Transportation Research Interdisciplinary Perspectives 38 (2026) 102108

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.sciencedirect.com/journal/transportation-research-interdisciplinary-perspectives

A general data-driven methodology for predicting future air traffic distributions around airports

Lorenzo Dorbolò, Marco Pretto *, Pietro Giannattasio

Polytechnic Department of Engineering and Architecture, University of Udine, Via delle Scienze 206, 33100 Udine, Italy

ARTICLE INFO

Keywords:
Air traffic distribution
Flight tracking
Trajectory clustering
Traffic scenario prediction
Airport noise

ABSTRACT

A reliable forecast of future low-altitude air traffic distributions is essential for the prediction of aircraft noise and pollutant emissions in airport areas. This work proposes a novel methodology for generating aircraft movements around airports based on historical flight tracking data. These data are exploited to derive representative flight paths via unsupervised machine learning and typical airport operations via statistical analysis. These outcomes feed a custom-made algorithm that generates future traffic distributions, requiring only traffic intensity and forecast timeframe as inputs. The methodology, tested at Stockholm Arlanda Airport, uses 2022 tracking data to predict the traffic distribution in 2024 and is validated comparing simulated and actual 2024 day-evening-night noise levels L_{DEN}. An accurate prediction is achieved, as moderate L_{DEN} differences involve regions under 45 dB(A) while $L_{DEN} \geq 45$ dB(A) contour areas are overestimated by less than 6 %. Errors come mostly from changes in airport operations, to be addressed in future developments.

1. Introduction

To this day, airport emissions remain a serious environmental concern, with well documented adverse effects on human health and life quality due to both pollutants (Bendtsen et al., 2021) and noise (Topriceanu et al., 2025) (Preisendörfer et al., 2022). This concern is bound to intensify as civil air traffic grows, with forecast indicating a 5.8 % worldwide increase for 2026 (IATA, 2025) and a 2.2 % yearly rise in European flights until 2031 (EUROCONTROL, 2025). The expected longer-term traffic growth is so strong as to offset even the environmental benefits of fleet upgrades (Huang et al., 2023), threatening even the ambition of a sustainable future aviation (European Commission, 2012). Challenges come also from commuting dynamics, as people move to cities during daytime and experience the highest airport activity and emission levels (Zhu et al., 2025). Therefore, accurate prediction of airport emissions and their impact on the population represents a task of major importance, at the core of which is the need for reliable modelling of the space–time distribution of air traffic within the TMA (Terminal Manoeuvring Area), especially when the impact of future air traffic scenarios is investigated (Simonetti et al., 2015).

A key step toward this goal is the identification of the main airport departure and arrival routes, which is critical in applications such as airspace design and monitoring (Gariel et al., 2011), as well as conflict (Madar et al., 2021) (Gaume et al., 2025) and anomaly detection (Olive & Basora, 2019) (Wang et al., 2025). Traditionally, these routes are derived from published aeronautical charts, which provide nominal procedures designed by airspace authorities (ECAC, 2016b). However, the actual aircraft behaviour often deviates from these prescribed tracks due to factors such as weather, traffic management, or operational constraints. A reliable way to quantify such deviations is leveraging large air traffic datasets through data-driven approaches, proved valid for both regional (Xu et al., 2023) and national analyses (Xu et al., 2024), whose effectiveness has been further enhanced by the availability of flight tracking data based on ADS-B (Automatic Dependent Surveillance-Broadcast) technology (Schäfer et al., 2014). Starting from historical ADS-B flight operations, Olive and Morio (Olive & Morio, 2019) applied the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm to arrival trajectories in the TMA of

Abbreviations: ADS-B, Automatic Dependent Surveillance-Broadcast; ANP, Aircraft Noise and Performance; ATC, Air Traffic Control; CDF, Cumulative Density Function; ECAC, European Civil Aviation Conference; EUROCONTROL, European Organization for the Safety of Air Navigation; HDBSCAN, Hierarchical Density-Based Spatial Clustering of Applications with Noise; ICAO, International Civil Aviation Organization; KDE, Kernel Density Estimator; MASA, Mixed Analysis-Synthesis Approach; METAR, Meteorological Aerodrome Report; PDF, Probability Density Function; SEL, Sound Exposure Level; SID, Standard Instrument Departure; STAR, Standard Terminal Arrival Route; TMA, Terminal Manoeuvring Area.

* Corresponding author.
E-mail address: marco.pretto@uniud.it (M. Pretto).

https://doi.org/10.1016/j.trip.2026.102108
Received 10 December 2025; Received in revised form 13 April 2026; Accepted 10 June 2026
Available online 15 June 2026

Toulouse Airport, obtaining clusters in good agreement with official STAR (Standard Terminal Arrival Route) tracks. In a similar application, DBSCAN was employed to identify nominal and rerouting traffic patterns in the New York metropolitan region (Condé Rocha Murça et al., 2016). More sophisticated approaches have also been proposed, such as a deep autoencoder and Gaussian mixture model method (Zeng et al., 2021) that captured successfully the complex traffic patterns around Guangzhou Baiyun International Airport.

Once the main traffic routes are determined, the next step is modelling the lateral dispersion of aircraft within each route, so as to capture the variability in how aircraft actually fly inside a given corridor. This aspect is critical for applications such as noise modelling (Guruprasad et al., 2023), safety assessment (Bao et al., 2024), airspace protection (Eerland & Box, 2016), and airspace design analysis (Zhu et al., 2023). Traditional methods (ECAC, 2016a) (ECAC, 2016b) assume a Gaussian distribution of the lateral dispersion, which is still a widely adopted solution for environmental noise studies (Xie et al., 2022), (Giladi & Menachi, 2024). However, an accurate characterization of traffic flows requires also information on the vertical dispersion, as it directly affects aircraft thrust settings and, consequently, noise emissions. The vertical dimension, known as flight profile, exhibits a certain degree of variability depending on aircraft type and weight, weather conditions, and ATC (Air Traffic Control) constraints. Therefore, a comprehensive air traffic flow model should consider the dispersion of the entire flight paths, each of which combines a ground track and a flight profile. These two components can be addressed either jointly or separately.

Regarding the path dispersion, Eerland and colleagues (Eerland et al., 2016) combined clustering with Gaussian process modelling to quantify the statistical uncertainty around nominal flight paths, addressing horizontal and vertical dimensions together. Krauth et al. (Krauth et al., 2023) proposed a deep generative model based on variational autoencoders and temporal convolutional networks, which managed to capture the path dispersion and synthesize realistic 4D trajectories accounting for ATC interventions. Similarly, Wu and colleagues (Wu et al., 2022) employed generative adversarial networks for long-term 4D trajectory prediction, learning the underlying space–time distribution of flights from historical data. Their solution effectively reproduced realistic variations in both horizontal and vertical directions, showing potential for capturing complex trajectory distributions. However, when considering the flight profiles, these studies do not account for the influence of the aircraft type, which is a critical aspect when environmental impact predictions are considered.

Finally, beyond identifying representative traffic flows and flight trajectories, it is equally important to obtain the operational statistics of an airport, which are essential for predicting the space–time distribution of future air traffic patterns and hence of aircraft emissions. For instance, Einicke and Kennedy (Einicke & Kennedy, 2025) used flight tracking data to derive fleet mix, runway configuration, and air traffic patterns for predicting noise levels at Dublin Airport in 2040, while Xie et al. (Xie et al., 2024) quantified the operational frequency of representative ground tracks at Baiyun International Airport to support the prediction of future noise exposure.

This analysis shows that many solutions leveraging flight tracking data to capture the features of air traffic in TMAs already exist, but neither a complete method for the prediction of TMA air traffic distributions nor a clear strategy to validate the environmental impact of such distributions have been proposed yet. These two issues are addressed in this work, which presents a general methodology for predicting future air traffic distributions on the basis of historical information and validates it through the analysis of airport noise levels. The traffic distribution in the TMA is modelled using large amounts of historical flight tracking data, which are processed with a modelling tool previously developed by the authors (Pretto et al., 2024). Focusing on the horizontal dimension of the flight trajectories, unsupervised machine learning is employed to cluster similar flights, and a novel method is proposed to model the lateral dispersion of trajectories and generate representative ground tracks. In the vertical dimension, representative flight profiles are derived taking into account the effect of different aircraft types. Moreover, airport operational statistics such as runway usage and fleet mix composition are derived directly from the tracking data, rather than relying on reports or literature sources. Finally, an original algorithm combines these elements into a coherent data-driven methodology for determining future air traffic distributions. The methodology is applied to Stockholm Arlanda Airport, where air traffic distribution and noise levels predicted for 2024 using tracking data from 2022 are compared to the results obtained directly from the tracking data of 2024.

The data-driven methodology proposed in this work presents a key advantage compared to official scenario prediction approaches (ECAC, 2016a), which is the generation of air traffic distributions using real-world local traffic patterns rather than stereotyped prescriptions. This leads to a more accurate simulation of aircraft operations that can be tailored to each airport regardless of location and size, while simultaneously ensuring widespread applicability thanks to the reliance on open data. However, this methodology disregards some operational aspects, such as evolution of runway usage and fleet composition, which are known to influence air traffic distribution and noise (Simons et al., 2026) (Pretto et al., 2020). The impact of these elements is quantified in the results, showing possible improvements towards a reliable and fully data-driven prediction of future air traffic scenarios.

The article is structured as follows. Section 2 presents the dataset enabling the implementation of the proposed methodology, which is described in detail in Section 3. The results are shown and examined in Section 4, while conclusions and future developments are presented in Section 5.

2. Dataset of historical air traffic and aircraft performance parameters

Stockholm Arlanda Airport, located around 37 km north of the capital of Sweden and serving over 22 million passengers in 2024, was selected as a case study, as it is a large three-runway airport whose typical air traffic is complex enough to illustrate the capabilities of the proposed methodology (Swedavia Airports, 2025). The layout of the airport is shown in Fig. 1, where the runway names are enclosed in white boxes.

The ADS-B data describing the air traffic at Arlanda in 2022 were retrieved from the OpenSky Network (Schäfer et al., 2014) using the Python *traffic* library (Olive, 2019), resulting in 76,839 departures and 77,868 arrivals. These flight operations were processed individually according to the authors' modelling tool (Pretto et al., 2024) outlined in the flowchart of Fig. 2. Firstly, the ADS-B flights were enriched with additional data from several external sources, including *OpenStreetMap* (The OpenStreetMap Foundation, 2025), the ANP database (EASA - ANP, 2025), METARs (Meteorological Aerodrome Reports), and an in-house aircraft database based on open information. Then, for each flight operation in the resulting dataset, the ground track was computed, and the mixed analysis-synthesis approach (MASA) was used to reconstruct the flight profile, and hence the aircraft performance (Pretto et al., 2022) (Pretto et al., 2024). Finally, a custom implementation of the ECAC noise engine (ECAC, 2016b) was employed to compute the acoustic impact of individual operations. As a result, for each flight in the 2022 dataset, the following outputs were made available:

- the reconstructed flight path resulting from the merging of ground track and flight profile, which holds information on aircraft position, altitude, speed, and engine thrust over time;
- the noise impact of that flight, quantified by the A-weighted Sound Exposure Level (SEL);

Fig. 1. Layout of Arlanda Airport and names of its three runways.

- details about the operation (i.e. metadata), including flight generalities, aircraft configuration, and weather information, as in the example of Table 1.

Focusing on the MASA, its purpose is minimizing the error between ADS-B data and the flight profile modelled following ANP procedural steps (EASA - ANP, 2025) and ECAC flight mechanics equations (ECAC, 2016b). This is done by varying the performance parameters in Table 2, assumed fixed by ANP, until the smallest error is achieved. In arrivals, the optimization targets initial descent angle γ_{in} and airspeed $V_{C,in}$, as well as level-flight length Δs_{lev} and height h_{lev}. In departures, it targets profile type, take-off and climb thrust reduction factors K_T and K_C, respectively, variations of initial and mid-climb target heights Δh_L and Δh_M, respectively, and energy share fraction factor f_e to modulate the aircraft acceleration. All details are available in the referenced work (Pretto et al., 2024).

Finally, for the purposes of this work not all flight operations were included in the following analysis. In particular, operations associated with aircraft types, identified by their ICAO four-character codes, that contributed less than 0.1 % to the total air traffic (e.g. small private jets) were disregarded, as they had negligible impact on it while massively increasing the complexity of the statistical analysis.

3. Methodology for the data-driven prediction of air traffic distributions

This section provides all the details that enable the proposed data-driven prediction of future air traffic distributions for a given airport. Firstly, the focus is on the generation of a set of representative flight paths to be used in place of ADS-B flight trajectories for modelling the air traffic flow. Following consolidated practices (ECAC, 2016a), ground

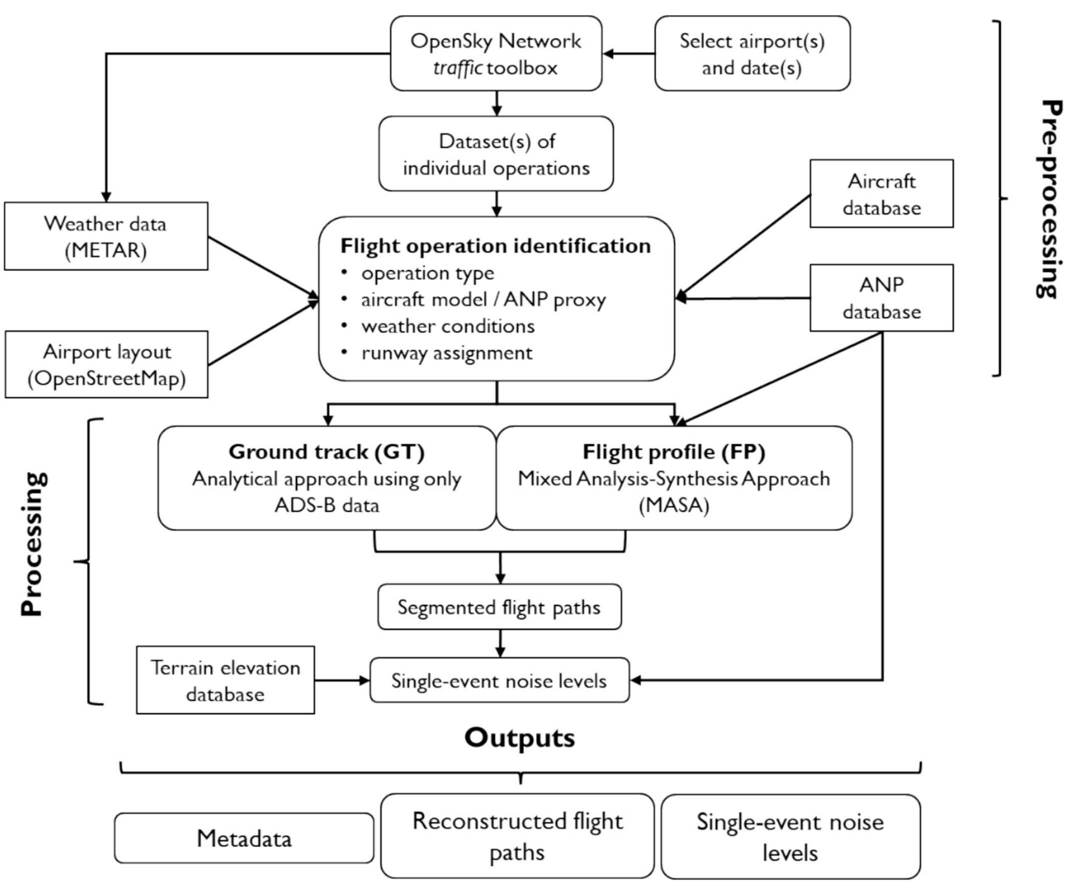

Fig. 2. Flowchart summarizing the authors' tool for processing individual flight operations from ADS-B data (Pretto et al., 2024).

Table 1
Example of additional information (metadata) available for single flight operations on 1 January 2022. ID: identifier; OP = operation: arrival (A) / departure (D); WD, WS = wind direction, speed; RH: relative humidity.

	Flight Generalities			Aircraft configuration				Weather information				
ID	**Time**	**OP**	**Rwy**	**ICAO**	**Type**	**Engine**	**Proxy**	**Temp [K]**	**Pres [kPa]**	**WD [°]**	**WS [kt]**	**RH [%]**
ETH714	04:40:49	A	01L	B789	787–9	TRENT 1000	7879	272.4	101.1	350	12.5	89
ETH714	05:56:29	D	01L	B789	787–9	TRENT 1000	7879	271.4	101.3	350	11.3	87
ETH715	19:23:40	A	01L	B789	787–9	TRENT 1000	7879	269.7	102.0	173	3.3	92
ETH715	21:09:26	D	19L	B789	787–9	TRENT 1000	7879	269.9	101.8	140	5.5	89
QTR59G	11:42:33	A	01L	B788	787–8	GENX-1B64	7878R	270.7	102.0	320	5.3	86
QTR75G	14:46:55	D	01L	B788	787–8	GENX-1B64	7878R	270.4	102.1	307	4.0	85
…	…	…	…	…	…	…	…	…	…	…	…	…

Table 2
Example of the results of the MASA applied to individual flight operations.

	MASA optimization variables										
	Arrivals				Departures						
ID	γ_{in} [°]	$V_{C,in}$ [kt]	Δs_{lev} [%]	h_{lev} [ft]	**Profile type**	K_T [-]	K_C [-]	**TOW [lb]**	Δh_L [ft]	Δh_M [ft]	f_e [-]
ETH714	1.76	248.9	0	0	–	–	–	–	–	–	–
ETH714	–	–	–	–	ICAO_A	0.82	1.00	397,141	498	0	0.87
ETH715	1.88	248.7	74	−424	–	–	–	–	–	–	–
ETH715	–	–	–	–	DEFAULT	0.89	0.90	413,318	594	1241	0.88
QTR59G	2.81	232.5	66	−312	–	–	–	–	–	–	–
QTR75G	–	–	–	–	DEFAULT	0.98	1.00	435,152	501	765	0.84
…	…	…	…	…	…	…	…	…	…	…	…

tracks and flight profiles, whose combination leads to flight paths, are treated separately in Section 3.1 and 3.2, respectively. Then, in Section 3.3 a set of statistical distributions that reflect typical airport operations is derived. Finally, Section 3.4 describes the actual generation of future air traffic distributions from the previous outcomes, while Section 3.5 outlines its current implementation for fast prediction of airport noise levels.

3.1. Representative ground tracks

In air traffic flow modelling, it is normal practice to represent specific air routes through a backbone track with a small number of laterally displaced sub-tracks. In absence of radar data, the backbone tracks are usually built from dedicated information, which typically consists of SID (Standard Instrument Departure) and STAR procedures extracted from aeronautical information publications. Then, their lateral dispersion is modelled assuming that the spread of tracks perpendicular to backbone follows a normal Gaussian distribution (ECAC, 2016b). However, the availability of ADS-B data has led to new, data-driven modelling opportunities for generating better representative backbones and sub-tracks. Also following ECAC (ECAC, 2016b), any association between ground track and aircraft type that flew it, very weak at most, is neglected.

3.1.1. Data pre-processing and clustering

The generation of representative ground tracks is based on clustering techniques, the application of which requires a higher data regularity than that granted by the ADS-B data. Therefore, the ADS-B tracks underwent two normalization steps. Firstly, each track was either truncated or extended to have the same cumulative distance from the starting point, 120,000 ft for departures and 250,000 ft for arrivals, and then all tracks were resampled so as to be composed of 50 equally spaced points.

From the normalized tracks, the HDBSCAN (Hierarchical DBSCAN) clustering algorithm (Campello et al., 2013) was employed for the identification of the preferential routes at airports. This algorithm does not require a predetermined number of clusters, because it was designed precisely for independent identification of clusters with different shape, size and density, distinguishing also noisy trajectories that do not fit any pattern. In this work, the implementation of HDBSCAN was conducted using the *scikit-learn* Python library (Pedregosa et al., 2011), and it required setting three key parameters, whose effects are explained below.

- Minimum samples s_{min}: it is related to the distance between a trajectory and its nearest neighbour. Setting s_{min} too low results in large clusters being fragmented into many smaller ones, while high values lead to small clusters being merged and more trajectories being classified as noise.
- Minimum cluster size c_{min}: it is the minimum number of samples in a group for that group to be considered a cluster. Groupings smaller than c_{min} are treated as noise.
- Distance threshold ϵ, whose value is useful for merging smaller clusters in certain regions.

Due to their intrinsic differences, arrivals and departures were processed separately. In both cases, a grid search was carried out to identify the best combination of the three clustering parameters mainly in terms of the silhouette coefficient (Rousseeuw, 1987), which ranges between -1 and $+1$ and indicates how similar an object is to its own cluster compared to other clusters. In addition, this identification was supported by visual inspection and the evaluation of both the number of clusters and the fraction of trajectories classified as cluster outliers. In fact, sole reliance on the maximization of the silhouette coefficient can lead to misleading outcomes, such as unrealistically small clusters coupled with many outliers, or an artificially low number of clusters with only a few outliers, which oversimplify the structure of the traffic flows. For departures, the optimal values of s_{min}, c_{min} and ϵ turned out to be 40, 35, and 0.05, respectively, while for arrivals they were 40, 37, and 0.06. About 88 % of all flight operations identified at Arlanda Airport in 2022 were grouped in the clusters shown in Fig. 3, while the remaining 12 % was classified as noise.

The reliability of the clusters generated using these parameter values is confirmed by a comparison with the recommended near-airport trajectories, which can be usually extracted from the SIDs and STARs for departures and arrivals, respectively. A relevant example is presented in

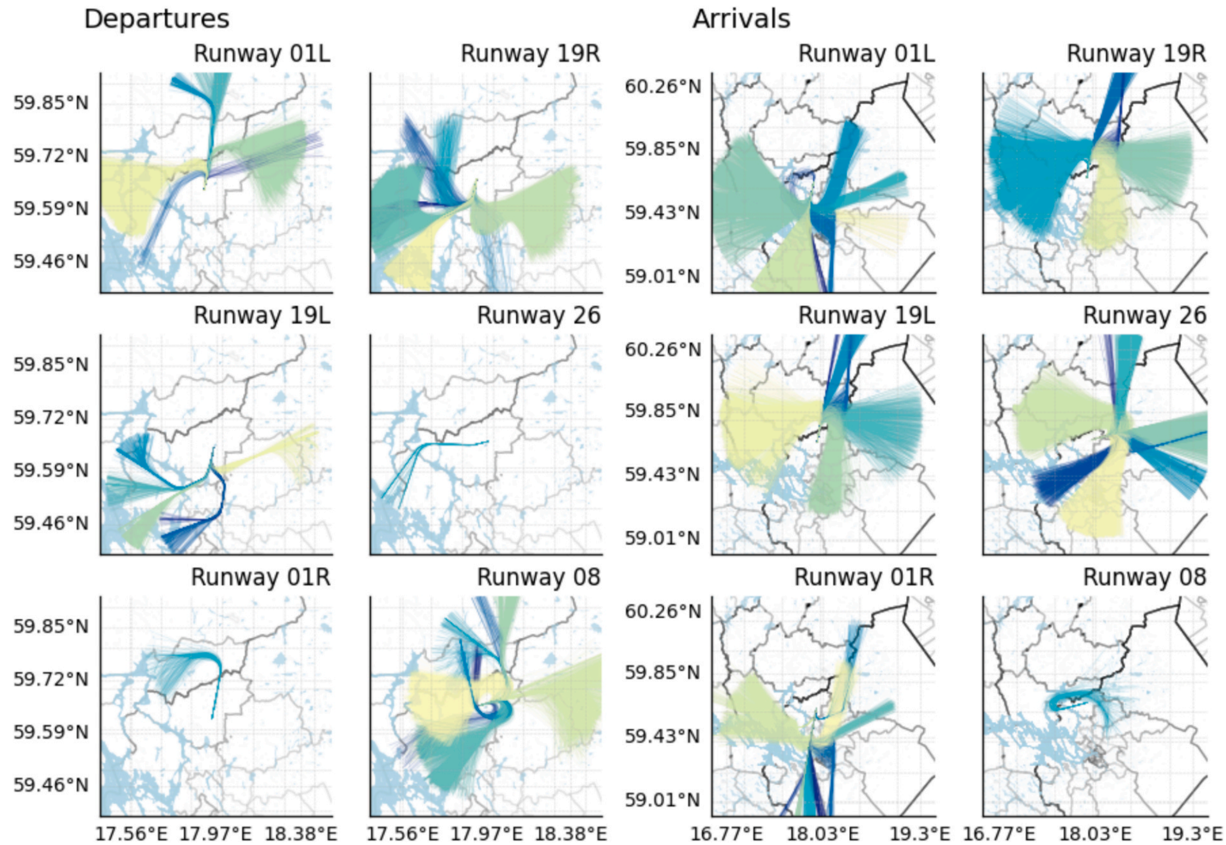

Fig. 3. For each runway, the departure and arrival clusters found at Arlanda Airport in 2022. In each subplot, the clusters are represented with different colours.

Fig. 4 using the departure cluster centroids, computed by averaging the trajectories in each cluster, and the SIDs of Arlanda Airport in 2026 (Luftfartsverket, 2026), which are the closest to 2022 available online, of three representative runways. The comparison is quite favourable, with a very close match reported between centroids and SIDs. As for the differences, just one centroid belonging to 01L (red line) has no associated SID, but this is because it was derived only from trajectories of small aircraft, which enjoy fewer restrictions thanks to their lower environmental footprint (ECAC, 2016a). Moreover, the lack of centroid counterparts for some SIDs is explained by the fact that although commercial flights must followed the SID tracks, not all of the latter actually must be used, which is even clearer looking at the sole clusters of runways 26 and 01R shown in Fig. 3. Finally, for the arrivals this comparison is impossible because the available STARs do not constrain the trajectories near the runways, but it is reasonable to expect that the good outcomes of departures can extend to approaching flights.

3.1.2. Lateral dispersion of flights inside each cluster

Accurately modelling the lateral dispersion of flights within each cluster is essential for generating synthetic tracks for future air traffic distributions aimed at replacing non-existent flight tracking data. For a single cluster, the lateral dispersion modelling follows the steps below.

1. The centroid is used as a reference path for modelling the lateral distribution of flights. Like all other trajectories, the centroid is composed of 50 points.
2. For each point of the centroid, along its perpendicular direction, the intersections with the tracks belonging to the cluster are found. This process is illustrated on the left-hand side of Fig. 5, where the intersections are denoted with blue dots.
3. The signed distances between centroid point and corresponding intersections are computed. The distances are negative for points located on the left-hand side of the centroid, and positive otherwise. Then, each centroid point is assigned an array of signed distances, which indicates how flights are distributed within that particular cluster.

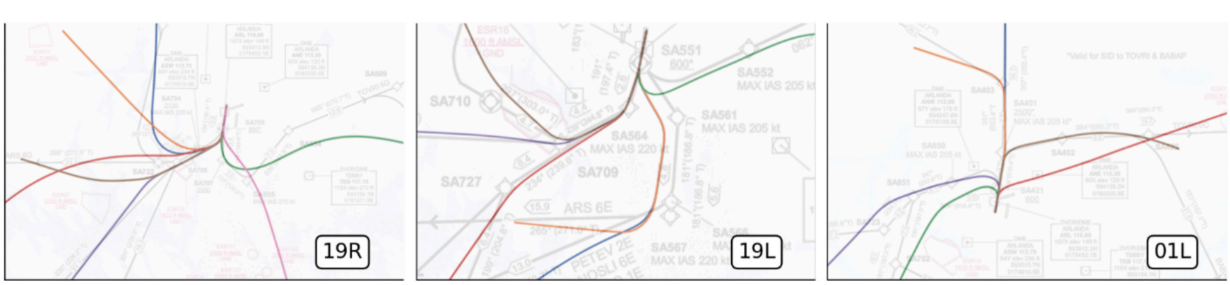

Fig. 4. Comparison between 2026 SIDs (grey lines in the background) and centroids of departure clusters based on 2022 flight tracking data (coloured lines) for three runways at Arlanda Airport.

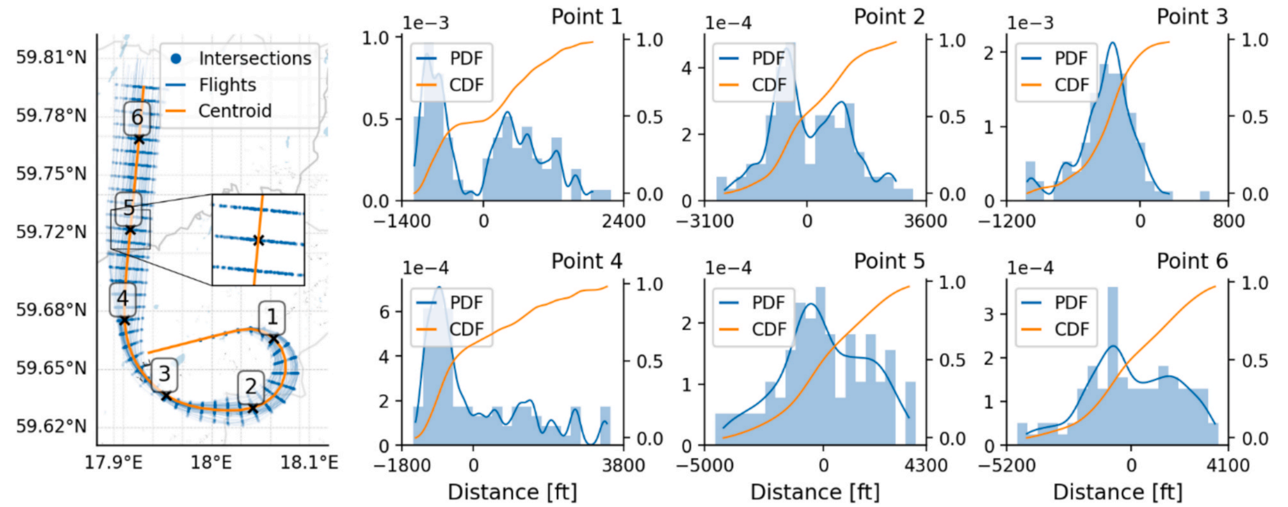

Fig. 5. An example of a departure cluster with its centroid (left), and PDFs and CDFs computed at the six sample points indicated on the map (right).

4. These discrete distance values are transformed into a Probability Density Function (PDF) using the Gaussian Kernel Density Estimation (KDE), defined in Equation (1):

$$\begin{cases} f(x) = \frac{1}{n \cdot h} \cdot \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \\ K(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}}, \end{cases} \quad (1)$$

where n is the number of values in the distance array, x_i is the i-th distance, and $K(u)$ is the Gaussian kernel function, while bandwidth h controls the smoothness of the estimated PDF by adjusting the kernel width. Selecting an appropriate value for h is crucial, as small bandwidths may result in noisy PDFs, while large values may yield over-smoothed curves. In this study, h is determined for each PDF using the improved Sheather-Jones algorithm (Sheather & Jones, 1991).

5. Finally, the corresponding Cumulative Density Function (CDF) is computed from each PDF.

Fig. 5 shows an example of PDFs and CDFs calculated at six sample points of the cluster centroid illustrated on the left-hand side of the same figure. The shape of the PDFs suggests that the trajectory swath cannot always be modelled accurately with a Gaussian distribution. In fact, in several locations the distributions exhibit asymmetric or multimodal behaviours, suggesting that the lateral dispersion is influenced by unknown flight operational factors. As a result, a simple Gaussian model appears insufficient to capture the variability in the lateral dispersion of flights around a backbone track in a satisfying manner. This modelling procedure was applied to all the clusters shown in Fig. 3.

3.1.3. Generation of representative ground tracks

The information obtained on the lateral dispersion of flights is used to determine seven representative ground tracks (i.e. sub-tracks) for each identified cluster. This number has been found sufficient by ECAC (ECAC, 2016b) to capture the full width of trajectory swaths with acceptable accuracy for noise emission estimation purposes.

Following again ECAC, each of the seven tracks is assigned predefined percentages of flight movements or traffic volume, whose values are given in Table 3. However, these percentages cannot be related directly to standard deviations as ECAC does, because the lateral dispersions do not follow Gaussian distributions. Instead, the percentages are mapped to fixed CDF quantiles, which provides a more robust representation that is independent of the distribution type. For example, the outermost sub-track to the left of the centroid (3A) is assigned 8 % of the total movements inside the cluster, and the corresponding quantile q is set to 4 % as the mid-point between 0 % and 8 %. Similarly, the next sub-track (2A) is assigned 12 % of the movements, with a corresponding quantile q of 14 % (i.e. 8 % + 12 % / 2). This procedure is repeated for all sub-tracks, so that each quantile corresponds to the mid-point of the cumulative probability interval defined by the associated percentage of movements.

Table 3
Percentage of movements and corresponding quantiles for the seven ground tracks representing each identified cluster.

Sub-track name	CDF quantile q [%]	Percentage of movements [%]
3A	4	8
2A	14	12
1A	28	16
0	50	28
1B	72	16
2B	86	12
3B	96	8

The process for building the sub-track associated with the quantile q is as follows. For each point along the cluster centroid, its associated CDF is selected, and q is used to obtain the corresponding distance d. The sub-track point is then positioned at the distance d from the related centroid point, along its perpendicular direction. After all sub-track points are generated, the resulting trajectory is filtered to remove outliers, ensuring that the final path is smooth and flyable. A visual illustration of this process is provided in Fig. 6.

The ground tracks generated for the air traffic at Arlanda Airport in 2022 are illustrated in Fig. 7 for both departures and arrivals. These trajectories are superimposed onto the original ADS-B-based tracks (in grey), showing that they indeed cover rather accurately the vast majority of the observed air traffic patterns, and thus they can be taken as reference ground tracks.

3.2. Generation of representative flight profiles

Unlike ground tracks, which depend primarily on airport layouts and ATC requirements on where aircraft are allowed to fly, the relationship between flight profile and aircraft type cannot be overlooked. This is because the vertical motion of an aircraft is mostly a function of weight, aerodynamic performance, and thrust output, the variability of which is massive between regional jets, single-aisle aircraft, and double-deck airliners (ECAC, 2016b). Moreover, after grouping the ADS-B flight

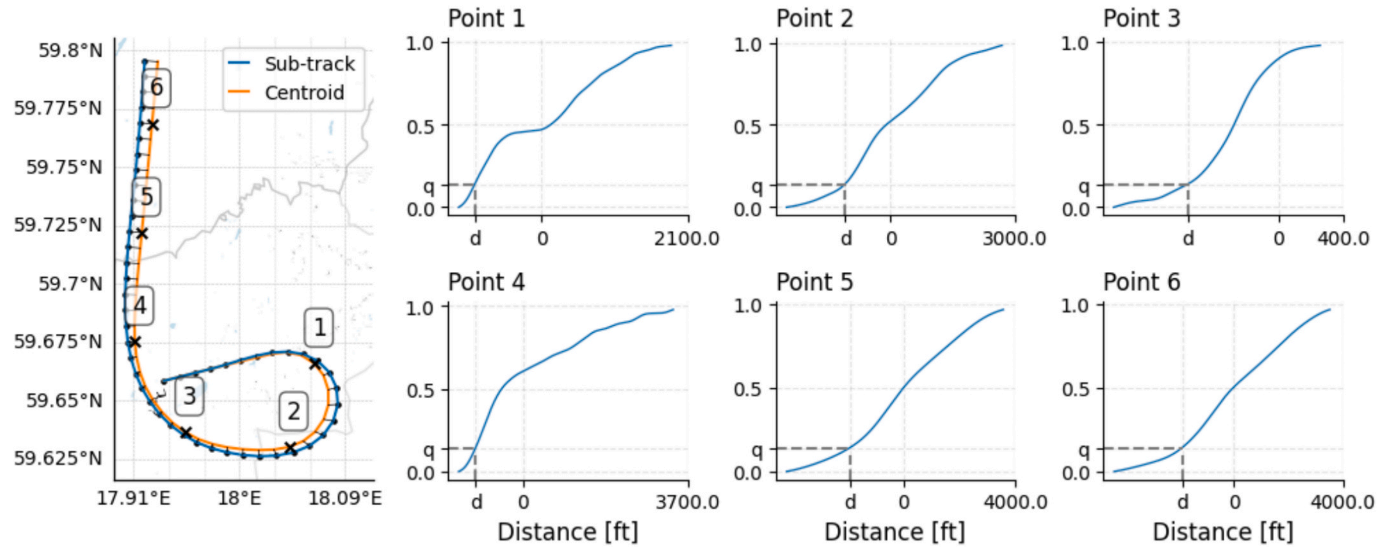

Fig. 6. Visual representation of how the sub-track associated with $q = 0.14$ (sub-track 2A) is generated.

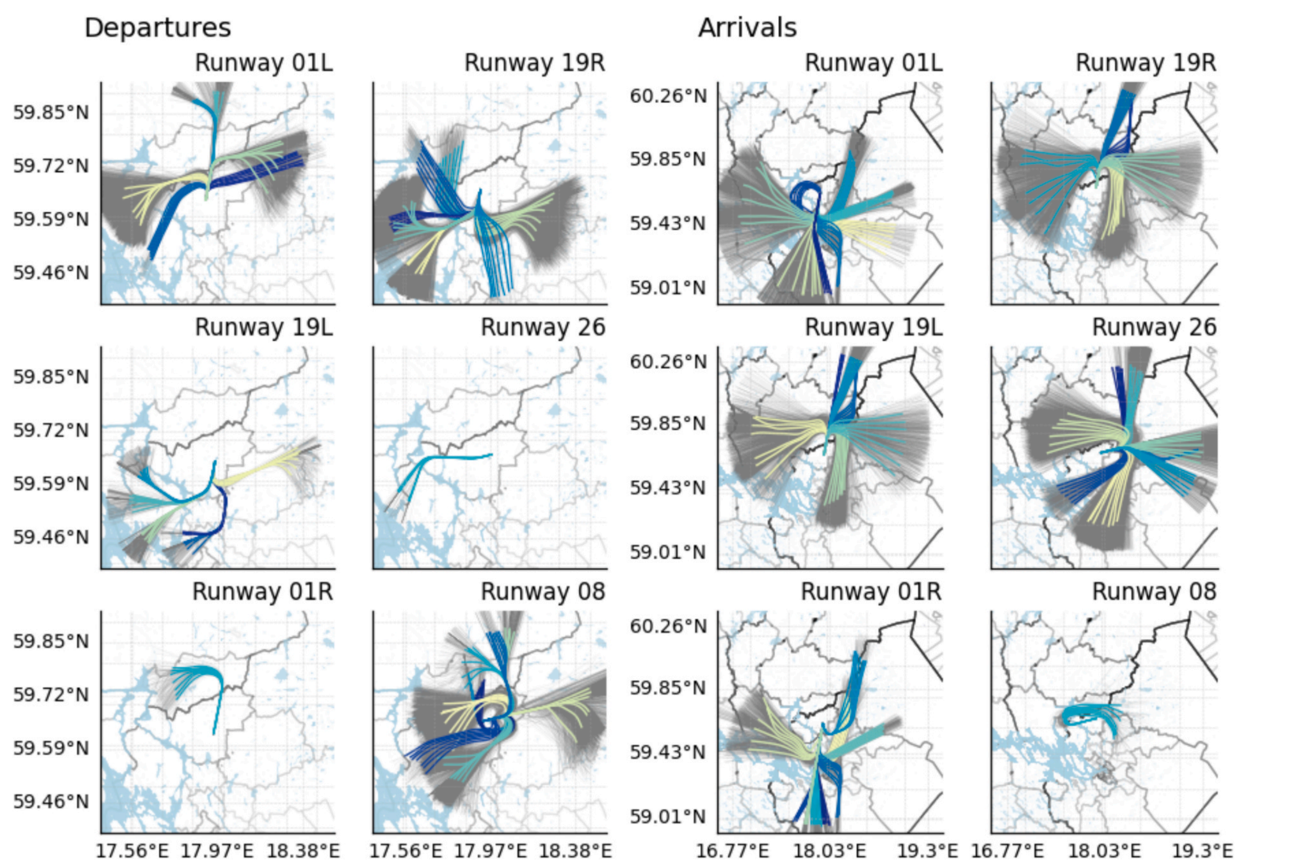

Fig. 7. Generated and actual departure and arrival ground tracks at Arlanda Airport for each runway in 2022. In each subplot, the seven-track groups are represented with different colours, while actual ADS-B tracks are in grey.

profiles by aircraft type, the analysis of their dispersion in terms of speed and altitude over time revealed a much lower spreading than that of ground tracks, which is due to the narrow performance envelope of a given aircraft type. Therefore, aiming to obtain representative air traffic patterns while keeping the computational costs low, it was decided to use only one representative flight profile for each aircraft type. At fixed type, this reference flight profile is generated as follows.

1. A representative ANP proxy is selected from the ANP substitution tables (EASA, 2025) using the information on the actual aircraft associated with the ICAO aircraft type.
2. The outputs of the MASA optimization for each flight operation are taken from the flight metadata (Table 2) to extract representative values. For the continuous variables the median is used, while the mode is picked for the discrete ones (profile type, K_C).

3. These MASA medians and modes are used to modify ANP procedures and thrust outputs (Pretto et al., 2024), which are included in the ECAC flight mechanics equations (ECAC, 2016b) to generate the representative flight profile. The annual median weather conditions at the airport are used in the calculation.

Application of the algorithm to Arlanda Airport in 2022 led to two sets (departures and arrivals) of representative profiles, and those of the four most frequent aircraft types are presented in Fig. 8. For illustrative purposes only the altitudes are shown, but each profile also holds information on aircraft speed (true / calibrated airspeed, ground speed), and engine thrust. Note that the representative profile is not the average of all ADS-B profiles, as per the calculation procedure described above.

3.3. Statistical distributions describing how the airport is operated

Representative ground tracks and flight paths constitute key elements of data-driven air traffic distributions, but they are insufficient for a full prediction of the traffic patterns, which depend also on how the airport is usually operated (e.g. slot allocation, runway usage). Information in this regard can be obtained from the analysis of statistical distributions computed from the flight metadata of Table 1, and it can be further refined by incorporating information about the cluster associated with each flight operation. The figures below summarize the key indicators used to generate representative air traffic patterns.

Firstly, the left-hand side of Fig. 9 shows the usage of runways at Arlanda Airport in 2022, distinguishing between departure and arrival operations. Aircraft typically took off from runways 19R, 08 and 01L, whereas the vast majority of arrivals occurred on runways 26, 01L, and 19R, with some runways used sparingly or not at all for certain operations (e.g. 01R and 26 for departures). The right-hand side of Fig. 9 shows how flight operations were generally allocated throughout the day. Unsurprisingly, around 70 % of operations occurred during the conventional daytime, with the rest taking place in the evening and at night.

Concerning the airspace utilization, the frequency heatmaps in Fig. 10 show an example of the frequencies associated with specific combinations of aircraft type and air route, the latter corresponding to a cluster, for take-offs and landings separately. Each heatmap considers on the x-axis only the air routes for one specific runway, namely *aa*_19R for departures and *bb*_26 for arrivals (*aa* and *bb* are arbitrary identifiers), while the y-axis hosts the ten most common aircraft types at Arlanda in 2022. The heatmaps illustrate the existence of preferential type-cluster combinations, the knowledge of which is particularly useful when trying to represent accurately the airspace usage near airports and is almost impossible without exploiting flight tracking data.

3.4. Generation of the future air traffic distribution

The algorithm developed for the generation of future air traffic distributions depends on two user-defined inputs, which are the number of departures and arrivals and the time window to be simulated. The algorithm is split into the following three steps, with all numbers in the tables below being valid for the air traffic at Arlanda in 2022, and treats all the frequencies reported in Fig. 9 and 10 as probability distributions to be used as inputs for assignment operations. All assignments are conducted pseudo-randomly using a permuted congruential generator (O'Neill, 2014).

1. The flight operations are split into departures and arrivals, which are assigned to runways and time of the day (daytime, evening, night) according to the frequency values in Fig. 9. The outcome is a preliminary traffic distribution such as that of Table 4, where each row indicates the number of flights departing from or arriving at a given runway during a specific part of the day.
2. This preliminary result is refined by incorporating air route (i.e. cluster) and aircraft type information. For each row of Table 4, the algorithm selects the appropriate cluster-aircraft distribution, structured as shown in Fig. 10, and a multinomial draw is carried out to allocate the total number of flights among these route-aircraft pairs. The result is a more refined distribution (Table 5) that specifies how many flights of a certain aircraft type are assigned to each cluster for each runway, type of operation, and time of the day.
3. Each type-runway-time-aircraft-cluster set structured as in Table 5 is further broken down into its seven representative sub-tracks, which is done by assigning to each set (i.e. each row of Table 5) the fraction of aircraft movements defined in Section 3.1.3.

The output of this algorithm is the final space–time distribution of air traffic (Table 6), according to which each flight operation is characterized by a specific set of operation type, runway, time of day, aircraft type, cluster, and sub-track.

3.5. Computation of flight paths and airport noise levels

Table 6 includes all information concerning each aircraft operation in terms of reference ground track (runway, air route, sub-track) and flight profile (operation type, aircraft type), merging which its flight path and the associated noise footprint can be calculated. However, for a much lower computational burden it is more convenient to perform first a preliminary computation of all the possible flight path-noise pairs, and to reconstruct later the future air traffic distribution and its acoustic impact according to the specific user inputs (number of total flight

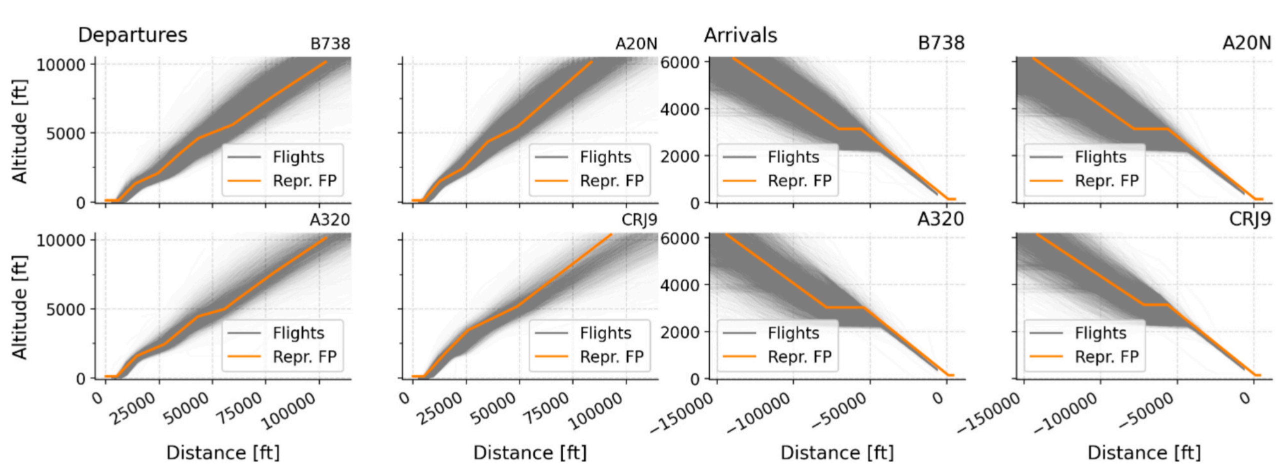

Fig. 8. Representative departure and arrival flight profiles (Repr. FP) as a function of the distance from the airport runway for the four most used aircraft types at Arlanda Airport in 2022.

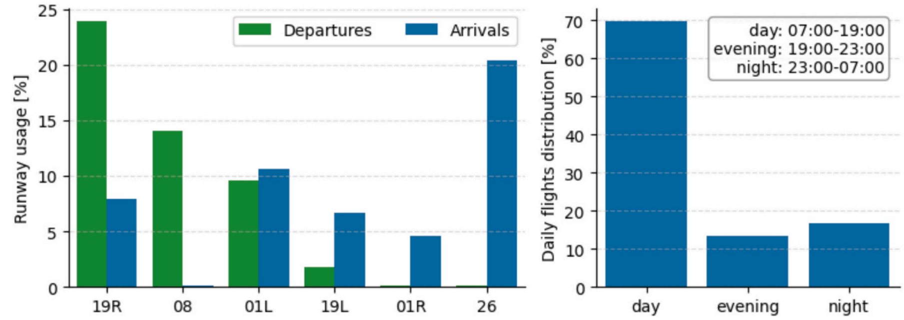

Fig. 9. Runway usage for departures and arrivals (left) and flight allocation as a function of the time of day (right).

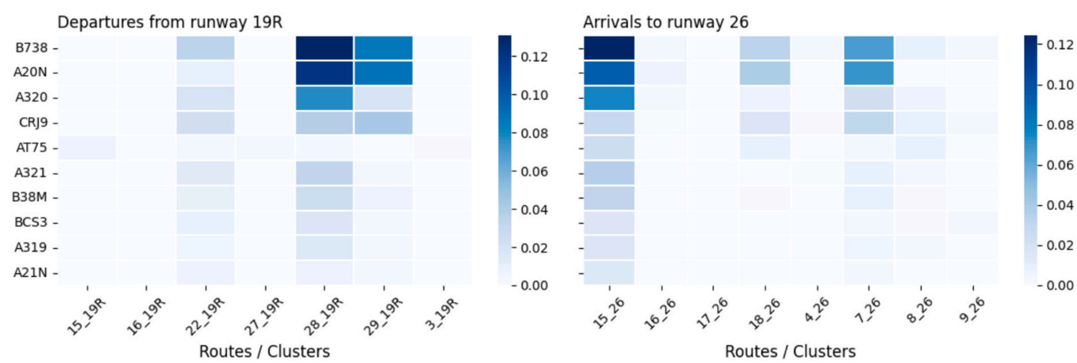

Fig. 10. Frequencies associated with specific combinations of air route and aircraft type.

Table 4
Overview of the preliminary traffic distribution, which covers type of operation, runway, and time of day.

Type of operation	Runway	Time of day	Number of operations
Departure	19R	daytime	27,835
Arrival	26	daytime	25,235
Departure	08	daytime	16,419
Departure	01L	daytime	11,295
Arrival	01L	daytime	10,562
Arrival	19R	daytime	10,118
Arrival	19L	daytime	8,706
Departure	19R	evening	7,703
Arrival	26	evening	6,846
…	…	…	…

Table 5
Overview of the refined air traffic distribution, accounting for type of operation, runway, time of day, aircraft type, and air route.

Type of operation	Runway	Time of day	Aircraft type	Air route (cluster)	Number of operations
Departure	19R	daytime	B738	28_19R	3,747
Departure	19R	daytime	A20N	28_19R	3,335
Arrival	26	daytime	B738	15_26	3,098
Departure	19R	daytime	A20N	29_19R	2,428
Departure	19R	daytime	B738	29_19R	2,428
Arrival	26	daytime	A20N	15_26	2,385
Departure	19R	daytime	A320	28_19R	2,100
Departure	01L	daytime	B738	20_01L	1,978
Arrival	26	daytime	A320	15_26	1,958
…	…	…	…	…	…

operations, simulation timeframe), which are independent of the reference flight paths. The noise output of each flight operation is computed in terms of SEL using the ECAC noise engine (ECAC, 2016b) customized by the authors (Pretto et al., 2019) (Pretto et al., 2020).

Once the preliminary computation is complete, the airport noise caused by the generated traffic distribution is calculated simply by cumulating the contributions of all operations. Using day-evening-night level L_{DEN} as the cumulative metric of choice, its value is derived as per Equation (2):

$$L_{DEN} = log_{10}\left(\frac{1}{T_0}\cdot\sum_{i=1}^{N} n_i\cdot 10^{(SEL_i+\Delta_i)/10}\right), \tag{2}$$

where SEL_i is the noise level of the i-th flight path, n_i the number of times that this path occurs (i.e. last column of Table 6), N the number of separate flight paths (i.e. length of Table 6), and T_0 the duration of the simulated time window in seconds. Finally, Δ_i is the L_{DEN} time-of-day level offset, which is 0 dB(A) for daytime, 5 dB(A) for evening, and 10 dB(A) night operations, and it can be assigned using the third column of Table 6.

4. Results

As mentioned in Section 1, the performance of the proposed methodology is assessed for Stockholm Arlanda Airport, the ADS-B data pertaining to which were collected for both 2022 and 2024, with the latter used for validation purposes. First, the 2022 dataset was processed to derive representative ground tracks (Section 3.1) and flight profiles (Section 3.2), as well as statistical distributions describing the frequency of airport operations (Section 3.3). Using these outcomes, the air traffic distribution for 2024 was generated by applying the algorithm described in Section 3.4 with the following inputs:

1. the actual operations in the 2024 ADS-B dataset (85,708 departures and 87,269 arrivals),
2. a time period of one calendar year (366 days) for the simulation.

Finally, the cumulative L_{DEN} values were calculated from the precomputed single-operation noise levels according to the procedure

Table 6
Overview of the final air traffic distribution, accounting for type of operation, runway, time of day, aircraft type, air route and sub-track.

Type of operation	Runway	Time of day	Aircraft type	Air route (cluster)	Sub-track	Number of operations
Departure	19R	daytime	B738	28_19R	0	1,054
Arrival	26	daytime	B738	15_26	0	934
Departure	19R	daytime	A20N	28_19R	0	916
Arrival	26	daytime	A20N	15_26	0	683
Departure	19R	daytime	A20N	29_19R	0	671
Departure	19R	daytime	B738	29_19R	0	664
Departure	19R	daytime	B738	28_19R	1A	608
Departure	19R	daytime	A320	28_19R	0	608
Departure	19R	daytime	B738	28_19R	1B	585
…	…	…	…	…	…	…

described in Section 3.5 over a 50x50 km^2 grid of receivers centred around the airport reference point with a spatial resolution of 500 m.

It is noted that the number of future flight operations is generally unknown and has to be derived from travel demand forecasts (EUROCONTROL, 2025), but in this work it was already available and it was used as such for the validation of the proposed methodology. Concerning its computational burden, on a common personal computer the preliminary noise computation took about 10 h, while the generation of the new air traffic distribution and the estimation of its acoustic impact required about one minute. Conversely, a direct computation of the noise footprint after generating the new traffic scenario would take several hours for each input change. This approach was chosen also with a view to future developments targeting longer-term future air traffic scenarios, whose higher uncertainty will require fast simulation of the impacts of variable fleet mix, airport operational patterns, and travel demand estimates on the airport emissions.

The quality of the generated air traffic distribution is examined by using L_{DEN} maps as the key means of assessment. This is done by comparing noise footprints based on the actual air traffic in 2024, computed using the modelling tool outlined in Section 2, against the results for the 2024 traffic distribution generated from the 2022 data. These noise maps are shown in Fig. 11, where on the left-hand side is the noise map for the 2024 annual L_{DEN}, while on the right-hand side is the map generated from the simulated traffic distribution. The analysis of the two noise contour maps reveals immediately a good agreement between actual and simulated scenarios: the noise regions with $L_{DEN} \geq 50$ dB(A) cover essentially the same area near and along the airport runways, and low-noise regions having $L_{DEN} < 50$ dB(A) show similar shapes. However, a few differences emerge for low L_{DEN} , such as those to the east of the airport along runway 08/26.

To carry out a quantitative assessment of these differences, the simulated noise map was compared directly against the actual one by taking the differences ΔL_{DEN} between the two for each grid point. The resulting differential contour plot is shown on the left-hand side of Fig. 12, which shows that across around 85 % of the area surrounding Arlanda Airport the $\Delta L_{DEN} = L_{DEN,simulated} - L_{DEN,actual}$ values remain below 2.5 dB(A), indicating that the proposed methodology manages to reproduce with good fidelity the spatial distribution of noise exposure. This is also confirmed by the dispersion in the upper right plot of Fig. 12, which indicates that in the area enclosed by the actual 45 dB(A) iso-level around 92.5 % of the grid points have ΔL_{DEN} values that fall into the $[-2.5, 2.5]$ dB(A) range. Deviations larger than 2.5 dB(A) exist, but they occur primarily in regions where the overall noise exposure is lower than 45 dB(A), as indicated by the dispersion in the lower right plot of Fig. 12. Therefore, such deviations, which originate mostly from the few actual trajectories classified as outliers in the clusters of Section 3.1.1 and thus were ignored in the simulated traffic distribution, are of limited practical relevance for airport noise impact assessments (ECAC, 2016a).

These findings are further supported by a quantitative comparison of the areas enclosed by five L_{DEN} iso-levels conducted through the relative area difference, ΔA, for each iso-level, as shown in Table 7. The comparison shows that ΔA drops from a rather low value, less than 6 %, at 45 dB(A), to essentially zero at 65 dB(A), which suggests that while the simulated traffic distribution provides an overall reliable reconstruction of actual airspace utilization, some differences could be related to minor changes in air route usage. Indeed, these changes would cause negligible differences near the airport, but far away from it the gap between the two noise footprints would widen, as it actually happens in the present case.

The low relevance of these contour area differences is supported by the analysis of their impact on the population. In Fig. 13 the actual and simulated 45- and 50-dB(A) L_{DEN} contours are overlaid onto the population density maps of Sweden in 2024, obtained as square tiles with 30 arc-second (about 1 km) resolution from WorldPop datasets (Bondarenko et al., 2025). The three cities near the airport from most to least populated are Upplands Väsby, Märsta, and Brunna, and the figure

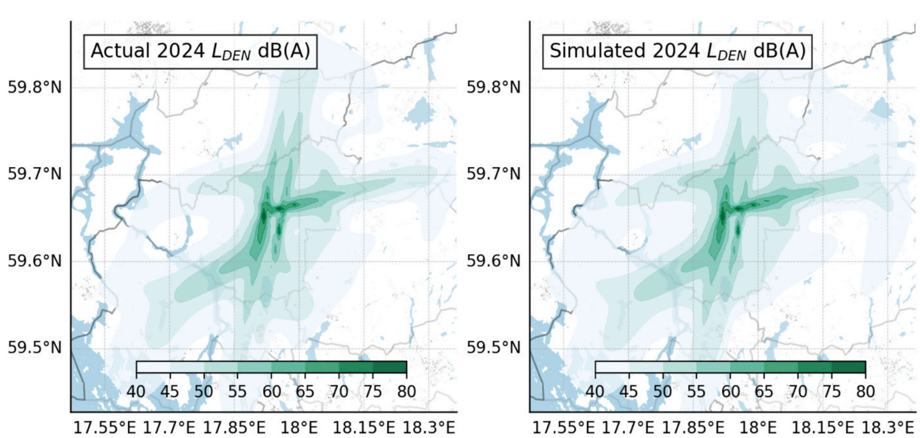

Fig. 11. Annual L_{DEN} noise maps at Arlanda Airport: based on actual ADS-B data for 2024 (left), and based on the simulated traffic distribution for 2024 derived from 2022 ADS-B data (right).

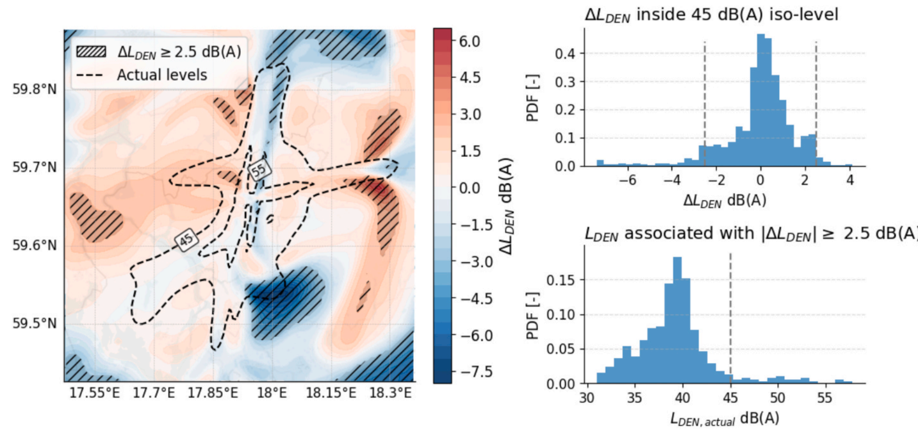

Fig. 12. Difference between simulated and actual 2024 L_{DEN} noise maps expressed as a contour plot of $\Delta L_{DEN} = L_{DEN,simulated} - L_{DEN,actual}$ (left), frequency of grid points with ΔL_{DEN} values inside the actual 45 dB(A) iso-level (upper right), and frequency of points with $L_{DEN,actual}$ values having $|\Delta L_{DEN}| \geq 2.5$ dB(A) (lower right).

Table 7
Contour areas associated with multiple L_{DEN} values for actual and simulated air traffic distributions, and relative area differences ΔA between each of them.

Level [dB(A)]	A_{actual} [km^2]	$A_{simulated}$ [km^2]	$\Delta A[\%] = \frac{A_{simulated} - A_{actual}}{A_{actual}} \cdot 100$
45	458.67	485.64	5.88
50	174.69	184.20	5.44
55	56.54	59.04	4.43
60	18.30	19.07	4.21
65	6.35	6.34	−0.11

shows that the slightly larger simulated areas cause visible yet small differences only for the two largest cities, with the 45-dB(A) contour enclosing around 3,000 more people (+12 %). Such differences become negligible for the 50-dB(A) contours, which include only the outskirts of Upplands Väsby and Märsta for both maps and result in just a minor overestimation (+2.6 %) of people exposed. Although this good result is specific to Arlanda Airport and cannot be extended to airports that, for instance, lie closer to large cities, it is also a strong indicator of the potential of the proposed methodology for reliable estimation of the detrimental impact of airport activity.

Having illustrated the minor practical impact of the estimated noise level discrepancies, insight into their possible relation with evolving operational patterns at the airport between 2022 and 2024 is cast in Fig. 14. Firstly, the bar plot on the left-hand side of Fig. 14 shows that the proportion of flights involving new-generation aircraft (e.g. Airbus A320 Neo, Boeing 737 Max 8) increased significantly, while the share of older types (e.g. Boeing 737–800) decreased. This shift in fleet mix, unaccounted for in the simulation, indicates a gradual renewal trend that may explain the slight overestimation of contour areas, since new aircraft are quieter than old ones (Simons et al., 2026). Secondly, the two bar plots on the right-hand side of Fig. 14 indicate that also runway usage patterns changed between 2022 and 2024. For departures, the overall distribution is rather consistent between the two years, with minor differences only for runways 01L and 19R, but for arrivals more significant variations are reported, with a notable increase in the use of runway 19L balanced by a less used runway 26. This change should affect primarily the spatial distribution of flight trajectories, which agrees with local $|\Delta L_{DEN}|$ peaks in the noise difference map of Fig. 12.

Finally, the analysis of air routes (i.e. clusters) is conducted in Fig. 15 by showing cluster centroids and their usage frequency for both actual and simulated 2024 traffic. On the one hand, between the two distributions the number of centroids is basically preserved, and each simulated one matches its actual counterpart rather well in both location and shape. This consistency is responsible for the strong agreement between actual and simulated noise maps, strengthening the validity of the proposed methodology. On the other hand, differences are observed in the centroid frequency values, which can in part be ascribed to the different runway usage (e.g. arrivals shifting from runway 26 to runway 19L), but cannot in other part (e.g. the increase in departures from runway 19R towards the north-west and their decrease towards the south). These changes may be explained by a shift in destinations covered by the airlines operating at Arlanda, but this methodology does not consider yet market demand forecasts that would enable such a complex prediction. Nevertheless, when focused on short-term scenarios, the

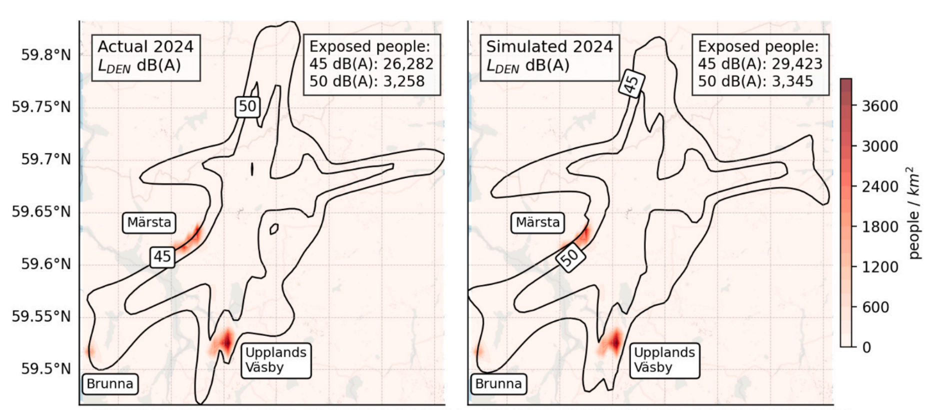

Fig. 13. Actual and simulated 45- and 50-dB(A) L_{DEN} contours for 2024 overlaid onto the population density map of the area surrounding Arlanda Airport in the same year, including estimates of people affected by such levels.

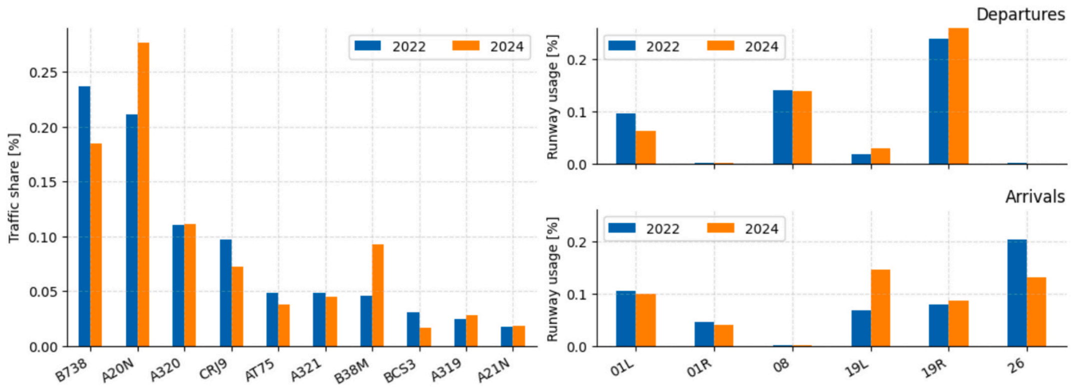

Fig. 14. In both 2022 and 2024, proportion of flights associated with different aircraft types (left), and departure and arrival runway usage (right).

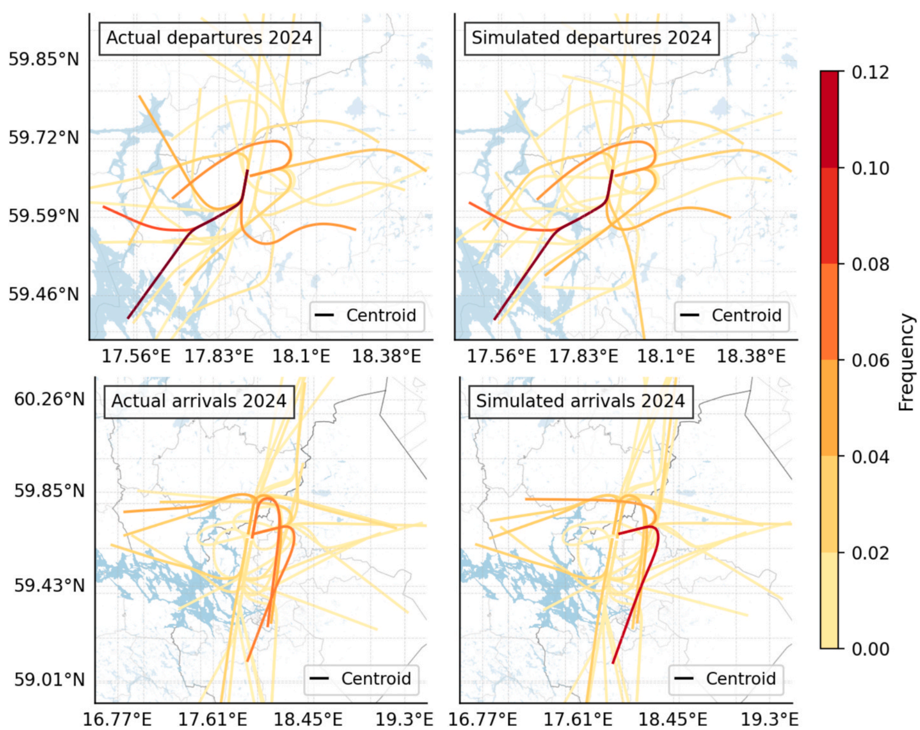

Fig. 15. Cluster frequency represented through ground track centroids in actual (left) and simulated (right) 2024 air traffic distributions, further split into departures (top) and arrivals (bottom).

proposed approach appears to work with satisfactory accuracy and reliability.

5. Conclusions and future developments

The detrimental impact of civil air traffic in airport areas has fostered multiple studies on aircraft operations in TMAs, aiming to investigate and possibly predict flight trajectories. In this context, and exploiting large amounts of flight tracking data using modelling tools previously developed by the authors, this work proposes a novel data-driven methodology for predicting and validating future air traffic distributions in airport TMAs. Firstly, the HDBSCAN clustering algorithm is applied to the tracking data to extract the clusters representing the air routes, and seven sub-tracks for each cluster are defined on a statistical basis. Secondly, flight profiles hosting the aircraft performance parameters are generated from the tracking data, selecting one representative profile for each aircraft type. Thirdly, statistical distributions of runway usage, traffic allocation, and air route selection by aircraft type are computed. These elements feed an algorithm that generates the future air traffic distributions after setting only number of departures and arrivals and forecast timeframe. Finally, after a preliminary computation of flight paths and single-operation noise levels, the new air traffic distribution and its noise impact are estimated very quickly.

Stockholm Arlanda Airport has been selected as the case study for evaluating the proposed methodology, which has been tested using ADS-B data from 2022 to simulate flight paths and resulting noise maps for 2024 and validated against the outputs of actual 2024 air traffic. The accuracy has been assessed by comparing actual and simulated L_{DEN} maps, which has shown that in areas affected by $L_{DEN} \geq 45$ dB(A) about 92.5 % of the time the deviation is contained within the ± 2.5 dB(A) range, with an overestimation lower than 6 % of contour areas enclosed by iso-levels at or beyond 45 dB(A). The small discrepancies are explained by fleet mix evolution, runway usage changes, and different air route frequencies.

Thanks to the fully data-driven approach, the proposed methodology seems promising for an accurate prediction of air traffic distributions around airports. Moreover, it does not require preliminary acquisition of procedural information (e.g. aeronautical charts), it is based on open data, which enables analysis of any airport worldwide, and it appears to provide good predictions of both air traffic and noise levels. These elements suggest a strong potential for reliable generation of alternative air traffic scenarios, especially for short-term predictions, that could also be embedded into tools for analysis and forecast of airport activity. However, limitations arise for longer-term predictions, where fleet renewal plays a larger role, traffic growth becomes hard to quantify, and procedural or infrastructural changes may occur at the airport. The assumption of stationary operational patterns will be relaxed in future developments, formulating data-driven approaches also to assess the impacts of fleet mix evolution, variations in runway usage, increasing traffic demands, potential airport layout modifications (e.g. new runways), and incorporation of airport activity data (e.g. gate usage) if available. The use of more advanced machine learning techniques will be also evaluated, aiming to capture additional non-stationary trends that could enhance the prediction of both short- and long-term future air traffic distributions.

Funding statement

The authors are grateful to the European Commission for supporting this work, performed within the NEEDED project, funded by the European Climate, Infrastructure and Environment Executive Agency (Grant Agreement no. 101095754). This publication solely reflects the authors' view and neither the EU, nor the Funding Agency can be held responsible for the information it contains.

CRediT authorship contribution statement

Lorenzo Dorbolò: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Marco Pretto:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Conceptualization. **Pietro Giannattasio:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used in this work are freely available from the sources referenced in the manuscript.

References

Bao, J., et al., 2024. Exploring the conflict risk characteristics of air weaving sections in Metroplex terminal areas with flight trajectory data and adaptive graph spatial-temporal transformer. *J. Air Transp. Manage.,* Volume 120, p. 102667. doi: 10.1016/j.jairtraman.2024.102667.

Bendtsen, K.M., et al., 2021. A review of health effects associated with exposure to jet engine emissions in and around airports. Environ. Health 20 (1), 10. https://doi.org/10.1186/s12940-020-00690-y.

Bondarenko, M., et al., 2025. *Constrained estimates of 2015-2030 total number of people per grid square at a resolution of 30 arc (approximately 1km at the equator) R2025A version v1. Global Demographic Data Project.* Funded by The Bill and Melinda Gates Foundation (INV-045237): WorldPop - School of Geography and Environmental Science, University of Southampton. doi:10.5258/SOTON/WP00840.

Campello, R. J. G. B. et al., 2013. Density-Based Clustering Based on Hierarchical Density Estimates. In: *Advances in Knowledge Discovery and Data Mining.* Berlin: Springer, pp. 160-172. doi: 10.1007/978-3-642-37456-2_14.

Condé Rocha Murça, M. et al., 2016. Trajectory Clustering and Classification for Characterization of Air Traffic Flows. In: *16th AIAA Aviation Technology, Integration, and Operations Conference.* Washington D.C., USA: AIAA, pp. 3760. doi: 10.2514/6.2016-3760.

EASA - ANP, 2025. *Aircraft Noise and Performance (ANP) Data.* [Online] Available at: https://www.easa.europa.eu/en/domains/environment/policy-support-and-research/aircraft-noise-and-performance-anp-data [Accessed 25 September 2025].

EASA, 2025. *ANP aircraft substitutions.* [Online] Available at: https://www.easa.europa.eu/en/domains/environment/policy-support-and-research/aircraft-noise-and-performance-anp-data/anp-aircraft-substitutions [Accessed 18 June 2025].

ECAC, 2016a. *Doc 29 4th Edition: Report on Standard Method of Computing Noise Contours around Civil Airports - Volume 1: Applications Guide,* Neuilly-sur-Seine: ECAC.

ECAC, 2016b. *Doc 29 4th Edition: Report on Standard Method of Computing Noise Contours around Civil Airports - Volume 2: Technical Guide,* Neuilly-sur-Seine: ECAC.

Eerland, W.J., et al., 2016. Modeling the dispersion of aircraft trajectories using gaussian processes. J. Guid. Control Dyn. 39 (12), 2661–2672. https://doi.org/10.2514/1.G000537.

Eerland, W. J. & Box, S., 2016. Trajectory Clustering, Modeling and Selection with the focus on Airspace Protection. In: *AIAA Infotech @ Aerospace.* San Diego, CA, USA: AIAA, pp. 1411. doi: 10.2514/6.2016-1411.

Einicke, K., Kennedy, J., 2025. Predicting airport noise impact to 2040: Traffic growth and technology uptake. Appl. Acoust. 227, 110229. https://doi.org/10.1016/j.apacoust.2024.110229.

EUROCONTROL, 2025. *EUROCONTROL Forecast Update 2025-2031 - Autumn 2025.* [Online] Available at: https://www.eurocontrol.int/publication/eurocontrol-forecast-update-2025-2031 [Accessed 7 November 2025].

European Commission, 2012. *Flightpath 2050 – Europe's vision for aviation,* Luxembourg City: Publications Office of European Union.

Gariel, M., et al., 2011. Trajectory Clustering and an Application to Airspace monitoring. IEEE Trans. Intell. Transp. Syst. 12 (4), 1511–1524. https://doi.org/10.1109/TITS.2011.2160628.

Gaume, K., et al., 2025. Extracting aircraft conflict-resolution situations from historical ADS-B data. Transp. Res. Interdiscip. Perspect. 34, 101669. https://doi.org/10.1016/j.trip.2025.101669.

Giladi, R., Menachi, E., 2024. Validating aircraft noise models: Aviation environmental design tool at Heathrow. J. Air Transp. Manage. 116, 102557. https://doi.org/10.1016/j.jairtraman.2024.102557.

Guruprasad, S. M. et al., 2023. A quantitative approach to density-based clustering of flight trajectories for efficient air traffic noise simulations. In: *INTER-NOISE and NOISE-CON Congress and Conference Proceedings.* Glasgow, UK: INCE-USA, pp. 5045-5056. doi: 10.3397/IN_2022_0729.

Huang, F., et al., 2023. CO2 emission change in China's aviation industry: a fleet-wide index decomposition and scenario analysis. Transp. Res. Part D: Transp. Environ. 119, 103743. https://doi.org/10.1016/j.trd.2023.103743.

IATA, 2025. *Global Outlook for Air Transport.* [Online] Available at: https://www.iata.org/en/iata-repository/publications/economic-reports/global-outlook-for-air-transport-june-2025/[Accessed 1 December 2025].

Krauth, T., et al., 2023. Deep generative modelling of aircraft trajectories in terminal maneuvering areas. Mach. Learn. Appl. 11, 100446. https://doi.org/10.1016/j.mlwa.2022.100446.

Luftfartsverket, 2026. *AROWeb – AIS MET and Voyage Planning – EAIP.* [Online] Available at: https://aro.lfv.se/Editorial/View/IAIP?folderId=79[Accessed 2 April 2026].

Madar, S., et al., 2021. *Application of Trajectory Clustering for Aircraft Conflict Detection.* San Antonio, TX, USA, 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC). doi: 10.1109/DASC52595.2021.9594344, pp. 1-9.

O'Neill, M.E., 2014. PCG: a family of simple fast space-efficient statistically good algorithms for random number generation. ACM Trans. Math. Softw. 204, 1–46.

Olive, X., 2019. Traffic, a toolbox for processing and analysing air traffic data. *J. Open Source Softw.,* 4(39), pp. 1518, 1-3. doi: 10.21105/joss.01518.

Olive, X. & Basora, L., 2019. *Identifying Anomalies in past en-route Trajectories with Clustering and Anomaly Detection Methods.* Vienna, ATM Seminar 2019. doi: /hal-02345597.

Olive, X., Morio, J., 2019. Trajectory clustering of air traffic flows around airports. Aerosp. Sci. Technol. 84, 776–781. https://doi.org/10.1016/j.ast.2018.11.031.

Pedregosa, F., et al., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830. https://doi.org/10.48550/arXiv.1201.0490.

Preisendörfer, P., et al., 2022. Annoyance due to residential road traffic and aircraft noise: Empirical evidence from two European cities. Environ. Res. 206, 112269. https://doi.org/10.1016/j.envres.2021.112269.

Pretto, M., et al., 2019. Web data for computing real-world noise from civil aviation. Transp. Res. Part D: Transp. Environ. 69, 224–249. https://doi.org/10.1016/j.trd.2019.01.022.

Pretto, M., et al., 2020. Forecasts of future scenarios for airport noise based on collection and processing of web data. Eur. Transp. Res. Rev. 12 (1), 4. https://doi.org/10.1186/s12544-019-0389-x.

Pretto, M., et al., 2022. Mixed analysis-synthesis approach for estimating airport noise from civil air traffic. Transp. Res. Part D: Transp. Environ. 106, 103248. https://doi.org/10.1016/j.trd.2022.103248.

Pretto, M., et al., 2024. Aircraft operation reconstruction and airport noise prediction from high-resolution flight tracking data. Transp. Res. Part D: Transp. Environ. 135, 104397. https://doi.org/10.1016/j.trd.2024.104397.

Rousseeuw, P., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Schäfer, M. et al., 2014. Bringing up OpenSky: A large-scale ADS-B sensor network for research. In: *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks.* Berlin, Germany: IEEE. doi: 10.1109/IPSN.2014.6846743, pp. 83-94.

Sheather, S.J., Jones, M.C., 1991. A Reliable Data-based Bandwidth selection Method for Kernel Density Estimation. J. r. Stat. Soc. B 53 (3), 683–690 https://www.jstor.org/stable/2345597.
Simonetti, I., et al., 2015. Air quality impact of a middle size airport within an urban context through EDMS simulation. Transp. Res. Part D: Transp. Environ. 40, 144–154. https://doi.org/10.1016/j.trd.2015.07.008.
Simons, D.G., et al., 2026. Contribution of aircraft types to noise levels across the NOMOS network of Amsterdam Airport Schiphol. J. Air Transp. Manage. 131, 102910. https://doi.org/10.1016/j.jairtraman.2025.102910.
Swedavia Airports, 2025. *Stockholm Arlanda Airport - About the airport.* [Online]. Available at: https://www.swedavia.com/arlanda/about-the-airport/[Accessed 18 September 2025].
The OpenStreetMap Foundation, 2025. *OpenStreetMap.* [Online] Available at: https://www.openstreetmap.org/about/[Accessed 26 September 2025].
Topriceanu, C.-C., et al., 2025. Higher aircraft noise exposure is linked to worse heart structure and function by cardiovascular MRI. J. Am. Coll. Cardiol. 85 (5), 454–469. https://doi.org/10.1016/j.jacc.2024.09.1217.
Wang, R., et al., 2025. Aircraft approach risk identification based on cluster analysis of flight surveillance data in terminal area. In: *8th International Conference on Artificial Intelligence and Big Data.* Chengdu, China: IEEE, pp. 93-98. doi: 10.1109/ICAIBD64986.2025.11081987.
Wu, X., et al., 2022. Long-term 4D trajectory prediction using generative adversarial networks. Transp. Res. Part C Emerging Technol. 136, 103554. https://doi.org/10.1016/j.trc.2022.103554.
Xie, J., et al., 2022. Novel methodologies for the development of large-scale airport noise map. Sustainability 14 (11), 6573. https://doi.org/10.3390/su14116573.
Xie, J., et al., 2024. Methodologies for the prediction of future aircraft noise level. Noise Mapp. 11 (1), 20240005. https://doi.org/10.1515/noise-2024-0005.
Xu, Q., et al., 2023. Air traffic density prediction using Bayesian ensemble graph attention network (BEGAN). Transp. Res. Part C Emerging Technol. 153, 104225. https://doi.org/10.1016/j.trc.2023.104225.
Xu, Q., et al., 2024. PIGAT: Physics-informed graph attention transformer for air traffic state prediction. IEEE Trans. Intell. Transp. Syst. 25 (9), 12561–12577. https://doi.org/10.1109/TITS.2024.3386128.
Zeng, W., et al., 2021. Aircraft trajectory clustering in terminal airspace based on deep autoencoder and gaussian mixture model. Aerospace 8 (9), 266. https://doi.org/10.3390/aerospace8090266.
Zhu, H., et al., 2025. Dynamic assessment of airport noise impact: a case study of Shenzhen airport. Transp. Res. Part D: Transp. Environ. 143, 104741. https://doi.org/10.1016/j.trd.2025.104741.
Zhu, X., et al., 2023. Predicting aircraft trajectory uncertainties for terminal airspace design evaluation. J. Air Transp. Manage. 113, 102473. https://doi.org/10.1016/j.jairtraman.2023.102473.