

# Major Air Traffic Flow Identification with Fractal-Based Graph Simplification

Chunhao Ma\*, Aditya Ishaan Talamudipi\*, Huijuan Yang<sup>†</sup>, Thinh Hoang-Dinh<sup>†</sup>, Sameer Alam\*  
, Daniel Delahaye<sup>†</sup>,

\*Air Traffic Management Research Institute, Nanyang Technological University, Singapore

<sup>†</sup>OPTIM Lab, Ecole Nationale de l'Aviation Civile, Toulouse, France

**Abstract**—Major air traffic flows are concentrated streams of flights that follow similar trajectories between specific geographical regions or airport pairs, and they play a key role in tasks such as workload evaluation and demand–capacity balancing. However, real-world flows exhibit strong interconnectivity, including merging, overlapping, and splitting, which makes flow boundaries difficult to distinguish. This complexity limits traditional methods based on trajectory similarity or predefined origin–destination pairs, which rely on a top-down perspective to directly extract flow segments from the global traffic structure. This paper reconceptualizes major flow identification as a bottom-up process. Major flows are identified by deriving and consolidating local flow structures, thereby capturing flow interconnectivity to offer a more interpretable and robust representation. First, flow-shaping areas are detected through density-based clustering of trajectory waypoints, capturing potential origins, terminations, and transit points. Second, local flow trees are constructed through fractal-based simplifications to preserve the essential flow hierarchy while smoothing small-scale deviations. Third, a stability metric is introduced to extract major flows that exhibit both high traffic volume and spatial persistence. Finally, an optimization framework consolidates these local major flows into a coherent set of global major flows across the airspace. The methodology is tested on northwest–southwest traffic in European airspace with flight plan data on 14th July 2023. Results show that the majority of flights can be organized into major flows, yielding stable representations of dominant traffic corridors. In addition, the framework highlights alternative routing options and abnormal trajectory patterns, supporting applications in traffic flow management, re-routing, and anomaly detection.

**Index Terms**—Air traffic flow; fractal simplification; major flow identification; network science.

## I. INTRODUCTION

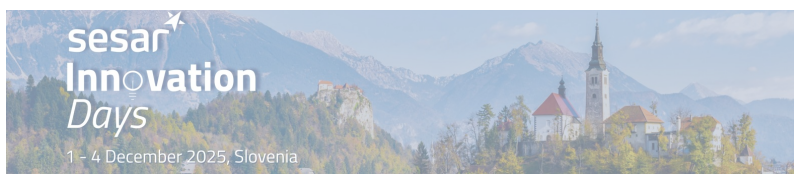
Major air traffic flows prevalently refer to concentrated streams of flights that follow similar trajectories between specific geographical regions or airport pairs [1], [2]. Identifying major flows is a foundational enabler for various downstream tasks such as anomaly detection, trajectory prediction, and demand–capacity balancing [3]. For instance, identifying major flows improves the accuracy of flow prediction models by constraining future trajectory predictions within learned flow corridors [4]. Establishing major flow patterns supports anomaly detection by distinguishing recurrent traffic structures from emerging or irregular deviations, such as those caused by weather disruptions, special events, or operational constraints [5]. In free-route airspace, where fixed airways are absent, the identification of major flows further enables the dynamic

construction of virtual routes and supports adaptive airspace reconfiguration in response to evolving traffic demand [6].

Given the fundamental role of major flow identification in efficient air traffic management, it has received considerable attention in the literature. Current approaches to identifying major air traffic flows generally fall into two categories: (1) grouping flights based on shared origin–destination (OD) pairs within a region, and (2) clustering flight trajectories based on spatial or spatiotemporal similarity to reveal commonly traveled routes. Grouping flights that share a common OD pair as an air traffic flow is effective and straightforward to implement. In [7], major flows within a sector are identified using density-based clustering of flight entry and exit points. Each pair of clustered entry–exit points defines a major flow, with all flights traversing between them grouped accordingly. In many air traffic flow prediction models, flows are commonly represented as flights traveling between spatial regions such as waypoints or sectors, serving as key inputs for short-term traffic demand forecasting [4], [8]. Alternatively, some studies define major flows based on origin–destination airport pairs. For instance, in [9] and [10], flows are modeled as sets of flights operating between two airports, with the diversity of routes within each flow providing flexibility for Air Traffic Flow Management (ATFM) strategies, including re-routing and congestion mitigation.

However, identifying major flows purely based on OD pairs is inherently limited, as it reflects predefined endpoints rather than the underlying structure of the traffic flow. As illustrated in the left panel of Fig. 1, flights with different destination airports can fundamentally belong to the same flow, sharing substantial portions of their trajectories. Categorizing flows solely by OD fails to capture critical flow characteristics such as interactions, merges, splits, and commonly shared segments. This limitation reduces the effectiveness and flexibility of OD-based flow definitions in supporting network-wide flow optimization and dynamic airspace management.

The second major category of approaches for identifying air traffic flows involves clustering flight trajectories based on spatial or spatiotemporal similarity. A common practice in current air traffic management is manual flow selection, guided by prior assumptions and expert judgment. While straightforward, this approach lacks the ability to dynamically detect and update flow structures in response to evolving traffic patterns. As such, it is limited in its capacity to support proactive



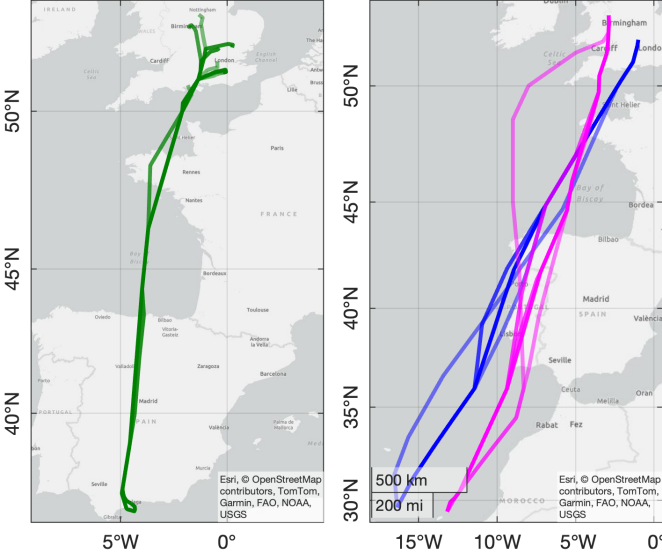


Figure 1. Examples of major flows. Left panel: flights from different origin-destination pairs can belong to the same flow. Right panel: flights with similar trajectories may not belong to the same flow, as they can represent intersecting flows.

and adaptive traffic management. Alternatively, data-driven trajectory clustering offers a more flexible solution to respond to changing traffic conditions by identifying flows from the evolving traffic data. There are trajectory clustering models which directly cluster the trajectory points using K-means or DBSCAN to identify major flows [11], [12], but they struggle when trajectories are long, noisy, or intersect frequently. Major flows often overlap or cross, making cluster boundaries ambiguous. Therefore, instead of clustering individual trajectory points, research has increasingly focused on clustering entire trajectories as holistic objects. The two mainstream approaches are: (1) computing trajectory distances, and (2) learning trajectory embeddings, both of which precede the application of clustering algorithms. In the distance-based approach, pairwise dissimilarities between trajectories are quantified using metrics such as Euclidean distance, Dynamic Time Warping (DTW) [13], or Fréchet distance [14]. The resulting distance matrix is then used with clustering algorithms that operate directly on pairwise distances—such as DBSCAN or Spectral Clustering—to identify trajectory clusters. For instance, [15] employed the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm to cluster entire trajectories. Trajectory similarity was evaluated using two different metrics—Euclidean Distance and Symmetrized Segment-Path Distance (SSPD) [16]—which led to varying clustering outcomes. Similarly, in [17], each trajectory was represented as a high-dimensional vector of 2D spatial positions sampled at uniform time intervals. Euclidean distance was used to compute similarity, and DBSCAN was applied for clustering. A comparable approach was adopted in [18] and [19] for Terminal Maneuvering Airspace (TMA) traffic, where similarity was computed using a Weighted Euclidean Distance or Fréchet distance.

Distance-based methods require computing a full pairwise distance matrix, which scales poorly with data size. These methods are sensitive to the choice of distance metric, which may only captures low-level geometric or temporal alignment. Trajectory embedding scales more efficiently to large datasets, as it avoids computing costly pairwise distance matrices. By representing each trajectory as a fixed-length vector, it enables the use of a wide range of machine learning algorithms, including K-means, GMM, and neural networks. In [5], an air traffic flow pattern identification framework was proposed for analysing the weather impact on the dominant flow patterns in the TMA. A deep autoencoder network was trained to reconstruct flight trajectories and extract low-dimensional representations of input trajectories. The K-means algorithm was then applied to these representations to obtain clusters of major TMA flows. Similarly, in [3] and [20], flight trajectories were featured through embeddings in an autoencoder network and organized into clusters of major flows through DBSCAN or Gaussian mixture models.

In summary, approaches in the literature adopt a top-down perspective that operates from the global structure of air traffic to extract the underlying major flow segments. However, real-world air traffic flows exhibit strong interconnectivity, which makes flow boundaries difficult to delineate. Therefore, this paper proposes a bottom-up approach: rather than extracting major flows directly from the global traffic structure, the method first identifies local flow structures and then integrates them into global flows, thereby reducing complexity while preserving the hierarchical and interconnected nature of air traffic. Fractal simplification is applied to preserve the essential flow hierarchy while smoothing small-scale deviations, representing local air traffic as a tree whose branches correspond to different flow paths. Major flows are then identified using a stability metric that balances traffic volume and persistence, ensuring that only flows that are both concentrated and persistent are retained. Finally, an optimization framework consolidates these local flows into a coherent set of global major flows across the airspace.

## II. METHODOLOGY

The air traffic flow network is shaped by the union of flows emanating from different locations, as illustrated in Fig. 2. To address the limitations of top-down approaches that operate from the global traffic structure, the proposed method reconceptualizes major flow identification as a bottom-up process. It decomposes the air traffic flow network into local flow structures and then consolidates them into global flows, thereby explicitly capturing flow interconnectivity that traditional methods are unable to represent effectively. The process involves four main steps as shown in Fig. 3: detecting flow-shaping areas through clustering of trajectory waypoints, constructing local flow trees with fractal simplification, identifying representative major flows using a stability metric that balances persistence and traffic volume, and consolidating these local major flows into a coherent set of global major

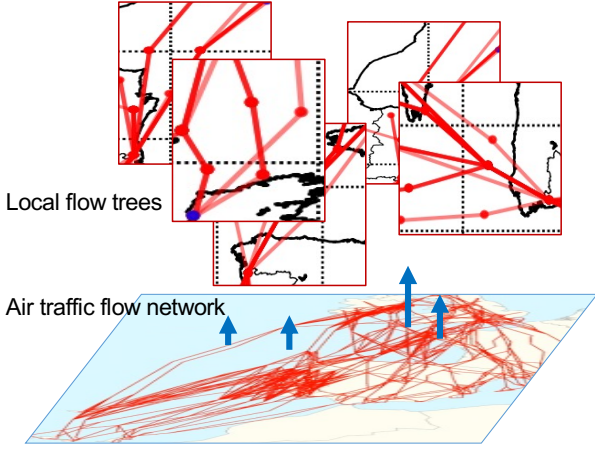


Figure 2. The air traffic flow network is shaped by the union of the hierarchical flow trees emanating from different locations.

flows. Each of these steps is described in detail in the following sections.

#### A. Data Preprocessing–Flow Shaping Areas Identification

Major flows in airspace are manifested as concentrated streams of flights, giving rise to areas of elevated traffic density along their transit paths. This section introduces a preprocessing methodology that employs density-based clustering to identify flow-shaping areas—potential origins, terminations, and transitions of major flows—which provide the foundation for downstream flow identification.

Aircraft trajectories tend to be arbitrarily shaped curves described by a dense sequence of waypoints. While this granularity provides fine-scale detail, it also increases representational complexity and hinders the ability of machine learning algorithms to detect macro-level traffic flow patterns. Denote the 4D trajectory of the flight  $f_i$  as  $(p_{i,j}, \phi_{i,j}, \theta_{i,j}, h_{i,j}, t_{i,j})$ , where  $p_{i,j}$  represents the  $j$ -th waypoint, and  $\phi_{i,j}, \theta_{i,j}, h_{i,j}, t_{i,j}$  denote its latitude, longitude, altitude, and time, respectively. To preserve the essential structural characteristics of the trajectories and reduce data complexity, a trajectory simplification step is first applied to extract the key waypoints shaping the flight trajectories. Specifically, the *Ramer–Douglas–Peucker (RDP) algorithm* is employed, which decimates a curve composed of line segments into a geometrically similar curve with fewer points [21]. For flight  $f_i$  with trajectory  $\mathcal{P}_i = \{p_{i,1}, \dots, p_{i,n_i}\}$ , the RDP with tolerance  $\varepsilon$  returns  $\tilde{\mathcal{P}}_i = \{p_{i,j} : j \in \mathcal{I}_i^\varepsilon\}$ .  $\mathcal{I}_i^\varepsilon \subseteq \{1, \dots, n_i\}$  denotes the subset of waypoint indices selected by RDP that always contains the endpoints  $\{1, n_i\}$  and ensures that the deviation between  $\tilde{\mathcal{F}}_i$  and  $\mathcal{F}_i$  does not exceed  $\varepsilon$ .

Air traffic flows follow a hierarchical organization connecting regional feeders to international hubs [22], [23]. Therefore, hierarchical clustering [24] is adopted to discover the flow shaping areas by aggregating the geographical locations of the key shaping waypoints across flights:

$$S = \bigcup_i \{x_{i,j} : j \in \mathcal{I}_i^\varepsilon\} \quad (1)$$

and applying HDBSCAN with parameters  $(d, m_s, m_c)$  to  $S$  to yields clusters:

$$\text{HDBSCAN}(S; d, m_s, m_c) \rightarrow \{C_r\}_{r=1}^R \quad (2)$$

where  $\{C_r\}$  are the identified clusters (flow-shaping areas) and  $x_{i,j} = (\phi_{i,j}, \theta_{i,j}) \in \mathbb{R}^2$  represents the geographical locations of waypoint  $p_{i,j}$ .

The three parameters  $d$ ,  $m_s$ , and  $m_c$  denote the cluster selection epsilon, the minimum cluster size, and the minimum samples, respectively. These parameters are selected through a grid-search procedure by identifying the transition point at which the number of datapoints labeled as noise begins to increase sharply, because such transition typically indicates that the clustering structure becomes unstable and overly fragmented. The distances among the waypoints during clustering are calculated based on the great circle distance.

#### B. Local Flow Tree Modeling

With the flow-shaping areas identified, this section describes the methodology for characterizing the air traffic flows emanating from each area through a tree-structured representation. The objective of this step is to smooth small-scale trajectory deviations while preserving the underlying core flow structures so that the major flow identification is not affected by local trajectory variations.

1) *Traffic Flow Graph Modeling*: For a flight  $f_i$  with original trajectory  $\mathcal{P}_i = \{p_{i,1}, \dots, p_{i,n_i}\}$ , each waypoint  $p_{i,j}$  is assigned to a cluster index  $r$  if  $p_{i,j} \in C_r$ , or 0 if it does not belong to any cluster. The trajectory of  $f_i$  is then represented by the ordered sequence of cluster indices:

$$\alpha_i = \text{compress}(\{r : p_{i,j} \in C_r, j = 1, \dots, n_i\}), \quad (3)$$

where  $\text{compress}(\cdot)$  removes consecutive duplicates and zeros. In this way, each flight is encoded as a sequence of flow-shaping areas.

The air traffic emanating from the flow-shaping area  $C_r$  can then be represented by a weighted directed graph  $G_r = (V_r, E_r, w)$ , where the node set  $V_r = C_1, \dots, C_R$  corresponds to the flow-shaping areas and an edge  $(C_r, C_s) \in E_r$  is created whenever at least one flight sequence  $\alpha_i$  originating from  $C_r$  contains the transition  $C_r \rightarrow C_s$ . The weight  $w(C_r, C_s)$  is defined as the number of such flights.

The left panel of Fig. 4 shows one example of the constructed traffic flow graph from one flow shaping area  $C_r$  indicated by the red circle. The red and blue segments represent the flight trajectories described by the sequences of flow-shaping areas (depicted by the centroid of the clusters). The red segments highlight the trajectory segments in the radius of  $r_s$  to  $C_r$ .

2) *Flow Tree Modeling – Fractal-Inspired Simplification*: Identifying major flows requires distinguishing global traffic structures from local trajectory variations. However, the appropriate level of detail depends on the spatial scale of analysis: when examining major flows in a small region, local flow features should be preserved, whereas at a broader airspace level, fine-scale deviations can obscure the underlying flow

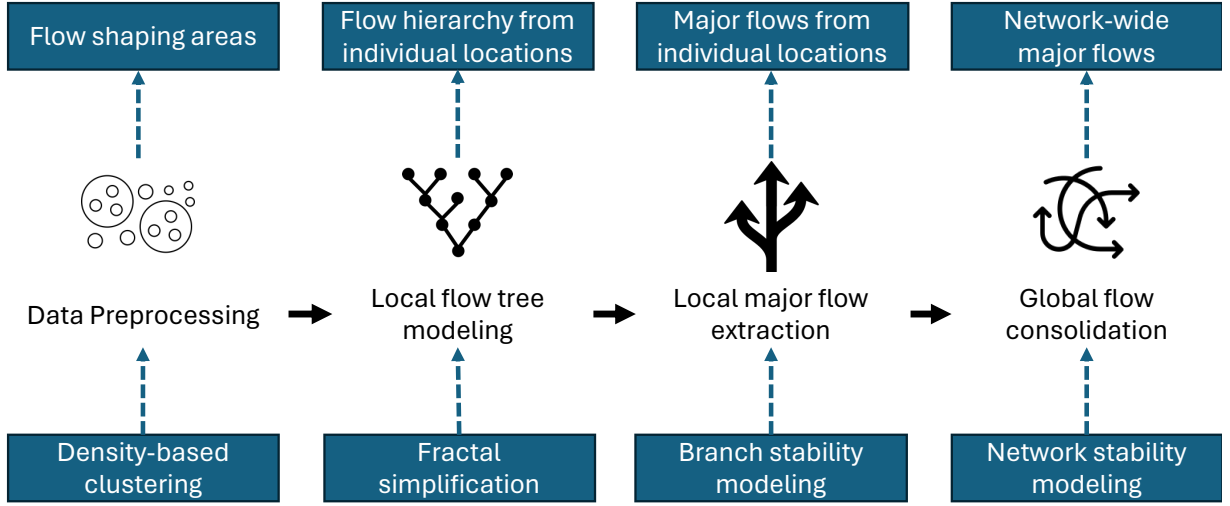


Figure 3. Concept diagram of the proposed methodology for major flow identification.

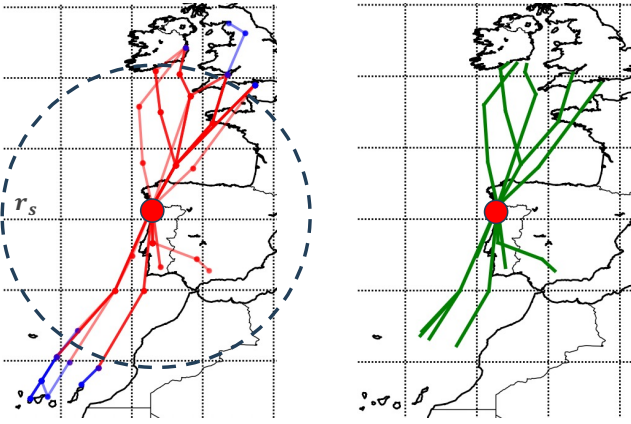


Figure 4. Local flow tree modeling. Left: traffic graph constructed from flow-shaping area  $C_r$  (red circle), with red and blue segments representing trajectories; red segments within radius  $r_s$  are used for fractal simplification. Right: resulting flow tree  $\mathcal{T}$  representing the core flow hierarchy at  $C_r$ , where branching denotes flow splits into subflows.

organization. This creates the need for controlled granularity in flow modeling, enabling small-scale variations to be selectively suppressed or retained depending on the analysis objective.

The traffic graph in Fig. 4 shows that air traffic is organized hierarchically: traffic flows branch into subflows and finer streams. This structure resembles the multi-scale, fractal-like organization observed in river networks and transportation systems, where large trunks subdivide into progressively smaller branches while maintaining a coherent structural pattern [25]. Although air traffic does not demonstrate formal fractal behavior in the strict mathematical sense [26], the hierarchical and multi-scale nature of these patterns provides a natural motivation for a fractal-inspired simplification. For example, in river networks, small tributaries are omitted when observing the system at a global scale but become essential when analyzed locally; a similar multi-scale treatment is desirable

for air traffic flows.

Building on this idea, a tree-structured representation of the air traffic flows originating from  $C_r$  is derived from the weighted graph  $G_r$  by preserving the shortest-path distances from the root node  $C_r$  to all leaf nodes within a radius  $r_s$ . The fractal-inspired parameter  $r_s$  governs the granularity of the simplification: smaller values reserve finer-scale flow organization around  $C_r$ , whereas larger values reveal broader, more global flow structures.

The simplification objective is formulated as identifying a subgraph  $\mathcal{T} \subseteq G_r$  that minimizes the total root-to-leaf path length within the radius  $r_s$  of  $C_r$ :

$$\min_{\mathcal{T} \subseteq G_r} \sum_{v \in \text{Leaves}(\mathcal{T}, r_s)} \text{Length}_{\mathcal{T}}(C_r, v), \quad (4)$$

where  $\text{Leaves}(\mathcal{T}, r_s)$  denotes the set of leaf nodes located within the radius  $r_s$  around  $C_r$ , and  $\text{Length}_{\mathcal{T}}(C_r, v)$  is the path length from  $C_r$  to  $v$  in  $\mathcal{T}$ . The resulting  $\mathcal{T}$  encodes the hierarchical organization of flows emanating from  $C_r$ . As shown in Fig. 4, the traffic graph within the radius  $r_s$  of  $C_r$  (highlighted by the red edges in the left panel) is simplified by preserving the shortest path distances from  $C_r$  to the leaf nodes. The right panel illustrates the resulting flow tree  $\mathcal{T}$  obtained through fractal simplification, where the root corresponds to the origin, internal branches represent flow splits, and leaves denote flow terminations.

### C. Local Major Flow Extraction

After constructing the local flow tree at  $C_r$ , the next step is to identify the branches that represent major traffic flows. Major flows should exhibit two key properties: (i) a concentrated number of flights, ensuring sufficient traffic volume, and (ii) a meaningful flow span, reflecting their role as flow corridors connecting different geographical regions. As illustrated in Fig. 5, the red circle denotes the root node  $C_r$ . Suppose that each flight is associated with a single air traffic flow. If the branch from the root to the pink node is

selected as the major flow, it captures a larger number of flights. Alternatively, if the two branches extending from the root to the orange nodes are identified as major flows, they together cover a longer portion of the flow hierarchy than the pink-branch flow. This raises the key question: which branch selection more appropriately represents the major air traffic flows emanating from  $C_r$ ? To address this, we propose the metric of flow stability that balances traffic volume and flow span to identify the most representative major air traffic flows.

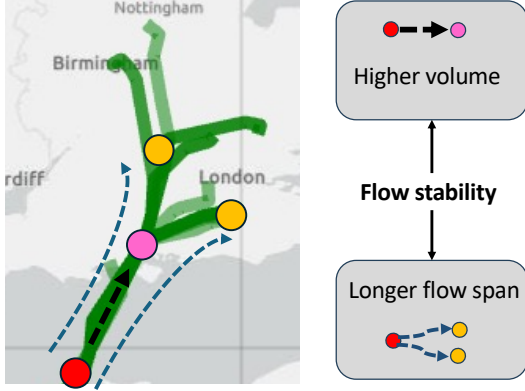


Figure 5. Illustration of branch selection in the local flow tree at  $C_r$ . The red circle marks the root node. The branch to the pink node captures more flights, whereas the two branches to the orange nodes span a longer portion of the flow hierarchy. This highlights the trade-off between traffic volume and flow span (length) in identifying major flows.

**Flow Stability:** Let  $\lambda$  denote the flow span parameter. A higher  $\lambda$  corresponds to a more persistent flow, while a lower  $\lambda$  indicates a less persistent flow. Formally, let  $\lambda(v)$  denote the path length from a node  $v$  in the flow tree  $\mathcal{T}$  to the root node  $C_r$ :

$$\lambda(v) = \text{Length}_{\mathcal{T}}(C_r, v), \quad (5)$$

A flow is considered to split into new flows at a branching node  $v$  of the flow tree when two or more child flows are formed after the split. Each resulting flow  $F$  is regarded as a valid new flow if it satisfies:  $|F| \geq N_{\min}$ , ensuring sufficient traffic volume. Branches that fail to meet the condition are classified as noise. For a flow  $F$ , let  $\lambda_b(F)$  be the  $\lambda$  value of node  $v$  when  $F$  forms after a split at  $v$ . For a flight  $f$ , let  $\lambda_d(f)$  be the  $\lambda$  value when flight  $f \in F$  leaves the flow (due to a split into new flows) or becomes noise. The stability of a flow  $F$  is defined as:

$$\text{Stability}(F) = \sum_{f \in F} (\lambda_d(f) - \lambda_b(F)), \quad (6)$$

which quantifies both the persistence of  $F$  across the flow hierarchy (through the span  $\lambda_d(f) - \lambda_b(F)$ ) and its traffic volume through the sum over the number of flights in  $F$ .

**Optimization Objective:** The objective is to select a subset of flows  $\mathcal{F} \subseteq \mathcal{H}$  that maximizes the total stability:

$$\max_{\mathcal{F} \subseteq \mathcal{H}} \sum_{F \in \mathcal{F}} \text{Stability}(F), \quad (7)$$

subject to the following constraints:

$$F_i \cap F_j = \emptyset, \quad \forall F_i \neq F_j, F_i, F_j \in \mathcal{F}, \quad (8)$$

$$\lambda_d(F_i) > \lambda_{\min}, \quad \forall F_i \in \mathcal{F}, \quad (9)$$

This optimization is subject to two conditions. First, flows must be non-overlapping, meaning that no two selected flows share the same flights. Second, each selected flow must have a minimum span  $\lambda_{\min}$  to ensure persistence and avoid selecting short-lived, unstable flows. In this way, the representative major flows with the highest flow stability are identified within the local flow tree rooted at  $C_r$ .

#### D. Global Flow Consolidation

To consolidate the local flows into a global flow network, we formulate a network flow stability optimization model. Let  $\mathcal{F}$  denote the set of local flows and define the global network as  $G = \bigcup \mathcal{F}$ . The objective is to integrate the local major flows identified at different flow-shaping areas into a unified set of global major flows by selecting those that maximize cumulative stability across the network. Formally, the optimization problem is defined as

$$\max_{z, a} \sum_{F_i \in G} \lambda_d(F_i) \cdot a_i \sum_{f_j \in \mathcal{D}} z_{i,j} b_{i,j}, \quad (10)$$

where  $\lambda_d(F_i)$  denotes the span, i.e., path length, of flow  $F_i$ .  $\mathcal{D}$  denotes the set of all flights in the dataset.  $b_{i,j} = 1$  if flight  $f_j$  belongs to flow  $F_i$  (and 0 otherwise), and  $z_{i,j} \in \{0, 1\}$  is the decision variable indicating whether flight  $f_j$  is assigned to  $F_i$ . The binary variable  $a_i$  indicates whether flow  $F_i$  is selected.

This problem is subject to the following constraints:

$$\sum_{F_i \in G} z_{i,j} \leq 1, \quad \forall f_j \in \mathcal{D}, \quad (11)$$

$$a_i = \begin{cases} 1, & \text{if } \sum_{f_j \in \mathcal{D}} z_{i,j} b_{i,j} \geq N_{\min}, \\ 0, & \text{otherwise,} \end{cases} \quad \forall F_i \in G. \quad (12)$$

Constraint (11) ensures that each flight can be assigned to at most one flow. Constraint (12) requires that each flow contains at least  $N_{\min}$  flights; otherwise, it is considered noise and not selected ( $a_i = 0$ ).

### III. EXPERIMENTAL STUDY

To verify the efficacy of the proposed method, this section conducts experiments on major flow identification using air traffic plan data on 14 July 2023, with a focus on north-west-southwest (NW-SW) traffic in European airspace. The maximum great-circle length of flight segments in this traffic is 1943 nm; therefore,  $r_s$  is set accordingly in the fractal simplification process to preserve the core flow structures at this level of granularity.

Fig. 6 presents the major flow identification results. With parameter settings of  $N_{\min} = 15$  and  $\lambda_{\min} = 500\text{nm}$ , a total of 22 major flows are identified. Red lines highlight the segments of flight trajectories belonging to major flows.

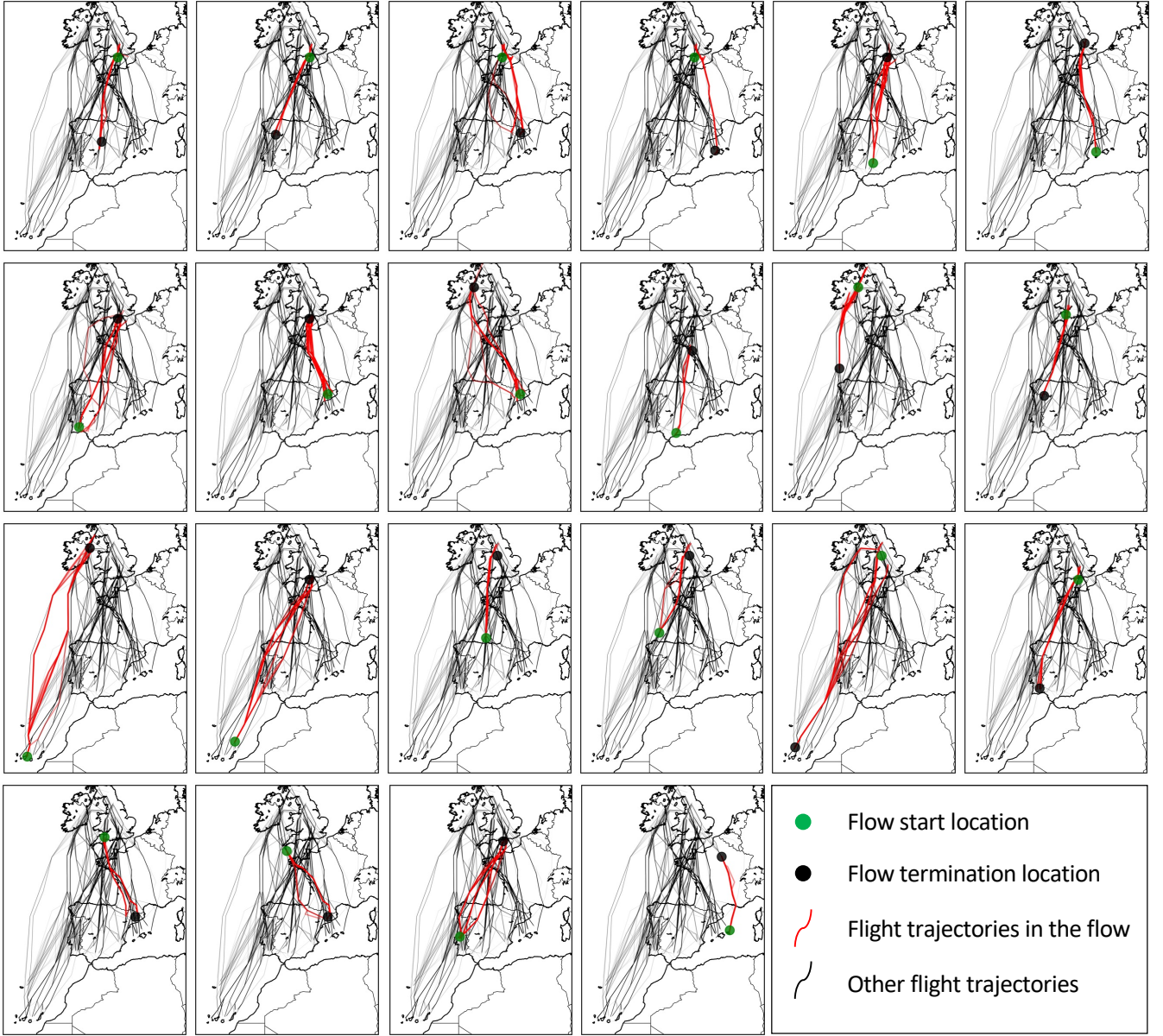


Figure 6. Major flow identification results with parameter settings of  $N_{\min} = 15$  and  $\lambda_{\min} = 500$  nm. Red lines highlight the segments of flight trajectories belonging to major flows, with green and black circles indicating the flow starting and termination locations, respectively. Black lines represent the trajectories of other flights.

Green and black circles mark the starting and termination locations of these flows, respectively. Black lines represent the trajectories of other flights. The average great-circle length of these flows is 647 nm, and they collectively include 826 flights out of the 1368 daily flights. This indicates that the majority of flights are captured within major flows, demonstrating the capability of the proposed method to organize most of the traffic into coherent flow structures. As shown in Figure 6, flights belonging to the same major flow between a given geographical pair may follow multiple alternative routes, providing flexibility for re-routing or flow regulation to alleviate congestion on a particular route. Furthermore, the identification results reveal that a subset of flights within

the same flow deviate significantly from the dominant routes, highlighting abnormal trajectory patterns.

The results of major flow identification are influenced by the parameter settings of  $N_{\min}$ ,  $\lambda_{\min}$ , and  $r_s$ . When  $N_{\min}$  is increased to 30, the number of identified flows decreases to 13, and the average flow length is reduced to 606 nm. When  $\lambda_{\min}$  is raised to 800 nm, 20 major flows are identified, encompassing 225 flights. Figure 7 illustrates the variations in average flow length and the volume of flights in the major flows under different values of  $N_{\min}$  and  $\lambda_{\min}$ . In general, a larger  $\lambda_{\min}$  yields longer major flows but reduces the number of flights they contain. On the other hand, a larger  $N_{\min}$  shortens the identified flows and excludes those with low

traffic, thereby reducing the total volume of flights represented as major flows. Therefore, the proposed method offers flexibility in major flow identification, enabling parameter settings to be adjusted in line with varying analytical requirements and operational objectives, and thereby allowing adaptation to diverse air traffic management contexts. For instance, a smaller  $\lambda_{\min}$  may be used to capture finer local structures for terminal area analysis, while a larger  $\lambda_{\min}$  is more suitable for identifying persistent long-haul flows across regions.

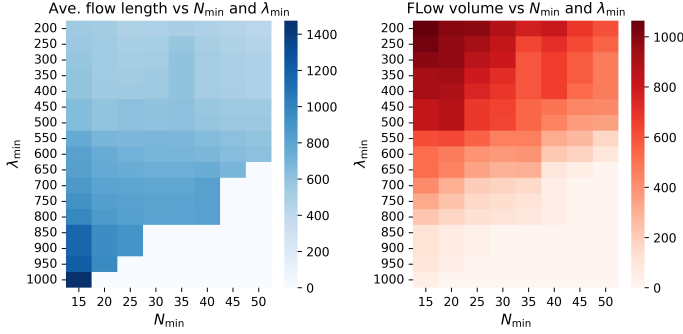


Figure 7.  $N_{\min}$  and  $\lambda_{\min}$  affects major flow identification: A larger  $\lambda_{\min}$  produces longer flows but with fewer flights, whereas a larger  $N_{\min}$  yields shorter flows and excludes low-traffic branches, reducing the total flight volume represented as major flows.

The proposed major flow identification method also enables systematic monitoring of flows over time. This includes tracking the number of flights within each flow, identifying when flows emerge and disappear, and detecting peak and low-demand periods. Fig. 8 shows an example of monitoring the identified major flows. The two subplots on the left panel show two identified major flows, while the two subplots on the right panel depict the corresponding number of flights involved in these flows over time. It can be observed that both flows exhibit multiple peaks and valleys, and that the timing of peak demand may coincide or differ between flows. Such information can provide insights for network-level management of air traffic, particularly in supporting demand forecasting, capacity planning, and congestion mitigation.

#### IV. CONCLUSIONS AND DISCUSSIONS

This paper proposed a bottom-up methodology for major flow identification that addresses the limitations of top-down approaches. By first deriving local flow structures and then integrating them into global flows, the framework captures the hierarchical and interconnected nature of air traffic. The key methodological components include the detection of flow-shaping areas through density-based clustering, the construction and fractal simplification of local flow trees, the introduction of a flow stability metric balancing traffic volume and persistence, and the optimization-based consolidation of local flows into a coherent global representation.

The experimental evaluation on European north-west-southwest traffic on 14th July 2023 demonstrated the efficacy of the proposed framework. Results showed that the majority of flights can be organized into major flows,

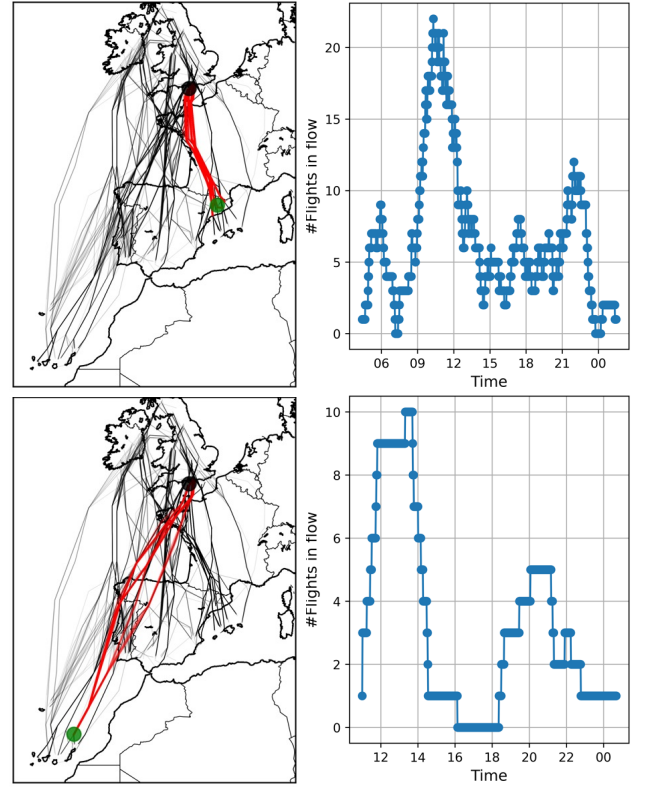


Figure 8. An example of major flow monitoring. Left: two identified flows. Right: the corresponding number of flights over time, illustrating multiple peaks and valleys with peak timings that may coincide or differ between flows.

yielding interpretable representations of dominant traffic corridors. The analysis also highlighted the flexibility of the approach: varying parameters such as  $N_{\min}$ ,  $\lambda_{\min}$ , and  $r_s$  allows the level of granularity to be tailored to different operational requirements. Furthermore, the framework not only identifies stable and representative flows but also provides a systematic basis for monitoring flow dynamics, such as emergence, disappearance, peak periods, and abnormal deviations. The monitoring example illustrated that flows may experience multiple peaks and valleys in demand, with peak timings that can align or diverge across flows. Such insights have direct implications for network-level demand-capacity balancing and flow regulation strategies.

Beyond major flow identification, the proposed framework opens several avenues for further research. Coupling flow identification with predictive models would enable proactive congestion management and support collaborative decision-making processes. The identified major flows can facilitate the analysis of congestion correlations and propagation at the network level. For example, by applying the proposed major-flow identification method to planned trajectories, a network manager can obtain a set of persistent, high-volume major flows together with their associated alternative routing structures. These outputs enable the manager to identify which flows are likely to generate high demand in specific regions

and to assess where capacity may be exceeded. With this information, targeted rerouting strategies can be designed to redistribute traffic and avoid the formation of hotspots.

In conclusion, the bottom-up methodology introduced in this paper provides a robust and interpretable means of identifying and monitoring major air traffic flows. By decomposing and reassembling flow structures, the framework not only improves the accuracy of flow detection in complex airspace but also lays a foundation for practical applications in traffic monitoring, re-routing, demand–capacity balancing, and anomaly detection.

#### ACKNOWLEDGMENT

This research is part of the enabled Dynamic Air Traffic Flow Configuration for Flow-Centric Airspace Management (DeepFlow) project, funded by the SESAR 3 Joint Undertaking (JU) under grant agreement No. 101166935. The views and opinions expressed in this publication are those of the authors and do not necessarily reflect the views of SESAR 3 JU, which cannot be held responsible for any use of the information contained herein.

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] Y. Chen, Y. Zhao, and Y. Wu, “Recent progress in air traffic flow management: A review,” *Journal of Air Transport Management*, vol. 116, p. 102573, 2024.
- [2] C. Ma, Y. Guleria, S. Alam, and M. Z. Li, “Deep reinforcement learning-based air traffic flow coordination in flow-centric airspace,” *Advanced Engineering Informatics*, vol. 65, p. 103342, 2025.
- [3] X. Olive and L. Basora, “Detection and identification of significant events in historical aircraft trajectory data,” *Transportation Research Part C: Emerging Technologies*, vol. 119, p. 102737, 2020.
- [4] Q. Xu, Y. Pang, and Y. Liu, “Air traffic density prediction using bayesian ensemble graph attention network (began),” *Transportation Research Part C: Emerging Technologies*, vol. 153, p. 104225, 2023.
- [5] W. Zhang, W. Pan, X. Zhu, C. Yang, J. Du, and J. Yin, “Identification of traffic flow spatio-temporal patterns and their associated weather factors: A case study in the terminal airspace of hong kong,” *Aerospace*, vol. 11, no. 7, 2024.
- [6] 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.
- [7] F. P. Moreno, V. F. G. Comendador, R. D.-A. Jurado, M. Z. Suárez, D. Janisch, and R. M. A. Valdés, “Methodology of air traffic flow clustering and 3-d prediction of air traffic density in atc sectors based on machine learning models,” *Expert Systems with Applications*, vol. 223, p. 119897, 2023.
- [8] C. Ma, S. Alam, Q. Cai, and D. Delahaye, “Text-enriched air traffic flow modeling and prediction using transformers,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, 2024.
- [9] Y. Xu, X. Prats, and D. Delahaye, “Synchronised demand-capacity balancing in collaborative air traffic flow management,” *Transportation Research Part C: Emerging Technologies*, vol. 114, pp. 359–376, 2020.
- [10] L. De Giovanni, C. Lancia, and G. Lulli, “Data-driven optimization for air traffic flow management with trajectory preferences,” *Transportation Science*, vol. 58, no. 2, pp. 540–556, 2024.
- [11] Y. Lin, L. Li, P. Ren, Y. Wang, and W. Szeto, “From aircraft tracking data to network delay model: A data-driven approach considering en-route congestion,” *Transportation Research Part C: Emerging Technologies*, vol. 131, p. 103329, 2021.
- [12] P. Han, W. Wang, Q. Shi, and J. Yue, “A combined online-learning model with k-means clustering and gru neural networks for trajectory prediction,” *Ad Hoc Networks*, vol. 117, p. 102476, 2021.
- [13] H. Li, “Time works well: Dynamic time warping based on time weighting for time series data mining,” *Information Sciences*, vol. 547, pp. 592–608, 2021.
- [14] K. Bringmann, M. Künnemann, and A. Nusser, “Discrete fréchet distance under translation: Conditional hardness and an improved algorithm,” *ACM Transactions on Algorithms (TALG)*, vol. 17, no. 3, pp. 1–42, 2021.
- [15] L. Basora, J. Morio, and C. Mailhot, “A trajectory clustering framework to analyse air traffic flows,” in *7th SESAR Innovation Days, Belgrade, Serbia*, 28-30 November, 2017.
- [16] S. O. Onyango, S. A. Owiredo, K.-I. Kim, and S.-L. Yoo, “A quasi-intelligent maritime route extraction from ais data,” *Sensors*, vol. 22, no. 22, p. 8639, 2022.
- [17] M. C. R. Murca and R. J. Hansman, “Identification, characterization, and prediction of traffic flow patterns in multi-airport systems,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 5, pp. 1683–1696, 2018.
- [18] W. Zhang, A. Payan, and D. N. Mavris, “Air traffic flow identification and recognition in terminal airspace through machine learning approaches,” in *AIAA SCITECH 2024 Forum, Orlando, FL, USA*, 8-12 January 2024, p. 0536.
- [19] S. Wang, J. Chu, J. Li, and R. Duan, “Prediction of arrival flight operation strategies under convective weather based on trajectory clustering,” *Aerospace*, vol. 9, no. 4, p. 189, 2022.
- [20] W. Zeng, Z. Xu, Z. Cai, X. Chu, and X. Lu, “Aircraft trajectory clustering in terminal airspace based on deep autoencoder and gaussian mixture model,” *Aerospace*, vol. 8, no. 9, p. 266, 2021.
- [21] J. M. Jayoma, G. P. Navales, F. A. M. Bustamante, and L. R. Tradío, “Enhanced ramer-douglas-peucker algorithm for efficient 2d line simplification,” in *2023 IEEE 15th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Coron, Palawan, Philippines*. IEEE, 19-23 November 2023, pp. 1–6.
- [22] 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.
- [23] N. Adler, E. T. Njoya, and N. Volta, “The multi-airline p-hub median problem applied to the african aviation market,” *Transportation Research Part A: Policy and Practice*, vol. 107, pp. 187–202, 2018.
- [24] 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.
- [25] F. Martinez, A. Ojeda, and H. Manríquez, “Application of fractal theory to describe the morphological complexity of large stream networks in chile,” *Water Resources*, vol. 49, no. 2, pp. 301–310, 2022.
- [26] C. Vu, T. Truong, and J. Kim, “Fractal structures in flexible electronic devices,” *Materials Today Physics*, vol. 27, p. 100795, 2022.

