Air Traffic Flow Representation and Prediction using Transformer in Flow-centric Airspace

Chunyao Ma*, Sameer Alam*, Qing Cai*, Daniel Delahaye†
*Air Traffic Management Research Institute, School of Mechanical and Aerospace Engineering
Nanyang Technological University, Singapore
†OPTIM Lab, Ecole Nationale de l'Aviation Civile, Toulouse, France
Email: *M180146@e.ntu.edu.sg, {*sameeralam| *qcai}@ntu.edu.sg,
†daniel.delahaye@enac.fr

***Abstract*—The air traffic control paradigm is shifting from sector-based operations to cross-border flow-centric approaches to overcome sectors' geographical limits. Under the flow-centric paradigm, prediction of the traffic flow at major flow intersections, defined as flow coordination points in this paper, may assist controllers in coordinating intersecting traffic flows which is the main challenge for implementing flow-centric concepts. This paper proposes to predict the flow at coordination points through a transformer neural network model. Firstly, the flow coordination points, i.e., the major flow intersections, are identified by hierarchical clustering of flight trajectory intersections whose location and connectivity characterize daily traffic flow patterns as a graph. The number of coordination points is optimized through graph analysis of the daily flow pattern evolution. Secondly, air traffic flow features in the airspace during a period are described as a "paragraph" whose "sentences" consist of the time and callsign sequences of flights transiting through the identified coordination points. Finally, a transformer neural network model is adopted to learn the sequential flow features and predict the future number of flights passing the coordination points. The proposed method is applied to French airspace based on one-month ADS-B data (from Dec 1, 2019, to Dec 31, 2019), including 158,856 flights. Results show that the proposed prediction model can approximate the actual flow values with a coefficient of determination (R^2) between 0.909 to 0.99 and a mean absolute percentage error (MAPE) varying from 27.4% to 11.7% with respect to a 15-minute to 2-hour prediction window. The sustainability of the prediction accuracy under an increasing prediction window demonstrates the potential of the proposed model for longer-term flow prediction.**

***Index Terms*—air traffic management, flow-centric operation, air traffic flow prediction, transformer neural networks.**

I. INTRODUCTION

The traditional Air Traffic Control (ATC) service is established in geographically partitioned sectors, organized and managed by the national Air Navigation Service Provider (ANSP) under expected traffic demand and airspace availability [1]. However, such local sector-based approaches prevent the regional level management and optimization of air traffic as it only manages individual flight paths within a sector. Facing the growth of air traffic demand and difficulties in subdividing heavily loaded sectors, researchers have started examining and testing the concept of sectorless ATC, which views the airspace as a whole instead of the current practice of dividing the airspace into small sectors [2].

One primary practice of sectorless airspace is flow-centric operation [3], which relies on controlling and monitoring flow-based formation and evolution of air traffic instead of a fixed number of geographical sectors. The flow-centric approach is based on the management of dynamic flow corridors. It opens the opportunity to distribute air traffic more efficiently in the airspace without being constrained by sector boundaries.

Despite the benefits of the flow-centric concept, its implementation has been rather limited. One primary challenge is the coordination between traffic flows to avoid potential conflicts [4]. The highly dynamic nature of air traffic flow brings uncertainties into the flight routes, traffic pattern, and airspace constraints, adversely affecting the effective, efficient, and safe coordination and regulation of air traffic. An accurate short-term prediction of air traffic flows, i.e., one to two hours look-ahead time [5], in the airspace can provide network-wide flow information to ANSPs and airspace users. It can be a vital support to flow coordination beforehand to avoid conflict situations involving a large number of flights. For instance, flow regulation measures, such as traffic re-routing, can be executed in advance to provide a smooth and efficient flow of traffic when the predicted flow exceeds available capacity [6].

In the literature, most research modeled air traffic flows as the number of flights at specific geographical locations in the airspace, such as airways, waypoints, or sectors [7]. one way of making air traffic flow prediction is to predict the individual flight trajectories over time and count the future number of flights in the airspace [8]. The prediction error of such trajectory-based methods increase exponentially when the forecast time horizon increases beyond 20 minutes, which may be caused by its sensitivity to the prediction accuracy of various flights under consideration [9]. Furthermore, the dimension of the models depend on the number of flights in scope, making such models untenable in real world practice due to limited computational resources.

An alternation to the trajectory-based approach is the aggregate air traffic flow prediction model focusing on predicting the overall distribution of the air traffic flows [10]. The aggregated prediction methods are less sensitive to the uncertainty factors related to individual flights, which can achieve a longer forecast time horizon with less prediction errors and requires significantly less computational cost comparing to the trajectory prediction based approaches.

One way to make aggregate flow prediction is to describe the number of future flights handled by a control center or a sector as a linear function of a collection of influencing observations. Examples include the Linear Dynamic System Models (LDSM) based on state transition matrices [11], [12] and the linear regression model [13]. The major shortcoming of those linear models is their incapability in describing complex relations other than linearity among the air traffic flow and the contributing variables. Considering the air traffic flow in the airspace is the consolidated operation of a variety of stakeholders over space and time, e.g., controllers, airspace users and ANSPs, simple linear models may not be able to capture the complex and dynamic behaviors of air traffic flow to produce accurate short-term predictions. Another widely discussed approach for aggregated flow prediction is the machine learning based model [14], such as time-series extraction and prediction of the number of flights passing a given waypoint using Extreme Learning Machine based models [15], prediction of air traffic demand between two airports using long Short-Term Memory (LSTM) and Support Vector Machine (SVM) [16], and end-to-end spatial-temporal time-series prediction based on the concatenation of Convolutional Neural Networks or Graph Convolutional Networks (GCN) with recurrent neural networks, such as LSTM, to respectively capture spatial and temporal features [5], [17]. Most of the time-series models use the number of flights passing areas or points in the airspace as the only input feature. Those models mainly analyze the temporal or spatial-temporal variation of the number of flights [18], which may fail to consider more detailed air traffic features, such as the flight sequences transiting through different flow coordination points, and therefore, limits the prediction accuracy.

In view of the above analysis, this paper proposes a transformer base model for predicting traffic flow at the major flow intersections.First, the flow coordination points are identified by clustering the flight trajectory intersections using a hierarchical clustering algorithm. The number of clusters is determined through graph analysis of the flow pattern evolution. Then, the complex and dynamic airspace flow features are described as a "paragraph" with "sentences" representing sequences of the time and callsigns of flights transiting through the coordination points. Finally, These sequential flow details are used as the input to a transformer neural network model to extract the contextual relations of elements in long sequences and predict future flow on the coordination points.

II. Methodology

A. Methodology Overview

Focusing on the flow-centric concept, this paper investigates air traffic flow prediction in two main steps. The first step is to develop a mechanism to identify the flow coordination points where major air traffic flows intersect from air traffic data. The second step is developing a flow prediction model to predict the number of flights passing the identified coordination points during different look-ahead periods. Fig. 1 presents a concept diagram of the proposed flow prediction method.

In the flow-coordination-points identification step, the daily flow patterns are represented as a graph whose nodes are extracted from clusters of daily flight trajectory intersection points and whose edges are constructed based on the connectivity between the nodes. To determine the optimal number of clusters for flow representation, this paper proposes a graph analysis approach that evaluates the homogeneity of nodes and edges across graphs constructed from different days of air traffic. By maximizing the homogeneity between the graph representations of the daily air traffic flow, the cluster number of trajectory intersection points is determined, and the centers of the clusters are identified as the flow coordination points.

In the air traffic flow prediction step, each flight trajectory is represented as a sequence of the identified coordination points through trajectory registration which minimizes the dissimilarity between the original trajectory and the trajectory formed from the coordination points. Then, the flow feature in the airspace during a period is modeled as sequences of flights transiting through the set of coordination points. It is forwarded to a transformer based flow prediction model to learn the contextual description of the complex dynamics of the air traffic flow and predict the behavioral dynamics of the flow components, i.e., the number of flights transiting through the coordination points during a future period.

B. Flow Regulation Points Identification

ADS-B data for aircraft surveillance can give highly accurate aircraft position and velocity information, providing a valuable source for analytical solutions to effective and efficient airspace usage [19]. This paper identifies the flow coordination points based on the flight trajectories analysis using ADS-B data, including intersection points clustering, daily flow representation, and graph analysis.

1) Intersection Points clustering: As mentioned in Section II-A, the flow coordination points in this paper are defined as the positions in the airspace where air traffic flows intersect. Therefore, this paper proposes to determine the coordination points clustering the intersection points of flight trajectories.

The intersection points between flight trajectories in this paper are calculated as the intersections of the mapped trajectories onto the earth's surface determined by latitudes and longitudes. Then clustering algorithms are applied to extract the natural groupings of the trajectory intersection points to identify the flow coordination points.

Hierarchical clustering relies on the hierarchical decomposition of the data based on group similarities to find a multilevel hierarchy of clusters, where clusters at one level are joined as clusters at the next level [20]. Air transportation networks are commonly a nodal hierarchy that follows the spoke–hub structure in which traffic flows are often concentrated around the traffic hubs, and flight routes are organized as a series of "spokes" that connect outlying areas to a hub area. There is a hierarchy of air traffic flows ranging from regional feeders to international hubs [21]. Therefore, this paper adopts the

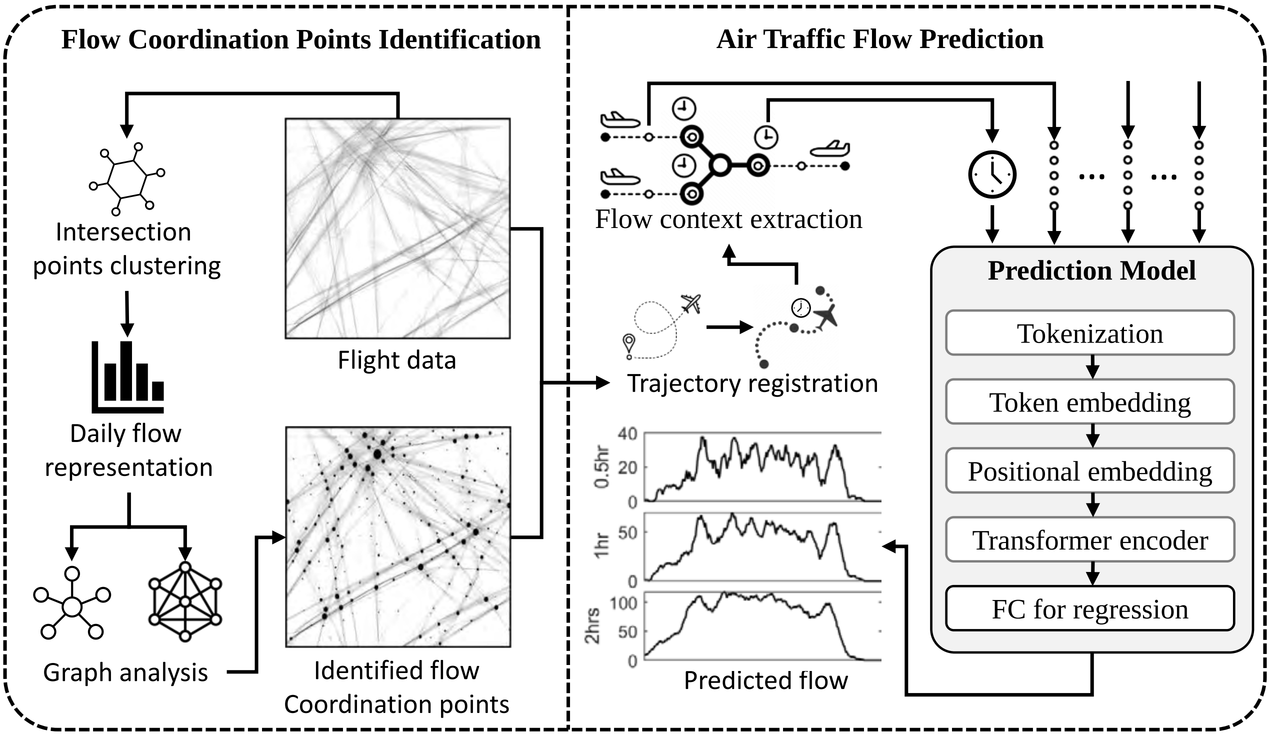


Figure 1: Conceptual diagram of the proposed method for coordination points flow prediction using transformer neural networks.

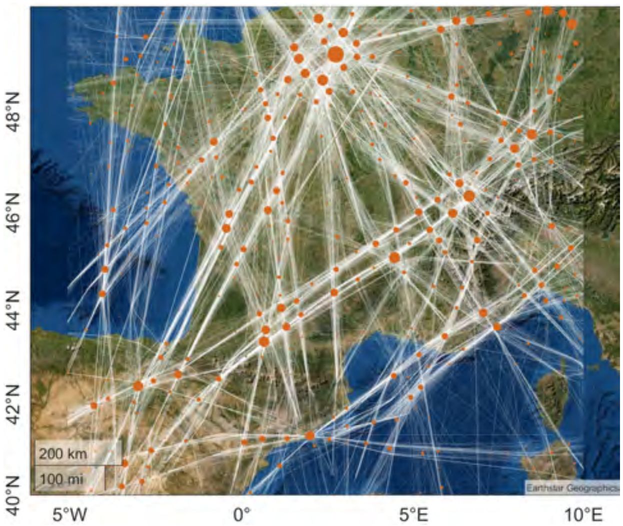


Figure 2: An example of the intersection point clustering.

hierarchical clustering algorithm to discover the representative flow coordination points.

Fig. 2 shows an example of the clusters of intersection points. The white curves depict the flight trajectories, and the solid red circles denote the major flow intersections, which are the centers of the identified trajectory intersection clusters. The size of each circle is proportional to the number of intersection points in the cluster.

The number of clusters to be identified by the hierarchical clustering algorithm is an important parameter, as the clustering outcome should be able to discover and represent the natural organization of traffic flow intersections in the airspace. On the one hand, if the number of clusters is too large, the clustering result will be susceptible to small fluctuations in air traffic flow, which may be unable to capture major flow patterns to produce credible flow prediction results. On the other hand, when the number of clusters is too small, the identified flow intersections can largely deviate from real flow paths, which dilutes the implementation significance as a supporting tool to controllers for flight coordination.

2) Daily Flow Pattern Representation: To determine the optimal number of clusters that should be identified by the clustering algorithm, this paper proposes a graph [22] based approach which represent the daily air traffic flow pattern as a weighted graph $G = (V, E)$, where V is the set of nodes denoting the air traffic flow components, i.e., the flow coordination points, which are represented as the centers of the identified clusters. The flow connections between the nodes can be described by the weighted edges E. If two nodes are passed by a flight consecutively, they will be considered connected, and there will be an edge between them. The weight of each edge is proportional to the air traffic volume transiting through it, i.e., the number of flights whose trajectories consecutively cross the two nodes connected by the edge.

C. Graph Analysis

Considering the necessity of consistent and dependable performance of airspace users for improvement of the ATM system predictability, [23], the representation of the traffic flow patterns should be able to describe the behavioral consistency of air traffic flow in the airspace as well as the daily alternations in the geographical positions of the intersection points and the variations in the flow structure. Thus, this section proposes to determine the optimal number of clusters by modeling the consistency of daily flow intersection patterns versus the changes in the number of clusters.

The flow patterns can evolve temporally following the daily alternations in the air traffic flow, geographically and structurally. Therefore, the flow pattern consistency is evaluated from two perspectives: 1) geographical consistency in node positions; 2) structural consistency in node connections.

a) Nearest neighbour analysis: To measure the geographical consistency in the daily node positions, a nearest neighbour based analysis is conducted to match the nodes on the temporal horizon. Concretely, given the traffic pattern graphs for two consecutive days, for each node in a graph, this analysis first searches in the other graph to find its nearest node based on the geographical coordinates. Considering the daily flow consistency, this pair of nodes is expected to represent the same flow coordination point in the airspace despite some minor positional shifts. Therefore, under a proper graphical flow pattern representation, the resulting nodes pair, a node with its identified nearest neighboring node, should be identical for the nodes in both graphs. Conversely, an unqualified graph representation may result in significantly different matches between the nodes. Therefore, the geographical consistency in node positions is quantified as the ratio of the mutually matched nearest node pairs to the total number of nodes in the graphs.

Given a number of nodes n in the graph G_k, let $V_k = \{v_k^1, v_k^2, ..., v_k^i, ..., v_k^n\}$ represent the set of nodes in the graph constructed for day D_k, and let $V_{k+1} = v_{k+1}^1, v_{k+1}^2, ..., v_{k+1}^j, ..., v_{k+1}^n$ represent the set of nodes in the graph G_{k+1} constructed for day D_{k+1}. For each node v_k^i in V_k, find its nearest node $v_{k+1}^{a_i}$ in V_{k+1}. By representing the $(latitude, longitude)$ of node v_k^i as $(\varphi_k^i, \lambda_k^i)$ and node v_{k+1}^i as $(\varphi_{k+1}^i, \lambda_{k+1}^i)$, the nearest neighbour of node v_k^i in V_{k+1} is identified according the great circle distance:

$$a_i = \underset{\substack{\{j\} \\ j=1,...,n}}{\arg\min} \quad D(ij) = R \cdot 2sin^{-1}(min(1, \sqrt{h})) \tag{1}$$

where

$$\begin{aligned} h &= sin^2\left(\frac{\Delta\varphi}{2}\right) + cos\left(\varphi_k^i\right) cos\left(\varphi_{k+1}^j\right) sin^2\left(\frac{\Delta\lambda}{2}\right) \\ \Delta\varphi &= \varphi_k^i - \varphi_{k+1}^j \\ \Delta\lambda &= \lambda_k^i - \lambda_{k+1}^j \end{aligned} \tag{2}$$

R is earth's radius (mean radius = 6,371km), and h is the haversine formula which determines the great-circle distance between two points on a sphere given the longitudes and latitudes.

By identifying the nearest neighbouring node $v_{k+1}^{a_i}$ of node v_k^i for $i = 1, ..., n$ and identifying the nearest neighbouring node $v_k^{b_j}$ of node v_{k+1}^j through the same calculation, two sets of matched node pairs between the two graphs can be obtained: $S_1 : (v_k^1, v_{k+1}^{a_1}), (v_k^2, v_{k+1}^{a_2}), ..., (v_k^n, v_{k+1}^{a_n})$ and $S_2 : (v_k^{b_1}, v_{k+1}^1), (v_k^{b_2}, v_{k+1}^2), ..., (v_k^{b_n}, v_{k+1}^n)$. Thus, the geographical consistency gc in node positions is quantified as:

$$gc_1 = \frac{|S_1 \cup S_2|}{n} \tag{3}$$

where $|S_1 \cup S_2|$ represents the number of node pairs in the union of S_1 and S_2, which will be denoted as l in the paper.

b) Graph structure analysis: Let $S = S_1 \cup S_2$ represent the set of paired nodes of graph G_k and graph G_{k+1}, while $C_k = \{c_k^1, c_k^2, ..., c_k^l\}$ and $C_{k+1} = \{c_{k+1}^1, c_{k+1}^2, ..., c_{k+1}^l\}$ denote the nodes from graph G_k and G_{k+1} in S respectively. Let e_{c-k}^{i-j} represent the edge connecting nodes c_k^i and c_k^j. Let w_{c-k}^{i-j} represent the weight on edge e_{c-k}^{i-j}.

The structural consistency in the air traffic flow pattern is measure by the divergence in the weighted edges between graphs. More specifically, with the set of paired nodes between graphs, the structural consistency is evaluated by the mutual flow connections of nodes in C_k and C_{k+1} versus all the flow connections in graph G_k and G_{k+1}, which is formulated as:

$$gc_2 = \frac{\sum_{i=1}^{l-1}\sum_{j=i+1}^{l} min(w_{c-k}^{i-j}, w_{c-k+1}^{i-j})}{0.5 \times (\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_k^{i-j} + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{k+1}^{i-j})} \tag{4}$$

By calculating the geographical consistency gc_1 and the structural consistency gc_2 versus varying number of clusters, a "saddle point" on the curve can be adopted as the optimal number of clusters for traffic flow representation.

D. Air Traffic Flow Prediction

Most traffic flow prediction model adopts recurrent neural networks, e.g., LSTM and GRU [24], to capture the temporal relations in a time series data. However, data in such sequential models need to be processed in order, i.e., the model architecture requires the output of the previous state to calculate the next state, which makes it difficult to capture long-term relations in the sequences, reducing the model's prediction power when the prediction window increases.

Transformers are the state-of-the-art neural networks that learn context by tracking relationships in sequential data like the words in the sentence. A transformer combines the attention mechanism (a mechanism that enables to highlight of relevant features of the input sequence dynamically) based representations and a parallel architecture in processing elements in a sequential input. Without the use of recurrent connections, the transformer model allows the ingestion of an entire sequence in parallel and has the potential to understand the relationship between sequential elements that are far from each other.

The seminal transformer uses an encoder-decoder architecture. The encoder extracts features from an input sequence, and the decoder uses the features as input to produce an output sequence, such as translating an English sentence into a French sentence. Therefore, the air traffic flow prediction model in this paper proposes to adopt a transformer neural network architecture to extract the air traffic flow sequence features and predict future air traffic flow.

Moreover, most existing air traffic flow prediction models are flow-density-based models that only analyze the temporal or spatial-temporal variations of the number of flights. They may not be a potent representation of the airspace's complex air traffic flow features. In this consideration, air traffic flow information in this paper is described by a detailed "paragraph" whose "sentences" are the sequences of flights passing the flow coordination points in a time sequential order.

In view of the above analysis, this section first registers the flight trajectories onto a sequence representation of the flow coordination points. The traffic flow context in the airspace during a period is then represented as a "paragraph" whose "sentences" are the sequences of flights (characterized by the flight callsigns) in the airspace. Then this contextual description of air traffic flow is used as input to a prediction model to learn from the flow context and predict the future air traffic flow. Details of each step will be illustrated in the following paragraphs.

1) Trajectory Registration: The objective of trajectory registration in this paper is to search for a sequence of flow coordination points that can optimally approximate the original trajectory. This objective is formulated as finding the minimum dissimilarity between the original trajectory and a representative trajectory constituted by a subset of the coordination points. Let T_o represent a flight trajectory which is depicted by a sequence of points $\{p_o^1 \to p_o^2, \to ..., \to p_o^M\}$ obtained from ADS-B data. Let $FRP = \{p_r^1, p_r^2, ..., p_r^k, ..., p_r^N\}$ represent the set of N identified flow coordination points. Let $T_r = \{p_r^{r_1} \to p_r^{r_2} \to, ..., \to p_r^{r_m}\}$ represent a sequence of m flow coordination points with r_1 being a numerical index referring to the r_1-th points in the original coordination points set. Therefore, the objective of trajectory registration is to find a representative trajectory T_r^*, which has the largest similarity to the original trajectory T_o, from all the possible T_r. The trajectory similarity in this paper is measured by the $Fre\acute{c}het\ distance$ [25], a measure of similarity between curves by taking into account the location and ordering of the points along the curves. Thus, the objective function of the trajectory registration is formulated as follows:

$$\begin{aligned} obj = \quad & \arg\min_{T_r} \delta_F(T_o, T_r) \\ & T_r = \{p_r^{r_1} \to p_r^{r_2} \to, ..., \to p_r^{r_m}\} \\ & \{p_r^{r_1}, p_r^{r_2}, ..., p_r^{r_m}\} \in FRP \end{aligned} \tag{5}$$

The function $\delta_F(T_o, T_r)$ denotes the $Fre\acute{c}het\ distance$ between T_o and T_r. After trajectory registration, the trajectory T_o will be represented as the sequence $T_r^* = \{p_r^{r_1^*} \to p_r^{r_2^*} \to, ..., p_r^{r_k^*} \to, ..., \to p_r^{r_m^*}\}$. The flight crossing time at point $p_r^{r_k^*}$ is approximated as the flight crossing time at the closest point to $p_r^{r_k^*}$ in T_o.

2) Flow Context Extraction: As mentioned before, the traffic flow context in the airspace during a period is then represented as a "paragraph" whose "sentences" are the sequences of flights (characterized by the flight callsign) in the airspace. Use t_0 and t_1 to denote the start and end times of a period. Let FP_r^k denote the traffic flow at coordination point p_r^k during t_0 and t_1. FP_r^k can be described as a sequence of flights (callsigns) according the time flights passing p_r^k, i.e., FP_r^k : $f_k^1, f_k^2, ..., f_k^i, ..., f_k^{m_k}$ with f_k^i denoting the i-th flight in the sequence and m_k denoting the total number of flights passing p_r^k. Analogizing FP_r^k as a "sentence" depicting the flow context at p_r^k, the combinations of "sentences" during t_0 and t_1, i.e., "$t_0 - t_1$: $f_1^1\ f_1^2\ ...\ f_k^{m_1}$; $f_2^1\ f_2^2\ ...\ f_2^{m_2}$; ...; $f_N^1\ f_N^2\ ...\ f_N^{m_N}$.", constitute a "paragraph" description of the flow context in the entire airspace. The paragraphs for various periods will be used as inputs to the transformer based prediction model, introduced in the next section, to learn the contextual relations between air traffic flow and make predictions about the future air traffic flow.

3) Transformer based Prediction Model: The structure of the transformer based prediction model is shown in Fig. 3. The main components are tokenization, embedding, transformer encoder, and a fully-connected layer to produce the prediction outcome.

a) Tokenization: Given an input sequence, i.e., the traffic flow context including the time and the flight callsign sequences at different flow coordination points, tokenization [26] is applied to convert the input into a list of integers that can be embedded into a vector space. The flight callsigns in the input sequence refer to specific flights in the airspace, which differ from the natural word composition, whose prefixes, suffixes, and infixes can change their inherent meaning. Therefore the tokenization in this study adopts word-tokenization, which splits the data based on natural breaks and meaning, such as time (number), callsigns (words), and sequence separations (delimiters).

b) Embedding: After tokenization, the token embedding layer converts the list of integers, i.e., tokenized flow sequence, into a list of vectors, as is the case in language processing applications in general. For the model to use the order of the elements in the sequence, positional embeddings, which contain information on the relative or absolute position of the elements, are added to the token embeddings as the input to the transformer encoders. The embeddings are trained jointly with the rest of neural network. Back-propagation is carried through all the network layers up to the embeddings that are updated as other parameters. After embedding the elements in the input sequence, each of them flows through the transformer encoders which will be described in the follow section.

c) Transformer encoder: The transformer encoder processes each element in the input sequence and compiles the information it captures into a context tensor, i.e., an array of numbers. The encoder sends the context tensor to the fully-connected layer to project the context to the predicted flows. As shown in Fig. 3, the transformer encoder is composed of a stack of N_e encoder blocks. Each encoder block consists of two sub-layers. The first layer is a multi-head self-attention mechanism, and the second layer is a position-wise fully connected feed-forward network.

The self-attention mechanism allows the encoder to look at other elements in the input sequence when encoding a specific element. The self-attention creates three vectors from each

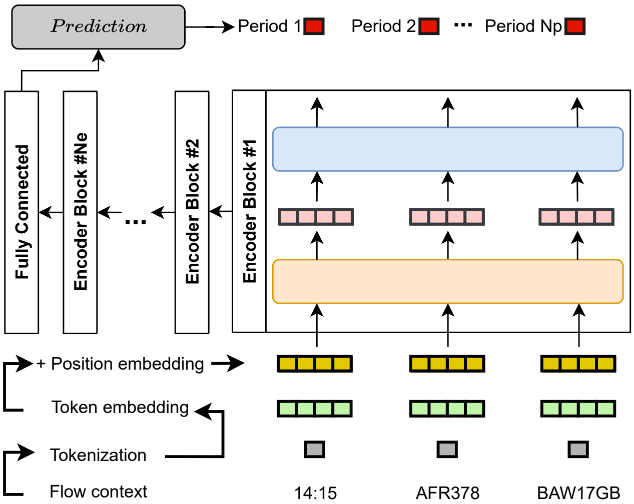


Figure 3: Proposed flow prediction model based on transformer neural networks.

of the encoder's inputs, Query vector, Key vector, and Value vector, by multiplying the embedding by three matrices trained during the training process. The three vectors are used to score the relevance of other elements in the sequence against the specific element. The intuition of self-attention is to keep the values of relevant elements intact and drown out irrelevant elements. Multi-headed attention refined the self-attention layer by adding a mechanism called "multi-headed" attention that runs through an attention mechanism several times in parallel. Intuitively, multi-head attention allows the model to jointly attend to information from different representation subspaces at different positions [27].

The output from the encoder is then forwarded to a fully connected layer to obtain the flow prediction results for different looking-ahead windows, i.e., future period 1, period 2,..., and period N_p.

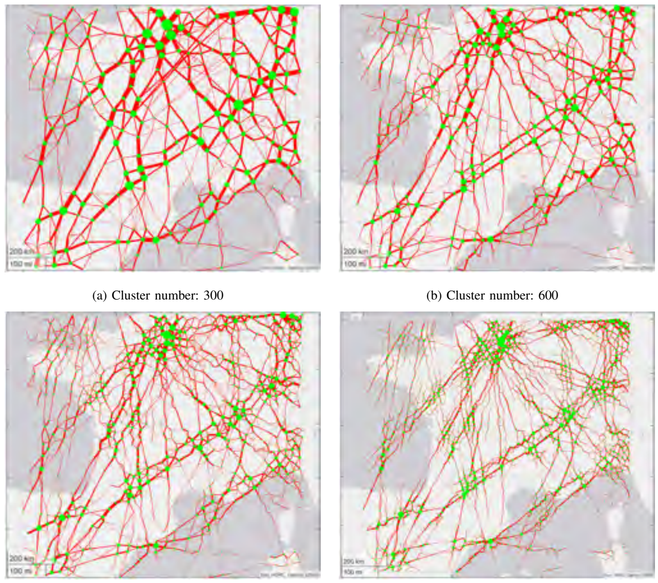
III. Experimental Study

To verify the efficacy of the proposed method, an experimental study has been carried out on the French airspace using one-month ADS-B data from December 1 to December 31, 2019, including a number of 158856 flights. This study focus on the en-route air traffic above 10,000 ft. The prediction target in this experimental study is set as: using the past one-hour traffic information in the airspace to predict the future number of flights that will transit through a flow coordination point in the coming 15 minutes, 30 minutes, 45 minutes, 1 hour, 1.25 hours, 1.5 hours, 1.75 hours, and 2 hours.

A. *Flow Regulation Points Identification*

By finding intersections of flight trajectories in the French airspace and clustering the intersection points to identify the major flow intersections on a daily basis, a graph representation of the daily air traffic flow pattern can be obtained. The nodes in the graph show the major flow intersections, and the edges describe the flow volume between these intersections.

Fig. 4 shows the constructed graphs for the air traffic flow in the French airspace on December 1, 2019. The sub-figures 4a, 4b, 4c, and 4d display the constructed graph when the number of clusters is set as 300, 600, 1000, and 2000 respectively. Each green node in the graphs denotes a flow intersection point, and the node's size is proportional to the number of trajectory intersections in the corresponding cluster. Each red line represents an edge in the graph, while the line thickness is proportional to the number of flights traveled between the edge's start/end nodes. It can be observed from Fig. 4 that, on the one hand, when the number of clusters is small, the graph depicts air traffic flow features between city pairs, as most of the identified flow intersections are featured by the geographical locations of the metropolis and small cities. Such a graph may fail to represent important intersecting flows in the en-route phase of flights. On the other hand, when the number of clusters is large, the graph structure becomes complex as it may be affected by trivial interactions between the air traffic flow.



(a) Cluster number: 300

(b) Cluster number: 600

(c) Cluster number: 1000

(d) Cluster number: 2000

Figure 4: The constructed graphs for the air traffic flow in the French airspace on Dec 1, 2019 with different nodes.

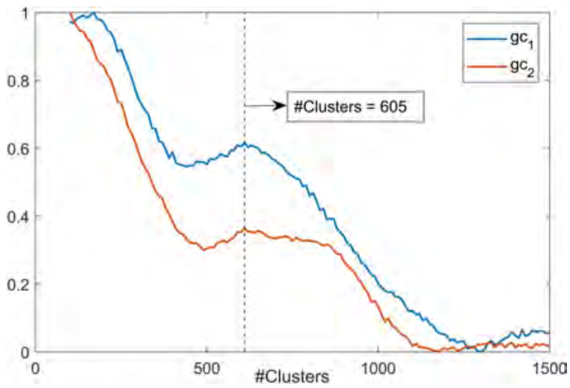


Figure 5: cluster Number determination.

In accordance with a set of different cluster numbers ranging from 100 to 1500, this paper calculated the geographical consistency gc_1 and the structural consistency gc_2. Fig. 5 shows the changes of gc_1 and gc_2 which are scaled between

0 an 1 versus the varying number of clusters. A “saddle point” is observed for gc_1 and gc_2 around the cluster number 605. Therefore, this paper takes the value 605 as the number of clusters to be identified by the hierarchical clustering algorithm. The centers of identified clusters are determined as the flow coordination points.

Figure 6: Comparison of the identified flow coordination points for two different days: 1st (red dots) and 2nd (green dots), December 2019.

Fig. 6 shows the identified flow coordination points for December 1 (red dots) and 2 (green dots), 2019. The sizes of the dots are proportional to the traffic volume passing the corresponding coordination points. It can be observed that the identified coordination points show consistent patterns in geographical distributions and traffic volumes, although there are some variations in their geographical locations due to the differences in daily air traffic flow organizations.

B. Air Traffic Flow Prediction

The last one-hour flight sequences at all flow coordination points are extracted from the air traffic data and fed to the transformer neural network flow prediction model to predict the traffic flow for up to two hours in the future. The transformer encoder module in this study adopts a widely used structure that stacks six encoder blocks. The mean square error (MSE), a commonly used metric for the evaluation of the performance of regression algorithms, between the predicted and true values is used to compute the model’s loss function. The batch size is sixteen during the training phase, and the learning rate is 0.00002. The training, validation, and test data consist of 60%, 20%, and 20% of the whole data set, respectively. Concretely, eighteen days of flow data are used for model training, the following six days of data for model validation, and the last six days of data for testing the model.

C. Result Analysis

Given that there are over 600 flow coordination points identified, this section uses the coordination point, which is over the Paris area control center and handles the highest traffic volume, as an example to present the prediction result on the test dataset.

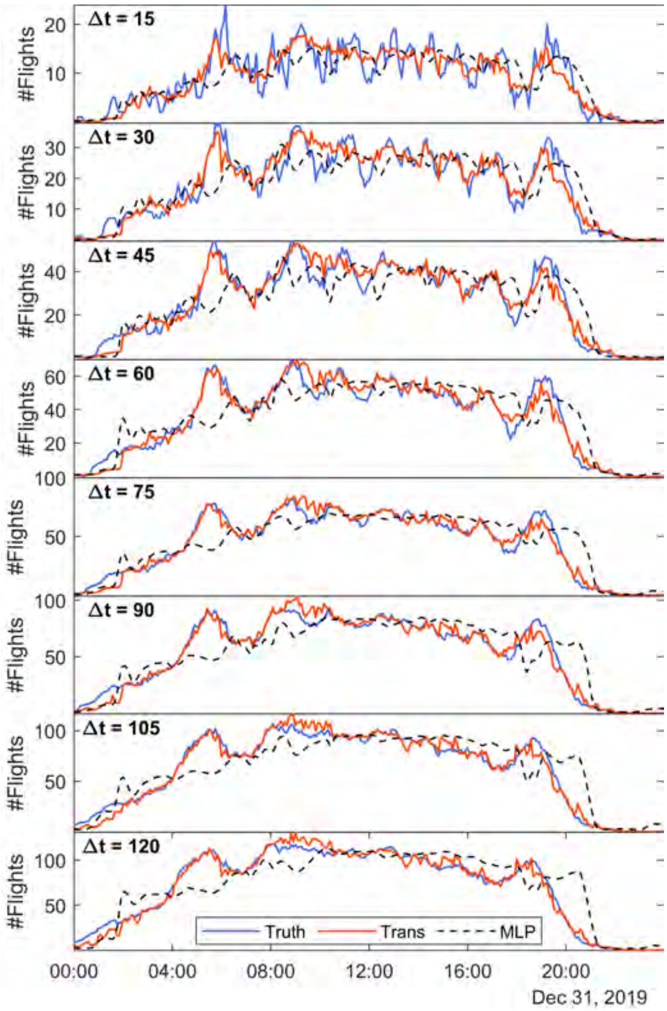


Figure 7: Flow prediction results for Dec 30, 2019 over different time horizons as shown by the y-axis.

Fig. 7 presents the traffic flow prediction result from 00:00 to 23:59 on Dec 30, 2019. The eight panels plot the prediction results for the future 15 minutes, 30 minutes, 45 minutes, 1 hour, 1.25 hours, 1.5 hours, 1.75 hours, and 2 hours respectively. The blue solid shows the actual number of flights passing the coordination point, while the red line shows the predicted value with the proposed transformer neural network prediction model. We can observe from Fig. 7 that the proposed method sustainably gives forecasts in close proximity to the actual flow value as the prediction window increases from 15 minutes to 2 hours. Furthermore, the model performs better in approximating the true flow value under a larger prediction window, as the prediction curve is more laminated on the actual curve for the 2-hour prediction window than the 15-minute window.

The proposed model has been compared with the canonical Multi-layer Perceptron (MLP) neural network. Blue dashes in Fig. 7 show the predicted values using MLP neural network. We can observe that the proposed model gives better predictions for all prediction windows.

Table I shows the quantified prediction performance of the proposed method, including four metrics: Mean Absolute

TABLE I: Prediction performance of the proposed method under different prediction windows in terms of MAE, MSE, MAPE, and R^2.

	MAE	MSE	MAPE	R^2
15 minutes	1.683	6.082	0.274	0.909
30 minutes	2.339	12.184	0.217	0.951
45 minutes	2.755	17.092	0.179	0.968
1 hour	3.102	21.819	0.155	0.977
1.25 hours	3.378	28.561	0.139	0.980
1.5 hours	3.794	34.200	0.129	0.983
1.75 hours	3.958	34.934	0.122	0.988
2 hours	4.129	35.837	0.117	0.990

Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2). It can be seen from Table I that the MAE, MSE, and R^2 values of the proposed method increase with the prediction window, while the MAPE value decreases with the prediction window.

Fig. 8 visualizes the changes in the prediction performance of the proposed method versus MLP for varying prediction windows. It can be observed that the proposed method outperforms MLP in all of the metrics. The MSE, MAE, and MAPE values of the proposed method are lower than MLP, indicating more minor prediction errors. At the same time, the R^2 value is higher than MLP, showing a better approximation to the actual flow curve. The growth in the MSE and MAE values shows that the gap between the predicted value and the true value increases as the prediction window increases. The reason for this may be the increase in the number of flights passing the coordination points in a 2-hour duration compared to 15-minute duration. As can be seen from Fig. 7, the maximum number of flights in a 15-minute duration is 24, while in a 2-hour duration is 120. In contrast, the decreasing MAPE value, as a percentage, and the increasing R^2, as a regression performance metric, indicate that the model's prediction accuracy improves as the prediction window increases. The reason may be that the number of flights transiting in the air space during the next 2 hours has less variation compared to a future 15-minute duration, which is more stable and predictable.

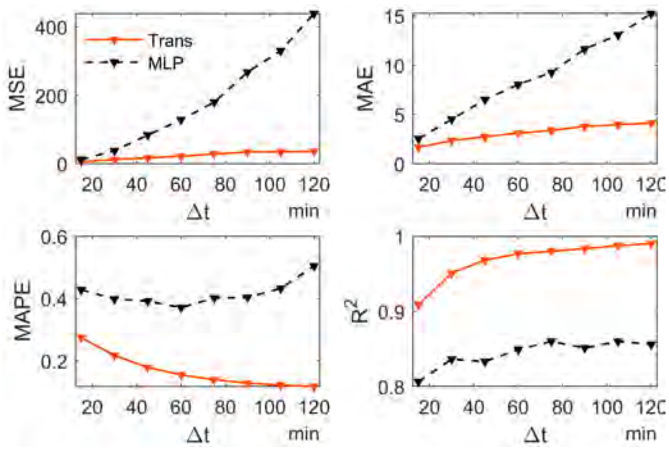


Figure 8: Prediction performance (MAE, MSE, MAPE, and R^2) of the proposed method versus MLP as the prediction window increases.

IV. Conclusions

Aimed at contributing to the future flow-centric ATC paradigm, this paper proposed a transformer neural network approach for air traffic flow prediction at flow coordination points. Firstly, the flow coordination points were identified through hierarchical clustering of the flight trajectory intersections and the number of clusters was identified by analysing the consistency of daily traffic flow patterns modelled as graphs. Secondly, this paper described the air traffic flow as a "paragraph" of the flights sequences passing the coordination points and a transformer neural network model was adopted to learn from the contextual descriptions of air traffic and predict future air traffic flow at the coordination points. An experimental study on French airspace using the flight data from December 1-31, 2019. Results showed that the proposed prediction model approximated the true flow values with a R^2 between 0.909 to 0.99 and a MAPE varying from 27.4% to 11.7% for a 15-minute to 2-hour prediction window. The sustainability of the prediction performance under the increasing prediction window demonstrated the potential of the proposed model for longer-term flow prediction.

The proposed air traffic flow prediction method provides a graph-based approach to modelling air traffic flow patterns under the flow-centric paradigm. Predicting the future flow on the coordination points can inform flow-centric operations of the anticipated intersecting flows so that flow coordination can be applied in advance to avoid complex and conflicting situations involving a large number of flights.

This paper's prediction of air traffic flow only considered flight numbers as the prediction target. However, the mixture of the intersecting flows can also affect the complexity and difficulty in flow management. A flow scenario with more crossing traffic is more challenging than a scenario with more parallel traffic, although the same number of flights are involved. Therefore, in the future, this traffic flow prediction method can be improved by considering the prediction of flow mixture to deliver a more meaningful prediction result.

Acknowledgment

This research is supported by the National Research Foundation, Singapore, and the Civil Aviation Authority of Singapore, under the Aviation Transformation Programme. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore and the Civil Aviation Authority of Singapore.

References

[1] D. Gianazza, "Learning air traffic controller workload from past sector operations," in *12th USA/Europe Air Traffic Management R&D Seminar, Seattle, United States*, 26-30 June, 2017.

[2] B. Korn, V. Mollwitz, T. Finck, and C. Edinger, "Validating sectorless atm in the hungarian airspace: Results of human in the loop simulations," in *2020 Integrated Communications Navigation and Surveillance Conference (ICNS), Herndon, VA, USA*. IEEE, 8-10 September, 2020, pp. 3B1–1.

[3] E. Zeki, P. L. de Matos, K. Purves, and M. Gibellini, "Business models for flight-centric air traffic control," *Competition and Regulation in Network Industries*, vol. 20, no. 4, pp. 319–332, 2019.

[4] I. Gerdes, A. Temme, and M. Schultz, "From free-route air traffic to an adapted dynamic main-flow system," *Transportation Research Part C: Emerging Technologies*, vol. 115, p. 102633, 2020.

[5] Y. Lin, J.-w. Zhang, and H. Liu, "Deep learning based short-term air traffic flow prediction considering temporal–spatial correlation," *Aerospace Science and Technology*, vol. 93, p. 105113, 2019.

[6] B. Sridhar, T. Soni, K. Sheth, and G. Chatterji, "Aggregate flow model for air-traffic management," *Journal of Guidance, Control, and Dynamics*, vol. 29, no. 4, pp. 992–997, 2006.

[7] Y. Zhang, R. Su, G. G. N. Sandamali, Y. Zhang, C. G. Cassandras, and L. Xie, "A hierarchical heuristic approach for solving air traffic scheduling and routing problem with a novel air traffic model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 9, pp. 3421–3434, 2018.

[8] E. C. Fernández, J. M. Cordero, G. Vouros, N. Pelekis, T. Kravaris, H. Georgiou, G. Fuchs, N. Andrienko, G. Andrienko, E. Casado *et al.*, "Dart: a machine-learning approach to trajectory prediction and demand-capacity balancing," in *SESAR Innovation Days, Belgrade, Serbia*, 28–30 November, 2017.

[9] I. Lymperopoulos, J. Lygeros, and A. Lecchini, "Model based aircraft trajectory prediction during takeoff," in *AIAA Guidance, Navigation, and Control Conference and Exhibit, Keystone, Colorado*, 21-24 August, 2006.

[10] P. Wei, Y. Cao, and D. Sun, "Total unimodularity and decomposition method for large-scale air traffic cell transmission model," *Transportation research part B: Methodological*, vol. 53, pp. 1–16, 2013.

[11] B. Sridhar, S. R. Grabbe, and A. Mukherjee, "Modeling and optimization in traffic flow management," *Proceedings of the IEEE*, vol. 96, no. 12, pp. 2060–2080, 2008.

[12] D. Chen, M. Hu, H. Zhang, J. Yin, and K. Han, "A network based dynamic air traffic flow model for en route airspace system traffic flow optimization," *Transportation Research Part E: Logistics and Transportation Review*, vol. 106, pp. 1–19, 2017.

[13] E. Gilbo and S. Smith, "A new model to improve aggregate air traffic demand predictions," in *AIAA Guidance, Navigation and Control Conference and Exhibit, Hilton Head, South Carolina*, 20-23 August, 2007, p. 6450.

[14] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 1–17, 2017.

[15] Z. Zhang, A. Zhang, C. Sun, S. Xiang, J. Guan, and X. Huang, "Research on air traffic flow forecast based on elm non-iterative algorithm," *Mobile Networks and Applications*, vol. 26, no. 1, pp. 425–439, 2021.

[16] G. Gui, Z. Zhou, J. Wang, F. Liu, and J. Sun, "Machine learning aided air traffic flow analysis based on aviation big data," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 4817–4826, 2020.

[17] C. Ma, S. Alam, Q. Cai, and D. Delahaye, "Sector entry flow prediction based on graph convolutional networks," in *International Conference on Research in Air Transportation, Tampa, Florida, USA*, 19-23 June, 2022.

[18] R. Dalmau-Codina, B. Genestier, C. Anoraud, P. Choroba, and D. Smith, "A machine learning approach to predict the evolution of air traffic flow management delay," in *2021 ATM Seminar virtual event*, 20-24 Sep 2021.

[19] C. Ma, Q. Cai, S. Alam, B. Sridhar, and V. N. Duong, "Airway network management using braess's paradox," *Transportation Research Part C: Emerging Technologies*, vol. 105, pp. 565–579, 2019.

[20] Q. Cai and J. Liu, "Hierarchical clustering of bipartite networks based on multiobjective optimization," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 421–434, 2018.

[21] F. Seabra, A. M. Valente, L. R. Silva, R. Assis, C. Acordi, A. F. Marcon, and M. M. Bauer, "Determinants of brazilian international flights: The role of hub-and-spoke and infrastructure variables," *Journal of Air Transport Management*, vol. 89, p. 101866, 2020.

[22] C. Ma, S. Alam, Q. Cai, and D. Delahaye, "Critical links detection in spatial-temporal airway networks using complex network theories," *IEEE Access*, vol. 10, pp. 27 925–27 944, 2022.

[23] M. C. R. Murça, R. J. Hansman, L. Li, and P. Ren, "Flight trajectory data analytics for characterization of air traffic flows: A comparative analysis of terminal area operations between new york, hong kong and sao paulo," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 324–347, 2018.

[24] M. S. Islam and E. Hossain, "Foreign exchange currency rate prediction using a gru-lstm hybrid network," *Soft Computing Letters*, p. 100009, 2020.

[25] C. Lyu, X. Wu, Y. Liu, and Z. Liu, "A partial-fréchet-distance-based framework for bus route identification," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[26] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral–spatial feature tokenization transformer for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

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