TRANSPORTES
revistatransportes.org.br
anpet
associação nacional de pesquisa e ensino em transportes
anpet.org.br

Data-driven modeling of urban airspace availability for air mobility operations in the São Paulo Metropolitan Region

Modelagem baseada em dados da disponibilidade do espaço aéreo urbano para operações de mobilidade aérea na Região Metropolitana de São Paulo

João Vitor Turchetti Ribeiro[1], Mayara Condé Rocha Murça[1]

[1]Aeronautics Institute of Technology, São José dos Campos, São Paulo, Brasil

Contato: joaojvtr@ita.br, (JVTR); mayara@ita.br (MCRM)

Submitted:
3 April, 2023
Accepted for publication:
3 February, 2024
Published:
16 April, 2024
Associate Editor:
Li Weigang, Universidade de Brasília, Brasil

Keywords:
Urban air mobility.
Air traffic management.
Airspace management.
Machine learning.

Palavras-chave:
Mobilidade aérea urbana.
Gerenciamento de tráfego aéreo.
Gerenciamento do espaço aéreo.
Aprendizado de máquina.

DOI: 10.58922/transportes.v31i2.2896

ABSTRACT

Urban Air Mobility (UAM) is an emerging form of transportation that is expected to introduce novel flight networks into already busy and complex airspace surrounding major cities and metropolitan regions. This work provides a data-driven approach to modeling the urban airspace availability for emerging UAM operations toward supporting their safe and efficient integration. Using historical flight tracking data, clustering analysis is first performed to learn the current patterns of urban airspace use by conventional traffic and identify the airspace volumes that are least constrained and best accessible for UAM flights. Meteorological data is then incorporated into the machine learning framework to create a probabilistic model of the spatiotemporal distribution of conventional traffic flows. This model enables the prediction of active airport arrival/departure patterns and the resulting airspace availability for UAM given dynamic operational conditions. The data-based approach is demonstrated for the São Paulo metropolitan area, which is the largest in Brazil and a promising market for UAM. It allowed for a high-fidelity characterization of the São Paulo urban airspace use patterns as well as for accurate predictions of the available airspace for UAM, bringing novel insights and capabilities in support of dynamic and efficient urban airspace management.

RESUMO

A Mobilidade Urbana Aérea (UAM) é uma modalidade emergente de transporte urbano que deverá introduzir novas rotas de voo no já ocupado e complexo espaço aéreo das principais cidades do mundo. Este artigo oferece uma abordagem baseada em dados para modelar a disponibilidade do espaço aéreo urbano para operações UAM, visando servir de suporte técnico à sua integração segura e eficiente. Utilizando dados históricos de rastreamento de voos convencionais, um processo de clusterização é realizado para aprender padrões de uso do espaço aéreo urbano pelo tráfego aéreo convencional, bem como para identificar as porções do espaço aéreo menos utilizadas por eles e, portanto, mais acessíveis para voos UAM. Em seguida, dados meteorológicos são incorporados para criar um modelo probabilístico da distribuição espaço-temporal dos fluxos de tráfego aéreo convencional. Este modelo possibilita a previsão de procedimentos ativos de chegada e de saída dos principais aeroportos da cidade e, consequentemente, a obtenção da disponibilidade resultante do espaço aéreo urbano para operações UAM. O modelo é aplicado à região metropolitana de São Paulo, que, por ser a maior do Brasil, será um mercado promissor para a implementação de UAM em um futuro próximo. Ademais, nossa abordagem permite uma caracterização de alta fidelidade dos padrões de uso do espaço aéreo urbano de São Paulo, bem como previsões precisas do espaço aéreo disponível para uso por novas operações UAM, permitindo a obtenção de insights e a compreensão de capacidades inovadoras em suporte ao gerenciamento dinâmico e eficiente do espaço aéreo urbano.

1. INTRODUCTION

Every day, billions of people around the world have the basic need of moving around cities to commute to work or accomplish personal duties. Some of them travel dozens of miles through clogged roads to reach their destination, spending more time than desired. This is a common reality especially in large metropolitan regions that face significant levels of transportation congestion on the ground. In São Paulo, for instance, research performed by the Brazilian Institute of Public Opinion and Statistics (IBOPE, 2014) estimated that the average citizen spends up to 167 minutes daily in road traffic while commuting, with 70% of the interviewees considering the traffic situation bad or very bad. As a result, these megacities have been increasingly affected by congestion externalities, such as reduced productivity, economic losses and environmental impacts.

Urban Air Mobility (UAM) is an emerging concept of transportation of people and cargo that is expected to alleviate congestion on the ground and reduce the costs of traveling within cities by leveraging novel types of aircraft that are safe, quiet and efficient, known as Electric Vertical Takeoff and Landing vehicles (eVTOL), to provide regular and on-demand transportation services. A 74 billion dollar UAM market is expected to rise in the next decade, with up to 23,000 eVTOL vehicles operating globally (EmbraerX, 2019).

High demand for UAM is expected to emerge in busy urban clusters (Goyal et al., 2021). This high volume of operations will lead to new flight networks, notably in low altitude layers (Vascik et al., 2019). However, the urban airspace above highly populated cities is already organized into dense route structures that properly accommodate existing conventional traffic, under both Instrument Flight Rules (IFR) and Visual Flight Rules (VFR). In large metropolitan regions served by multi-airport systems (metroplexes), the terminal airspace structure can be particularly complex due to the need to prevent conflicts between aircraft flying the procedures from different closely located airports. This often results in longer departure and arrival routes, which ultimately translates into delays and lower fuel efficiency. Hence, the introduction of novel UAM operations to those busy airspace becomes a great challenge.

Several stakeholders have worked towards the definition of a Concept of Operations (ConOps) for UAM integration into urban airspace. Since legacy ATM systems are already operating near capacity in busy terminal areas and are unlikely to be able to handle such a rise in traffic in the short term, UAM operations are expected to be initially integrated into urban airspace without interfering with conventional operations and compromising current safety and efficiency levels (EmbraerX, 2019; FAA, 2020; Lascara et al., 2020). To minimize the impacts on current ATM, one possible approach is to identify airspace volumes that are procedurally separated from existing published flight procedures, i.e., for which the minimum standard separation requirements are planned to be automatically met without requiring Air Traffic Control (ATC) services from the legacy ATM system (Lascara et al., 2020; Vascik et al., 2019). Given the high variability in how published procedures are actually flown due to tactical ATC interventions, aircraft performance, among other factors, such identification of available airspace volumes ideally should be backed by operational data. The increasing availability of detailed aircraft surveillance data from open sources is certainly an ally in this process, allowing for precise mapping of current patterns of airspace use.

Another potential approach for UAM integration is to adopt a dynamic airspace management solution, which leverages the dynamics of conventional traffic flows, as proposed in the ConOps documents (EmbraerX, 2019; FAA, 2020; Lascara et al., 2020). To allow for increased accessibility of UAM operations in controlled airspace and enable its scalability, the allocation of the urban airspace and the activation and deactivation of routes for UAM should be done dynamically, in

response to changes in conventional air traffic patterns due to operational conditions. In that case, the ability to predict airspace availability given the dynamics of conventional air traffic operations is required for decision-making. Yet, limited studies have attempted to address this problem towards providing such operational decision support.

This study presents a data-driven approach to modeling the urban airspace availability for emerging air mobility operations. We leverage machine learning methods to perform descriptive analysis and predictive modeling of the current patterns of urban airspace use by conventional air traffic. The resulting data-driven models allow for the identification of the airspace volumes that are least constrained and best accessible for UAM flights. Moreover, they provide the basis for predictive tools in support of dynamic urban airspace management.

In order to demonstrate the use of the modeling proposed, we study the case of São Paulo, which is the main financial and economic center of Brazil and one of the largest cities in the world, with a population of about 12 million residents (IBGE, 2023). The economic strength and the routine high levels of ground traffic congestion have made the use of helicopters very common for urban mobility within the São Paulo region. The Brazilian Association of Helicopter Pilots (Abraphe, 2019) estimated that approximately 1,300 helicopter landings and takeoffs are performed daily in the city of São Paulo. Therefore, it is expected to be a promising market for UAM.

2. LITERATURE REVIEW

The topic of urban airspace design and management for integration of emerging UAM operations has been the subject of numerous studies. Different authors have addressed the matter using different techniques, from geometric approaches supported by published procedures and topographic data, to flight trajectory data analytics based on the application of machine learning algorithms on Automatic Dependent Surveillance-Broadcast (ADS-B) data.

Vascik et al. (2019) developed a geometric framework that used only static, publicly available information, evaluated in multiple scenarios. The authors defined four ConOps scenarios and studied seven airspace constructs, such as terrain clearance, airports airspace clearance, special flight rules areas, and others. Then, geometrically combining all constructs, they were able to estimate the airspace availability for UAM under their specific hypotheses and scenarios.

The fashion of how topography and urban infrastructure affect airspace availability is a theme of great importance because complex and large cities are exactly the environment in which eVTOLs are expected to operate. This line of research has also been addressed by Cho and Yoon (2018) and Mohamed Salleh et al. (2018). The first proposed a topological analysis framework to identify free and usable airspace in a 3D environment filled with geometric elements; the 3D environment was the urban environment, and the geometric elements were buildings, skyscrapers or any other physical structures. The second had similar objectives, but more oriented towards the establishment of urban route networks.

Regarding data-driven approaches, most of them resort to flight tracking data gathered from surveillance systems. Vascik and Hansman (2019) developed a statistical approach based on percentile metrics defined over data to evaluate airspace availability. The authors used commercial, rotary-wing and general aviation flight data from Airport Surface Detection Equipment - Model X (ASDE-X) to build the concept of containment boundaries: regions around multiple observed flight trajectories delimiting keep-out regions where UAM must not enter. The containment boundaries defined in Vascik and Hansman (2019) were used by Vascik and Hansman (2020) as

inputs to create airspace cutouts especially designed to procedurally segregate UAM operations from conventional traffic with a special focus on ATC workload.

Burke (2019) used Automatic Dependent Surveillance-Broadcast (ADS-B) to evaluate manned air traffic below 500 ft AGL within a 5 miles radius from John Wayne Airport (SNA). The author's objective was to offer a new paradigm for UAS safety near airports. He concluded that most of the volumes of highly populated airspace near airports were located at some distance from the traffic patterns, suggesting that regions very close to airports might be available for UAS.

Murça (2021) addressed the problem of identifying available volumes of urban airspace for UAM operations using ADS-B data of arriving and departing traffic at Congonhas airport, the most central airport in São Paulo, Brazil. The author also analyzed the impact of adopting different lateral separation buffers (ATC-assumed minimum lateral separation) to procedurally separate UAM from conventional traffic, and the differences between integrating UAM dynamically or statically.

It is evident that understanding current air traffic flows is a major priority to properly assess urban airspace availability. To this end, we can leverage a significant line of research dedicated to exploiting large-scale aircraft tracking data to detect, better understand and anticipate traffic flow patterns in the airspace.

Clustering has been a prominent Machine Learning (ML) technique used to approach this pattern recognition problem. A very promising type of clustering algorithm commonly applied to flight trajectory datasets are density-based. Many authors have used the renowned Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm with different purposes (Gariel, Srivastava and Féron, 2011; Liu et al., 2017; Murça et al., 2018, 2020; Murça, 2021; Eerland, Box and Sóbester, 2016; Olive and Morio, 2018; Olive et al., 2021; Vascik and Hansman, 2019). DBSCAN works well with noise and is not limited to globular shapes, being able to discover trajectory clusters of any form. In a busy TMA environment, plenty of operational artifices generate noise (ATC vectoring, holding patterns, go-around procedures, and so on), turning DBSCAN suitable for the identification of trajectory patterns in the presence of these noise observations. The algorithm can also deliver satisfactory results without the need for the user to inform the number of clusters beforehand.

The analysis of the existing scientific literature reveals that while numerous studies have succeeded in identifying spatial and temporal patterns from flight trajectory data, only a handful have leveraged this knowledge to predict traffic distribution. For instance, knowledge about spatial trajectory patterns was used to predict route choices as well as operationally acceptable reroute options by Marcos, García-Cantú and Herranz (2018) and Evans and Lee (2019). Murça and Hansman (2018) developed a trajectory data analytics framework to identify and predict traffic flow patterns, resulting from coupled airport runway and terminal airspace configurations, in a metroplex environment. The study did not consider the forecast of the spatial distribution of air traffic within the predicted patterns, however.

This paper advances upon this particular research direction by introducing data-driven models focused on identifying and forecasting the spatiotemporal distribution of conventional air traffic in busy urban areas and the resulting airspace availability for emerging UAM operations.

3. METHODOLOGY

We developed a machine learning framework to identify and predict urban airspace availability for UAM operations. First, we process raw flight tracking data to get it prepared for a clustering analysis aimed at identifying the main conventional traffic patterns flown in the urban terminal

airspace. Then, the knowledge of actual traffic patterns is combined with weather data to learn a probabilistic model capable of predicting the traffic patterns that are active, their spatial confidence regions and the resulting airspace availability for UAM at any given time given a weather forecast. The framework is summarized in Figure 1 and detailed in this section.

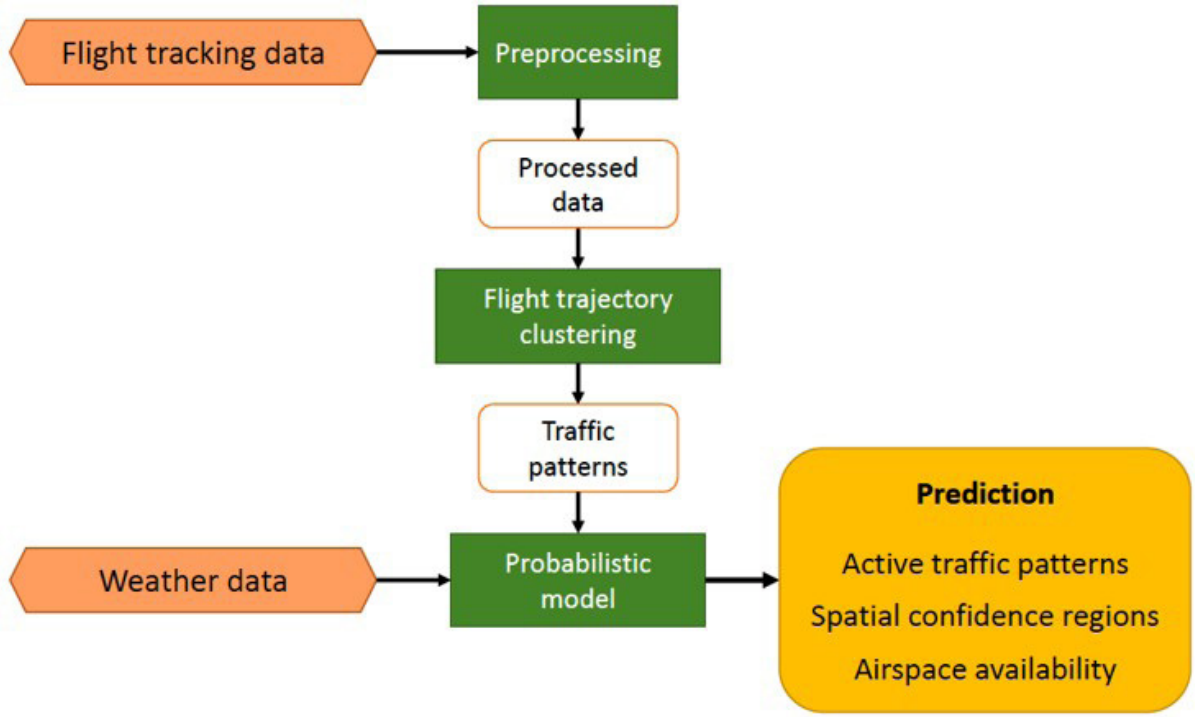


Figure 1. Machine learning framework for urban airspace availability modeling.

3.1. Region of interest

In this work, we study the case of São Paulo, which is the largest city in Brazil and one of the largest in the world, with a population of about 12 million residents (IBGE, 2023). It is part of the São Paulo Metropolitan Region, which consists of 39 municipalities and has approximately 23 million inhabitants, being the main financial and economic center of Brazil.

The São Paulo Metropolitan Region is served by two major commercial airports, São Paulo/Congonhas Airport (CGH) and São Paulo/Guarulhos International Airport (GRU). CGH is the city airport, located 6 miles from downtown São Paulo. GRU is located in the neighboring city of Guarulhos and is 17 miles from downtown São Paulo. GRU and CGH are the busiest airports in Brazil, with an average of 817 and 610 operations per day, respectively (CGNA, 2019).

Our Region of Interest (ROI) is therefore the urban airspace surrounding these airports, which is contained in the São Paulo TMA. We define the ROI as two overlapped cylinders of 40 nautical miles (NM) radius respectively centered GRU and CGH airports' centroids. Figure 2 shows a map of the ROI and its surroundings and indicates the location and the orientation of CGH and GRU runways. CGH runways are oriented northwest to southeast and are denominated 17L/35R and 17R/35L, being the latter more frequently used due to its dimensions and available IFR procedures. GRU runways are oriented west to east and are denominated 10L/28R and 10R/28L, the first being used mainly for takeoffs and the latter for landings.

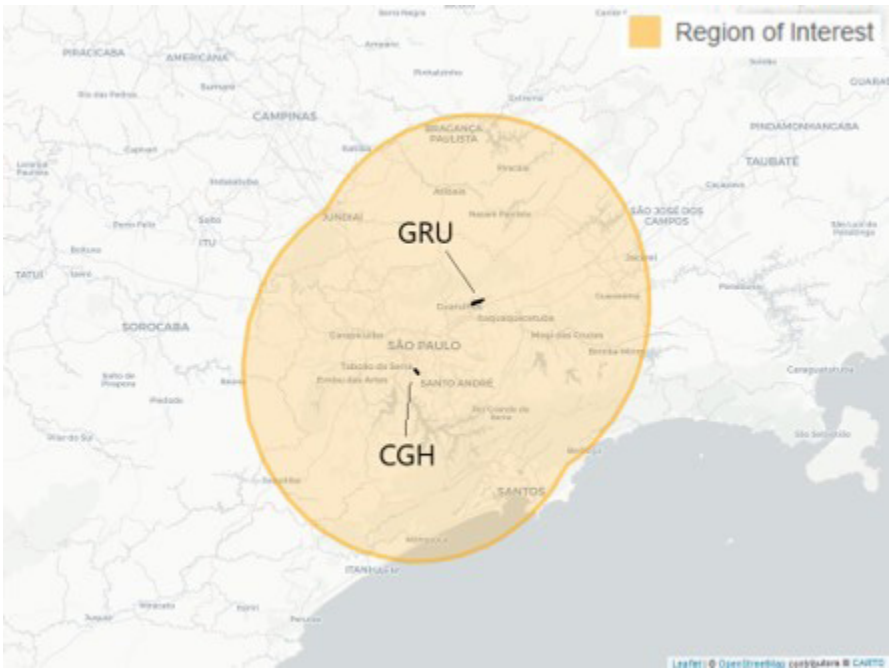


Figure 2. Region of interest.

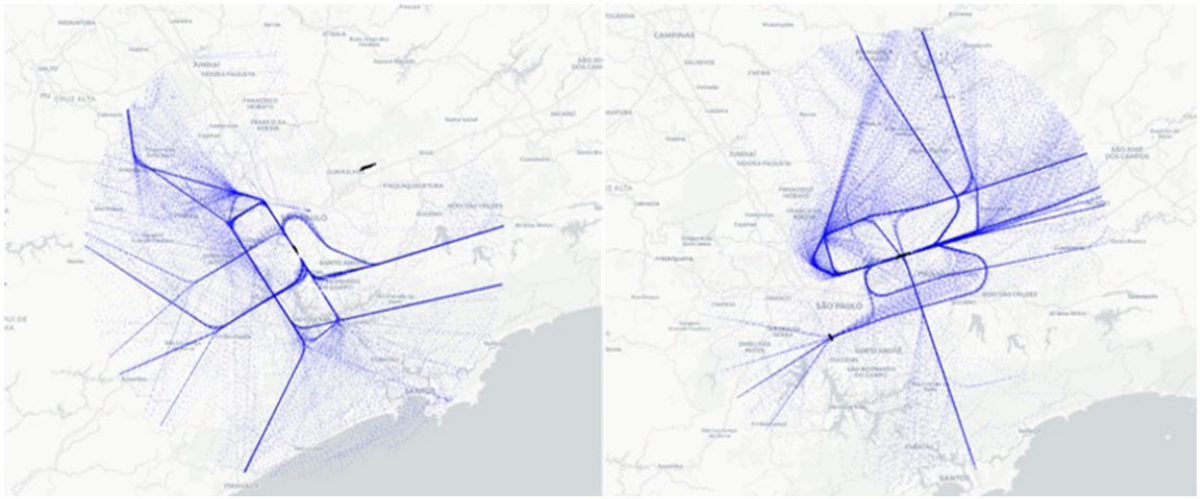
3.2. Data

3.2.1. Flight tracking data

The emergence of ADS-B turned incredibly large amounts of flight tracking data easily accessible to air traffic managers, flight planners, researchers and others. ADS-B is a surveillance technology incorporated into regular transponder units in which an aircraft automatically broadcasts its navigation data (speed, altitude, heading etc.) in regular intervals. This work uses ADS-B data gathered inside the São Paulo TMA to map arrival and departure flows at CGH and GRU and understand the behavior of commercial air traffic inside the urban airspace. For this, we rely on the OpenSky Network (Schäfer et al., 2014) databases. The OpenSky Network is a non-profit association that provides open access to real-world air traffic data to the public for scientific purposes. Volunteers and other supporters around the world collect raw flight tracking data from their own ADS-B receivers and share it with the Network.

In this study, we use 30 days of flight tracking data gathered via OpenSky's REST API. The dataset covers the period of November 2019 -- last month with regular demand levels right before the holiday season and the Covid-19 pandemic -- and consists of almost 1 million observations, corresponding to around 15,000 flights. The dataset contains, among other attributes, information about aircraft callsigns, destination and origin airports and, most importantly, trajectory data (latitude, longitude and altitude) for each timestamp. The acquisition rate is approximately 6 records per minute.

Data pre-processing was conducted to filter the observations inside the terminal area (all trajectory observations not further than 40 NM from the airport of operation and not higher than 19,500 ft were filtered) and to eliminate flights with incomplete or noisy trajectory data based on sanity checks. Figure 3 presents a visualization of the filtered flight tracking data gathered from OpenSky Network plotted over the ROI, separately for CGH and GRU. The black straight lines represent the airports' runways.



(a) Arrival/departure flight tracks at CGH (b) Arrival/departure flight tracks at GRU

Figure 3. Filtered flight tracking data in the ROI.

The aircraft tracking dataset was also preprocessed to be structured as a flight trajectory dataset. In the raw dataset, each object (line) is a single observation of some flight at a given timestamp. We changed the structure so that each flight trajectory was represented by a single line. Thereby, each object of the new dataset is a trajectory and the attributes are the sequential position data (latitude, longitude and altitude). Moreover, the number of observations per flight can vary a lot, depending on the time spent inside the ROI. Thus, in order to make the new dataset suitable for clustering, it was necessary to resample all data so that each object was represented with the same number of attributes.

3.2.2. Meteorological data

Meteorological data for the same 30-day period of the flight tracking dataset was used to learn the probabilistic model. The dataset contains information about wind, ceiling, visibility and precipitation conditions. These meteorological parameters are listed in Table 1, along with their source of acquisition and the area covered. Incorporating these inputs into our model allows for understanding the distribution of air traffic under variable weather conditions, which is crucial for developing an accurate and reliable prediction model.

Table 1: Meteorological parameters of interest.

Meteorological Parameter	Data Source	Area Covered
Wind [knots]	METAR	Airport
Ceiling [f t]	METAR	Airport
Visibility [m]	METAR	Airport
Convective precipitation rate [kg/m^{2}s]	NOAA-NCEP	Any lat/lon pair

The first three parameters were obtained from historical Meteorological Aerodrome Report (METAR) messages for CGH and GRU, which were collected via the REDEMET's API (REDEMET,

2022). METAR are coded and standardized messages issued by local aviation authorities to inform aircraft about an airport's actual meteorological situation, including wind, ceiling and visibility. Hence, we have this information for each airport of interest (separately for CGH and GRU) at each timestamp of the flight tracking database.

While the visibility parameter is used just as reported, the ceiling parameter is defined as the height of the lowest layer of Broken (BKN) or Overcast (OVC) clouds. The wind is broken down into two different parameters: headwind and crosswind. Since CGH and GRU have runways oriented in the magnetic bearings of 169º/349º and 095º/275º, respectively (as of May 2023), some computation is needed before feeding the flight tracking database with the wind components.

Although accurate and widely available, METAR messages only provide information about the weather observed at a particular airport and fail to represent a wider spatial region. Those messages may be enough to understand patterns of runway use, but give poor information when one needs to study the weather along large regions as our ROI, which is about 40 NM wide. In order to fill this gap, we resort to a numerical forecast model called National Centers for Environmental Prediction (NCEP) Reanalysis (Kanamitsu et al., 2002).

The National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Laboratory (PSL) conducts weather, climate and hydrologic research to advance the prediction of water availability and extremes. The NCEP Reanalysis is a project that combines historical observations from various sources, including satellites, weather balloons, ships, and buoys, with computer modeling to create a comprehensive record of past weather and climate conditions. The Reanalysis data covers the entire globe and ranges decades of observations, which provides a consistent and reliable dataset useful to study climate variability and change. It contains data from various numerical forecast models, including the renowned Global Forecast System (GFS) (NOAA, 2021).

The NCEP Reanalysis data includes a wide range of variables, such as temperature, pressure, wind, humidity, and precipitation. In our case, we were especially interested in the variable called *cprat.sfc*, which is the convective precipitation rate, given in kg/m^2s. The reason is that the convective precipitation rate combines both the amount of rain and the amount of convective/cumuliform clouds, which are two parameters related to microweather and which have the utmost importance for determining whether some airspace volumes should be open or closed to small aircraft as UAM.

Since the convective precipitation rate data is provided for every 1.75 degrees of latitude and longitude, we made a bivariate interpolation in order to find values of *cprat.sfc* for any possible lat/lon pair, especially for those observed in the flight tracking database. As an example, Figure 4 presents values of *cprat.sfc* in kg/m^2s as a plot on a map for a single execution of the NCEP Reanalysis, specifically November 15th at 1200Z. The nine circles on the map are the lat/lon pairs that the model actually computes, hence bivariate interpolation was necessary to determine the values of convective precipitation rate inside the entire ROI, as mentioned. Since the NCEP Reanalysis provides four executions per day for *cprat.sfc*, we were able to feed its outputs to our flight tracking database adequately at the timestamps needed and to interpolate its results to determine accurate values of convective precipitation rate for each spatial position in the database.

Being fed with all the meteorological information detailed in this subsection, along with the results of the clustering phase, the probabilistic model is capable of learning the spatial distribution of air traffic given the weather conditions.

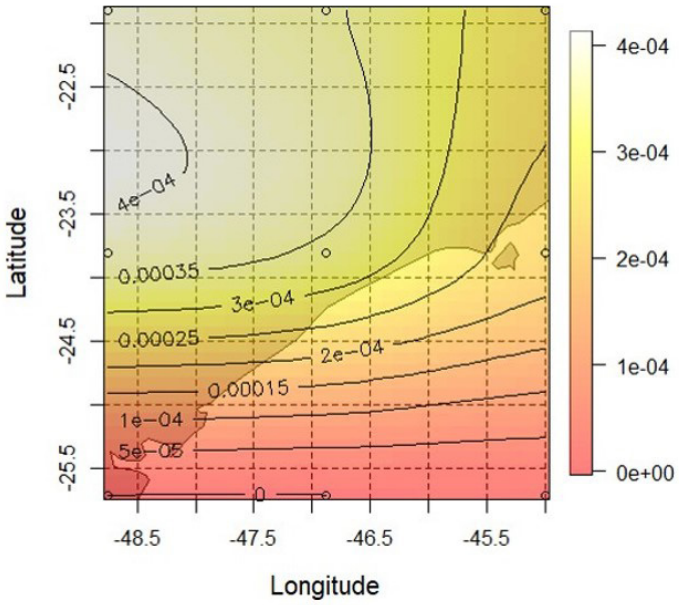


Figure 4. Visualization of the precipitation parameter cprat.sfc for a single execution of NCEP, in kg/m^{2}s.

3.3. Trajectory clustering analysis

After preprocessing the flight tracking data, we performed trajectory clustering in order to identify the most common spatial traffic patterns in the ROI. The terminal airspace is structured with several arrival and departure standard procedures that should be accomplished by flight operators to navigate between the airport and the enroute airspace. These procedures organize the traffic within the airspace surrounding the airport in a more manageable way for air traffic controllers, ensuring appropriate safety separations between aircraft coming from different directions and between aircraft and terrain/obstacles. Yet, for many reasons, standard procedures are not performed exactly as defined in aeronautical charts. A busy TMA usually makes the actual flight trajectories prone to deviations from the planned flight procedures since the high traffic volume results in the constant use of ATC vectoring, holding patterns and direct heading clearances to avoid traffic conflicts and to increase global efficiency. Therefore, air traffic flows tend to exhibit a natural variation.

To model such spatial variation from operational data, the first step of our framework was to perform a clustering analysis of arrival/departure flight trajectories in order to identify the main traffic patterns in the ROI. Clustering is an unsupervised learning method used to identify groups of similar observations in a dataset without any prior knowledge. After careful study, as detailed in section 2, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was chosen for the clustering task. The ability to handle noise and non-convex clusters and the robustness when working with an unknown number of clusters were key factors that led to this choice, besides the success achieved by other authors when using DBSCAN for clustering trajectory datasets. It is worth mentioning that the ATC artifices described above sometimes can result in significant trajectory deviations and abnormal profiles that can be considered noise, which highlights the importance of using a clustering algorithm that can deal with it.

DBSCAN is a clustering algorithm that seeks data samples of high density and expands clusters from them. Naturally, it is a density-based algorithm and may not work well with very sparse trajectories, such as an unusual route (which would be labeled as noise). DBSCAN works with two input parameters: *eps* and *min_samples*. *Eps,* or *epsilon,* or ϵ is the maximum distance between two

samples for one to be considered inside the neighborhood of the other. *Min_samples* is the minimum number of samples in a neighborhood required for a point to be considered as a core point and start a cluster. After running tests with different parameter values and performing visual validation of the clustering results, we selected the input parameters shown in Table 2. The visual validation aimed to ensure that trajectories from distinct procedures were not merged into one cluster and that trajectories following the same procedure were not split into multiple different clusters.

Table 2: DBSCAN parameters used for clustering.

	CGH DEP	CGH ARR	GRU DEP	GRU ARR
eps	1.0	1.2	1.5	2.0
min_samples	10	10	15	20

3.4. Probabilistic air traffic model

Once the clustering process is done and the air traffic flow patterns are identified, the next step is to learn a probabilistic model of the spatiotemporal distribution of air traffic that can be used to make predictions of active traffic patterns and the resulting airspace availability for UAM based on dynamic operational conditions such as weather.

For this, the trajectory clusters identified with DBSCAN were modeled in the form of a Gaussian Mixture Model (GMM) (McLachlan and Basford, 1988). We assume that the variability within traffic patterns can be modeled with a probability density function given by a weighted sum of Gaussian densities. Each Gaussian density is referred to as a component of the mixture and it models a particular trajectory cluster, i.e., the distribution of the traffic within a particular procedure. The GMM is mathematically defined as follows from Equation 1 to Equation 7:

$$p\left(X\right)=\sum_{y=1}^{K}\pi_y p(X\mid Y=y)=\sum_{y=1}^{K}\pi_y \mathcal{N}\left(X;\mu_y,\Sigma_y\right) \tag{1}$$

X is a multivariate random variable that represents the aircraft trajectory and a specific weather condition; it results from concatenating two vectors: X_T, which contains the trajectory information, and X_W, which contains the weather information; π are the mixture weights; K is the number of clusters (patterns, Gaussian components); μ is the mean vector of the Gaussian density that models the y^{th} procedure; Σ_y is the covariance matrix of the Gaussian density that models the y^{th} procedure.

Fundamentally, Gaussian component weights cannot be individually greater than one and sum up to one:

$$0\leq\pi_y\leq 1 \tag{2}$$

$$\sum_{y=1}^{K}\pi_y=1 \tag{3}$$

The multivariate variable X is composed of the vectors X_T and X_W. The former contains flight trajectory information (a set latitude and longitude values), and the latter contains weather information (specifically wind, ceiling, visibility and a numerical forecast of precipitation). We can expand them as follows:

$$X = \left[X_T \quad X_W \right] \tag{4}$$

$$X_T = \left[X_{lat} \quad X_{lon} \right] \tag{5}$$

$$X_W = \left[X_{hw} \quad X_{cw} \quad X_{ceil} \quad X_{vis} \quad X_{convp} \right] \tag{6}$$

X_{lat} and X_{lon} are the trajectory vectors describing the time series of latitude and longitude, respectively; x_{hw} and x_{cw} are meteorological variables describing headwind and crosswind component values at the time of landing/takeoff for arrival/departure flights, respectively; x_{ceil} and x_{vis} are meteorological variables describing ceiling and visibility at the time of landing/takeoff for arrival/departure flights, respectively; X_{convp} is a meteorological vector describing the convective precipitation rate along the trajectory.

The weather vector X_W contains meteorological features that are expected to cause aircraft to follow a particular flight procedure and to impact the spatial variability within the procedure. Headwind and crosswind components are the primary determinants of the active runway configuration and procedures in use in a given period of time. Moreover, the weather condition in terms of ceiling, visibility and precipitation often affects the conformance of flight trajectories, potentially impacting the spatial distribution of the traffic within the procedure. Therefore, these variables are intended to capture the influence of weather on how flight procedures are actually flown.

After the model estimation, we apply it to make predictions of active procedures, their spatial traffic distribution and the resulting airspace availability given the meteorological conditions. This is accomplished by the computation of marginal and conditional densities. Equation 4 expresses the probability of occurrence of the y^{th} procedure given known weather conditions. Bayes's theorem is used to calculate that from the marginal distribution of weather conditions XW given the y^{th} procedure, which is Gaussian with mean vector $\mu_{y,W}$ and covariance matrix $\Sigma_{y,WW}$, as in Equation 8. This allows us to express the probability of occurrence of the y^{th} procedure given input weather conditions as in Equation 9.

$$p\left(Y = y|X_W\right) = \frac{p\left(X_W|Y = y\right)p\left(Y = y\right)}{p\left(X_W\right)} = \frac{\pi_y \; p\left(X_W|Y = y\right)}{\sum_{y=1}^{K} \pi_y \; p\left(X_W|Y = y\right)} \tag{7}$$

$$\left(X_W|Y = y\right) \sim \mathcal{N}\left(\mu_{y,W}, \Sigma_{y,WW}\right) \tag{8}$$

$$\pi_y^* = \frac{\pi_y \; \mathcal{N}\left(X_W; \mu_{y,W}, \Sigma_{y,WW}\right)}{\sum_{y=1}^{K} \pi_y \; \mathcal{N}\left(X_W; \mu_{y,W}, \Sigma_{y,WW}\right)} \tag{9}$$

To model the spatial distribution of air traffic for a known weather condition, we need to compute the conditional probability distribution of aircraft trajectories X_T given input meteorological conditions X_W for the y^{th} procedure. Bishop and Nasrabadi (2006) shows that this probability distribution is Gaussian with mean vector μ_y^* and covariance matrix Σ_y^*, as expressed by Equations 10 to 14.

$$\left(X_T|X_W Y = y\right) \sim \mathcal{N}\left(\mu_y^*, \Sigma_y^*\right) \tag{10}$$

$$\mu_y^* = \mu_{y,T} + \Sigma_{y,TW}\,\Sigma_{y,WW}^{-1}\left(X_W - \mu_{y,W}\right) \tag{11}$$

$$\Sigma_y^* = \Sigma_{y,TT} - \Sigma_{y,TW}\,\Sigma_{y,WW}^{-1}\,\Sigma_{y,WT} \tag{12}$$

$$\mu_y = \begin{bmatrix} \mu_{y,W} \\ \mu_{y,T} \end{bmatrix} \tag{13}$$

$$\Sigma_y = \begin{bmatrix} \Sigma_{y,WW} & \Sigma_{y,WT} \\ \Sigma_{y,TW} & \Sigma_{y,TT} \end{bmatrix} \tag{14}$$

Finally, we can use the marginal and conditional densities to forecast active procedures and their spatial confidence regions. Equation 15 defines the set S of procedures forecast to be active for a given probability threshold γ. Equation 16 expresses the confidence region R_y of the spatial distribution of air traffic for a significance level α for each active procedure y.

$$S = \left\{ y : \pi_y^* \geq \gamma \right\} \tag{15}$$

$$R_y = \left\{ X_T : p\left(X_T | X_W, Y = y\right) = 1 - \alpha \right\} \tag{16}$$

4. RESULTS AND DISCUSSION

4.1. Identification of current air traffic patterns in the urban airspace

The first step of the machine learning framework was the identification of the current patterns of urban airspace use by conventional air traffic. The flight trajectory clustering analysis was executed with the DBSCAN algorithm separately for each São Paulo airport and type of operation (departures/arrivals). Table 3 summarizes the results of the clustering analysis, presenting the number of clusters identified and the percentage of data noise encountered for each airport arrival/departure traffic flow. As an example, six different trajectory clusters were identified for flights departing from CGH, as shown in Figure 5. Figure 6 shows the percentage of flights in each cluster identified. The departure patterns were found to be mainly concentrated to the west and to the east, with an empty space in the northern and southern sectors. This is a result of the procedural separation between departing and arriving traffic. The clustering analysis for CGH arrival trajectories revealed that the arrival patterns are mainly concentrated to the north and to the south. Overall, the clustering results for CGH departures were clean and did not present much dispersion. Lastly, it is worth noting that clusters 3 and 5 were mainly composed of flights performing the busy São Paulo - Rio de Janeiro route, together with some flights bound to the northeast.

Table 3: Clustering results.

Flow	# clusters	Noise
CGH DEP	6	10.3%
CGH ARR	6	10.2%
GRU DEP	6	12.1%
GRU ARR	7	14.9%

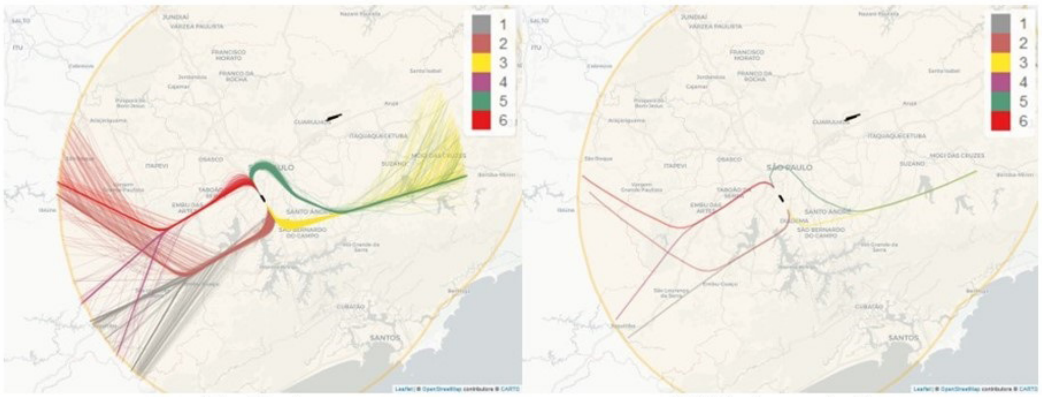


Figure 5. Trajectory clusters identified for CGH departures.

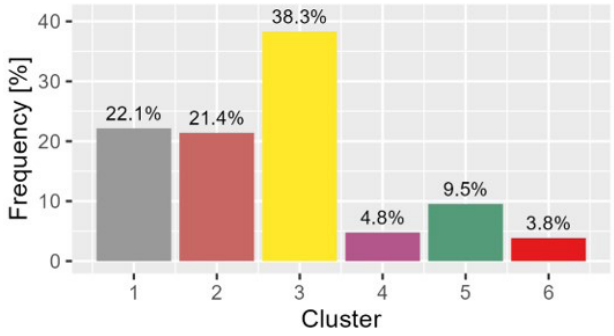
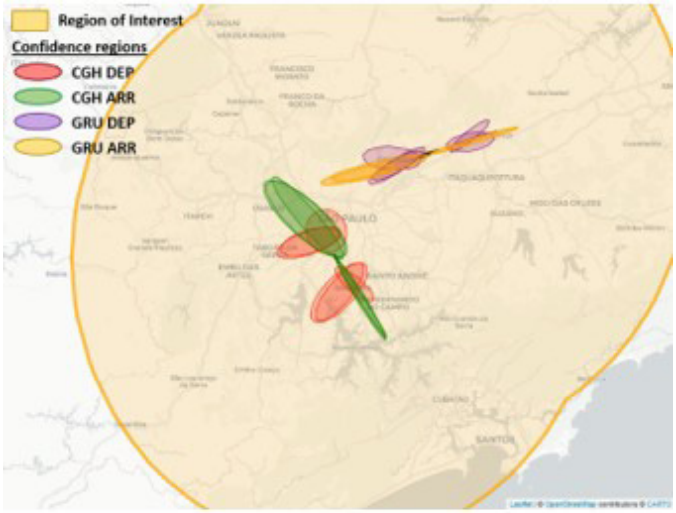


Figure 6. Distribution of flight trajectories by cluster for CGH departures.

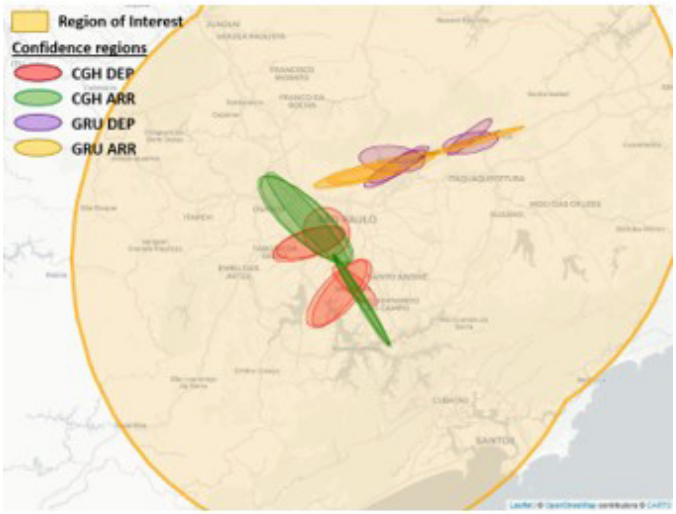
4.2. Descriptive analysis of the air traffic spatial distribution

After the clustering analysis, a GMM was estimated for each airport and type of operation based on the traffic patterns identified. The GMM allowed us to examine the spatial distribution of air traffic inside each cluster; for a given altitude layer, a spatial confidence region was determined for each traffic pattern. These spatial confidence regions indicate the airspace volumes where conventional air traffic is likely to be present. We analyzed three different confidence levels – 90%, 95% and 99% – and two different altitude layers – ground to 3,000 ft AGL and ground to 1,500 ft AGL.

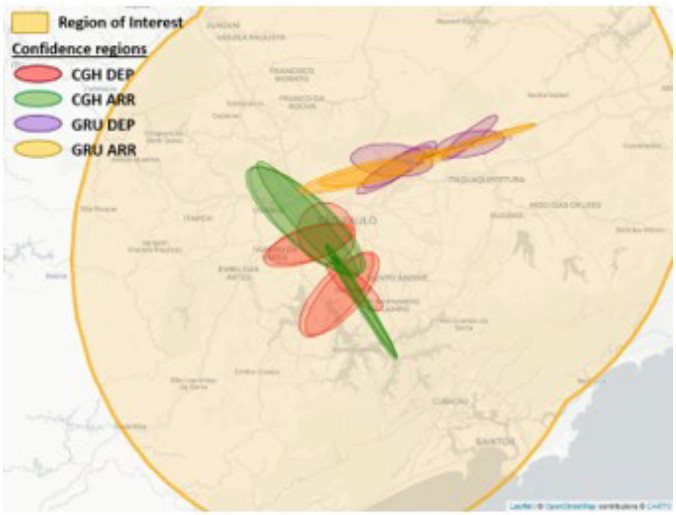
Figures 7 and 8 display the confidence regions of each trajectory cluster for the three different confidence levels considered. The confidence regions for CGH departure patterns are drawn in red color, CGH arrival patterns are shown in green color, GRU departure patterns are displayed in purple color and GRU arrival patterns are shown in orange color. As expected, the higher the confidence level, the larger the spatial confidence regions. This is clearly noticeable when we compare the three maps displayed in Figures 7 and 8. In addition, when comparing the two altitude layers, it is evident that the higher the altitude, the larger the confidence regions, due to the increased spatial dispersion of air traffic with altitude. The results help us to better understand how the TMA is actually structured, for instance, to understand at which regions and distances from the airports the planned procedures converge into a single path or split into many others. A visual inspection of the spatial confidence regions suggests that they have different levels of compactness and the spatial distribution of trajectories varies among the arrival/departure patterns.



(a) Confidence level: 90%

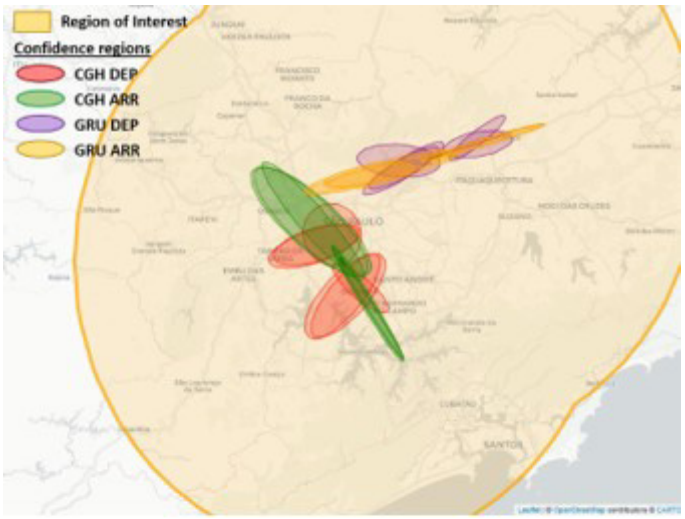


(b) Confidence level: 95%

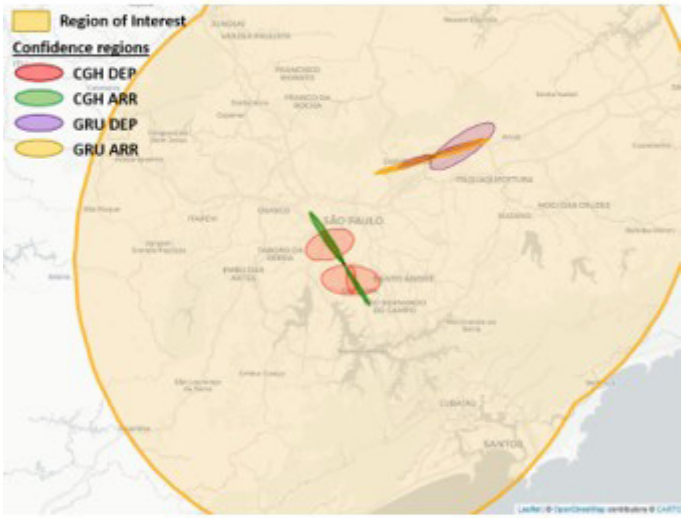


(c) Confidence level: 99%

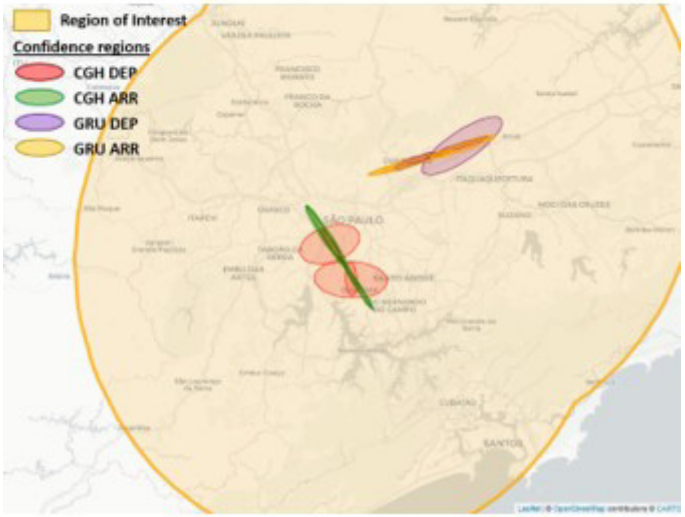
Figure 7. Spatial confidence regions below 3,000 ft AGL for GRU and CGH arrival/departure patterns.



(a) Confidence level: 90%



(b) Confidence level: 95%

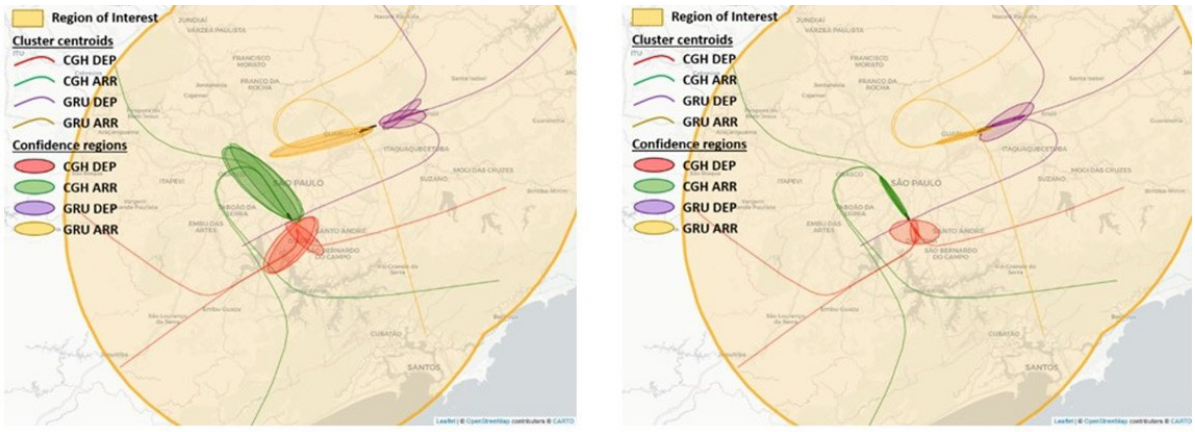


(c) Confidence level: 99%

Figure 8. Spatial confidence regions below 1,500 ft AGL for GRU and CGH arrival/departure patterns.

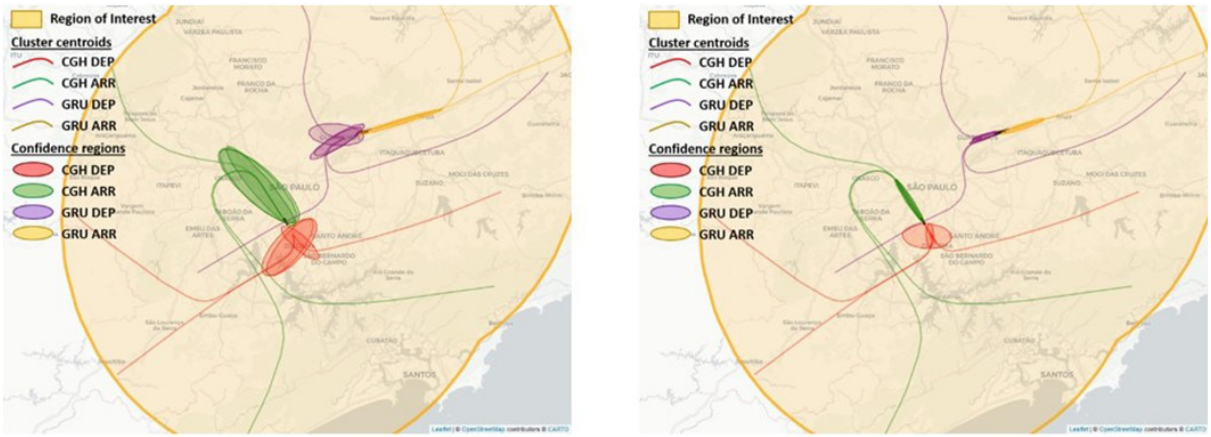
4.3. Predictability of the urban airspace availability

Once the probabilistic model is estimated, it can be used to make predictions of the traffic patterns that are likely to be active, their spatial confidence regions and the resulting urban airspace availability for UAM for any given meteorological condition. As an example, Figure 9 shows the results of this prediction, based on the application of GMM with a confidence level of 95% and a probability threshold of 5%, for the following meteorological condition, which is very common in São Paulo during regular sunny days: VMC (ceiling 10,000 ft, visibility 10 km), winds coming from the southeast at 8 knots and no rain. For comparison purposes, Figure 10 shows the results for weather conditions observed in a typical cold front day: marginal VMC (ceiling 1,500 ft, visibility 5 km), winds coming from the south at 15 knots and 0.004 kg/m^{2}s convective precipitation rate.



(a) Below 3,000 ft AGL (b) Below 1,500 ft AGL

Figure 9. Predicted active patterns, 95% confidence level, 5% probability threshold and airspace availability for the following meteorological condition: VMC, winds coming from the southeast at 8 knots, without rain.



(a) Below 3,000 ft AGL (b) Below 1,500 ft AGL

Figure 10. Predicted active patterns, 95\% confidence regions and airspace availability for the following meteorological condition: marginal VMC, winds coming from the south at 15 knots, with rain.

An analysis of the GMM's predictive performance was done to evaluate the quality of the model. As expressed by Equation 15, a certain procedure is labeled as active if the probability of its occurrence is equal to or greater than a threshold value γ. Therefore, by testing different values for this parameter, we evaluated the accuracy of the predictive model. The accuracy was defined

by the mean of the model's success rate (whether the predicted active procedures contained the trajectories actually observed for the period), measured day by day, for a dedicated test dataset, defined as one third of the whole dataset (randomly selected). The results are displayed in Table 4. A comparison between the performance of the predictive model with and without the implementation of the NCEP numerical meteorological model, i.e., with and without microweather features, is provided.

Table 4: Probabilistic model performance for predicting active traffic patterns.

		Number of predicted active traffic patterns (mean)		Accuracy (mean)	
Flow	γ	Without microweather features	With microweather features	Without microweather features	With microweather features
CGH DEP	0.00	6.0	6.0	100.0%	100.0%
	0.01	4.7	4.8	98.0%	98.4%
	0.05	3.5	3.5	96.8%	97.1%
CGH ARR	0.00	6.0	6.0	100.0%	100.0%
	0.01	4.5	4.6	98.3%	98.7%
	0.05	3.4	3.5	97.0%	97.5%
GRU DEP	0.00	6.0	6.0	100.0%	100.0%
	0.01	5.1	5.2	98.8%	99.0%
	0.05	3.9	4.1	98.1%	98.6%
GRU ARR	0.00	7.0	7.0	100.0%	100.0%
	0.01	5.0	5.0	98.2%	98.6%
	0.05	3.7	3.8	97.6%	97.9%

It is evident that for a probability threshold γ of 0%, every single learned procedure would be active regardless of the input meteorological condition. By increasing the threshold to 1% and 5%, we find that the average number of predicted procedures decreases, indicating that the model is able to reasonably distinguish between active and inactive procedures and the airport runway configuration resulting from a given meteorological condition. Naturally, such prediction is not perfect, and the model's accuracy decreases with the use of higher probability thresholds. However, we observed an accuracy greater than 95% for a threshold of 5% that predicts nearly half of the patterns as active, which is very promising.

Microweather is an important meteorological feature to be considered during UAM emergence. Thus, we are interested in investigating the gains of using the NCEP numerical meteorological model in our predictive tool. The results from Table 4 make evident that, although slightly, the accuracy of the GMM predictive model with microweather features is greater in all cases. This is a promising result and suggests that other more specialized numerical models could be tested to try to increase the model's performance even further.

Next, we verify the accuracy of the predictive model in forecasting the spatial distribution of air traffic for the procedures predicted as active. For each confidence level and altitude layer, we computed the percentage of flights for which the actual position was contained by the predicted confidence region (capture percentage). Tables 5 and 6 display the capture percentage of the predicted confidence regions, averaged over all traffic patterns and altitude layers, for the same two meteorological conditions

considered previously in this section, respectively. A comparison between the performance of the predictive model with and without the implementation of the NCEP numerical meteorological model, i.e., with and without microweather features, is also provided. The results show that the capture percentage closely matches the confidence levels, especially for higher confidence levels. For lower confidence levels, the model resulted in a little more conservative confidence regions. The differences between the models without and with microweather features were slight.

Table 5: Capture percentage of the predicted confidence regions for the following meteoro- logical condition: VMC, winds coming from the southeast at 8 knots, without rain.

		Capture Percentage (mean)	
Flow	Confidence Level	Without microweather features	With microweather features
CGH DEP	90%	94.6%	94.7%
	95%	96.3%	96.4%
	99%	98.1%	98.1%
CGH ARR	90%	96.9%	96.9%
	95%	98.1%	98.1%
	99%	98.8%	98.8%
GRU DEP	90%	94.4%	94.4%
	95%	96.3%	96.4%
	99%	98.5%	98.6%
GRU ARR	90%	98.1%	98.1%
	95%	98.4%	98.4%
	99%	98.8%	98.8%

Table 6: Capture percentage of the predicted confidence regions for the following meteoro- logical condition: marginal VMC, winds coming from the south at 15 knots, with rain.

		Capture Percentage (mean)	
Flow	Confidence Level	Without microweather features	With microweather features
CGH DEP	90%	94.6%	94.7%
	95%	96.3%	96.4%
	99%	98.1%	98.1%
CGH ARR	90%	96.9%	96.9%
	95%	98.1%	98.1%
	99%	98.8%	98.8%
GRU DEP	90%	94.3%	94.2%
	95%	96.4%	96.1%
	99%	98.6%	98.1%
GRU ARR	90%	98.0%	97.9%
	95%	98.6%	99.5%
	99%	98.9%	99.9%

5. CONCLUSIONS

In this study, we presented a data-driven approach to modeling the urban airspace availability for emerging UAM operations toward supporting their safe and efficient integration. The approach

is based on a machine learning framework for descriptive analysis and predictive modeling of the current patterns of urban airspace use by conventional traffic. A density-based clustering algorithm is first applied to learn existing traffic flow patterns in the terminal airspace from historical aircraft tracking data. This knowledge is then combined with meteorological data to create a probabilistic Gaussian Mixture model of the spatiotemporal traffic distribution, which enables the prediction of active procedures, their spatial confidence regions and the resulting airspace availability for UAM flights.

The data-driven approach was demonstrated for the São Paulo metropolitan region, which is recognized as a promising market for UAM. With the clustering analysis, we identified and characterized the main arrival and departure patterns for GRU and CGH, the two major commercial airports that serve the São Paulo metropolitan area. We observed that the spatial dispersion of air traffic significantly increases with altitude. Moreover, regardless of the airport, air traffic was found to be much more dispersed in departure patterns than in arrival patterns for low-level altitude layers. This suggests that a wider region of airspace is more constrained and least accessible for UAM in the sectors dedicated to takeoffs. We also noted that the air traffic dispersion was higher for traffic departing from CGH than from GRU, while the opposite was verified for arriving traffic.

The probabilistic model showed good predictive performance, with an accuracy higher than 95%, and was useful to provide a better comprehension of two factors regarding airspace occupation: how many individual procedures are used and how widely air traffic distributes itself. Since the model is capable of delivering these outputs after being fed with meteorological information, one can forecast in advance the traffic patterns most likely to be active, and hence determine the probable urban airspace availability for UAM operation. This allows for smarter and more efficient planning by both ATM agents and UAM operators.

ACKNOWLEDGEMENTS

The authors thank Atech, an Embraer group's company, and the National Council for Scientific and Technological Development (CNPq) for the support.

REFERENCES

ABRAPHE (2019) *São Paulo tem 1,3 mil decolagens e pousos de helicópteros por dia*. Available at: <https://istoedinheiro.com.br/capital-paulista-tem-13-mil-decolagens-e-pousos-de-helicopteros-por-dia> (accessed: 02/03/2023).

Bishop, C.M. and N.M. Nasrabadi (2006) *Pattern Recognition and Machine Learning.* New York: Springer.

Burke, P.J. (2019) Small Unmanned Aircraft Systems (sUAS) and Manned Traffic Near John Wayne Airport (KSNA) spot check of the SUAS facility map: towards a new paradigm for drone safety near airports. *Drones*, v. *3*, n. 4, p. 84. DOI: 10.3390/drones3040084.

CGNA (2019) *Anuário Estatístico de Tráfego Aéreo*. Available at: <http://portal.cgna.decea.mil.br/files/uploads/anuario_estatistico/anuario_estatistico_2019.pdf> (accessed: 02/03/2023).

Cho, J. and Y. Yoon (2018) How to assess the capacity of urban airspace: a topological approach using keep-in and keep-out geofence. *Transportation Research Part C, Emerging Technologies*, v. *92*, p. 137-49. DOI: 10.1016/j.trc.2018.05.001.

Eerland, W.J.; S. Box and A. Sóbester (2016) Modeling the dispersion of aircraft trajectories using gaussian processes. *Journal of Guidance, Control, and Dynamics*, v. *39*, n. 12, p. 2661-72. DOI: 10.2514/1.G000537.

EMBRAER X (2019) *Flight Plan 2030: an Air Traffic Management Concept for UAM.* São José dos Campos: EMBRAER X; ATECH; L3HARRIS.

Evans, A. and P. Lee (2019) Using machine learning to dynamically generate operationally acceptable strategic reroute options. In *Air Traffic Management Research and Development*. Vienna: ATM R&D.

FAA (2020) *Concepts of Operations: Urban Air Mobility*. Washington, DC: NextGEN.

Gariel, M.; A.N. Srivastava and E. Féron (2011) Trajectory clustering and an applicationto airspace monitoring. *IEEE Transactions on Intelligent Transportation Systems*, v. *12*, n. 4, p. 1511-24. DOI: 10.1109/TITS.2011.2160628.

Goyal, R.; C. Reiche; C. Fernando et al. (2021) Advanced air mobility: demand analysis and market potential of the airport shuttle and air taxi markets. *Sustainability*, v. *13*, n. 13, p. 7421. DOI: 10.3390/su13137421.

IBGE (2023) *Census Indexes.* Available at: <https://ftp.ibge.gov.br> (accessed: 02/03/2023).

IBOPE (2014) *Apresentação Dia Mundial Sem Carro*, Rio de Janeiro: IBOPE.

Kanamitsu, M.; W. Ebisuzaki; J. Woollen et al. (2002) NCEP–DOE AMIP-II Reanalysis (r-2). *Bulletin of the American Meteorological Society*, v. *83*, n. 11, p. 1631-44. DOI: 10.1175/BAMS-83-11-1631.

Lascara, B.; A. Lacher; M. Degarmo et al. (2020) *Urban Air Mobility Airspace Integration Concepts: Operational Concepts and Exploration Approaches.* McLean: The MITRE Corporation.

Liu, Y.; M. Hansen; D.J. Lovell et al. (2017) *Causal analysis of enroute flight inefficiency – The U.S. Experience, 2017*. In *12th USA, Europe Air Traffic Management Research and Development Seminar*. Seattle.

Marcos, R.; O. García-Cantú and R. Herranz (2018) *A Machine Learning Approach to Air Traffic Route Choice Modelling.* New York: Cornell University.

Mclachlan, G.J. and K.E. Basford (1988) *Mixture Models: Inference and Applications to Clustering*. New York: M. Dekker.

Mohamed Salleh, M.F.B.; C. Wanchao; Z. Wang et al. (2018) Preliminary concept of adaptive urban airspace management for unmanned aircraft operations. In *2018 AIAA Information Systems-AIAA Infotech@ Aerospace*. Kissimmee: AIAA, p. 2260.

Murça, M.C.R. (2021) Identification and prediction of urban airspace availability for emerging air mobility operations. *Transportation Research Part C, Emerging Technologies*, v. *131*, p. 103274. DOI: 10.1016/j.trc.2021.103274.

Murça, M.C.R. and R.J. Hansman (2018) Identification, characterization, and predictionof traffic flow patterns in multi-airport systems. *IEEE Transactions on Intelligent Transportation Systems*, v. *20*, n. 5, p. 1683-96. DOI: 10.1109/TITS.2018.2833452.

Murça, M.C.R.; M.X. Guterres; M. Oliveira et al. (2020) Characterizing the Brazilian airspace structure and air traffic performance via trajectory data analytics. *Journal of Air Transport Management*, v. *85*, p. 101798. DOI: 10.1016/j.jairtraman.2020.101798.

Murça, M.C.R.; R.J. Hansman; L. Li et al. (2018) Flight trajectory data analytics for characterization of air traffic flows: a comparative analysis of terminal area operations between New York, Hong Kong and São Paulo. *Transportation Research Part C, Emerging Technologies*, v. *97*, p. 324-47. DOI: 10.1016/j.trc.2018.10.021.

NOAA (2021) Global Forecast System. Available at: <https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast> (accessed: 02/03/2023).

Olive, X. and J. Morio (2018) Trajectory clustering of air traffic flows around airports. *Aerospace Science and Technology*, v. *84*, p. 776-81. DOI: 10.1016/j.ast.2018.11.031.

Olive, X.; J. Sun; M.C.R. Murça et al. (2021) A framework to evaluate aircraft trajectory generation methods. In *14th USA/Europe Air Traffic Management Research and Development Seminar.* New Orleans, p. 20-24.

REDEMET (2022) *API-REDEMET – METAR*. Available at: <https://ajuda.decea.mil.br/base- de-conhecimento/api-redemet-mensagem-metar/> (accessed: 02/03/2023).

Schäfer, M.; M. Strohmeier; V. Lenders et al. (2014) Bringing up OpenSky: a large-scale ADS-B sensor network for research. In *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*. Berlin: IEEE, p. 83-94. DOI: 10.1109/IPSN.2014.6846743.

Vascik, P.D. and R.J. Hansman (2019) Assessing integration between emerging and conventional operations in urban airspace. In *AIAA Aviation 2019 Forum*. Dallas: AIAA, p. 3125. DOI: 10.2514/6.2019-3125.

Vascik, P.D. and R.J. Hansman (2020) Allocation of airspace cutouts to enable procedurally separated small aircraft operations in terminal areas. In *AIAA Aviation 2020 Forum*. Reston: AIAA, p. 2905. DOI: 10.2514/6.2020-2905.

Vascik, P.D.; J. Cho; V. Bulusu et al. (2019) Geometric approach towards airspace assessment for emerging operations. *Journal of Air Transportation*, v. *28*, n. 3, p. 124-33. DOI: 10.2514/1.D0183.