Swiss International Airlines* flight from Barcelona to Zurich, which climbs to flight level 300 (FL300), maintains this level, and subsequently climbs to FL360, with each transition characterised by its own pair of durations.

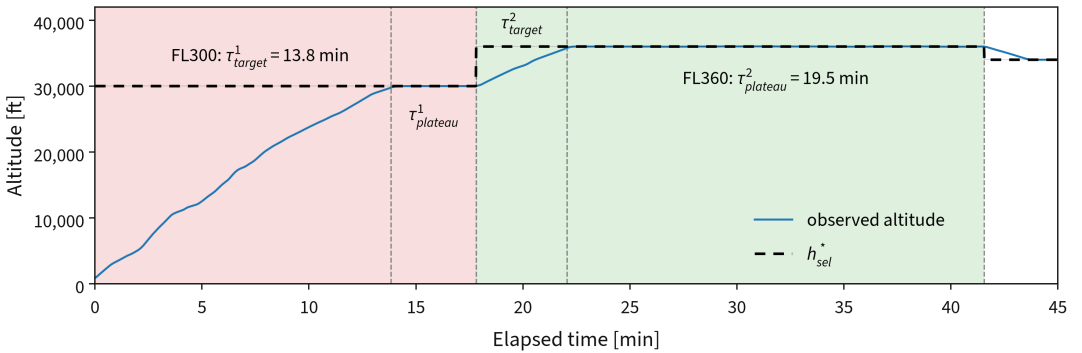


Figure 1. *Swiss International Airlines* flight SWR8ZF from Barcelona to Zurich on 12 April 2024, with the observed altitude in blue and the inferred h_{sel}^* in dashed black. Two command events are highlighted: the climb to FL300 (red) and the climb to FL360 (green).

¹QNH: barometric atmospheric pressure adjusted to mean sea level

By construction, the inferred target identifies the levels that the aircraft actually reached, rather than every altitude that may have been selected in the cockpit. Since the decoded BDS 4,0 selected altitude is not used to infer h_{sel}^* , it provides an independent check, with which 90.8 % of the inferred targets agree within 100 ft (see the supplementary material).

2.2.2 Inferred Speed Schedule

Each flight is assigned a single inferred speed schedule, expressed in calibrated airspeed (CAS) and Mach, following the operating pattern of the Airbus A320 family described by Airbus [22]. During climb, the schedule specifies a CAS target below FL100, consistent with the common 250 kt speed limitation, followed by a climb CAS target between FL100 and the crossover altitude, above which a constant Mach is maintained. During descent, the schedule transitions from the constant Mach target to separate descent CAS targets above and below FL100. The complete schedule therefore consists of four CAS values, two for climb and two for descent, and one Mach value.

Each CAS value is the median observed CAS within its phase and altitude band. The observed CAS is the indicated airspeed (IAS) where received, used as a proxy for CAS. Otherwise, it is derived from the received Mach or, failing that, from the ERA5 wind-corrected TAS. Above FL100, samples flown at the cruise Mach are excluded. The Mach value is the median received Mach on the cruise plateau, the first plateau at the highest altitude target. CAS values are rounded to 5 kt and the Mach value to 0.01. The speed schedule is determined once per flight, rather than being re-estimated for each command event.

At every second, the schedule is converted into a TAS target using the International Standard Atmosphere at the observed altitude [23], so that it does not depend on the ERA5 temperature. On the cruise plateau, the cruise Mach is flown. Elsewhere, the climb CAS applies until the end of the cruise plateau and the descent CAS after it, each with its value below or above FL100, and whichever of this CAS and the cruise Mach gives the lower TAS is flown. The crossover from CAS to Mach is therefore where both give the same TAS, rather than a separately estimated parameter. Figure 2 illustrates the three segments of the inferred climb speed schedule for a *British Airways* flight to Lisbon, expressed in TAS: the CAS target below FL100, the climb CAS target between FL100 and the crossover altitude, and the constant Mach target above the crossover altitude.

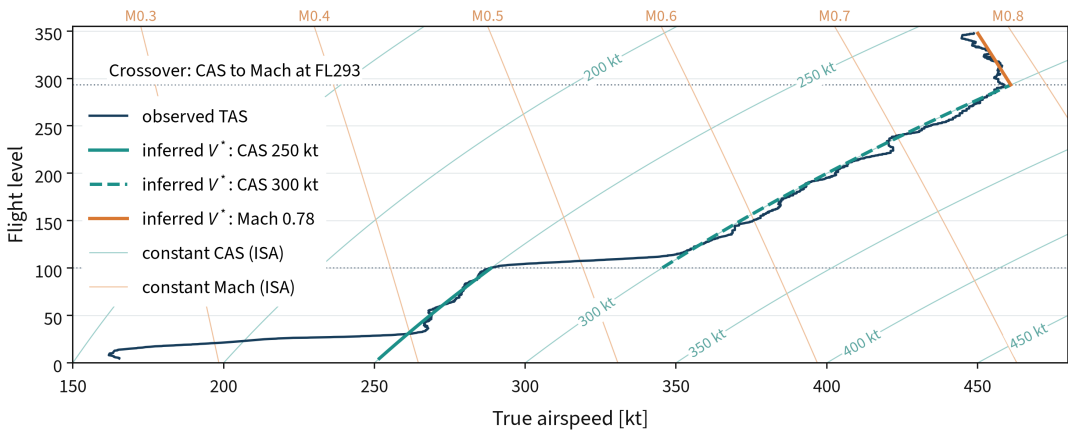


Figure 2. *British Airways* flight BAW5LI from London Heathrow to Lisbon on 19 May 2024, showing the observed TAS and the inferred speed target V^* during climb. The inferred schedule comprises three segments: a CAS target below FL100 (teal, solid), a climb CAS target between FL100 and the crossover altitude (teal, dashed), and a constant Mach target above the crossover altitude (orange). Thin lines show constant CAS (light teal) and constant Mach (light orange) under the ISA standard atmosphere, the same conversion used for V^* . Dotted lines indicate FL100 and the crossover altitude.

The resulting TAS target $V^*(t)$ is smoothed by a moving average over 150 s centred on each time, computed so that it does not depend on the sampling interval. This removes the steps at FL100 and at the start of the cruise plateau. The crossover needs no smoothing, since both speeds are equal there.

2.2.3 Total Specific Energy Plan

The total specific energy plan describes how the aircraft's energy changes over time and how these changes are distributed between altitude and speed [22]. To infer this plan, the observed altitude $h_K(t)$, obtained using the Kalman filter of Olive et al. [24], is combined with the inferred speed target $V^*(t)$. The total specific mechanical energy commanded along the flight, divided by the gravitational acceleration g , is expressed as an equivalent altitude:

$$H_E(t) = h_K(t) + \frac{V^*(t)^2}{2g}. \quad (2)$$

The observed TAS $V(t)$ is not part of the commands. It is derived from the ADS-B ground speed and ERA5 wind, replaced by the Mode S TAS where the two differ by more than 20 kt, and is used only to initialise the replay and in the operational pool.

To support subsequent recombination, the inferred specific energy plan is represented in a compact form using the Ramer–Douglas–Peucker (RDP) algorithm [25, 26]. The algorithm approximates $H_E(t)$ by a piecewise-linear function $H_E^*(t)$, whose breakpoints represent changes in the inferred energy rate. The approximation is performed separately for climb, level flight, and descent to prevent individual linear segments from spanning different operational phases. These phases are identified from a running median of the vertical rate over 15 s (15 samples for observed trajectories, four for trajectories propagated at four-second intervals), with thresholds of ± 200 ft/min. The same classification procedure is used throughout the evaluation.

The RDP tolerance ϵ_E is the largest vertical distance, in feet of H_E , allowed between a sample and its linear segment, so the scale of the time axis does not matter. It controls the trade-off between the compactness of the energy representation and its fidelity to the observed profile. We selected ϵ_E from a sweep of five values between 30 ft and 500 ft across the 4237 flights, evaluating the approximation error of H_E independently of the replay results. This trade-off is shown in the supplementary material, from which a tolerance of 125 ft was selected.

The slope of each segment j of the piecewise-linear approximation defines a constant total specific energy rate,

$$p_j^* = g \frac{H_E^*(t_{j+1}) - H_E^*(t_j)}{t_{j+1} - t_j}, \quad (3)$$

expressed in W kg^{-1} . Given the inferred speed target $V^*(t)$ and energy plan $p^*(t)$, the corresponding vertical rate $\dot{h}^*(t)$ and flight path angle $\gamma^*(t)$ follow from the energy balance,

$$\dot{h}^*(t) = \frac{p^*(t) - V^*(t) \dot{V}^*(t)}{g}, \quad \gamma^*(t) = \arcsin\left(\frac{\dot{h}^*(t)}{V^*(t)}\right), \quad (4)$$

neglecting vertical atmospheric motion. The derivative $\dot{V}^*$ is obtained by central finite differences of the smoothed V^* . At a constant energy rate, acceleration reduces the implied climb rate, whereas deceleration increases it. The flight path angle command $\gamma^*(t)$ is therefore not inferred independently but is derived from the energy plan and speed schedule. Since H_E uses the inferred speed target, the vertical motion derived from the energy plan and the speed schedule reproduces the observed altitude to within the RDP tolerance, while the speed follows the inferred schedule rather than the observed speed.

2.3 Replay Evaluation

The replay evaluation assesses whether the inferred command sequences can reproduce the observed vertical flight profiles when applied to the physics-guided Neural ODE flight dynamics model (NODE-FDM) of Jarry et al. [15]. The model propagates the aircraft's altitude, TAS, flight path angle, and heading at four-second intervals. At each step, a neural network receives the differences between the current state and the commanded targets, together with ERA5 temperature and longitudinal and lateral wind, and returns the state derivatives, which are integrated through embedded kinematic relations. We use the NODE-FDM v2 checkpoint of Jarry and Olive [16], without retraining it.

For each flight, the inferred command sequence is supplied to NODE-FDM. Over each four-second step, the speed and flight path angle commands are averaged, while the altitude target is held at its value at the start of the step. ERA5 meteorological fields are interpolated in time. The replay is initialised using the observed altitude, heading, TAS, and vertical rate at its first step. After initialisation, no observed altitude, speed, or vertical rate is supplied to the model, so that the replay evolves solely from the inferred commands and meteorological conditions. A successful replay demonstrates that the inferred command representation can reproduce the observed flight profile, but does not establish that the inferred quantities correspond to the commands actually selected in the cockpit.

The primary measure of replay fidelity is the altitude error between the replayed and observed trajectories. For each flight i , the replayed and observed altitudes, denoted by $\hat{h}_i(t_{i,m})$ and $h_i(t_{i,m})$, respectively, are compared at the M_i airborne time steps $t_{i,m}$ ($m = 1, \dots, M_i$), four seconds apart. The state predicted over each step is compared with the observation at the end of that step, without temporal alignment. The observed altitude is h_K , and ground samples are excluded. The mean absolute error (MAE) and root mean square error (RMSE) are defined as

$$\text{MAE}_{h,i} = \frac{1}{M_i} \sum_{m=1}^{M_i} \left| \hat{h}_i(t_{i,m}) - h_i(t_{i,m}) \right|, \quad \text{RMSE}_{h,i} = \left(\frac{1}{M_i} \sum_{m=1}^{M_i} (\hat{h}_i(t_{i,m}) - h_i(t_{i,m}))^2 \right)^{1/2}. \quad (5)$$

RMSE is more sensitive to large deviations than MAE, while the signed mean error, calculated as replayed minus observed altitude, quantifies systematic altitude bias. To assess whether replay accuracy varies across flight phases, the MAE is also calculated separately for climb, level, and descent.

Altitude-error metrics alone do not show whether the replay reproduces individual level-offs. We therefore evaluate plateau closure, which determines whether the replayed altitude lies within ± 250 ft of the inferred target at a representative instant during each plateau. The evaluation includes all targets held for at least 32 s for which the observed aircraft entered the target band. The final target of each flight is excluded. For each eligible plateau k of flight i , the evaluation instant t_{ik}^{mid} is the midpoint between the first entry of the observed altitude into the target band and the end of the plateau. A plateau is closed when its closure discrepancy c_{ik}^α does not exceed 250 ft. For the replayed and observed trajectories, denoted by $\alpha \in \{\text{replay}, \text{obs}\}$, this discrepancy is defined as

$$c_{ik}^\alpha = \left| h_i^\alpha(t_{ik}^{\text{mid}}) - h_{\text{sel},i}^*(t_{ik}^{\text{mid}}) \right|. \quad (6)$$

Since both the inferred altitude target h_{sel}^* and the evaluation instant t_{ik}^{mid} are derived from the observed trajectory, observed plateau closure serves as a consistency check of the inference rather than an independent reference. Comparing observed and replayed closure nevertheless distinguishes plateaus where the observed aircraft itself deviated from the inferred target from those where only the replay did. In addition, the capture timing error $\Delta t_{ik}^{\text{capture}}$ is defined as the difference between the first entries of the replayed and observed trajectories into the target band. Positive values indicate that the replay reaches the target later than the observed aircraft. The metric is calculated only when both trajectories enter the band.

2.4 Conditional Command Sampling

Conditional sampling generates new command sequences that reproduce the operational characteristics of a given route without replicating individual observed flights. To this end, we build a route-specific empirical library from the inferred command representations of the flights in the *library set*. Rather than fitting parametric distributions to h_{sel}^* , V^* , and p^* , the generator samples command events, speed schedules, and plateau durations from this library and recombines them into new sequences. Every event, speed schedule, and level-time allocation thus comes from an observed flight, while the elements of a generated sequence generally come from different flights.

The basic element of the empirical library is the command event, representing a transition from an initial altitude target h_{from}^* to a new target h_{to}^* . Each event stores the target-reaching duration τ_{target} , the subsequent plateau duration $\tau_{plateau}$, and an energy profile consisting of RDP segments with durations τ_j and constant energy rates p_j^* over τ_{target} . Only completed events, for which the observed altitude entered the tolerance band around the new target, are retained. The net energy change must also be non-zero and have the sign of the altitude change, so that the profile can be rescaled. Figure 3 illustrates an example of a command event from the empirical library.

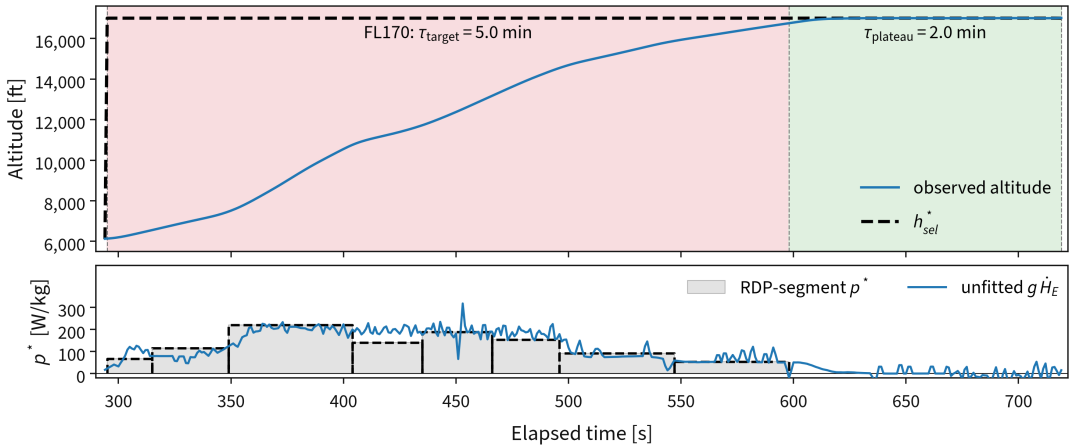


Figure 3. Command event inferred from *TAP Air Portugal* flight TAP51PM from London Heathrow to Lisbon on 1 May 2024. Upper panel: observed altitude (solid) and inferred h_{sel}^* (dashed), which steps to FL170 at the start of τ_{target} (red region) and is held through $\tau_{plateau}$ (green region). Lower panel: energy profile of the event, with one constant rate p^* per RDP segment (dashed steps), over the unfitted total specific energy rate $g \dot{H}_E$ computed from h_K and the observed TAS V .

Within each observed flight, command events form a chronological sequence, typically consisting of climb events from departure to cruise followed by descent events towards arrival. Less common sequences may include intermediate descents during cruise or go-arounds. In addition to these command events, the empirical library stores two flight-level components. The first is the inferred speed schedule. The second is the level-time allocation, which divides the total time spent at altitude targets into three components: the cumulative plateau duration before cruise, the duration at cruise altitude (defined by the highest altitude target of the flight), and the cumulative plateau duration after cruise. The libraries of the five routes are summarised in the supplementary material.

The generation of a new command sequence proceeds in five stages. The first stage constructs the target path, a sequence of selected altitude targets determined independently of timing, speed, and energy. The path begins with the first command event of a randomly selected library flight, with all flights being equally likely. Each subsequent event is drawn uniformly from the route-specific library events whose starting altitude target matches the current target. Two constraints preserve the temporal structure of observed flights. First, the relative progress of an event, defined as the

fraction of its source flight elapsed when the event begins and discretised into twenty equal bins, cannot be smaller than that of the preceding event. Second, the path cannot enter descent before reaching its cruise level. Once an event originating after the cruise level of its source flight is selected, all subsequent events must also originate after the cruise level of their respective source flights. At each step, the path may terminate with a probability determined by the frequency with which library flights ended at the current target at the same or a later relative progress. This procedure allows intermediate descents and go-arounds to occur in generated target paths. A path is discarded if no admissible continuation exists or if its number of events exceeds that of the longest library flight.

Figure 4 illustrates the construction of a target path for the London Heathrow to Lisbon route. The generated path begins with FL330, which is the first altitude target of 8 % of the library flights. It subsequently climbs to FL370, a transition represented by 17 % of the admissible events starting at FL330, before descending to FL110, a transition represented by only 5 % of the admissible events starting at FL370. The path then continues to FL80 before landing. Since successive events may originate from different library flights, the sampling procedure can produce target sequences that are not present in any individual observed flight. To quantify this, we generated 1,000 target paths per route and compared their altitude-target sequences with those in the library. Across the five routes, 51 % to 71 % of them contained sequences that did not occur in any library flight.

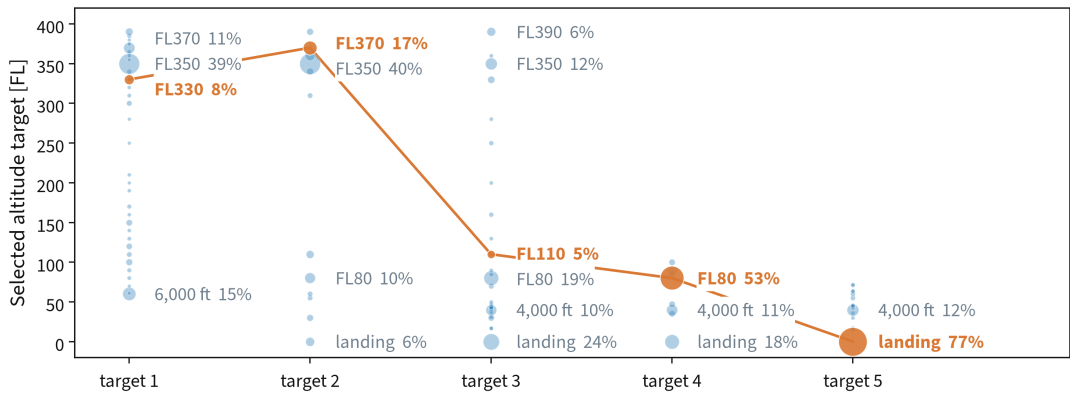


Figure 4. Target path of one draw on the London Heathrow to Lisbon route. At each step, the circles show the targets that can follow the current one, with area proportional to their probability. Targets below 1500 ft are grouped as landing.

The second stage assigns a command event to each altitude transition in the generated target path. For each transition, an event is drawn uniformly from all events in the route-specific library with matching start and end altitude targets. The selected event provides the target-reaching duration τ_{target} , the plateau duration τ_{plateau} , and the associated energy profile. These components are sampled together, preserving the relationship between the transition duration and the temporal structure of the energy-rate profile observed in the source flight.

The third stage adjusts the plateau durations across the generated sequence. Sampling command events independently may produce unrealistic combinations of plateau durations, particularly at cruise, where the available level-flight time is constrained by the remaining flight distance. To account for this dependency, a level-time allocation is drawn from a library flight with a target path similar to the generated one. Similarity requires the same number of altitude transitions before and after cruise, a cruise altitude matching to the nearest 5000 ft, and a total altitude change matching to the nearest 10 000 ft. If fewer than five library flights meet all three, the criteria are relaxed in turn, first to the transition counts and cruise altitude, then to the transition counts only, then to the climb transitions and cruise altitude, then to the climb transitions only, and finally to any flight

of the route. Plateaux before and after cruise keep the durations of their command events and are shortened proportionally only when their total exceeds the allocation for that phase. The cruise plateau takes the rest of the allocation's total level time.

The fourth stage assigns a speed schedule from the library. The schedule is applied as described in Section 2.2.2, with the cruise plateau of the generated path flown at the cruise Mach. To check the compatibility of a schedule, the altitude profile is approximated by linear interpolation between successive altitude targets over each τ_{target} and held constant during plateaux. Not every speed schedule is compatible with the generated target path. For each altitude transition, the required change in total specific energy depends on both the altitude difference and the commanded speeds at the beginning and end of the transition. Since the energy profile assigned in the second stage can only be rescaled without reversing its direction, the required energy change must have the same sign as that of the original event. The schedule is therefore drawn at random from the library schedules compatible with every transition of the path, which is equivalent to drawing from all schedules and rejecting incompatible ones. Once the schedule is drawn, the FL100 speed transitions, the crossover, and the conversion of the commanded speeds to TAS are evaluated at the altitude given by the assigned energy profile, $h = H_E - V^{*2}/2g$, and the energy profile of each transition is rescaled to the resulting speeds. Sequences with no compatible schedule are discarded, as are paths without an admissible continuation and sequences without an eligible context flight (Section 2.5).

The fifth stage rescales the energy profile of each command event to ensure consistency with the selected speed schedule. The required change in total specific energy depends on the altitude difference and the commanded speeds at the beginning and end of the transition, which may differ from those of the source flight. To match this energy change while preserving the temporal structure of the original profile, all segment rates p_j^* are multiplied by a common scaling factor κ , defined as

$$\kappa = \frac{g(h_{to}^* - h_{from}^*) + \frac{1}{2}(V_{end}^{*2} - V_{start}^{*2})}{\sum_j \tau_j p_j^*}, \quad (7)$$

where V_{start}^* and V_{end}^* denote the commanded TAS at the beginning and end of the transition, respectively. The rescaled profile delivers the required energy change over the unchanged target-reaching duration τ_{target} while preserving the relative magnitudes of the original segment rates. The commanded flight path angle γ^* is then obtained from Equation (4). During plateaux, both p^* and γ^* are set to zero. The resulting command sequence $u(t)$ has the same structure as in the replay evaluation.

2.5 Synthetic Trajectory Generation and Evaluation

The generated command sequences are propagated through NODE-FDM to obtain synthetic aircraft trajectories for subsequent evaluation against observed operations. Each sequence is propagated at four-second intervals, as in the replay evaluation, initialised with the starting altitude h_{from}^* of its first command event and the initial values of the generated V^* and γ^* . Since the generated command sequences contain neither meteorological information nor heading information, these inputs are obtained from a context flight operating on the same route. The context flight is selected only after the command sequence has been generated, ensuring that meteorological conditions do not influence command sampling. Eligible context flights are flights of the final dataset on the same route, from either the library or the held-out set. They must last at least as long as the generated sequence, have ERA5 temperature and wind throughout, and report a pressure setting wherever the sequence holds a QNH target. The context flight supplies the initial heading and the meteorological conditions required at each propagation step. Temperature and wind are taken from the context flight at the altitude of the generated target path, from its climb before the generated cruise plateau

and from its descent after it. On the cruise plateau, the context flight's own cruise conditions are used in time order, stretched to the generated cruise duration. Altitude targets referenced to a local pressure setting are converted to pressure altitude using the setting reported by the context flight during its climb or its descent, respectively. The observed trajectory of the context flight is not used as a reference for evaluating the generated trajectory.

For each route, the propagated trajectories form a synthetic pool $\mathcal{S}$, which is compared with an operational pool $\mathcal{O}$ consisting of the observed trajectories from the *held-out set*. The operational pool is constructed directly from the surveillance data rather than from replayed trajectories, ensuring that the synthetic trajectories are evaluated against observed operations rather than against the NODE-FDM itself. Each observed trajectory is represented by its smoothed altitude h_K , ADS-B vertical rate, TAS V , and flight path angle derived from the vertical rate and TAS. The two pools are compared using the distributions of altitude, TAS, vertical rate, and flight path angle, both over entire flights and separately for climb, level flight, and descent. For synthetic trajectories, these phases are identified from the propagated trajectories rather than from the underlying command sequences. Samples classified as ground are excluded from both pools. Distributional differences are quantified using a quantile-based approximation of the Wasserstein-1 distance,

$$W_1(\mathcal{O}, \mathcal{S}) = \frac{1}{|Q|} \sum_{q \in Q} |F_{\mathcal{O}}^{-1}(q) - F_{\mathcal{S}}^{-1}(q)|, \quad (8)$$

where $F_{\mathcal{O}}^{-1}$ and $F_{\mathcal{S}}^{-1}$ denote the empirical quantile functions of the operational and synthetic distributions, respectively, and Q contains nineteen equally spaced percentiles from the 5th to the 95th. The resulting distance is expressed in the units of the corresponding variable. In addition to these distributional comparisons, the flights are evaluated using structural characteristics: total flight duration, number of level-flight segments, proportion of time spent in level flight, and cruise altitude, defined as the highest median altitude among the level-flight segments lasting at least 60 s. Level-flight segments more than 1000 ft below cruise altitude are classified as intermediate levels.

To interpret the differences between the synthetic and operational pools, we establish a reference distance representing the variability expected between two samples of observed operations. For each route, the flights in the *held-out set* are randomly divided into two equally sized subsets twenty times. For each split, W_1 is calculated between the two subsets, and the median of the twenty distances is retained. This median is then multiplied by $\sqrt{n_{\mathcal{O}}(1/n_{\mathcal{O}} + 1/n_{\mathcal{S}})/4}$, where $n_{\mathcal{O}}$ and $n_{\mathcal{S}}$ denote the numbers of flights in the operational and synthetic pools, respectively, to account for their potentially different sample sizes. The resulting value serves as the reference for $W_1(\mathcal{O}, \mathcal{S})$.

While the preceding metrics compare aggregate distributions, they do not establish whether individual synthetic trajectories resemble observed flights. To assess flight-level similarity, each observed and synthetic trajectory is represented by a profile of altitude, vertical rate, and TAS, with each variable averaged over 200 equally spaced bins of the flight's duration. For each variable and bin, the profiles are standardised using the mean and standard deviation of the corresponding observed profiles. To avoid excessive weighting of bins with little observed variability, the standard deviations are bounded below by 250 ft for altitude, 250 ft min⁻¹ for vertical rate, and 5 kt for TAS. The distance between two profiles is then calculated as the root mean square difference across their standardised values. For each observed flight, the distance to its nearest other observed flight is determined, and the 95th percentile of these distances defines a reference threshold for flight-level similarity. Operational inclusion is defined as the proportion of synthetic trajectories whose nearest observed flight lies within this threshold, while coverage is the proportion of observed flights whose nearest synthetic trajectory lies within it. To ensure comparable nearest-neighbour searches, each search considers $n - 1$ randomly selected candidates, where n is the number of observed flights. The candidate selection is repeated twenty times, and inclusion and coverage are averaged.

3. Results

The results address the two research questions. Section 3.1 evaluates how accurately the inferred command sequences reproduce the observed trajectories, while Section 3.2 compares synthetic trajectory pools generated from the *library set* with operational trajectories from the *held-out set*.

3.1 Same-Flight Reconstruction Fidelity

Same-flight reconstruction was evaluated on all 4237 flights in the final dataset. Table 1 summarises altitude reconstruction errors and altitude-plateau capture performance by route and overall.

Table 1. Same-flight reconstruction fidelity by route and overall. Altitude errors are reported as median MAE [95th percentile], median RMSE, and phase-specific median MAE, all in feet. Plateau capture is the proportion of evaluated plateaux for which the replay lies within ± 250 ft of h_{sel}^* at t_{ik}^{mid} . Absolute capture timing errors are reported as median [95th percentile].

Route	Flights	MAE median [p95]	RMSE	Climb MAE	Level MAE	Descent MAE	Plateaux	Closed (%)	Timing (s) median [p95]
EGLL-LEBL	949	143.7 [270.3]	187.1	151.0	89.0	206.8	3,175	73.7	12 [64]
EGLL-LPPT	908	144.0 [281.6]	187.3	182.9	96.9	207.8	2,685	70.7	16 [183]
LEBL-LSZH	729	132.3 [307.0]	164.9	136.8	119.6	120.3	3,139	79.3	8 [36]
LSZH-LEBL	789	147.3 [326.5]	184.5	147.0	116.4	156.6	2,706	75.8	12 [64]
LSZH-LPPT	862	145.8 [285.0]	191.8	209.9	105.6	204.9	2,688	73.3	16 [212]
Overall	4237	143.4 [290.5]	183.7	163.7	103.2	177.1	14 393	74.7	12 [80]

Altitude errors are consistent across routes. The largest median phase MAE occurs during descent on three routes and during climb on the other two. The maximum per-flight MAE reaches 1882 ft. The replay shows a positive altitude bias on all five routes, with route medians of the signed mean error ranging from 4.1 ft on LSZH-LPPT to 71.5 ft on EGLL-LEBL (median 35.1 ft).

The median mid-plateau discrepancy is 127.6 ft for the replay and 5.7 ft for the observed aircraft. The replay lies within the ± 250 ft band at the midpoint of 74.7 % of evaluated plateaux, compared with 99.97 % for the observed aircraft. This share falls from 80.9 % at or above FL100 to 60.2 % below FL100. On 93.8 % of missed plateaux below FL100, the replay remains above the target. The replay never enters the band on 1249 plateaux (8.7 %). Where both trajectories enter the band, the replay arrives earlier in 36 % of cases and later in 41 %, with the remainder within the same 4 s step.

Below FL100, the commanded vertical rates of Equation (4) match the observed rates within a median of 5 ft min⁻¹ in descent and 3 ft min⁻¹ in climb. In commanded descents, however, the replay descends more slowly than prescribed, by a median of 12 ft min⁻¹ above FL100 and 9 ft min⁻¹ below, on 82 % and 81 % of flights. This leaves the replay a median of 55 ft and 220 ft above the commanded profile. Under level-flight commands, the replay sinks at a median rate of 3 ft min⁻¹.

Figure 5 shows a flight with an altitude MAE below 250 ft but a replay discrepancy of 274 ft from the inferred target at the FL100 plateau midpoint, compared with 20 ft for the observed aircraft.

3.2 Generated Trajectory-Pool Fidelity

For each route, 500 command sequences were sampled from the library, propagated through the NODE-FDM, and compared with the operational pool of the *held-out set* (364 to 474 flights per route). Eleven of the 2,500 draws were discarded, ten because their target path had no admissible continuation or more steps than the longest library flight (seven on LSZH-LEBL, three on EGLL-LEBL) and one on LEBL-LSZH because no context flight was long enough. Each sequence combined events from a median of four library flights. Although 39 % of the sampled target paths matched that of a library flight, all but two of these sequences combined events from several flights.

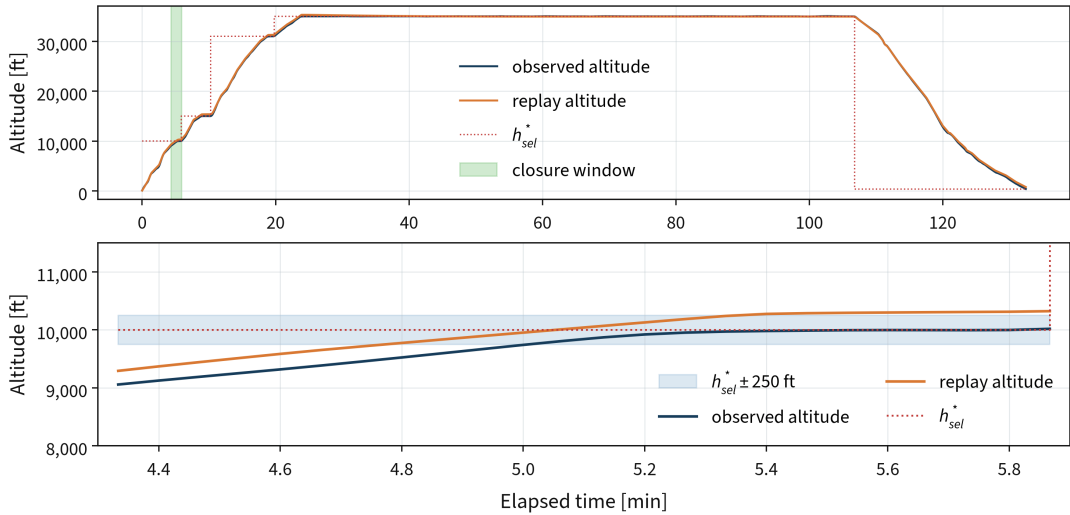


Figure 5. Replay of TAP Air Portugal flight TAP51PM from London Heathrow to Lisbon on 12 April 2024, which has a low altitude MAE but a large closure discrepancy. The upper panel shows the whole flight, and the lower panel enlarges the plateau.

Figure 6 compares the altitude profiles and timing of the synthetic and operational trajectory pools across all five routes. Aligned on departure, the median climb profiles closely overlap; the same holds for the median descent profiles when aligned by arrival. The 5–95th percentile bands are also comparable in width. Median flight durations differ by about 2 min or less, from 83.6 min for synthetic trajectories versus 85.2 min for operational flights on LEBL–LSZH to 151.4 min versus 153.5 min on LSZH–LPPT. Individual trajectories of both pools are shown in the supplementary material.

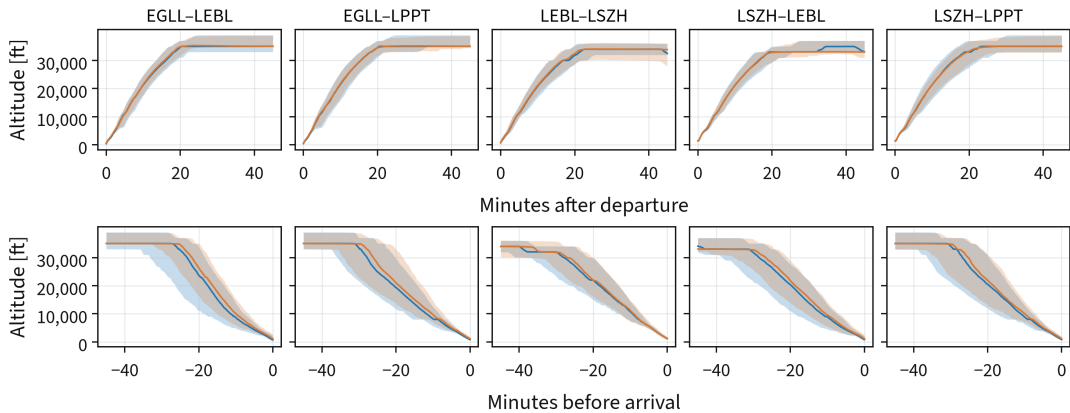


Figure 6. Altitude profiles of the operational (blue) and synthetic (orange) trajectory pools for each route, aligned by departure (top) and arrival (bottom). Solid lines indicate the median, and shaded bands represent the 5th–95th percentiles.

Table 2 compares synthetic and operational trajectory pools by variable and flight phase using a quantile-based approximation of the Wasserstein-1 distance (W_1). Across all variables, the synthetic–operational distances exceed the operational reference distances. For altitude, the largest relative discrepancy occurs during climb. On EGLL–LPPT and LSZH–LEBL, the whole-flight altitude distance exceeds that of every individual phase. Reweighting synthetic samples to match the operational phase proportions reduces the whole-flight altitude W_1 by 33 % to 53 %, to a median of 229 ft.

Table 2. Quantile-based approximation of the Wasserstein-1 distance (W_1) between operational and synthetic trajectory pools, reported as the median across five routes. Values in parentheses give the corresponding reference distances between two random subsets of the operational pool, adjusted for pool size.

Variable	Whole flight	Climb	Level	Descent
Altitude (ft)	376 (120)	348 (85)	370 (142)	244 (90)
True airspeed (kt)	4.9 (0.8)	10.2 (0.8)	2.8 (1.0)	10.6 (1.4)
Vertical rate (ft min ⁻¹)	48 (6)	76 (10)	19 (0.6)	63 (11)
Flight path angle (°)	0.06 (0.01)	0.16 (0.02)	0.02 (0.00)	0.11 (0.01)

In climb, synthetic TAS exceeds observed TAS at the 5th percentile (207 kt to 260 kt versus 170 kt to 212 kt) and is slightly lower at the median (398 kt to 409 kt versus 404 kt to 413 kt). The NODE-FDM reproduces the commanded TAS within a median of 0.9 kt to 1.3 kt. On EGLL-LPPT, LSZH-LEBL, and LSZH-LPPT, the synthetic climb above FL300 is faster than observed (median 960 ft min⁻¹ to 1100 ft min⁻¹ versus 900 ft min⁻¹ to 990 ft min⁻¹), so that it accounts for 18 % to 27 % of the climb time, compared with 21 % to 29 % for observed flights. During level flight, ADS-B vertical rates are exactly zero in 61 % to 70 % of observed samples.

Synthetic trajectories contain 1.94–3.23 level segments per flight, compared with 2.17–3.22 for observed flights, including 0.92–2.21 versus 1.07–2.16 intermediate levels. The median proportion of time spent in level flight differs by at most 3 percentage points across routes. On EGLL-LPPT, LSZH-LEBL, and LSZH-LPPT, cruise levels split almost evenly between two levels 2000 ft apart, so that the median synthetic cruise altitude lies 2000 ft below the observed median, although the shares of the cruise levels differ by at most 8 percentage points. On EGLL-LPPT, the corresponding interquartile ranges are FL349 to FL369 and FL350 to FL370. Of the 2489 synthetic trajectories, 9 terminate below 0 ft pressure altitude and none below -500 ft. Across routes, 0 % to 2.4 % terminate below the first percentile of observed final altitudes.

Table 3 reports the nearest-neighbour comparison of individual trajectories using standardised altitude, vertical-rate, and TAS profiles. Observed-to-observed distances provide the reference, while the remaining distances compare synthetic and observed trajectories in both directions. Operational inclusion denotes the proportion of synthetic trajectories within the observed operational spread, defined by the 95th percentile of observed-to-observed nearest-neighbour distances. Coverage denotes the proportion of observed flights with a synthetic neighbour within the same threshold. Across all routes, median draw-to-observed distances exceed the observed-to-observed reference. LSZH-LPPT achieves the highest inclusion and coverage, while EGLL-LEBL has the lowest coverage. The corresponding distance distributions are shown in the supplementary material.

Table 3. Nearest-neighbour comparison of individual profiles per route. Distances are root mean square differences of standardised altitude, vertical rate, and TAS profiles over normalised time, given as medians. The spread of observed operations is the 95th percentile of the distance from an observed flight to its nearest other observed flight.

Route	Observed	Draws	Observed to nearest observed	Draw to nearest observed	Observed to nearest draw	Operational inclusion (%)	Coverage (%)
EGLL-LEBL	474	497	0.68	0.92	0.96	65	47
EGLL-LPPT	454	500	0.62	0.86	0.86	72	66
LEBL-LSZH	364	499	0.77	1.02	0.99	48	54
LSZH-LEBL	394	493	0.77	0.93	0.95	76	74
LSZH-LPPT	431	500	0.66	0.79	0.78	82	79

4. Discussion

The results demonstrate that inferred altitude targets, speed schedules, and total specific energy plans provide a representation from which individual vertical flight profiles can be reconstructed and new profiles generated. Recombining these elements produces synthetic trajectories that reproduce the broad vertical structure of the observed routes, although discrepancies remain at the level of individual trajectories.

The inferred representation reconstructs its source flight to a median altitude error of 143 ft over 4237 flights, with no observed state supplied after the first sample. This replay follows the physical evaluation of Olive et al. [4], who replay trajectories in a simulator from decoded selected altitudes and speeds. Here, the simulator is the NODE-FDM and the targets are inferred rather than decoded. As they note, replay evaluates the targets and simulator together. A comparison with the selected altitude decoded from Mode S BDS 4.0, which is not used in the inference, supports the inferred targets: 90.8 % agree within 100 ft. Jarry and Olive [16] reported gaps and spurious steps in decoded selected altitude and suggested re-estimating it from the trajectory. The inferred targets provide such an estimate, and their agreement with independently decoded selector values supports their interpretation as operationally meaningful levels rather than artefacts of the fit. Agreement falls to 74.0 % below FL100, where most disagreements are levels held short of a selector set further on.

Plateau closure is weaker, despite the overall altitude error of the replay being small. During commanded descents, the NODE-FDM descends more slowly than prescribed on the vast majority of flights, leaving the replay above the commanded profile when it reaches lower altitude plateaux. Since the derived vertical rate below FL100 matches the observed one at the median, the accumulated replay drift plausibly contributes substantially to missed low plateaux, although time-varying errors in the inferred commands may also contribute. This descent drift is consistent with the positive altitude bias of the replay. Under a level command, the replay instead sinks slightly, as Jarry and Olive [16] observed for flat altitude segments.

The synthetic trajectory pools agree most closely with the operational altitude distributions during level flight and descent, while the largest altitude and TAS discrepancies relative to the reference occur during climb. TAS agreement is strongest in level flight, whereas the largest absolute vertical-rate discrepancy occurs during climb. The synthetic draws reproduce the approximate number of level segments, the proportion of time spent in level flight, and the range of cruise levels observed. The sampler therefore combines events from different flights while retaining the broad vertical structure characteristic of each route. In climb, the TAS discrepancy appears to arise primarily from the speed command rather than from the NODE-FDM response, which follows the commanded TAS closely. The single CAS value held below FL100 produces higher speeds than observed in the lower tail. Pepper and Thomas [13] suggested that their learned thrust compensated, around the CAS–Mach transition, for a nominal speed schedule or mass that did not match the routes. Here, the schedule is inferred per flight. The climb altitude distance on EGLL–LPPT, LSZH–LEBL, and LSZH–LPPT arises mainly near the top of climb, where the sampled commands already prescribe a faster climb than observed and the NODE-FDM climbs slightly faster still. On the other two routes, it is below 100 ft. The whole-flight altitude distance also reflects differences in the shares of time that the two pools spend in each phase, which change the weighting of samples across tens of thousands of feet. Matching the phase shares reduces this distance by a third to a half, bringing it below the climb distance on three routes.

The nearest-neighbour comparison reveals discrepancies that are less apparent from the pooled distributions. Following the similarity evaluation of Olive et al. [4], distances are also computed between observed flights to provide an operational reference. Synthetic inclusion and operational coverage range from 48 % to 82 % and 47 % to 79 %, respectively. Thus, a substantial proportion of

synthetic trajectories lie outside the observed operational spread, while many observed flights have no sufficiently close synthetic counterpart. Both measures reveal discrepancies in their respective directions of comparison. However, they cannot establish whether these arise from systematic shifts in trajectory characteristics or from insufficient coverage of particular operational patterns.

Several limitations apply to these results. First, command sampling is conditioned on the route and aircraft family only, without accounting for weather conditions. Although the propagation context is selected subsequently, operational choices such as cruise altitude and speed depend partly on wind. Second, the generated sequences place no constraint on the ground, so that nothing prevents a draw from descending past the arrival altitude, and up to 2.4 % of the draws per route end lower than the lowest 1 % of the observed flights. Third, progress is measured as a fraction of the flight duration. On a single route, the relation between time and distance depends mainly on wind and speed. Distance-based progress would be more faithful to the route. It cannot be used directly because the distance covered by a generated sequence is only known after propagation. Fourth, the evaluation has no competing generator, since the split-half reference measures the sampling noise of the operations, not the performance of an alternative method. Fifth, each flight holds a single speed schedule with one CAS value below FL100 in each phase, which limits how closely the climb TAS near the ground can be reproduced. Finally, a single NODE-FDM checkpoint trained on the A320 family is used throughout, so the results do not yet extend to other aircraft types.

5. Conclusion and Outlook

This study investigated whether variability in vertical flight profiles can be represented and generated through inferred operational commands rather than through trajectories directly. Altitude targets, speed schedules, and total specific energy plans were inferred from historical ADS-B and Mode S data. These commands were replayed through a physics-guided Neural ODE to reconstruct individual flights and subsequently recombined to generate synthetic trajectories. The resulting trajectory pools were evaluated against held-out operational flights.

The inferred representation reconstructs observed flights with a median altitude error of 143 ft, but the replay lies within ± 250 ft of the inferred target at the plateau midpoint on only 74.7 % of plateaux, compared with 99.97 % for the observed aircraft. The sampled pools retain the broad altitude structure and timing of the operations, with level-flight and descent altitude distances of 2.6–2.7 times the distance between two samples of the same operations. They differ most in climb, where the altitude and speed distributions remain several times further from the operations than the reference, and between 48 % and 82 % of the individual draws lie within the spread of real operations.

Beyond reproducing the broad characteristics of operational trajectories, the proposed approach represents variability through explicit altitude, speed, and energy targets. Unlike approaches that generate trajectories from latent representations [10], each generated profile can be traced to the explicit commands that produced it. This provides a basis for inspecting and modifying the inputs to synthetic trajectories, although the inferred commands do not reveal the operational circumstances that originally motivated them.

Future work should first separate errors in the inferred energy-rate profiles from those arising during NODE-FDM propagation to improve altitude-plateau closure. Generated-profile fidelity could be improved by refining the energy profiles near the top of climb and the low-altitude speed schedules, conditioning commands on weather, and constraining descent to the arrival altitude. Extending the command representation to lateral motion could also allow distance-based progress measures. Finally, comparisons with trajectory-level generators and evaluations across additional aircraft types would help establish the relative benefits and generalisability of command-based generation.

Acknowledgement

The authors acknowledge the contributions of three reviewers that greatly enhanced the value of this study. No potential conflict of interest was reported by the authors. No funding was received for this research.

Author contributions

Conceptualization (A.F, M.W, X.O), Methodology (A.F), Software (A.F, G.J), Validation (A.F), Formal analysis (A.F), Data Curation (A.F), Writing – Original Draft (A.F), Writing – Review & Editing (A.F, M.W, X.O), Visualization (A.F), Supervision (M.W, X.O, G.J), Project administration (M.W).

Open Data Statement

The software code used to download the OSN-data employed in this study is published on the following repository: https://github.com/alexfustagueras/aircraft_command_variability

Reproducibility Statement

The software code used to generate the results presented in this paper is published on the following repository: https://github.com/alexfustagueras/aircraft_command_variability. The NODE-FDM v2 code is available at <https://github.com/eurocontrol-asu/node-fdm/tree/node-fdm-v2>.

Supplementary material, including data cleaning and quality control, the empirical command libraries, the check of the inferred targets against the Mode S selected altitude, the selection of the RDP tolerance, the nearest-neighbour distance distributions and individual vertical profiles, is available at https://alexfustagueras.github.io/aircraft_command_variability/.

References

- [1] J. K. Kuchar and L. C. Yang. “A review of conflict detection and resolution modeling methods”. In: *IEEE Transactions on Intelligent Transportation Systems* 1.4 (2000), pp. 179–189. DOI: 10.1109/6979.898217.
- [2] Arnab Majumdar, Washington Y. Ochieng, James Benthams, and Martyn Richards. “En-route sector capacity estimation methodologies: An international survey”. In: *Journal of Air Transport Management* 11.6 (2005), pp. 375–387. DOI: 10.1016/j.jairtraman.2005.05.002.
- [3] Junzi Sun and Irene Dedoussi. “Evaluation of aviation emissions and environmental costs in Europe using OpenSky and OpenAP”. In: *Engineering Proceedings* 13.1 (2021), p. 5. DOI: 10.3390/engproc2021013005.
- [4] Xavier Olive, Junzi Sun, Mayara Condé Rocha Murça, and Timothé Krauth. “A framework to evaluate aircraft trajectory generation methods”. In: *14th USA/Europe Air Traffic Management Research and Development Seminar, ATM 2021*. Online, France, Sept. 2021. URL: <https://hal.science/hal-04808408>.
- [5] Angela Nuic, Damir Poles, and Vincent Mouillet. “BADA: An advanced aircraft performance model for present and future ATM systems”. In: *International Journal of Adaptive Control and Signal Processing* 24.10 (2010), pp. 850–866. DOI: 10.1002/acs.1176.
- [6] Junzi Sun, Jacco M. Hoekstra, and Joost Ellerbroek. “OpenAP: An open-source aircraft performance model for air transportation studies and simulations”. In: *Aerospace* 7.8, 104 (2020). ISSN: 2226-4310. DOI: 10.3390/aerospace7080104.

- [7] Matthias Schäfer, Martin Strohmeier, Vincent Lenders, Ivan Martinovic, and Matthias 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*. 2014, pp. 83–94. DOI: 10.1109/IPSN.2014.6846743.
- [8] Mayara Condé Rocha Murça and McWilliam de Oliveira. “A data-driven probabilistic trajectory model for predicting and simulating terminal airspace operations”. In: *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*. 2020, pp. 1–7. DOI: 10.1109/DASC50938.2020.9256644.
- [9] Timothé Krauth, Jérôme Morio, Xavier Olive, Benoit Figuet, and Raphael Monstein. “Synthetic aircraft trajectories generated with multivariate density models”. In: *Engineering Proceedings* 13.1, 7 (2021). ISSN: 2673-4591. DOI: 10.3390/engproc2021013007. URL: <https://www.mdpi.com/2673-4591/13/1/7>.
- [10] Timothé Krauth, Adrien Lafage, Jérôme Morio, Xavier Olive, and Manuel Waltert. “Deep generative modelling of aircraft trajectories in terminal maneuvering areas”. In: *Machine Learning with Applications* 11 (2024), p. 100446. ISSN: 2666-8270. DOI: 10.1016/j.mlwa.2022.100446.
- [11] Soyeon Jung, Amelia Hardy, and Mykel J. Kochenderfer. “Inferring traffic models in terminal airspace from flight tracks and procedures”. In: *Journal of Aerospace Information Systems* 22.8 (2025), pp. 730–740. DOI: 10.2514/1.I011503.
- [12] Matthias Poppe and Jörg Buxbaum. “Clustering climb profiles for vertical trajectory analysis”. In: *Proceedings of the 10th SESAR Innovation Days*. Dec. 2020. URL: https://www.sesarju.eu/sites/default/files/documents/sid/2020/papers/SIDs_2020_paper_66red.pdf.
- [13] Nick Pepper and Marc Thomas. “Learning generative models for climbing aircraft from radar data”. In: *Journal of Aerospace Information Systems* 21.6 (2024), pp. 474–481. DOI: 10.2514/1.I011359.
- [14] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. “Neural ordinary differential equations”. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. NIPS’18. Montréal, Canada: Curran Associates Inc., 2018, pp. 6572–6583.
- [15] Gabriel Jarry, Ramon Dalmau, Xavier Olive, and Philippe Very. “A neural ODE approach to aircraft flight dynamics modelling”. In: *SESAR Innovation Days 2025*. Dec. 2025. URL: <https://hal.science/hal-05640853>.
- [16] Gabriel Jarry and Xavier Olive. “Generation of vertical profiles with neural ordinary differential equations trained on open trajectory data”. In: *Journal of Open Aviation Science* 4.2 (Mar. 2026). DOI: 10.59490/joas.2026.8457.
- [17] Xavier Olive, Junzi Sun, Ryota Mori, and Junichi Naganawa. “Improving BDS type detection in passive Mode S surveillance: A multi-stage inference framework”. In: *Proceedings of the 2026 International Workshop on ATM/CNS*. Tokyo, Japan, Nov. 2026.
- [18] Hans Hersbach et al. “The ERA5 global reanalysis”. In: *Quarterly Journal of the Royal Meteorological Society* 146.730 (2020), pp. 1999–2049. DOI: 10.1002/qj.3803.
- [19] Junzi Sun and Esther Roosenbrand. “Fast contrail estimation with OpenSky data”. In: *Journal of Open Aviation Science* 1.2 (2023).
- [20] Junzi 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.
- [21] Carlo Tomasi and Roberto Manduchi. “Bilateral filtering for gray and color images”. In: *Proceedings of the Sixth International Conference on Computer Vision*. 1998, pp. 839–846. DOI: 10.1109/ICCV.1998.710815.
- [22] Airbus. *Getting to Grips with Aircraft Performance*. Airbus. Blagnac, France, Jan. 2002. URL: <https://ansperformance.eu/library/airbus-aircraft-performance.pdf>.
- [23] *Manual of the ICAO standard atmosphere: Extended to 80 kilometres (262,500 feet)*. 3rd ed. Doc 7488. International Civil Aviation Organization. 1993.

- [24] Xavier Olive, Jan Krummen, Benoit Figuet, and Richard Alligier. “Filtering techniques for ADS-B trajectory preprocessing”. In: *Journal of Open Aviation Science* 2.2 (2025). DOI: [10.59490/joas.2024.7882](https://doi.org/10.59490/joas.2024.7882). 653 654 655
- [25] Urs Ramer. “An iterative procedure for the polygonal approximation of plane curves”. In: *Computer Graphics and Image Processing* 1.3 (1972), pp. 244–256. DOI: [10.1016/S0146-664X\(72\)80017-0](https://doi.org/10.1016/S0146-664X(72)80017-0). 656 657 658
- [26] David H. Douglas and Thomas K. Peucker. “Algorithms for the reduction of the number of points required to represent a digitized line or its caricature”. In: *Cartographica: The International Journal for Geographic Information and Geovisualization* 10.2 (1973), pp. 112–122. DOI: [10.3138/FM57-6770-U75U-7727](https://doi.org/10.3138/FM57-6770-U75U-7727). 659 660 661 662