WHITE PAPER

APRIL 2026

Kendall Gilbert, Systems Engineering, Aireon LLC
Dr. Giuseppe Sirigu, Systems Engineering, Aireon LLC
John Dolan, Systems Engineering, Aireon LLC
Joseph Canlas, Systems Engineering, Aireon LLC
Dr. Michael A. Garcia, Systems Engineering, Aireon LLC

Strategic flight route prediction for air traffic management using transformers

Strategic planning in air traffic management (ATM) depends on anticipating how flights will traverse an airspace under evolving demands, procedures, and constraints. With air traffic growing globally, novel ATM strategies are increasingly essential to enhance efficiency, safety, and predictability. Aireon has developed a transformer-based sequence model designed to predict the Flight Information Region (FIR) trajectory an aircraft will take, and the associated departure and destination aerodrome (ADEP/ADES).

The model, trained on historical flight trajectories, exploits attention to capture long range dependencies among flight attributes, historical route choices, and context such as schedule, airline behavior, aircraft type, and day of operation signals. We evaluate performance on held out flights using standard classification metrics, exact sequence accuracy, and the completion of an ATM use case analyzing three FIRs in the

Southeast Asia / Pacific Region. By providing accurate forecasts hours before departure, this model aims to improve forecast demand, sector configuration, and collaborative decision-making in modern ATM.

With air traffic growing globally, the need for enhanced and reliable air traffic management (ATM) is clearer than ever.[1] ATM aims to ensure the safe and efficient flow of air traffic. The goal of safety and efficiency grows progressively more complex as the number of flights and aircraft in operation continues to rise. This complexity is compounded by gradual and sudden changes to traffic flow patterns resulting from weather events and global conflicts.

The modernization of ATM is essential to address the limitations of legacy systems. Systems around the world are based on outdated technology unable to meet the demands of the continuously growing aviation sector.[2] Integrating advanced management technologies into ATM has emerged as a central strategy to address increasing operational demands and mitigate risk. ATM research highlights the need for optimization in delay minimization, congestion, and scheduling.[3] A fundamental part of modern ATM systems is the prediction of aircraft intent up to several hours prior to the entry of an aircraft into the airspace of interest. These predictions power metrics and insights such as airspace sector loads and congestion which serve as the basis of policies to manage the flow of inbound and outbound traffic. Different approaches have been developed to solve this problem, ranging from forecasting based on aircraft physics and performance models to advanced machine learning. Methods based on performance models use a priori information, like flight intent and aircraft performance assessments to predict the behavior in specific conditions.[4] [5] Such models require a wealth of prior knowledge and assumptions that are difficult to obtain and maintain when looking at applying them at a global scale; furthermore, they do not easily adapt to changing environment conditions and airspace constraints. Machine learning models provide levels of abstraction and generalization that allow them to be employed in the prediction of aircraft trajectory under uncertainty yielding good accuracy. Georgiou et al.[6] validated this through a multi-stage hybrid approach combining semantic clustering with hidden Markov models and regression-based predictors, achieving spatial accuracy on the order of 2–3 km for complete flights in real airspace data. Concurrently, Zhang and Mahadevan[7] explored deep learning architectures for enroute trajectory prediction, proposing a multi-fidelity blended model that combined deep feedforward neural networks (DNN) for single-step accuracy with Long Short-Term Memory (LSTM) networks for longer-horizon forecasting, further augmented by a Bayesian uncertainty quantification framework to support flight safety assessment. Building on the representational limitations of scalar-concatenation regression inputs, Guo et al.[8] introduced FlightBERT, a Transformer-based framework that reframes trajectory prediction as a multi-binary classification problem through a binary encoding representation, incorporating an Attribute Correlation Attention (ACoAtt) mechanism to explicitly model inter-attribute dependencies and demonstrating state-of-the-art performance on large-scale real-world Air Traffic Control data. More recently, Zhang et al.[9] proposed a wavelet transformer-based encoder-decoder architecture that performs time-frequency decomposition of flight trajectories, enabling simultaneous modeling of global flight trends and local motion details: This approach is particularly effective during complex maneuver phases such as climb and descent. Together, these works reflect a clear progression from classical statistical models toward sophisticated deep learning architectures capable of capturing the nonlinear, multi-scale spatiotemporal dynamics of aircraft motion.

Aireon developed a transformer-based model that does not use a priori information (e.g., flight plans) to forecast the route an aircraft will follow hours before a flight departs. One of the motivations for this development is that flight plan data is known to have errors, gaps, and deviations from the original plans. Transformers were introduced in 2017[10] as an alternative to recurrent and convolutional sequence models, replacing recurrence with self-attention so tokens can directly reference one another regardless of distance. They became popular as they are highly parallelizable (unlike recurrent neural networks [RNNs]), scale predictably with data and compute, and capture long-range dependencies efficiently. A decoder-only architecture is a Transformer model that consists solely of stacked decoder blocks with causal (masked) self-attention. At each position t, the model can attend only to tokens $1,...,t-1$, which makes it an autoregressive next-token-predictor: given a context $x_{1:t-1}$, it models $p(x_t | x_{<t})$, and generates sequences by repeatedly sampling or selecting the next token. For sequence-to-sequence tasks, the "source" and "target" are typically concatenated into a single stream using a format like prompt/source + separator + target, so the same causal mechanism learns to condition on the source portion and then produce the target continuation. Decoder-only models are strong for sequence-to-sequence prediction because they reduce the problem to a single, uniform objective — next-token prediction — which is simple to train at scale and matches how the model is used at inference time. They also avoid an explicit encoder–decoder split, letting the model flexibly allocate capacity across "understanding the input" and "producing the output" within one network, and they naturally handle variable-length inputs/outputs (including multi-turn or multi-step contexts) without architectural changes. Practically, this design supports efficient generation with key/value caching and works well in zero-shot and few-shot settings, where demonstrations in the prompt effectively "program" the sequence-to-sequence behavior.

The model introduced in this paper utilizes Aireon's global space-based Automatic Dependent Surveillance Broadcast (ADS-B) data in combination with temporal and aircraft data to produce the Flight Information Region (FIR) and Aerodrome of Departure (ADEP) and Aerodrome of Destination (ADES) airports. The output of Aireon's model provides a simple and accurate solution to assist traffic managers in identifying potential imbalances between demand and capacity. Additionally, this supports earlier and effective interventions while improving overall efficiency and safety. The model can predict flight plans globally across 60 different airlines for over 2,000 airports.

The paper provides an overview of the Aireon system, highlighting how the historical Aireon space-based ADS-B dataset is used to develop the transformer model (II), an overview of the transformer model created to generate flight predictions (III), and an experiment and results analysis to show how the model works and how it can improve ATM technology in three different flight information regions around the Southeast Asia / Pacific Region. This paper also highlights the possibilities for on-going work and future improvements (IV).

Aireon System

In April of 2019, the Aireon space-based ADS-B system went operational and began providing service to air navigation service providers (ANSPs) around the world. As the world's first space-based air traffic surveillance system, Aireon delivers real-time aircraft tracking across the entire planet, including remote oceans, polar regions, and other airspace previously beyond the reach of traditional ground-based technology. The system operates using 66 cross-linked Low Earth Orbit (LEO) satellites, along with 14 in-orbit spares, all flying approximately 780 km above Earth as part of the Iridium constellation.[11]

By leveraging this advanced satellite network, Aireon has access to a historical record of all aircraft ADS-B data in the entire world since April 2019. If an aircraft is operational and equipped with the necessary avionics, it will continuously broadcast its position and other relevant flight information. Aireon's ADS-B data is utilized as one of the main components of the dataset used to train the transformer-based model.

Methodology

In this section, we introduce the transformer-based sequence model that treats routes as structured token sequences.

A. Tokenizer

To prepare the text data for the transformer model, we employ a sub-word tokenizer to convert the text into tokens. The method used to preprocess the text data for the model was Byte Pair Encoding (BPE). BPE iteratively merges the most frequent pairs of characters in a corpus until a desired vocabulary size is reached. The algorithm learns common sub words that efficiently represent frequent words while transforming rare characters into smaller components. The representation of rare and unseen words as a sequence of sub word units gained popularity around 10 years ago when it was proven to have the ability to improve model performance.[12]

Due to the nature of our research and the data being worked with, examples of common tokens include characters like "aal," the International Civil Aviation Organization (ICAO) code for American Airlines or "a320-200" a common aircraft type being flown around the world. Many of the input sequences used in this model are transformed from around 200 characters to just 20 tokens. Fig. 1 shows how common vocabulary in the dataset is split into tokens.

B. Model Architecture and Training

The chosen model architecture is a decoder-only transformer, designed to generate tokens autoregressively based on previous tokens. Instead of bidirectional attention, used in encoder-decoder architectures, these transformers use causal attention to guarantee that the model does not see future tokens during training. Fig. 2 shows the decoder-only architecture.

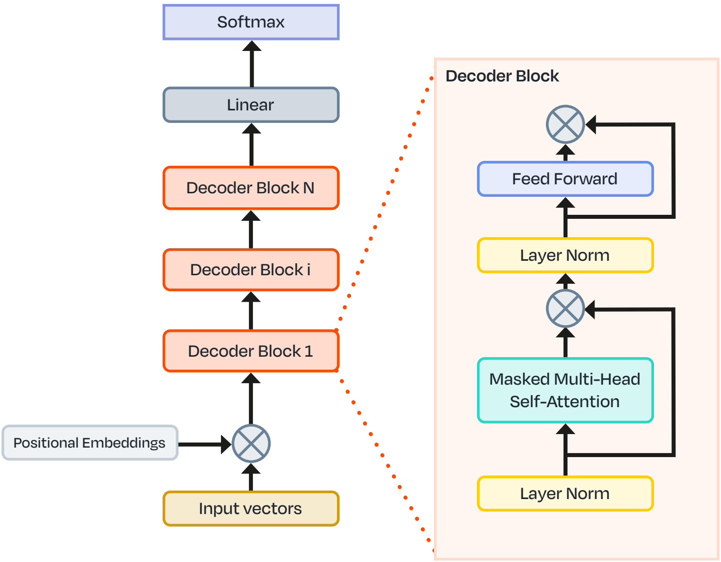
Several experiments have been performed to identify the best set of metadata to add to the model prompt (e.g., callsign, current location, date), the structure of the model, and the training parameters. For each experiment, the per-token accuracy and the per-sequence accuracy were monitored for both training and validation datasets.

FIGURE 1

BPE Example

FIGURE 2

Decoder-Only Transformer Architecture



A grid search approach was used to identify the best set of parameters for architecture and training, using the sets shown in Table I.

TABLE I

Training Parameters

Parameter	Value Set
Number of layers	[2, 3]
Number of heads	[8]
Batch size	[16, 32, 64]
Learning Rate	[1e-5, 2e-5]

Each combination of the above parameter was trained on a representative dataset composed of 300,000 inputs; the best parameters were identified as:

- Learning rate: 1e-5
- Number of layers: 2
- Batch size: 32

Experiment and result analysis

A. Data Set / Data Preprocessing

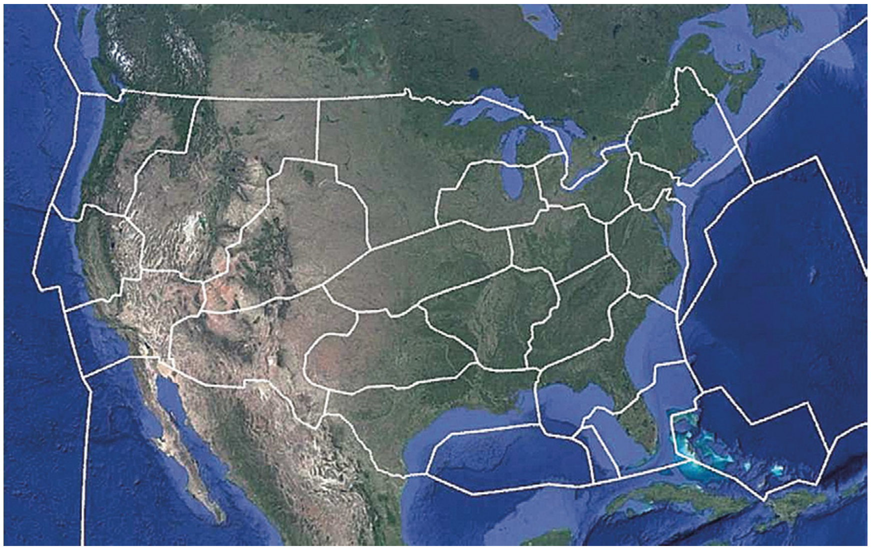
Four months of Aireon global space-based ADS-B data, augmented with FIR airspace, aircraft, and temporal features was utilized to train this transformer model. The features from nearly six million commercial flights from 2025-04-01 through 2025-07-31 were used to train, evaluate, and analyze the model.

The training set included commercial flights from 60 airlines around the world traveling to and from 2,696 unique airports. The specific airlines were chosen by looking at the airlines that had the most flights in seven regions around the world. The regions included North America, South America, Europe, Africa, Central Asia, Eastern Asia, and Australia. The most frequent routes are operated by airlines like Ryanair, Qatar, China Eastern, Southwest, American, IndiGo, United, and Delta.

FIR sequences range from a length of one to thirty. For instance, an aircraft traveling from Newark International Airport (KEWR) to Indira Gandhi International Airport (VIDP) will traverse 30 FIRs whereas an aircraft traveling from John F. Kennedy International Airport (KJFK) to Boston Logan International Airport (KBOS) will traverse two FIRs (KZBW, KZNY). The United States alone has 28 FIRs as shown in Fig. 3.
The quantity and complexity of the boundaries introduce a layer of complexity when training a transformer model to predict the correct sequences, in the correct order.

FIGURE 3

USA FIR Boundaries



Tables II and III show the input and output features with a detailed description of each feature.

TABLE II

Input Features

Feature	Description
Target Address	Unique six-character hexadecimal string assigned to the aircraft (e.g. 850EDC, 75862C)
Target ID	Flight Identifier assigned to the aircraft (e.g. AAL113, CEB4872)
Departure Time	Departure Day and Hour (e.g. Monday @ 2)
Aircraft Type	Aircraft Type (e.g. B777, A321)
Starting Position	Starting H3 Index - hexagonal spatial indexing system by Uber (e.g. 832a10fffffffff)[13]

TABLE III

Output Features

Feature	Description
ADEP	Aerodrome of Departure ICAO Code
FIR Sequence	FIR sequence aircraft travels
ADES	Aerodrome of Destination ICAO Code

The output sequence is a list of the ADEP, FIR Sequence, and ADES. For example, a flight traveling from Atlanta International Airport (KATL) to Boston International Airport (KBOS) would return a sequence like 'KATL KZTL KZDC KZNY KZBW KBOS,' where KATL and KBOS are the departure and arrival airports, and all other tokens are the FIRs in the order the model thinks the aircraft will traverse.

B. Evaluation Metrics

To evaluate the model's performance, we used masked accuracy and masked sequence accuracy on the training and validation set. Masking was utilized to exclude the padded tokens from the evaluation metrics. Sequence accuracy was used to evaluate how well the model predicts all the tokens in the correct order. The test set was used to evaluate the model's performance after all the training was completed.

Table IV presents the values of the chosen evaluation metrics.

TABLE IV

Evaluation Metrics

Metric	Value
Masked Accuracy	95.07%
Masked Sequence Accuracy	77.19%
Masked Val Accuracy	91.64%
Masked Val Sequence Accuracy	69.14%

The training set had a masked accuracy of 95.07% and a masked sequence accuracy of 77.19%. The validation set had a masked accuracy of 91.64% and a masked sequence accuracy of 69.14%. The sequence-level accuracy is comparingly lower due to the increased difficulty of predicting each token in the precise order, however, the high overall accuracy confirms that the model performs effectively. The model frequently predicts FIRs that are adjacent to the correct regions or sometimes omits minor FIRs that have little impact on the trajectory. Although these predictions still produce operationally reasonable flight paths, they are counted as errors under the sequence-level evaluation. The following section provides examples of routes the model has returned.

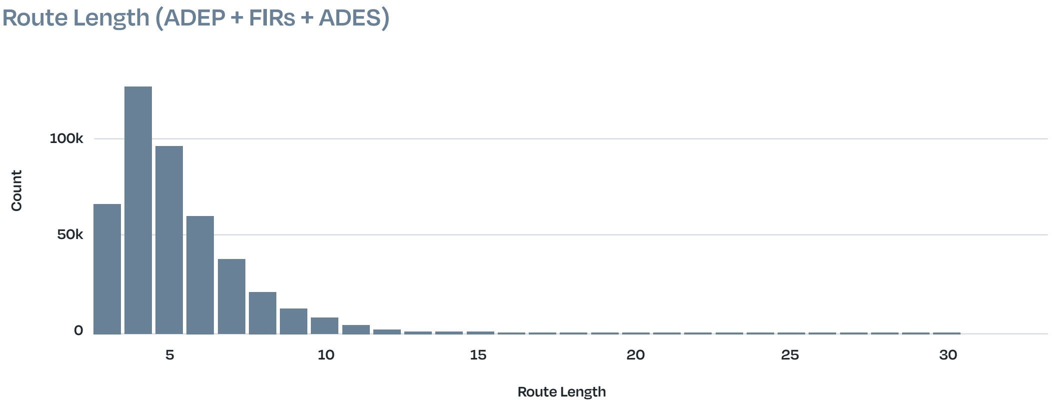
C. Analysis

The test set is made up of nearly 500,00 flights around the world, ensuring that model evaluation reflects the broad range of operations and traffic patterns globally. The test set data, never exposed to the model before, has an accuracy of 92.12% and a sequence accuracy of 70.12%.

The model can predict a variety of sequence lengths. The minimum sequence length is three and the maximum sequence length is 30. Most of the flights in the test dataset travel across two FIRs, making the most common route length four. Fig. 4 shows the distribution of route lengths across the test dataset. The right-tailed distribution illustrates that sequences are concentrated toward lower values.

FIGURE 4

Sequence Length

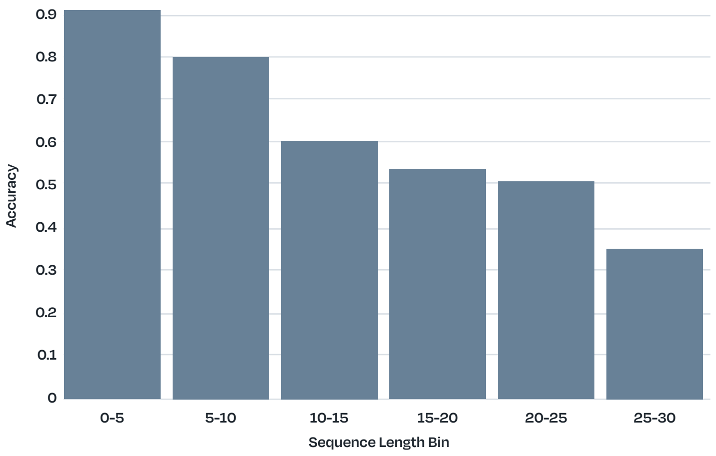


Route Length (ADEP + FIRs + ADES)

The model performs better on shorter routes due to the large quantity of flights that traverse those shorter flight paths as shown by the distribution. Fig. 5 bins a sample of 25,000 flights based on their length to show the accuracy of flights within a specified sequence length. Most notably, flights that fall within the first bin have the highest accuracy.

FIGURE 5

Accuracy Binned by Length



Flights in the 0-5 sequence length bin have an accuracy around 92% and a full sequence accuracy around 88%. This is the bin where most flights operated around the world fall. The accuracy drops significantly with flights that have a sequence length greater than 25.

The model can capture short-range and long-range routes. An example of a long-range flight route is United Arab Emirates flight UAE51N on 2025-07-08 traveling from Dubai International Airport (OMDB) to San Francisco International Airport (KSFO). This flight traverses 19 FIRs. The model accurately provides the ADEP, ADES, and FIR sequence given the input features. Fig. 6 shows the FIRs this aircraft crosses. The green dot represents the ADEP, and the red dot represents the ADES. The black dots are the exact path the aircraft took. In this example, the model accurately predicts all the correct FIR sequences in the correct order and the arrival and departure airports.

FIGURE 6

Flight UAE51N from OMDW to KSFO

Fig. 7 shows an example of a short-range route, flight FFT3216, traveling from Los Angeles International Airport (KLAX) to Hartsfield-Jackson Atlanta International Airports (KATL) operated by Frontier Airlines on 2025-07-07. The flight traverses five FIRs as it travels across the United States. Another example of a short-range flight is flight ETH825 from Addis Ababa Bole International Airport (HAAB) in Ethiopia to Prince Said Ibrahim International Airport (FMCH) in Comoros operated by Ethiopian Airlines on 2025-07-08. The flight, shown in Fig. 8, travels across four FIRs. The model maintains a high accuracy when forecasting less common flight paths, indicating strong prediction capabilities beyond common traffic patterns like the flight in Fig. 6.

FIGURE 7

Flight FFT3216 from KLAX to KATL

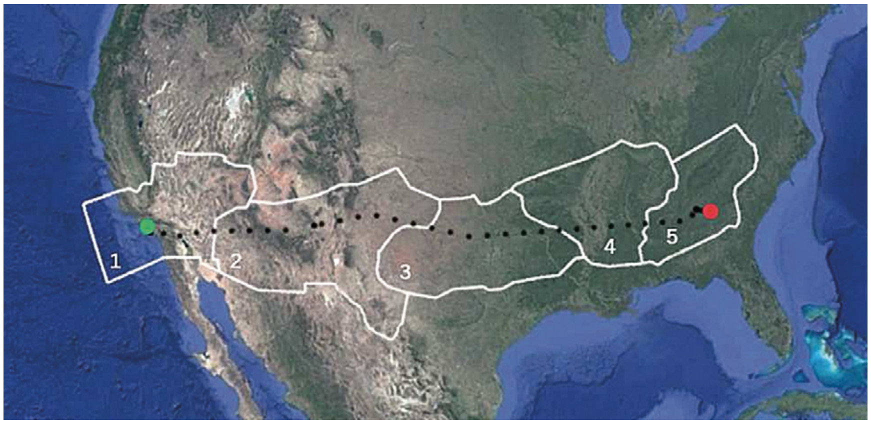
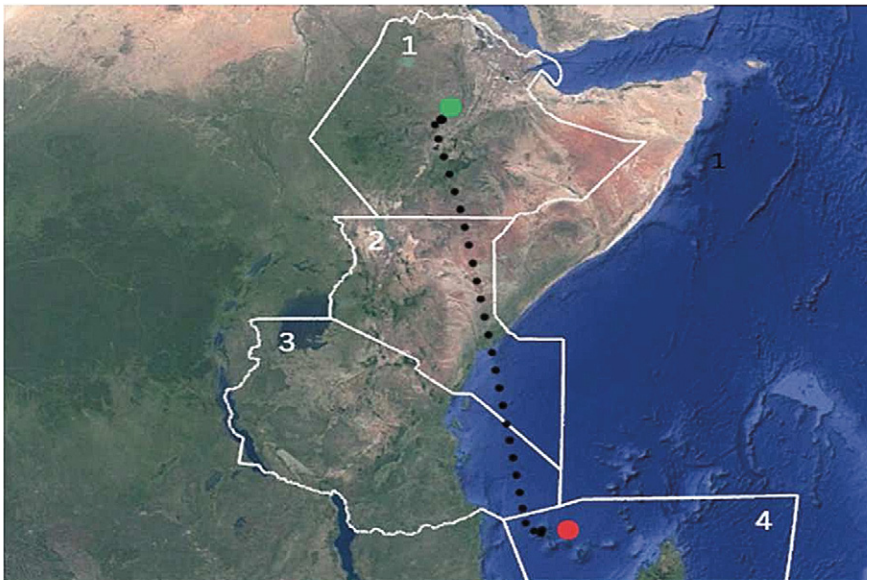


FIGURE 8

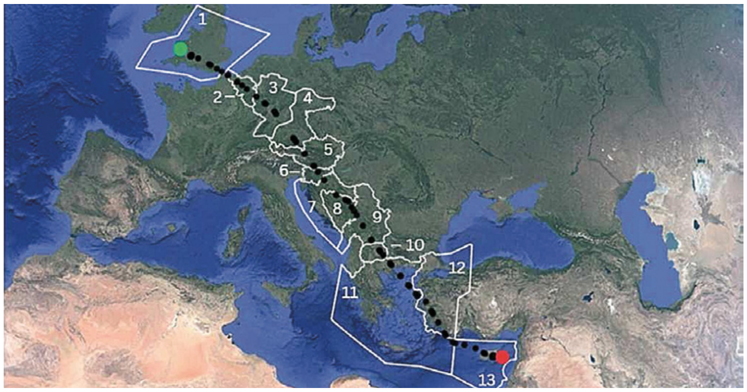
Flight ETH825 from HAAB to FMCH



Additionally, this model can predict exact routes even when the FIR sequences are complex. An example of this is flight EZY81FM operated by EasyJet from Bristol Airport in England (EGGD) to Larnaca Airport in Cyprus (LCLK) on 2025-07-08. This flight traverses 13 FIRs in a span of five hours, shown in Fig. 9. Even though the aircraft travels in and out of FIRs quickly, the model can predict the correct FIRs.

FIGURE 9

Flight EZY81FM from EGDD to LCLK

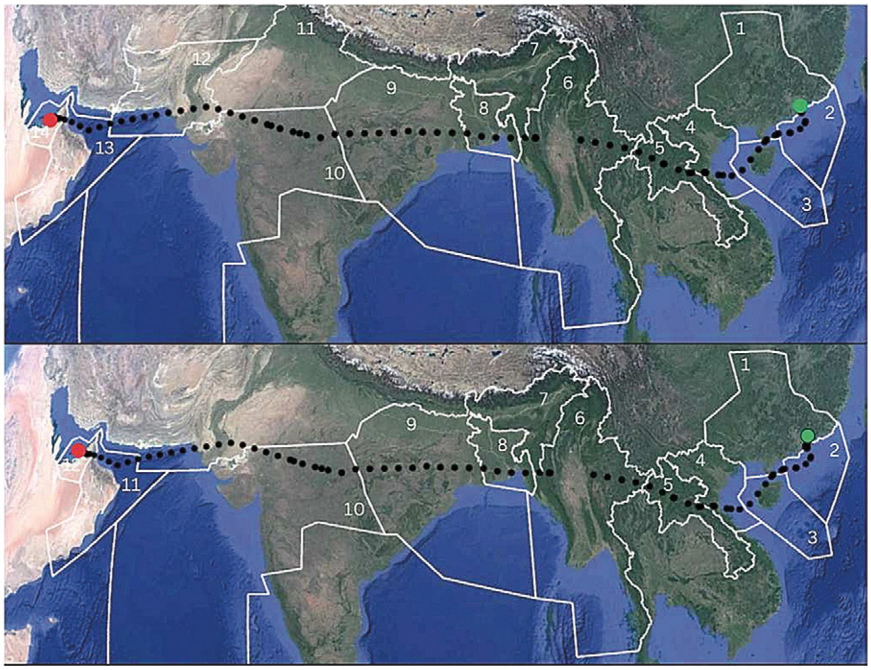


The lower sequence level accuracy is due to the variability in routes an aircraft will take. Variability in flight routes is normal and will occur due to things like weather, conflict, or restrictions. This causes the model to either predict extra FIRs, less FIRs, or the wrong FIRs. For instance, a flight traveling from Hong Kong (VHHH) to Dubai (OMDW) will not take the same route each time. Fig. 10 highlights an instance where the model failed to predict two FIRs the aircraft traveled through on 2025-07-08.

In Fig. 10, the top image overlays the aircraft path and the true FIRs the aircraft on flight UAE9787 traveled through. The bottom image shows the predicted FIRs. Although the model does not predict two FIRs (VIDF and OPKR), the sequence predicted still obtains the correct ADEP, ADES, and most of the FIRs while successfully providing a route to Dubai. It is instances like this that drag down the sequence accuracy. Even when a sequence is considered incorrect due to a token not being in the correct spot or the wrong token, the generated route is still largely realistic and usable.

FIGURE 10

Flight UAE9787 from VHHH to OMDW

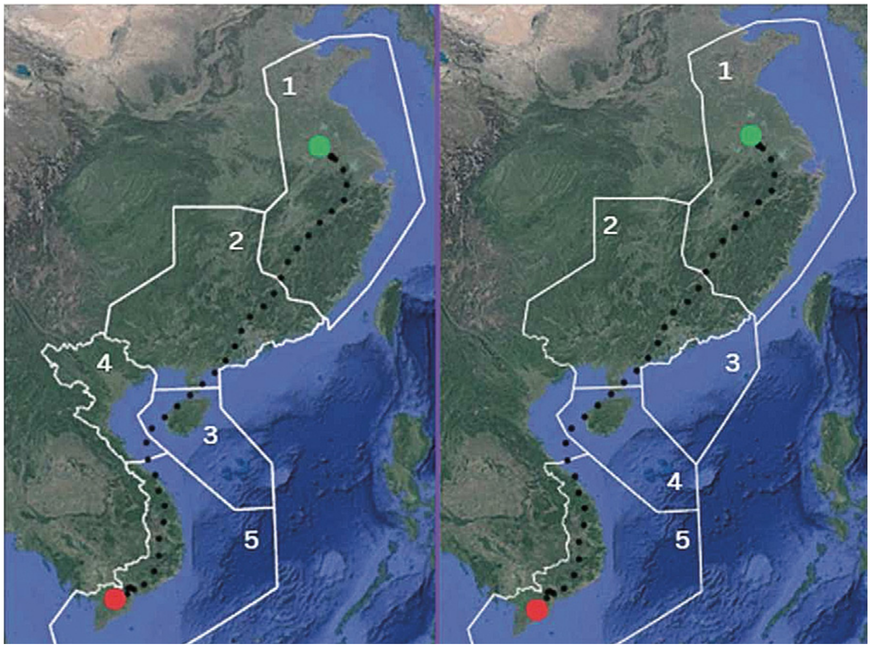


Additionally, the model often predicts flight paths for FIR sequences adjacent to the correct FIRs. Fig. 11 shows China Eastern Airways flight CES2839 traveling from Nanjing Lukou International Airport (ZSNJ) in East China to Tan Son Nhat International Airport (VVTS) in Ho Chi Minh City, Vietnam on 2025-07-27. Fig. 11 shows the true flight path (left) and the predicted flight path (right). The model incorrectly predicts the FIR in spot three and fails to predict the correct FIR at spot four. A prediction like this brings down the accuracy and sequence accuracy due to incorrect order and missing FIRs even if the predicted path deviates slightly from the true path.

Overall, this analysis demonstrates how the model can predict routes globally given minimal information. The next section continues the results and experiment section by conducting a case study on three FIRs.

FIGURE 11

Flight CES2839 from ZSNJ to VVTS

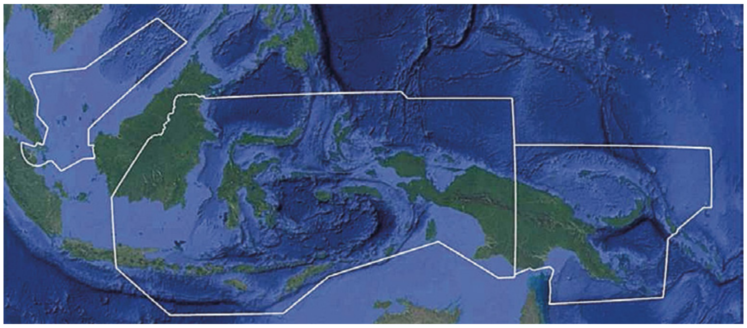


Use case

To illustrate how this model can improve Air Traffic Flow Management (ATFM), a one-week case study was conducted looking at flights flying into the Ujung Pandang (WAAF), Singapore (WSJC), and Papua New Guinea (AYPM) FIRs from 2025-10-25 through 2025-10-31. These Southeast Asia / Pacific regions were selected due to both the need for critical advancements in ATFM technology and the high volume of domestic and international flights operated within them. There are concerns that ATFM operations throughout this region may be untenable due to system inefficiencies.[14] With accurate knowledge of incoming traffic from a model like ours, these airspaces can proactively manage flow and optimize routing. Fig. 12 provides a visual of the three FIRs analyzed.

FIGURE 12

From Left to Right WSJC, WAAF AYPM



Out of the 60 commercial airlines included in the dataset, Aireon has records of 1014 flights flying through WAAF, 240 flights through AYPM, and 2083 flights flying through WSJC. These three airspaces provide a strong combination of flights coming from all directions, as shown by Fig. 13.

FIGURE 13

Flights in WAAF, WSJC and AYPM

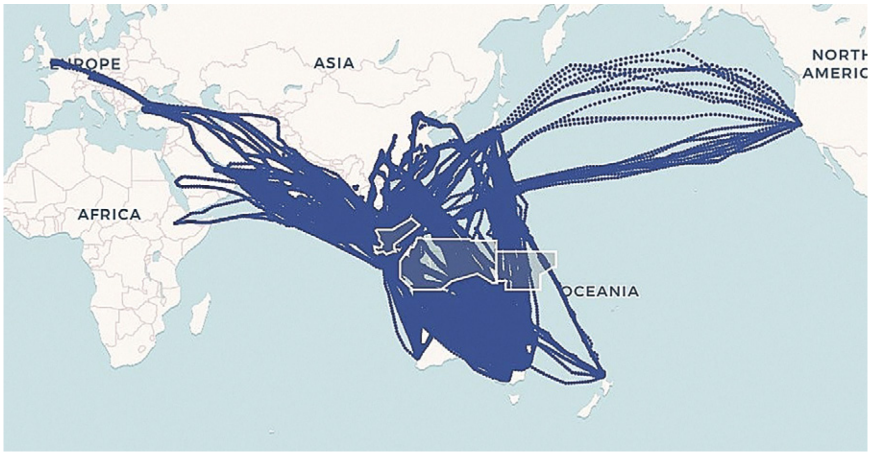


Table V highlights the accuracy of predicted flight plans for the recorded flights flying into the chosen FIRs and the percentage of flights correctly placed within that FIR.

TABLE V

FIR Matched Flights

FIR	Flights	Accuracy	Correctly Placed
WAAF	1014	91.06%	86.29%
WSJC	2083	86.36%	86.94%
AYPM	240	97.02%	96.25%

The accuracy is highest for flights flying into the Papua New Guinea airspace (AYPM) followed by Ujung Pandang (WAAF), then Singapore (WSJC). The model correctly places 86.29% percent of the 1014 flights in WAAF, 86.94% of the 2083 flights in WSJC, and 96.25% of the 240 flights in AYPM. Ujung Pandang and Singapore accuracy struggles due to the greater quantity of flights entering their airspace that traverse long sequences of FIRs.

Conclusion

Growing traffic levels and more complex patterns due to climatic and geopolitical factors are increasing the load on several airspaces across the world. Modern ATFM solutions are being deployed across the globe to help manage these new traffic patterns and maintain efficiency across the network.

This paper proposes a machine learning approach based on a decoder-only transformer architecture to predict the most likely high-level route followed by the aircraft. The model was applied to commercial flights only and yielded per-token accuracy around 92% and sequence-based accuracy around 70% on unseen data. The model reported lower performance on routes where the variability in the input data is very high, due to environmental factors (e.g., weather).

In the future, the approach will be expanded beyond commercial flights; this introduces new challenges as these flights exhibit fewer consistent patterns, making their trajectories difficult to predict without additional context. Additionally, the model will be refined to predict more detailed waypoint-level flight plans. This can be done with more context incorporated into the transformer, like weather and adaptation data. Furthermore, the prediction mechanism can be run iteratively during the flight to improve accuracy as the aircraft approaches its destination and seeks to maximize prediction accuracy where the value of the prediction is at its highest (which may be closer to mid-flight than the takeoff location). The integration of these enhancements is believed to substantially elevate the model's predictive abilities for all air travel globally.

Acknowledgment

The authors of this paper would like to thank Valerie Cox for her technical review and contributions to this paper.

References

[1] A. Leveine, A. Gordon, A. Wulz, C. Murphy, C. Shi, and K. Chang, "Air Travel Demand Outlook 2026: Revenues – And Costs – Are Rising," BCG, 2026.

[2] GAO, "Air Traffic Control FAA Actions are Urgently Needed to Modernize Aging Systems," [Online]. Available: GAO-24-107001, Air Traffic Control: FAA Actions Are Urgently Needed to Modernize Aging Systems

[3] V. Aditya, D. Aswin, S, Dhaneesh, et al. "A review on air traffic flow management optimization: trends, challenges, and future directions," Discover Sustainability, 2024.

[4] E. Gallo, J. López-Leonés, M. A. Vilaplana, F. A. Navarro, and A. Nuic, "Trajectory Computation Infrastructure Based on Bada Aircraft Performance Model," IEEE/AIAA 26th Digital Avionics Systems Conference, 2007, 10.1109/DASC.2007.4391830.

[5] D. P. Thipphavong, C. A. Schultz, A. G. Lee, and S. H. Chan, "Adaptive algorithm to improve trajectory prediction accuracy of climbing aircraft," Journal of Guidance, Control, and Dynamics, vol. 36, no. 1, pp. 15–24, 2013.

[6] H. Georgiou, N. Pelekis, S. Sideridis, D. Scarlatti, and Y. Theodoridis, "Semantic-aware aircraft trajectory prediction using flight plans," International Journal of Data Science and Analytics, 2020, 10.1007/s41060-019-00182-4.

[7] X. Zhang and S. Mahadevan, "Bayesian neural networks for flight trajectory prediction and safety assessment," Decision Support Systems, Volume 131, 2020, 113246, ISSN 0167-9236.

[8] D. Guo, E. Q. Wu, Y. Wu, J. Zhang, R. Law and Y. Lin, "FlightBERT: Binary Encoding Representation for Flight Trajectory Prediction," in IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 2, pp. 1828-1842, 2023.

[9] Z. Zhang, D. Guo, S. Zhou, J. Zhang, and Y. Lin, "Flight trajectory prediction enabled by time-frequency wavelet transform," Nature, volume 14, issue 1, pages 1-15, 2023.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need.," Advances in neural information processing systems 30, 2017.

[11] M. Garcia, "A Compilation of Measured ADS-B Performance Characteristics from Aireon's On-Orbit Test Program," ESAVS Conference 2018.

[12] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units." Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: long papers. 2016.

[13] Uber, "H3: Ubers Hexagonal Hierarchical Spatial Index," [Online]. Available: H3: Uber's Hexagonal Hierarchical Spatial Index | Uber Blog

[14] ICAO, "Ffteenth Meeting of the Asia/Pacific Air Traffic Flow Management Steering Group," [Online]. Available: WP14-Establishment-of-ATFM-Concept-Design-Ad-Hoc-Group.pdf