

# Normative Trajectory Modeling for Anomaly Detection at Airports: A Proof of Concept

Aaron Kushner, CEO, Quest Flight, Inc.  
[with GPT-4o/o3 Pro, OpenAI]

---

## 1. Introduction

Air traffic safety management spans a vast range of operational environments, from heavily monitored Class B airports like Los Angeles International (LAX) to completely untowered general aviation (GA) fields governed by informal community procedures. In this paper, we explore the feasibility of using a self-supervised machine learning approach to detect anomalous flight behavior by learning the "normative" patterns within these environments.

We found that the two domains present fundamentally different challenges. In the case of small untowered GA airports, where the expected flight envelope is compact and violations are statistically significant, a simple Masked Autoencoder (MAE) proved effective. In contrast, the complexity and phase-dependent structure of flight patterns around a major airport like LAX proved too demanding for an out-of-the-box MAE solution without further segmentation and contextual refinement.

This paper presents a focused proof-of-concept study of the GA airport case, demonstrating the viability of normative modeling using open-source trajectory data and discussing the limitations encountered in applying the same approach to large commercial airports.

---

## 2. Related Work

Anomaly detection in aviation has traditionally relied on rule-based systems or thresholding against known safe corridors. Recent efforts have explored supervised learning for conflict detection and trajectory forecasting, especially in high-density airspace. Meanwhile, the application of unsupervised or self-supervised learning to raw ADS-B or trajectory data remains nascent.

Masked autoencoders have emerged as a powerful tool in other domains for learning the underlying distribution of complex temporal sequences. Applications in cybersecurity, health monitoring, and finance have shown that such models can effectively flag novel behavior. In aviation, similar ideas have been proposed, but few studies have focused on small-airport general aviation traffic, and even fewer have compared performance across environments.

---

## 3. Methodology

### 3.1 Data Acquisition

We acquired and processed open-source trajectory data from two very different settings:

- **TrajAir Dataset:** High-quality general aviation trajectories from an untowered airport (Pittsburgh-Butler Regional Airport), featuring approach and pattern traffic within a 4 km bounding circle, already cleaned and divided into training and test scenes.
- **OpenSky Network:** Approximately 60 full days of complete data spanning June 2017 to June 2018 were downloaded from the OpenSky free scientific data set.

### 3.2 Preprocessing and Cleaning

- **TrajAir Dataset:** The training set consisting of X unique aircraft and Y distinct trajectories was used as downloaded. The test set required some additional cleaning (`clean_and_export_parquets.py`) to remove trajectories including “teleportations,” below-threshold frame counts, and altitudes above 1 km indicating non-interaction with the immediate airport airspace, retaining 116 aircraft complete trajectory “scenes” for scoring.
- **OpenSky Dataset:** The downloaded files were extracted and pruned of all trajectories with less than 10 points, maximum time gaps between time points of 120 seconds, duplicate entries, and > 5 km jumps between sequential points. A typical day of data retained between 3 and 15% of its data following this operation. A departure-heavy LAX bounding box was then isolated and split into training and test parquet files that were used without further processing, containing X and Y unique trajectories, respectively.

### 3.3 Normative Model Architecture

We used a Masked Autoencoder (MAE) with:

- **Input:** Time-series sequences of flight features.
- **Masking strategy:** Fixed-mask ratio across timesteps.
- **Model:** Lightweight Transformer encoder-decoder.
- **Loss:** Mean squared error across masked scenes (TrajAir) or points (OpenSky).

This model was trained to reconstruct missing trajectory segments based on learned patterns from the training set.

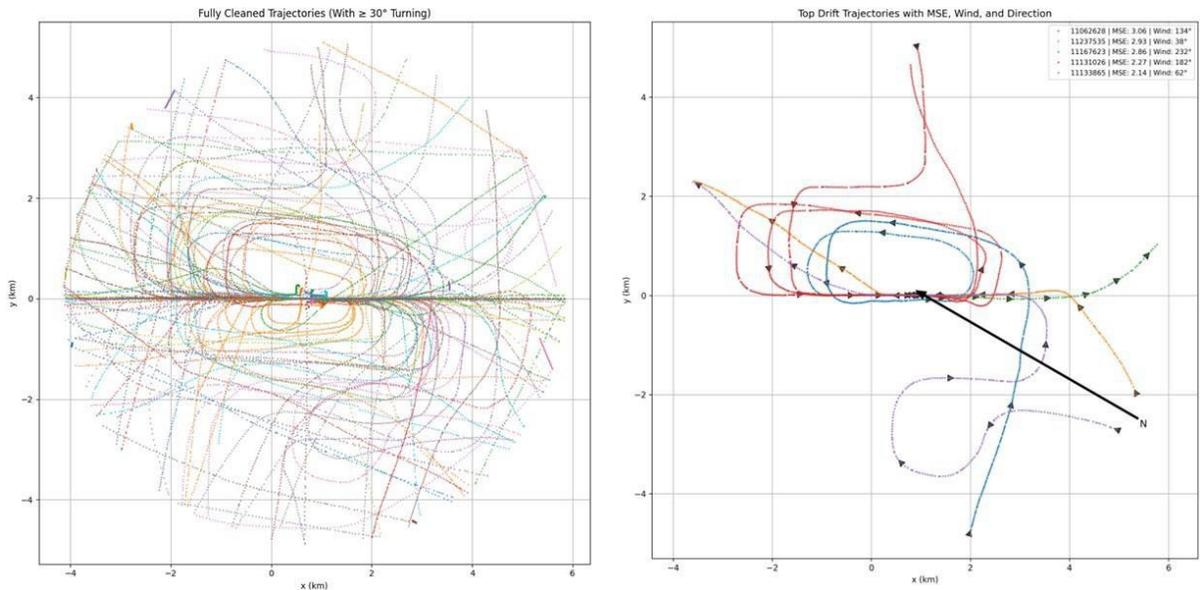
### 3.4 Drift Scoring and Anomaly Extraction

For each test trajectory, reconstruction error was computed point-wise, then aggregated to yield a **drift score**.

## 4. Results

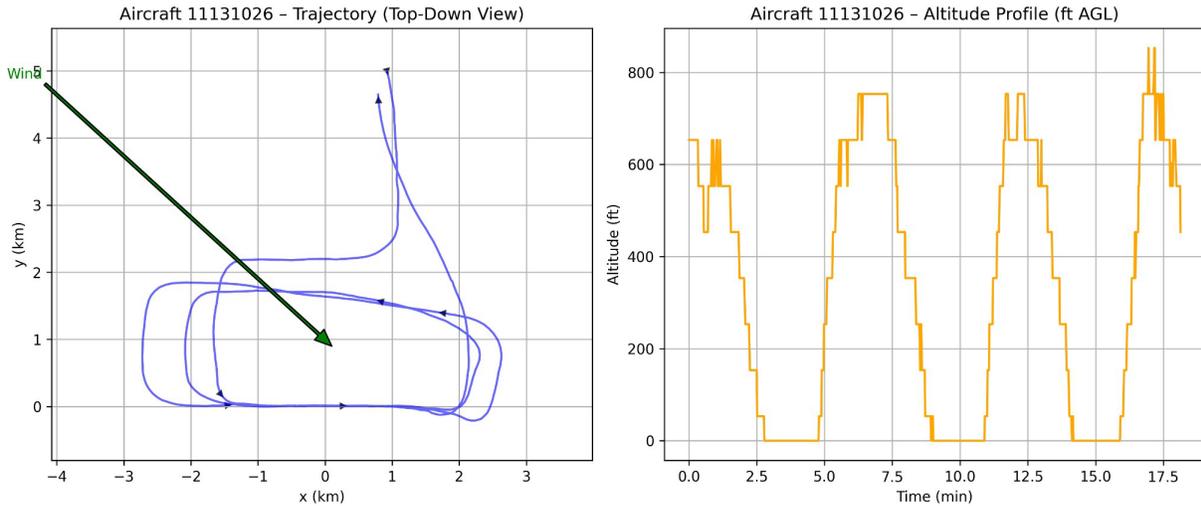
### 4.1 TrajAir

One can see from the top view plot of the entire test set in Figure 1A that while the number of trajectories are relatively few, many of them portray accurate and expected real world behavior of general aviation flight in the vicinity of an untowered airport. Five of the top scoring “abnormal” trajectories flagged by the scoring algorithm are plotted in Figure 1B, illustrating their flight direction along the trajectory, along with their MSE drift score, and recorded wind direction during the flight, with a magnetic north indicator for comparison. Three of these cases are further isolated for one by one “anomaly-type” analysis in the subsequent figures. For simplicity aircraft velocity is assumed to be consistent with the norm.



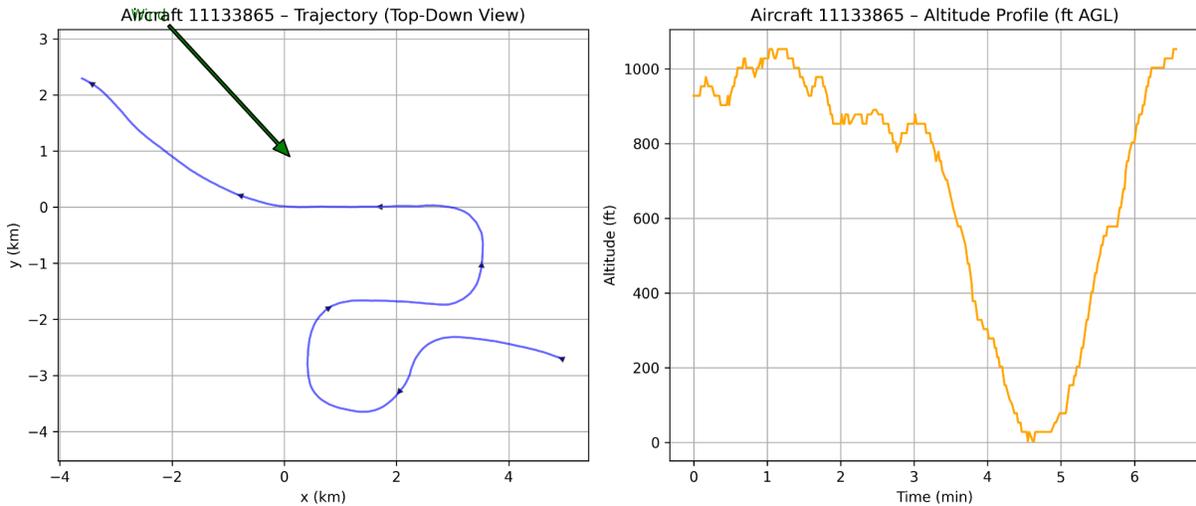
**Figure 1A:** The complete scene content of the TrajAir test dataset. **1B:** Top scoring anomalous scenes.

Aircraft 11131026 - 2D View, Altitude, and Wind

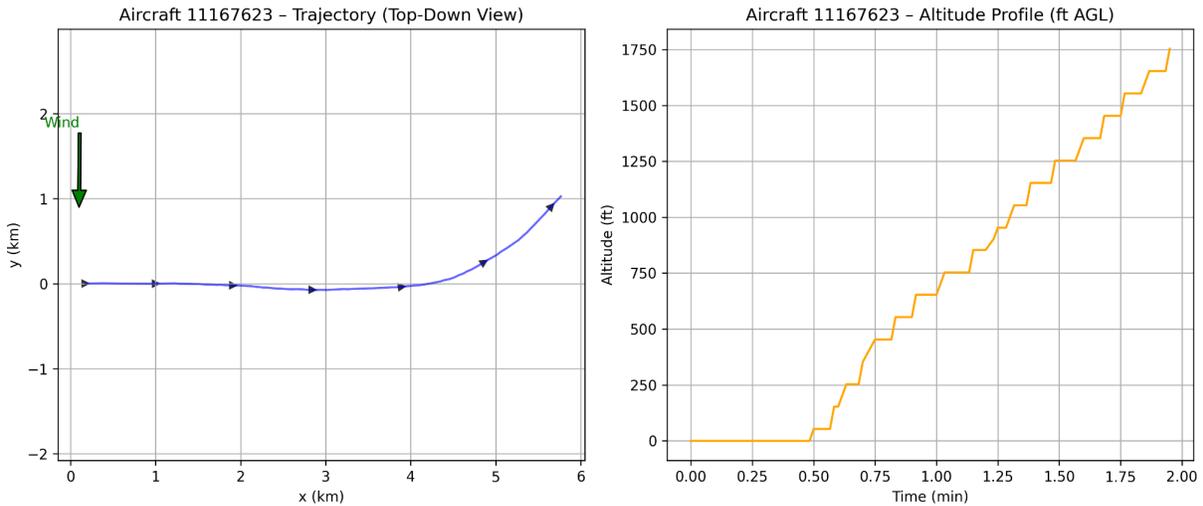


**Figure 2.** Aircraft 11131026 appears to enter the pattern on the upwind leg, although at a 90 degree angle rather than the preferred 45, then proceeds to enter a series of touch and goes. Both of these behaviors could be responsible for the anomaly flag for this scene.

Aircraft 11133865 - 2D View, Altitude, and Wind



**Figure 3.** Aircraft 11133865 follows a highly anomalous winding path directly into the downwind base leg of the standard pattern, before descending rapidly to briefly touch the runway surface, before climbing out again.

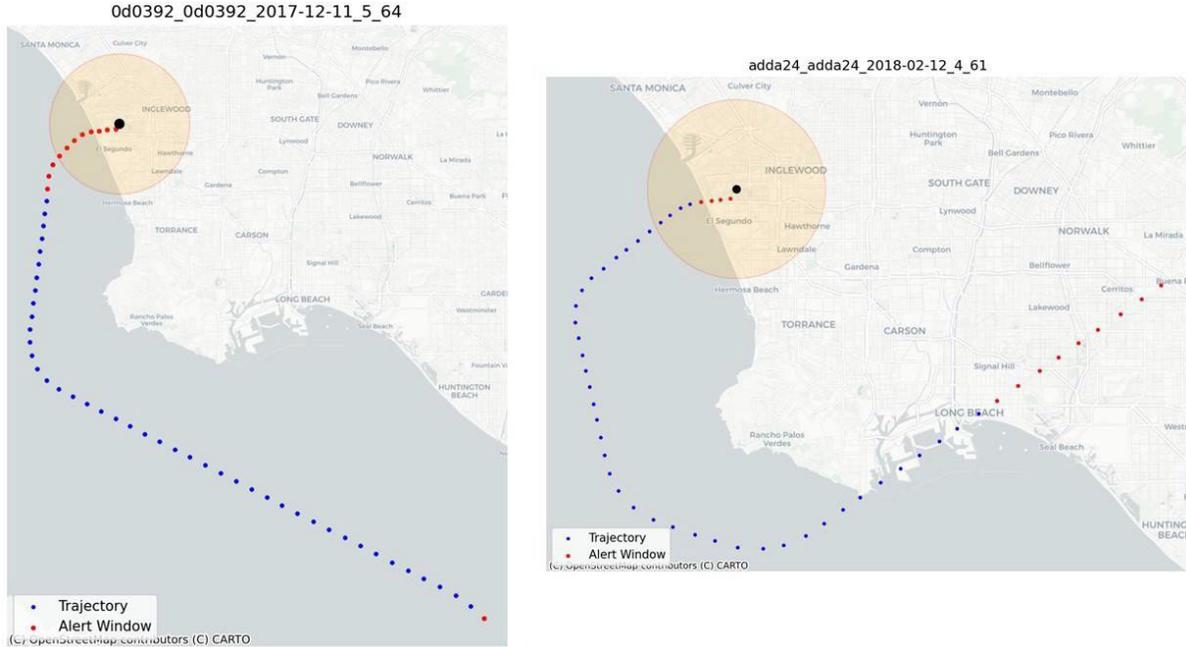


**Figure 4.** Aircraft 11167623 appears to follow a very standard takeoff trajectory. Possibly the directly perpendicular crosswind, resulting in a slight deviation to the downwind on takeoff before turning gradually into the wind, resulted in the anomaly flag for this scene.

## 4.2 OpenSky (LAX departures)

Two of the top 100 scoring trajectories in terms of number and magnitude of point alerts are plotted in Figure 5. In trajectory 0d0392\_2017-12-11 (Figure 5A), we see a series of flags on takeoff, possibly due to a jump in climb rate to above 15 m/s, higher than most median LAX profiles, but still a fairly standard steep LAX departure turning south-southeast over the ocean. In trajectory adda24\_2018-02-12 (Figure 5B), we see alerts occurring well away from the departure phase. The climb profile, headings, and altitude gain rate all look consistent, realistic, and safe. One possible reason for the alert could be due to an artifact: the alert extends to the end of the training distribution, where MAE is good at interpolation within a data distribution, but

bad at extrapolation, especially in slow drift segments like this.



## 5. Discussion

### 5.1 General Aviation: A Promising Application

In the GA airport setting, normative modeling worked surprisingly well. The expected flight behavior (standard pattern, entry/exit points, altitude ranges) is simple but well-defined. Deviations from this envelope are rare and usually meaningful — an aircraft departing without entering the pattern, flying straight through final, or making a sharp turn at low altitude stands out clearly. The MAE was able to learn these expectations without supervision and flag potential issues with little tuning.

### 5.2 Limitations at Scale: The LAX Case

By contrast, applying the same approach to LAX proved difficult. Despite LAX traffic having apparent order (standard procedures, published SIDs, and STARs), its complexity posed several challenges:

- Multiple flight phases with different distributions.
- Heavy routing variation based on runway assignment and ATC instructions.

- High density and interleaving of arrivals and departures.

Without flight phase segmentation or temporal context conditioning, the MAE struggled to learn a coherent normative baseline. Therefore, while the method is valid in principle, LAX-scale application would require:

- Phase-aware pre-segmentation or multi-task learning.
  - Tolerance modeling within each flight phase.
  - Refined scoring methods (e.g., slope of residual, uncertainty quantification).
- 

## 6. Future Work

- **Extend model evaluation across more untowered and lightly towered GA airports using FlightAware or ADS-B exchange.**
  - **Formalize statistical significance tests** for observed deviations.
  - **Develop real-time versions** of the model with onboard or mobile inference.
  - **Explore phase-aligned or multi-stage modeling** for high-complexity airports.
  - **Investigate multimodal extensions**, including weather (METAR), NOTAM, or airspace constraints.
- 

## 7. Conclusion

This study demonstrates a successful proof of concept for using masked autoencoders to detect anomalous behavior in general aviation traffic at untowered airports. The structured but simple flight envelope of these environments makes them well-suited for normative modeling. In contrast, large-scale applications like LAX present additional complexity that requires future modeling innovations.

Importantly, we show that **free and open-source air traffic data** is sufficient for building useful anomaly detection systems — a critical step toward democratizing airspace safety tools.

---

