Dynamic sector characterisation model with the application of machine learning techniques

F Pérez Moreno¹*, V F Gómez Comendador¹, R Delgado-Aguilera Jurado¹, M Zamarreño Suárez¹, D Janisch² and R M Arnaldo Valdés¹

¹Universidad Politécnica de Madrid (UPM), 28040 Madrid, Spain
²ATM Research and Development Reference Centre (CRIDA), 28022 Madrid, Spain

*Corresponding author: francisco.perez.moreno@alumnos.upm.es

Abstract. The ATC service has the objective of controlling airspace operations safely and efficiently. This control is becoming more and more difficult due to the increasing complexity of airspace. With the objective of collaborating and facilitating the provision of the control service, FLUJOS project aims to develop a methodology to characterise ATC sectors according to their complexity. This paper shows the first results obtained in this project. A methodology is proposed that first performs a statistical analysis of the data present in the flight plans of individual aircraft. The statistical analysis will be used to estimate the impact of air traffic flows. With this, the complexity of ATC sectors will finally be determined. In addition, a machine learning tool will be added to develop a dynamic methodology. After evaluating the methodology with data from Spanish airspace in 2019, it can be said that the results obtained are logical from an operational point of view, and that they allow a fairly accurate classification of the sectors based on their complexity. However, the proposed methodology is still a preliminary version, so more work will have to be done to add variables to achieve the objective of obtaining an even more accurate and realistic classification.

1 Introduction and objectives

The Air Traffic Management (ATM) System aims to enable efficient and safe operations for airspace users [1]. The ever-increasing number of aircraft in the airspace and the increasingly complex interactions between controllers and pilots make airspace increasingly complex [2]. This has led to the fact that the Air Traffic Control (ATC) service does not have sufficient capacity to handle all the foreseen demand.

But airspace complexity does not only rely on the number of aircraft or the interactions between controllers and pilots. Factors including the number of descending/escalating aircraft, the variation in aircraft ground speeds or the number of potential crossings of aircraft trajectories all play an important role in making airspace a more complex environment to be controlled. [3].

Airspace complexity is understood as the difficulty and efforts required in managing air traffic in a safe and orderly manner [4]. Although this is a globally spread concept, there is no current universal indicator of how to measure complexity [5]. However, over the last 25 years, numerous international programmes and projects have attempted to study the complexity of the airspace [6], [7], [8].

The conclusions drawn from these projects are important for the ATC service, as an increasing airspace complexity directly leads to an increase in controllers' workload. [9]. Thus, many benefits can be subtracted from the establishment of a specific complexity measure through allowing a more direct
assessment of the controllers’ workload. Due to this relationship between the sector operation complexity (SOC) and controllers’ workload, SOC will have an extraordinary role in air traffic management, e.g., airspace reconfiguration, air traffic flow management, and allocation of the ATC service resources [9]. These reasons place the study of complexity as a topic of great interest [10].

Added to the research interest on this topic worldwide, the above-mentioned reasons lead to the airspace complexity and unpredictability for the ATM system being a main cause for the deployment of technical and operational solutions to balance the capacity and demand of the System [11].

The FLUJOS project, funded by ENAIRE and CRIDA, aims to provide a technical solution and to complement the research conducted and the projects related to this topic. This solution is thought to be the design of a methodology to characterise ATC sectors according to their complexity, based on the behaviour of the main air traffic flows in the ATC sector. In addition, this project aims to develop a tool that will help the ATC service to increase its capacity thanks to the automatic application of the previously proposed methodology. Thanks to this tool, the ATC service will be able to anticipate possible conflict areas and manage its resources in an optimal way.

This methodology is special because of the introduction of machine learning to make the tool automatic and data-driven. Some of the research performed has limitations. Airspace complexity models are usually based on expert opinion and will therefore be subject to bias. In addition, the results generated by these models will be limited as they are subject to confirmation by air traffic controllers. [12]. To overcome these limitations, this paper proposes a methodology based on real operating data, and which is adaptive through the use of machine learning techniques. The advantages of this model will be:

- Possibility of using massive data. Real aircraft operation data from all Spanish airspace will be used in this ATC sector complexity characterisation model.
- Adaptive results. Thanks to the use of machine learning tools, the model will do away with the subjectivity of expert opinion and will be mainly data-driven. This will make the model capable of capturing patterns and considering different structural aspects in the ATC sectors.
- No active and continuous involvement of air traffic controllers or other stakeholders is required. As a process regulated by machine learning techniques, results will be adapted based on changes in data patterns.

Besides all this advantages, for a simpler use of the tool, the application of the methodology through the application of the characterisation tool will try to lead to intuitive results for the ATC service.

The proposed methodology is presented in section 2, outlining each step of the methodology. Section 3 shows the results obtained in an example with real data. Finally, section 4 discusses the conclusions obtained in this first work and how the sector characterisation tool will be further developed.

2 Methodology

The main objective of FLUJOS project and this paper is to develop a methodology to characterise ATC sectors according to their complexity and based on data from individual operations. This characterisation methodology is intended to assist the ATC service.

Specifically, the characteristics of individual operations that can determine the complexity of an ATC sector are, among others, the number of aircraft, mix of aircraft models, meteorology, separation between aircraft, aircraft speed or regulations affecting traffic [13]. These are aspects of individual aircraft.

Within the most relevant aspects of these individual operations, four fields of interest have been identified:

- Traffic density
- Vertical density
- Time distribution
- ATFCM regulations

These fields will be the one used in the statistical analysis which is the first step of the methodology. A statistical indicator will be created from each of the four fields, based on flight plan data. In order to
have a complete analysis, the indicators will be divided into average values and variability coefficients (see Eq. (1)).

\[
\text{Var. Coef.} = \frac{\sigma}{\mu}
\]

\[
\sigma = \text{Standard deviation}
\]

By taking the mean values and the variability coefficients separately, two complex variables are created from the combination of the statistical indicators of the four fields. These complex variables will be known as average impact and impact variability. The average impact and impact variability will be calculated as a weighted sum of the statistical indicators of the four fields:

Average Impact = \( x \cdot \) Average traffic density + \( y \cdot \) Average vertical density + \( z \cdot \) Average temporal distribution + \( w \cdot \) Average regulations
\( x + y + z + w = 1 \)

Impact Variability = \( x' \cdot \) Traffic density variability + \( y' \cdot \) Vertical density variability + \( z' \cdot \) Temporal distribution variability + \( w' \cdot \) Regulations ATFCM
\( x' + y' + z' + w' = 1 \)

Where \( x, y, z, w \) are the relative weights in the calculation of the average impact and \( x', y', z', w' \) are the relative weights in the calculation of the impact variability.

The second step of the methodology is the calculation of the impact of the air traffic flows in each of the ATC sectors. The definition of the impact will be based on the combination of the average impact and the impact variability into a table model. The impact classification is a five-level classification, with level one being the lowest impact and level five being the highest impact.

The third and final step of the methodology is the definition of the complexity of the ATC sectors. This definition is proposed based on two variables related to the air traffic flows of the sector: the number of flows within the sector, and the percentage of these flows that have an impact level of five. The complexity variable will be discretised into five classes, from one to five. Complexity level one is for the simplest sectors and level five for the most complex ones.

This is the proposed methodology for ATC sectors complexity characterisation. Machine learning techniques will be introduced in each of the three steps of the methodology with the objective of developing an automatic, adaptive methodology.

First, the relative weights of the four fields, which allow the calculation of the average impact and the variability of the impact, are determined based on experts’ opinion. These experts also determine the level distribution of the tables that calculate the impact of the flows and the complexity of the sectors. This is where machine learning will be introduced. The aim of machine learning will be to modify the relative weights of the variables in the different steps according to the initial data. Machine learning is added to study the large difference between the existing ATC sectors and their different structural aspects.

In the first step of the methodology, the weighted sums are considered regression problems. By using machine learning techniques, the relative weights in a regression problem can be subtracted [14]. Thus, the relative weights of the indicators of the four fields can be determined. This will modify the weighted sum depending on the input data. With this variation in the relative weights of the weighted sum, the average impact and the impact variability will also change according to the data provided.

In the second and third steps, machine learning techniques can also be used. In a general classification problem, the relative weights of the variables involved can be obtained [15]. Considering the tables that calculate impact and complexity variables classification problems, the relative weights can also be subtracted. The aim changes, and it is now to obtain balanced tables by the application of machine learning techniques. With this also learning from the data, the results will be differentiated.

Machine learning has the additional objective of helping the characterisation methodology to evolve. With more historical data, the overall traffic trends vary. Machine learning must be able to capture this evolution of air traffic and transmit it to the characterisation methodology. This would be very difficult.
to achieve with a general, invariant model. Therefore, the addition of machine learning is a fundamental point.

3 Results

This section shows the results obtained by testing the validity of the sector characterisation methodology in a real scenario. For this case, data from five ATC sectors in the Spanish airspace from 2019 have been used. As the characterisation methodology is divided into three steps, this section will also be divided into three steps.

1. Calculation of average impact parameters and impact variability parameters through statistical analysis of the data provided.
2. Calculation of the impact of the air traffic flows based on the average impact and the impact variability.
3. Final calculation of the complexity of the ATC sectors based on the number of flows in each sector and the percentage of flows of five-level impact.

3.1 First step results

The first step of the methodology is the calculation of the average impact and impact variability. These complex variables are calculated based on four statistical indicators, belonging to the four relevant fields of study. These indicators, differentiated by mean values and variability coefficients, will be combined into a weighted sum (see equation (2) and equation (4)).

Initially, the relative weights of these sums are established by experts’ opinion. A machine learning tool is implemented in order to capture the structural aspects of the different sectors. The machine learning will make an analogy between a regression problem and the weighted sum. The machine learning will then be able to estimate the relative weights of the weighted sum. This will be established by following an iterative process until stability in the estimated relative weights is reached.

Figure 1. shows the process of determining the relative weights of the average impact (see figure 1.a) and the impact variability (see figure 1.b). In this case, with six iterations, the required stability has been reached.

![Figure 1](image_url)

**Figure 1.** Evolution of relative weights with the machine learning model.

After six iterations of the iterative process carried out by the machine learning tool, the relative weights have reached a sufficient tendency to be considered the definitive relative weights. In both cases, the relative weight of ATFCM regulations is the highest. For the calculation of impact variability, traffic density is also important. Vertical density and time distribution do not have relevant relative weights in either case.

These results make sense from an operational point of view. Regulations are directly associated with an excess of demand to be assumed by the ATC service, and therefore with high complexity. A high number of aircraft will also be a major cause of high complexity. Vertical density and time distribution
are related to the number of aircraft, so their importance does not have to be as high as that of traffic
density.

With these results, the trends correspond to the reality of the operation in the sectors studied. But the
values of the relative weights calculated are too unbalanced. Both vertical density of aircraft and the
time distribution are variables that should at least be considered in the characterisation of the impact of
air traffic flows. Both the theoretical study prior to the development of the methodology and the experts’
opinion have thus taken this into account.

For this reason, it is worth revisiting the variables in this analysis in future iterations of the project,
in order to have a better balance between the four indicators.

Despite the problems identified in the process of calculating relative weights, the ones of the sixth
iteration of the process will be used in this case of study. The conclusion obtained from the application
of this first step of the methodology is that a revision of the statistical indicators is needed, but the
machine learning tool operates correctly. Therefore, under the assumption of validity of this preliminary
model, the methodology will continue to be applied systematically.

3.2 Second step results
The second step of the methodology aim is to calculate the impact of the air traffic flows within an ATC
sector. For this purpose, a five-level classification of the impact is performed by means of a table model.

Figure 2.a shows the table that is first used to calculate the impact. This initial table model is
proposed by experts’ opinion. The idea of this table model is to classify the air traffic flows based on
both average impact and impact variability and doing this in the most balanced way possible.

By studying the relative weights of the average impact and the variability of the impact with
the machine learning tool, it can be seen that the average impact is more important. The relative importance
of average impact is approximately 0.65, and the relative importance of impact variability 0.35. In this
example application, the table is unbalanced, and will not be suitable for classifying the data. So, the
table model will have to change in order to adjust the classification to the data. This new table model
will be selected from a set of table models designed to have sufficient variability in the scenarios
proposed. Figure 2.b shows the table that is balanced, and that will be used this time for the impact
calculation. It can be seen how the relative weights of the mean impact and the variability of the impact
are now very similar.

\[\text{VALUES OF RELATIVE IMPORTANCE WHEN CALCULATING IMPACT}\]

\[\begin{array}{c}
\text{Mean impact} \\
\text{Impact variability}
\end{array}\]

\[\text{VALUES OF RELATIVE IMPORTANCE WHEN CALCULATING IMPACT}\]

\[\begin{array}{c}
\text{Mean impact} \\
\text{Impact variability}
\end{array}\]

\[(a)\]

\[(b)\]

**Figure 2.** Table models for the calculation of impact.

Compared to the table model in figure 2.a, the table model in figure 2.b will have more combinations
of level-one, level-three and level-four impact. In return, it will have fewer five-level impact
combinations. This will mean that in the overall results obtained, fewer air traffic flows will be relevant
for controllers.

Once the table has been decided, the impact of air traffic flows in the sector can be calculated. The
flows of the Pamplona control sector (LECMPAU) on a representative day are shown in figure 3.
Figure 3. Example of traffic flows with their associated level of impact.

There are two main types of traffic flows. The first is made up of flows crossing LECMPAU from north to south, and the second stream is composed of flows crossing from east to west. These flows are the ones that actually operate in the sector, so the results obtained are correct. It is not possible to prove that the classification by impact levels is correct, as no previous references are available.

LECMPAU is known to be a very structured traffic sector. Around this sector, Madrid airport is to the south and Barcelona airport to the east. LECMPAU flights will have these two airports as their origin/destination airports. Figure 3. shows an accurate map of the operation in LECMPAU. So far, the results seem to be correct.

As the operation is very ordered and stationary, the ATC service will have no capacity problems to control operations in this sector. This makes the presence of many one-level and two-level impact flows meaningful. There will always be certain traffic flows that are more complex to be controlled. As the average impact and impact variability will be mainly conditioned by ATFCM regulations and traffic density, it is logical that five-level impact air traffic flows will be regulated or have a great number of aircraft.

The results obtained are correct from an operational point of view. This leads to believe that this step of the methodology is consistent despite the strong influence of the ATFCM regulations and traffic density in the calculation of the average impact and the impact variability.

3.3 Third step results
The last step of the proposed methodology aims to calculate the complexity of an ATC sector based on the behaviour of the air traffic flows within it. This will be done by a classification based on a table model. This classification will be divided into five levels and will take into account the number of flows in the sector and the percentage of these flows that have five-level impact.

Figure 4.a shows the initial table that will try to calculate the complexity. When calculating the relative weights by machine learning, it is seen that the number of flows is much more important (relative importance of 0.82) than the percentage of five-level impact flows (relative importance of 0.18). This first table is not suitable for the data entered in this example, so another table will have to be used. This new table model will be selected again from the same set of table models designed to have sufficient variability in the scenarios proposed. Figure 4.b shows a table whose relative weights are roughly balanced. This second table is the one used in the example.
Figure 4. Table models for the calculation of complexity.

The table model in figure 4.b will have more one-level and two-level impact combinations in exchange for having fewer five-level impact combinations. The amount of two-level and three-level impact combinations remains the same in the two table models, but they are differently distributed. With this new table model, it is more difficult to have complex sectors although this is the model that balances the number of air traffic flows and the percentage of five-level impact flows in the case studied.

Once it has been decided that the table model in Figure 4.b is the one that will classify the complexity of the sectors. Figure 5. shows results of the application of the third step of the methodology in five sectors of the Spanish airspace.

Figure 5. ATC Sector complexity classification.

El sector Castejón (LECMCJI) tiene una five-level complexity, mientras que Santiago (LECMSAN) y Barcelona Central (LECBCCC) tienen una four-level complexity. Los sectores Baleares (LECBBAS) y Pamplona (LECMPAU) son sectores más sencillos para el ATC service. Having three-level and two-level complexity respectively.

These results make sense from an operational point of view. Sector Castejón is one of the most complex control sectors in the Spanish airspace. Madrid airport is close to this sector, so that most of the departures and arrivals operations around this airport will fly through Castejón. This will cause the sector to have a large number of operations with high variability, and a large number of regulations due to the insufficient capacity of the ATC service. The Barcelona Central sector faces the same difficulties. Barcelona airport is within LECBCCC, so all flights to/from this airport will fly over the sector. Despite
this, LECBCCC has four-level complexity, being less complex than LECMCJI. The Santiago sector is also a very complex control sector. This sector has very complex operations, as it is the only sector in Spain that has free-route operations. This meaning that there are many interactions between aircraft. In addition, many aircraft from North America will go across LECMSAN, making the number of operations very elevated. These reasons make the methodology predict LECMSAN as a complex control sector.

Sector Baleares is a highly seasonal control sector. During the summer it will receive lots of operations due to holiday flights to Mallorca airport. But in winter it will be a sector with very little traffic. This leads to the methodology classifying it as a sector of less complexity. Sector Pamplona is actually the least complex sector in this example. The traffic in this sector is very structured and stationary.

The results obtained make sense from an operational point of view. This leads to the conclusion that the methodology is adequate, and that after some improvements, it has the potential to assist the ATC service by providing a fairly accurate classification of ATC sectors.

4 Conclusions and future work

The methodology developed to facilitate the labour of the ATC service seems to have quite realistic results in this preliminary version, and the most important conclusions obtained will be discussed below.

The most interesting aspect of this model is the addition of machine learning techniques. Machine learning helps determining the importance of the variables that define the methodology. This support tool learns according to the data presented, which means that the final classification itself will be different depending on the sectors analysed. Because the reason of this is machine learning from the data and being able to capture the structural aspects of the different ATC sectors, which would be very difficult without the presence of machine learning. Machine learning will also help to continuously and automatically update the methodology. If historical data is added, machine learning will learn from the trends in the data. This will cause the machine learning to change the results obtained. This helps to make the methodology dynamic.

Certain conclusions can also be drawn from the application of the methodology in a real scenario, using the sector characterisation tool:

- When calculating the average impact and impact variability on an air traffic flow, the importance of ATFCM regulations stands out above the rest. This is logical since regulations appear in situations of complexity when the demand of aircraft exceeds the capacity of the ATC system. Even so, the methodology intends to make all the fields influential in the calculation of the average impact and the impact variability, so it is convenient to review the definition of statistical indicators of the four fields.

- The results obtained in the calculation of the impact of the air traffic flows are quite accurate. The air traffic flows identified as main flows are indeed important operational air traffic flows. The predicted impact levels, although unverifiable, seem to be in line with the reality of the operations in LECMPAU sector.

- The results obtained in the calculation of the complexity of the ATC sectors also seem to be correct. The classification according to the complexity seems to capture the specific characteristics of the different ATC sectors and to differentiate between which sectors will be complex to control for the ATC service and which sectors will not.

These conclusions can be summarised in a good performance of the characterisation tool, giving realistic and intuitive results to the ATC service. The methodology seems to fulfil the objectives for which it was initially proposed. But this is preliminary work, and the methodology needs to be improved for its general implementation. Therefore, its development will continue. Certain lines of future research are proposed for this purpose:

- Application to more ATC sectors. This will provide a better assessment of the validation of the model and more solid conclusions on the scope and limitations of this characterisation methodology.
Study of new variables in the definition of sector complexity. Complexity is currently defined based on the number of air traffic flows and the percentage of impact five flows. But it is interesting to rethink this definition of complexity, as the addition of new variables can make this definition much more realistic.

- Add more table model with different distributions of impact/complexity levels. With this, the adjustment made by machine learning to balance the models would be more accurate.

5 References

[1] Verdonk Gallego CE, Gómez Comendador VF, Saez Nieto FJ and García Martínez M. Discussion on density-based clustering methods applied for automated identification of airspace flows. 37th Digital Avionics Systems Conference; 2018. 584-93.

[2] Lee KM, Feron E and Pritchett A. Describing Airspace Complexity: Airspace response to disturbances. Journal of Guidance Control and Dynamics 32, 2009. 210-22.

[3] Cao X, Zhu X, Tian Z, Chen J, Wu D and Du W. A knowledge-transfer-based learning framework for airspace operation complexity evaluation. Transportation research part C-Emeging Technologies 95, 2018. 61-81.

[4] Prandini M, Pirolli L, Puechmorel S and Brázdilová SL. Toward air traffic complexity assessment in new generation air traffic management systems. IEEE Transactions on Intelligent Transportation Systems 12, 2011. 809-18

[5] Gianazza D. Airspace configuration using air traffic complexity metrics. 7th FAA/Europe Air Traffic Management Research and Development Seminar. 2007.

[6] Pejovic T, Netjasov and Crnogorac D. 2020. Risk assessment in air traffic management. ed JA Pérez Castán chapter 2.

[7] Zhu X, Cao X and Cai K. Measuring air traffic complexity based on small samples. Chinese Journal of Aeronautics 30, 2017. 1493-505

[8] Radisic T, Novak D and Juricic B. Reduction of air traffic complexity using trajectory-based operations and validation of novel complexity indicator. IEEE Transactions on Intelligent Transportation Systems 18, 2017. 3038-48

[9] Xie H, Zhang M, Ge J, Dong X and Chen H. Learning Air Traffic as Images: A Deep Convolutional Neural Network for Airspace Operation Complexity Evaluation. Complexity 2021, 2021.

[10] Gómez comendador VF, Arnaldo Valdés RM, Vidosavljevic A, Sánchez Cidoncha M and Zheng S. Impact of Trajectories' Uncertainty in Existing ATC Complexity Methodologies and Metrics for DAC and FCA SESAR Concepts. Energies 12, 2019.

[11] SESAR JU. European ATM master plan 2015. 2015.

[12] Gianazza D and Guittet K. Selection and evaluation of air traffic complexity metrics. 2006 IEEE/AIAA 25TH DIGITAL AVIONICS SYSTEMS CONFERENCE 1-3,2006, 254-65.

[13] Oktal H and Yaman K. A new approach to air traffic controller workload measurement and modelling. Aircraft Engineering and Aerospace Technology 83, 2011. 35-42.

[14] Bi J and Chung J. Identification of drivers of overall liking – determination of relative importances of regressor variables. Journal of Sensory Studies 26, 2011. 245-54.

[15] Holz HJ, Loew MH. Relative feature importance - a classifier-independent approach to feature selection. Machine intelligence and pattern recognition 16, 1994. 473-87.