Expert Systems With Applications 252 (2024) 124146

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Prediction of ATFM impact for individual flights: A machine learning approach

Sergi Mas-Pujol [a], Luis Delgado [b,*]

[a] Department of Computer Architecture, Technical University of Catalonia (UPC), Castelldefels, Barcelona, Spain
[b] Centre for Air Traffic Management Research, School of Architecture and Cities, University of Westminster, London, United Kingdom

ARTICLE INFO

Keywords:
ATFM delay
Machine learning
Pre-tactical
Support tool
Airlines operations

ABSTRACT

Air Traffic Flow Management regulations consist of issuing pre-departure ground delays to deal with demand-capacity imbalances in the airspace, smoothing out the demand at congested infrastructures. This results in discrepancies between the initially planned and executed (regulated) flights, negatively impacting operational efficiency and economic performance. To mitigate downstream effects, airspace users need to assess the severity and impact of these regulations when defining their operational plan. It is crucial to anticipate these potential network issues and plan for mitigation measures during the pre-tactical phase (the day before operations (D−1)). This article presents four independent machine learning models to estimate the impact of Air Traffic Flow Management regulations for individual flights during this pre-tactical phase. An integrated view is proposed to combine the results of the different models, displaying the appropriate level of information. Two models are designed to provide information about the probability of a given flight being affected by Air Traffic Flow Management regulations and the reason for this regulation (airport or airspace congestion). These models have reported accuracy levels between 82% and 87%. The remaining models estimate the impact of the delay (amount of Air Traffic Flow Management delay issued to the flight if regulated), with a Mean Absolute Error of 9.5 min when predicting this amount of delay. The SHapely Additive exPlanations analysis is used to identify the most important factors for detecting Air Traffic Flow Management regulations for individual flights during the pre-tactical phase from a data-driven perspective. An integrated view is proposed to create a system which combines the results of the four different models, displaying the appropriate level of information and avoiding an overload of information. Finally, the results are compared against models considering only static or post-operational data.

1. Introduction

Air Traffic Flow Management (ATFM) measures are implemented by Air Navigation Service Providers (ANSPs) to deal with demand-capacity imbalances. The most common approach is to set ATFM regulations to limit the rate at which aircraft enter congested traffic volumes[1] during a given period. Concretely, ATFM delays are issued to flights that plan to use the congested resources considering a first-planned-first-served basis calculated by the Computer Assisted Slot Allocation (CASA) algorithm (EUROCONTROL, 2022a, 2022b). As the name indicates, the system assigns slots to use the congested infrastructure, which translates into a Calculated Take-Off Time (CTOT) requiring the flight to take-off at that time with a given tolerance window (−5, +10 min) (EUROCONTROL, 2022b). Therefore, the regulation of a flight, assigning a CTOT, generates pre-departure ground delay, smoothing the rate of arrival of the flows to the congested region. Rather than airborne delay, ground delay is widely accepted due to their reduced operational costs and environmental footprint (Delgado, Gurtner, Cook, Martín, & Cristóbal, 2020).

According to EUROCONTROL (2019), in the European Civil Aviation Conference (ECAC) area,[2] the average ATFM delay due to airport congestion remained stable at around 1.24 min per flight between 2015 and 2018. However, en-route delay showed a different pattern,

* Corresponding author.
E-mail addresses: sergi.mas.pujol@upc.edu (S. Mas-Pujol), l.delgado@westminster.ac.uk (L. Delgado).

[1] A traffic volume is associated with a single geographical entity (*e.g.* an aerodrome, a set of aerodromes, an airspace sector), considering all or specific traffic flows crossing it (EUROCONTROL, 2022b).

[2] European Civil Aviation Conference (ECAC) composed by 44 member states: all EUROCONTROL members (European Union states plus Albania, Armenia, Bosnia and Herzegovina, Georgia, Moldova, Monaco, Montenegro, North Macedonia, Norway, Serbia, Switzerland, Turkey, Ukraine and the United Kingdom), Azerbaijan, Iceland and San Marino.

https://doi.org/10.1016/j.eswa.2024.124146
Received 25 January 2024; Received in revised form 23 April 2024; Accepted 29 April 2024
Available online 3 May 2024

significantly increasing by 104% in 2018, reaching 19M minutes, while traffic increased by just 3.8% over the same period. Considering an average ATFM delay cost of 100 €/min (Cook & Tanner, 2015), this leads to an expected total cost close to 2B€. However, COVID-19 forced a major reduction in air traffic; all forecasts estimate a complete traffic recovery in the near future (EUROCONTROL, 2022c, 2023; Gudmundsson, Cattaneo, & Redondi, 2021). This demand growth, combined with the lack of capacity at several European airports, will likely result in substantial future performance degradation, leading to an increase in ATFM delay (EUROCONTROL, 2019).

Significant efforts are being made to include airspace users in demand–capacity balancing processes through Collaborative Decision Making (CDM) to minimise these negative impacts on their operations and revenue (Andribet, Baumgartner, Garot, & of the Air, 2022). However, there is still room for improvement. Two main reasons for such regulations are insufficient information sharing between stakeholders and capacity generally managed during the tactical phase (Bolić, Castelli, Corolli, & Rigonat, 2017). The European Air Traffic Management (ATM) Master Plan aims to improve early information sharing between stakeholders to improve the system's predictability (SESAR, 2020).

Previous research has focused on optimising, improving, and minimising ATFM delays. From a general point of view, Vossen, Hoffman, and Mukherjee (2012) provides an extended review of the ATFM concept. Kistan, Gardi, Sabatini, Ramasamy, and Batuwangala (2017) review the ATFM field trying to determine the ATFM research and development efforts that hold the best promise for practical technological implementations. Dalmau, Zerrouki, Anouraud, Smith, and Cramet (2021) studied whether all issued ATFM regulations were necessary and how this affects network congestion. de Arruda Junior, Weigang, and Milea (2015) presented a new collaborative decision-making process to improve sharing data between stakeholders, which could reduce the number of issued ATFM regulations. García-Heredia, Molina, Laguna, and Alonso-Ayuso (2021) proposed an integer programming solution to solve some ATFM regulations.

Recent studies focus on anticipating ATFM regulations using Machine Learning (ML) techniques (Bishop & Nasrabadi, 2006), which benefit from the data availability in aviation and the fast response to new complex scenarios. Rather than using a complex system to model the air transport network, ML can be used to identify the patterns behind ATFM regulations as they are the result of extensive planning procedures by the Network Manager (NM). For instance, Gui et al. (2019) and Sridhar, Chatterji, and Evans (2020) explore various factors that might influence these regulations. Schultz, Reitmann, and Alam (2021) show that the application of ML is an appropriate approach to quantify the correlation between decreased airport performance and weather events, leading to ATFM regulations. Dalmau, Genestier, Anoraud, Choroba and Peter and Smith (2021) tries to estimate the evolution of already assigned ATFM delays using embedded systems. Sanaei, Pinto, and Gollnick (2021) studied the usage of Convolutional Neural Networks (CNNs) to predict the total network delay and the number of regulated flights in the European Air Traffic Management Network (EATMN) area. Garrigó et al. (2016) explored the potential of visual analytics and ML to improve the understanding of ATFM regulations. A more general approach was presented in (Yousefzadeh Aghdam, Kamel Tabbakh, Mahdavi Chabok, & Kheyrabadi, 2021) where the authors proposed to use Long-Short Term Memorys (LSTMs) to improve accuracy in ATM management problems, including all the necessary activities such as handling ATFM regulations.

Other research has focused on specific types of ATFM regulations or presented models with a more narrow scope. Jardines et al. (2024), Jardines, Soler, and García-Heras (2021) used ML algorithms to predict network performance during adverse weather conditions. Many different algorithms are tested, and the results show that complex models (*e.g.* deep neural networks) tend to overfit, while the Random Forest algorithm reports the best performance. Also related to specific regulations reasons, Schultz et al. (2021) presented a machine learning-based approach to assess the strategic flight schedules regarding potential arrival/departure flight delays and cancellations when facing capacity regulations. Mas-Pujol, Salamí, and Pastor (2022) used a hybrid model based on combining a time-distributed LSTM and CNN to predict the regulated period due to capacity en-route regulations for specific traffic volumes over the Maastricht Upper Area Control Centre (MUAC) region. Dalmau and Gawinowski (2024) used supervised clustering techniques to identify possible flight diversions due to weather, highlighting situations where predictions require careful consideration.

From a more flight-centric approach, Gopalakrishnan and Balakrishnan (2017) presented a comparative analysis of models predicting ATFM delays for specific Origin-Destination (OD). The authors compared linear systems, classical ML techniques like classification and regression trees, and artificial Neural Network (NN). Results show that classical ML techniques outperform the rest, and classification seems to be the best approach. Similarly, Gui et al. (2019) shows how random forest-based models report higher performance than complex artificial NN predicting the ATFM delay assigned to flights. Concretely, it is shown that the best approach is to use a classification algorithm splitting the issued ATFM delay in different classes. Rebollo and Balakrishnan (2014) used Random Forest models to estimate future departure delays between 2 and 24 h for the 100 most delayed links in the USA system, investigating different classification thresholds and predicting delay values. Wang et al. (2022) provides an alternative tool for airports and airline managers to estimate flight delays. The results show how Random Forest and LightGBM provide better results than MultiLayer Perceptrons (MLPs) (*i.e.*, artificial NN).

As shown, most of the previous research focuses on optimising ATFM delay and resource allocation across the entire ATM network, or specific OD pairs, with a particular emphasis on the tactical phase (day of operation) when more information becomes available. This often involves addressing specific regulatory issues or introducing new paradigms of behaviour. However, it is important to note that the main stakeholders affected by ATFM regulations are the airspace users, whose fleet management is directly impacted. To address this issue, this article proposes an approach utilising ML algorithms and techniques to anticipate all types of ATFM regulations during the day before operations day prior to operations (D-1). Specifically, a flight-centric approach is used to predict the likelihood of ATFM events and estimate the expected final delay. The selected models provide a high level of explainability to gain trust in their predictions. Moreover, the system can be used as a what-if tool to identify the less disrupting actions. By adopting this approach, the proposed system enables better anticipation of the impact of ATFM regulations on individual flights, allowing airspace users (tactical planners) to plan their fleet management strategies while designing their operational plan with a longer prediction horizon and highlighting which flights should be closely monitored during the day of operations by the duty managers. This expert system can significantly improve the efficiency and effectiveness of the ATM network and airline operations.

This article further develops previously conducted research, wherein the idea was first introduced in (Mas-Pujol, De Falco, & Delgado, 2022), and preliminary results using post-operation data were subsequently presented in (Mas-Pujol, De Falco, Salamí, & Delgado, 2022). The work presented in this article uses data available at the prediction horizon instead, ensuring that the airlines can use the models operationally. This work is part of the Dispatcher3 initiative (Dispatcher3 Consortium, 2020), a Clean Sky 2 innovation action which aims at using machine learning techniques to support the airlines' processes before departure (including D-1 preparatory activities). Also, the visualisation of results has been improved to ensure that the system only displays the level of information required by the end user.

Our major contributions are summarised as follows:

- We propose a flight-centric approach to better anticipate the impact and severity of ATFM regulations the day before the operation when the levels of uncertainty are higher, and little information is available on the ATFM regulations that will be issued. These predictions will help the tactical planners and duty managers of airlines to figure out the impact of these regulations on their fleets;
- We have demonstrated that it is possible to anticipate the characteristics of ATFM regulations using simple and well-known ML algorithms with a high level of explainability. This is crucial due to the high levels of safety in aviation and the need to trust the prediction of the models by understanding the reason behind their outcome;
- The longer prediction horizon improves the airlines' planning capabilities, overcoming some of the limitations found in the state-of-the-art. Previously available tools focus on minimisation and reduction of delay during the day of operations; the system developed in this article aims, however, to be used pre-tactically;
- This expert system can also be used as a what-if tool to see the benefits of possible mitigation actions. This allows the tactical planner and the duty manager to plan for the actions that are more beneficial or effective if the delay situation materialises;
- The proposed approach can improve the efficiency and effectiveness of the ATM network and airline operations. The predictability of the operations at a network level could improve, reducing last-minute actions.

The article is organised as follows: Section 2 shows the rationale behind selecting the different ML models developed. Section 3 presents the features analysis conducted, the modelling approach followed, the evaluation metrics, and the eXplainable Artificial Intelligence (XAI) techniques applied. Section 4 shows the modelling approach followed to define the input/outputs of the models, the conducted feature correlation analysis, and the model selection and training. Section 5 describes the performance of the models and compares them to models which use only *static* information. The section also presents the feature importance analysis for the final models and the possible operational usage. Finally, the article closes with the conclusions in Section 6.

2. ATFM modelling rationale

Despite the effort for anticipation and the flexibility provided to the airlines, mitigating possible downstream effects of ATFM delay can be complex. If the ATFM delay assigned is large, some delay might be onward propagated, as even if the flight is ready to depart, it cannot take off until its CTOT window. On the other hand, the reactionary delay propagated by previous legs could be absorbed by the imposed delay due to the ATFM regulation, *i.e.*, some 'buffer' is generated in the schedule due to the imposed delayed departing window.

A small delay (or even zero minutes of delay, *i.e.*, Estimated Take-Off Time (ETOT) within the CTOT window) could still harm airline operations. If the flight cannot depart within its allocated slot due to other delays, *e.g.* prior reactionary, the ATFM slot will be missed, and a new slot will be required. This reassignment of a CTOT generally is performed close to the planned departing time (once the airline realises the slot will be missed). This tends to lead to significant extra delays in the flight as early slots might already not be available, and a later CTOT is then imposed. Airlines need to anticipate these situations to request a new slot to the NM as early as possible so that a new CTOT is obtained as close as possible to their ETOT. In this case, the CTOTs acts as a *threshold* during flight planning.

Airlines can sometimes respond to the ATFM regulations. For example, if the regulation issuing the delay is in the airspace, a new flight plan that avoids the congested airspace, *e.g.* re-routing, or maintaining a lower altitude (flight level capping) to avoid entering the congested airspace, could reduce (or eliminate) the issued delay. A trade-off between the cost of delay on-ground and extra required fuel and flying time to avoid the regulation will then be considered (Cook, Delgado, Tanner, & Cristóbal, 2016). Moreover, if the aircraft is ready (crew and passengers boarded), messages can be exchanged with the NM to try to benefit from potential new early slots made available due to delays or cancellations (EUROCONTROL, 2022b).

Accurate identification and prediction of ATFM regulations being issued to flights in D-1 – when there is almost no ATFM information – can reduce their impact on airspace users' operations. They can be incorporated into expert systems used by the airlines' tactical planners to draw the operational plan for the day of operations.

Four machine learning models to predict different ATFM characteristics 24 h before departure are developed and discussed in this article. These models have been defined considering the input from the members of the Dispatcher3 project's advisory board, composed, among others, of airlines, dispatchers, pilots, air transport experts and the network manager:

- **Probability of regulated flight**: Identification of flights likely to be regulated, therefore, inferring the flights that may have a ground delay and a reduced *departing window* and should be closely monitored on the day of operation. If the flights composing the airline's fleet is F. The group of flights F can be divided between expected regulated and non-regulates flights, where $R = \{f \in F \mid f \; is \, regulated\}$ and $N = \{f \in F \mid f \; is \, not - regulated\}$;
- **Protected ATFM location**: For regulated flights, whether the regulation imposing the delay is due to aerodrome or airspace restriction. Anticipating the congestion entity/region can aid in defining and implementing preventive measures to minimise disruptions. For example, if the regulation impacting the flight is in the airspace, the dispatcher could consider modifying the trajectory, *i.e.*, flight plan, *e.g.* applying level-capping or alternative routes; however, if the regulation imposing the delay is located at the airport, these approaches are not applicable, and other measures could be considered, *e.g.* flight swapping or even cancellations. Using set notation, for aerodrome ATFM regulations $A = \{f \in R \mid f \; delayed \; by \; aerodrome \; regulations\}$. For airspace regulations $S = \{f \in R \mid f \; delayed \; by \, airspace \; regulations\}$;
- **Zero VS Non-zero ATFM delay**: If the flight is regulated, the probability that the delay assigned is zero, *i.e.*, ETOT within CTOT. These flights should be closely monitored during the day of operations to avoid large unplanned delays due to missed slots, *e.g.* due to prior reactionary delays. The analysis of historical data indicates that around 1/3 of regulated flights received a delay equal to zero (see Section 4.1 for further details); this highlights the importance of these types of flights and the need for this classifier. Operational considerations, therefore, drive the development of this model. In this case, zero ATFM regulations corresponds to $Z = \{f \in R \mid f \; has \; delay \; = \; zero\}$. Non-zero regulations belong to $D = \{f \in R \mid f \; has \; delay \; > \; zero\}$.;
- **Distribution ATFM delay**: Expected value and distribution of ATFM delay, if regulated and non-zero delay. Similar to the previous model, it allows a better estimation of the potential impact and severity of regulation on current and future rotations by estimating the CTOT. To plan for the disruption due to ATFM, not only an accurate prediction of the expected delay is required, but an estimation of the uncertainty associated, as, for example, low probabilities of high delay could lead to significantly high expected costs for the airline (Delgado et al., 2020). For this last model, the delay that is predicted is for flights belonging to D.

Fig. 1 shows the interaction between models and which models trigger the subsequent ones for predicting the ATFM characteristics of a particular flight. Therefore, the figure represents the functionalities provided by the proposed system.

Positive or uncertain predictions decide if the downstream models are triggered. For example, if the *ATFM delay prediction* model indicates

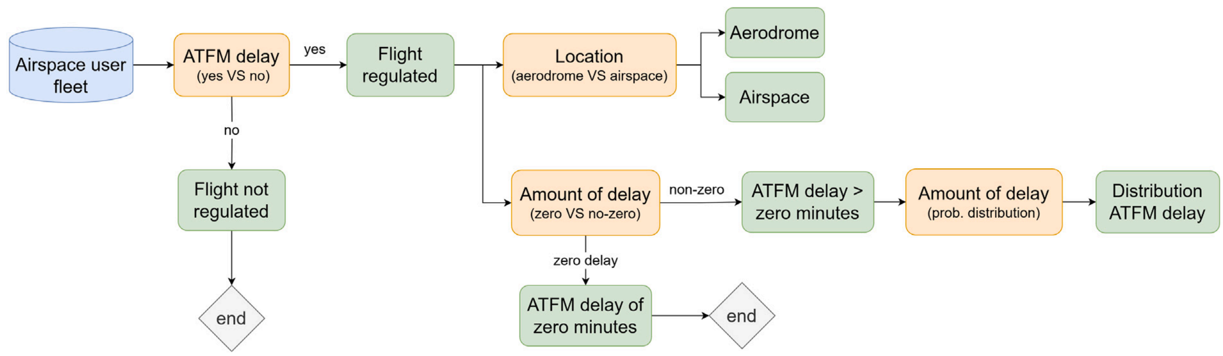


Fig. 1. Interaction between the proposed models. In orange, the ML models, and in green, their outcome.

that the flight is likely to be regulated or the model is uncertain, the model deciding where the regulation will be located is triggered. As presented before, that second model is trained with the subset of regulated flights (R). Therefore, the location of the regulation is subject to the flight being regulated, following a Bayesian approach. The end-user must consider that this second model describes the location if the flight ends up being regulated. These aspects are further described in Section 5.3, where execution examples are provided.

3. Background

The feature selection is presented in Section 3.1. The different modelling approaches implemented are introduced in Section 3.2. Section 3.3 details the evaluation metrics used. Section 3.4 reviews previous work using tools for model explainability and justifies the selected approach.

3.1. Feature selection analysis

The success of ML models primarily depends on the quality and quantity of information used to obtain the desired predictions. Good examples of the effect that the selected features can have on the performance of the models are found in (Pamplona & Alves, 2019; Rebollo & Balakrishnan, 2014). Feature selection is identifying and selecting a subset of relevant features from a larger set of features to improve the performance of a ML model. It involves analysing the importance of each feature and selecting only those most informative for the given task.

In this article, a *univariate feature selection method* with a scoring method based on an ANalysis Of VAriance (ANOVA) (Judd, McClelland, & Ryan, 2017) between labels and the features for classification and regression tasks is used to calculate a statistical score or a measure of importance for each feature in isolation, without assuming linearity in the relationship between the target variable and the features. The idea is to compute the F-statistic for each feature, which is the variance ratio between the groups to the variance within the groups. A high F-statistic indicates that the means of the groups are significantly different, and thus, the feature is more relevant for the classification task. In regression analysis, the F-test determines whether a set of independent variables (*i.e.*, features) are statistically significant in explaining the dependent variable (*i.e.*, target variable).[3]

3.2. Modelling approach

Two families of models are used to predict information related to ATFM regulations. *Binary classifiers* are used to predict the probability of events: if a flight will be regulated, which location is protected by the regulation (airspace or airport), and whether the issued delay will be zero. On the other hand, the expected issued delay and the *probability distribution* of this ATFM delay is estimated for regulated flights with an assigned delay greater than zero.

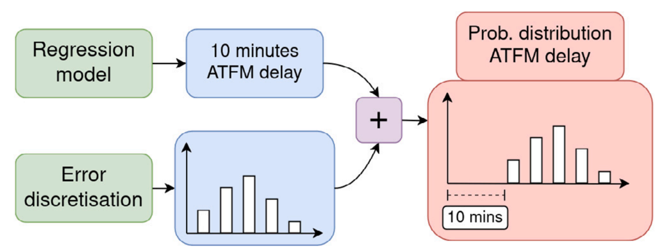


Fig. 2. Probability distribution ATFM delay approach.

A binary classifier in machine learning is an algorithm that predicts the likelihood of an input being one of two possible classes or categories. For instance, in a random forest classifier, the predicted class probabilities of an input sample are computed as the mean predicted class probabilities of the trees in the forest.

As mentioned in Section 2, to train the model that predicts the flights that will be regulated, label the flights according to: $R = \{f \in F \mid f\, is\, regulated\}$ and $N = \{f \in F \mid f\, is\, not - regulated\}$, being F is the flights from the airline's fleet.

For the model that predicts the protected location, the model considers only the flights which belong to the subset R labelling them according to the subsets $A = \{f \in R | f\, delayed\, by\, aerodrome\, regulations\}$ and $S = \{f \in R \mid f\ delayed\ by\ airspace\ regulations\}$.

Finally, to identify whether the issued delay will be zero or non-zero, the two possible classes come from the subsets $Z = \{f \in R \mid f\ has\ delay\ =\ zero\}$ and $D = \{f \in R \mid f\ has\ delay\ >\ zero\}$ are used, training the model with the flights of subset R.

For the expected ATFM delay of regulated flights, the goal is to provide a probability distribution of possible minutes of delays. To do so, the work done in (De Falco & Delgado, 2021) predicting the expected block time and in (Falco et al., 2023) for the prediction of turnaround and target off-block times have been adapted to the estimation of the expected ATFM delay. This approach transforms a regression problem into a combination of regression and classification problems. First, a regressor model predicts the target variable (*i.e.*, the amount of ATFM delay). Then, the error of this model on the dataset is discretised (from a minimum and maximum value) using a binning process and used to train a probabilistic multi-output classifier. The combination of the two models (regression and distribution of probability errors) produces the probability distribution of ATFM delay. Fig. 2 depicts the interaction between the two models.

For this model, the subsets of observations used to train the model come from regulated flights where a non-zero delay was issued (D).

[3] Python Scikit-learn tools have been used for this process. https://scikit-learn.org/ (Accessed May 2024).

Note that providing a probability distribution of possible ATFM delays, even though when it is expected to be small, instead of a single value is paramount when assessing the impact of the regulation as uncertainty plays a significant role when translating delay into costs due to the non-linearity of cost of delay (Cook & Tanner, 2015). For this purpose, a novel approach has been selected to study its potential, even though more consolidated methods exist in the literature.

3.3. Evaluation metrics

Performance metrics suited for the model type are used to measure their quality. Section 3.3.1 presents the metrics for the binary classifier. Section 3.3.2 evaluates the delay probability distributions of the ATFM delay.

3.3.1. Binary classifiers

Metrics based on thresholds, which compare the labels predicted by the models and the ground truth, are used in this article for the binary classifiers. The selected evaluation metrics are:

- **Accuracy**: proportion of correct predictions among all predictions made by the model (positives and negatives). A high accuracy score indicates that the model is making correct predictions overall;
- **Recall**: proportion of true positive instances the model correctly identified. A high recall score indicates the model correctly identifies positive instances;
- **Precision**: proportion of instances predicted as positive that are true positive instances. A high precision score indicates that the model is making accurate positive predictions;
- **F1-Score**: measure of accuracy for unbalanced datasets. The weighted average of precision and recall. A high F1-Score indicates the model has a good balance of precision and recall and correctly identifies both classes.

3.3.2. Probability distribution

In this case, the goal is not just to provide a real number from a regressor model about the minutes of ATFM delay but to consider the inherited uncertainty present in the ML models and the system. As the end goal is to estimate the uncertainty in the prediction, the evaluation metrics are designed with the same intention.

The **accuracy** of these models is computed as the difference between the expected value of the distribution and the actual ATFM delay (see red line in Fig. 3). Therefore, the Mean Absolute Error (MAE) can be computed to quantify how close the expected value of the distribution is to the actual ATFM delay.

As the classifier is trained to capture a continuous variable in a range of possible values after a binning process, it is possible to define a **measure of uncertainty** considering the range covered by a given distribution percentile (De Falco & Delgado, 2021). The average minutes required to cover 90% of the probability in the distribution is used to measure this uncertainty. The lower the uncertainty, the narrower the distribution; therefore, fewer minutes are required to cover 90% of the probability, which ensures that any possible uncertainty from the model is captured (see blue bins in Fig. 3). Measuring the uncertainty as the average time required to cover 90% of the probability in a discrete probability distribution is often referred to as a "90th percentile" or "P90" measure, which is a well-known approach. The 90th percentile represents the value below which 90% of the data falls, so it provides a good indication of the upper bound of the expected delay, capturing worst-case scenarios within the probability distribution.

Although the probability distribution provides a range of possible values, there may still be cases where there are significant discrepancies between the predicted distribution and actual ATFM delay. To better understand these extreme cases, *i.e.*, when the actual ATFM delay is much larger or smaller than the values predicted by the distribution,

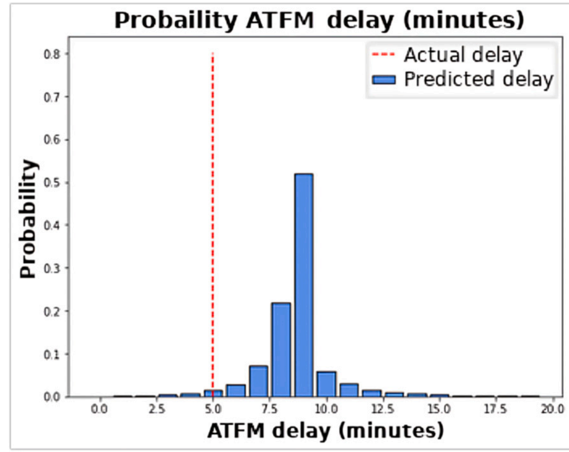


Fig. 3. Evaluation probability distribution ATFM delay.

the **Prediction Interval Coverage Probability (PICP)** is calculated as a measure of how well the computed prediction distributions cover the actual outputs (targets). Based on the definition presented in (Khosravi, Nahavandi, & Creighton, 2010), we define the PIPC as follows:

$$PICP = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{1}$$

where

$$C_i = \begin{cases} 1 & t_i \in [LPI_i, UPI_i] \\ 0 & t_i \notin [LPI_i, UPI_i] \end{cases} \tag{2}$$

LPI_i and UPI_i are, respectively, the lower and upper bounds of the prediction distribution constructed for the ith sample. t_i is the ith target value the model aims to estimate. N is the number of samples of the validation test.

Note that the classifier used to characterise the distribution is bounded by the discretisation of the error of the regressor as described in Section 3.2.

3.4. Explainable machine learning

In many real-life applications, especially those with high safety levels, the performance of the models is as important as its interpretability. That is, obtaining theoretical guarantees on the expected behaviour of machine learning-based systems during operation.

To understand the factors impacting the predictions, SHapley Additive exPlanations (SHAP) is used as it can explain the output of many machine learning models (see Lundberg and Lee (2017) for further details about SHAP). This technique is widely used in ML applied to ATM; for instance, Mas-Pujol, Salamí, and Pastor (2022) employed SHAP to study the influence of both scalar and image-based input features predicting the likelihood of traffic volumes to be regulated. Dalmau (2022) used it to understand the outcome of the proposed ensemble method to predict the likelihood of re-routing to mitigate ATFM regulations. Lambelho, Mitici, Pickup, and Marsden (2020) utilised SHAP to explain the models used for strategic slot flight assignment at London Heathrow Airport. Xie, Pongsakornsathien, Gardi, and Sabatini (2021) used it in a more general manner to explain ML solutions in ATM.

SHAP calculates the contribution of individual input features to the predictions made by the model. This quantification is expressed using numerical values, where each value represents the degree of influence that a specific feature has on a particular prediction. SHAP values can be positive or negative and around zero:

- **Positive** indicates that the presence or elevated value of that feature increases the model's prediction or output, driving the prediction in a favourable direction.

Table 1
Data sources used to predict ATFM regulations for individual flights.

Data source	Period time	Usage	Comment
Airline data	2018	Labelling	ATFM information (probability, location, and delay)
DDR2	2018	Features	Flight intentions (origin, destination, and Scheduled Off-Block Time (SOBT)) for the airline of interest
Airports data	Static	Features	Size (small, medium, big) and hub information (no, yes)
PREDICT	2018	Features	Airport demand (number departures and arrivals)
	2018	Features	Network demand (entry and occupancy count)
NOAA	2018	Features	Weather (origin/destination airport)

- **Negative** signifies that its presence or elevated value has contributed to reducing the model's prediction or output, steering the prediction in an adverse direction.
- **Zero** suggests that the presence or value of that feature did not influence the prediction for that specific observation. In essence, the feature played no role in determining the model's output.

Luo, Wu, Wang, Wang, and Meng (2021) found that SHAP provides a more intuitive and interpretative way of understanding the relationships between input features and the model's predictions than other model-agnostic techniques such as Local Interpretable Model-Agnostic Explanations (LIME).

4. Modelling

Section 4.1 introduces the data sources used. Both to label the observations and to engineer the input features. Section 4.2 summarises the engineered input features that train the models. Section 4.3 specifies the output of each model — the targets or labels. Section 4.4 presents the feature correlation analysis, determining the more relevant features to train the models. The supervised ML models are described in Section 4.5.

4.1. Data sources

The models presented in Section 4.5 aim to forecast ATFM regulations from a flight-centric perspective. Data from a European airline with a high volume of operations have been used to label the observations. This guarantees accurate data for the labelling process and their availability. Concretely, we have used data from 2018, composed of approximately 200,000 flights (F) from which 66,000 were regulated (R) – around 33% of all operated flights.

From the regulated flights, 61% were due to airspace ATFM regulations (S) while 39% have as protected location type the aerodrome (A). 31% of the regulated flights received a delay equal to zero (Z), 13% a delay greater than zero but smaller than 5 min, and only 10% a delay greater than 20 min. It is worth mentioning that cancelled regulations are omitted, and no filtering has been applied according to the regulation's reason.

The datasets for the binary classifiers are not balanced, *i.e.*, the distribution of classes is uneven. However, using evaluation metrics such as the F1-Score will help us determine if the models are reporting a balanced performance or whether the models have a preference between the two possible classes.

To create the input features, the Data Demand Repository 2 (DDR2) from EUROCONTROL is used to extract the Flight Intentions (FIs) (origin, destination, and SOBT). Static airport information specifies the size of the departure/arrival airports and whether they are used as a hub for the airline. To estimate the demand of the crossed airspace sector, it is necessary to have access to the flight plans. However, they are not normally available at the selected prediction horizon. The PREDICT software is the NM support tool intended to estimate the flight plans when those still need to be submitted. The implementation utilised in this article is purely based on historical flight plans, while the official implementation has additional features such as a catalogue or the shortest router finder for those flights not available in the historical data. National Oceanic and Atmospherics Administration (NOAA) is the selected source of weather forecast (National Oceanic and Atmospherics Administration (NOAA), 2024).

Table 1 collects all the data sources and their usage.

4.2. Input features

When developing machine learning models, ensuring that the data used to train the final models are available during execution time is crucial (Delgado, Mas-Pujol, Skorobogatov, Argerich, & Gregori, 2022). However, one of the challenges in the aviation field is that historic datasets tend to contain snapshots of released data, making it very difficult to know what data were available at a given moment. This is particularly relevant for the network and weather data. Therefore, the same label can be predicted with the information available at different horizons. As this article focuses on predictions at D-1, we need to ensure that data sources used are available 24 h before the SOBT. Note that as we are estimating characteristics of ATFM, if the prediction horizon is too close to the operation of the flight, *e.g.* a few hours before departure, the information on the regulations would already be known, and no benefit for planning purposes could be obtained from the models.

It is helpful to differentiate between *static* and *dynamic* data (and features). *Static data* (and features) do not vary as a function of the prediction horizon. Examples of these data are origin and destination, aircraft type, or time of the day when the flight is scheduled. *Dynamic data* might change at different horizons. For example, the weather forecast might be updated over time, or the expected demand at the arrival airport might differ depending on the available flight information. This distinction allows us to estimate the importance of the dynamic features on the overall performance of the algorithms and, therefore, the relevance of the prediction horizon for a particular problem.

Table 2 collects all the engineered features grouped by topic, their definition, and the feature type (static or dynamic). Hour, day, and month of planned departure characterise the day of operations and the seasonality. The characteristics of departure and arrival airports are also considered. We used the size of the airports and if it is used as a hub by the airline as defined in (Gurtner, Delgado, & Valput, 2021). Also related to the airports, the normalised number of departures and arrivals in the same hour as planned by the flight is computed using the OD pairs and ETOT. The normalisation of these features has been done using two techniques. First, by the average number of departures/arrivals at the same flight hour in the previous thirty days. Second, by the average number of flights in the previous thirty days at the same hour and day of the week. Note that the names of the airports are not explicitly used, even if state-of-the-art machine learning techniques can handle high-cardinal features, as the goal is to create models which are as general as possible. The specific airports will, indeed, be correlated with some of the features already considered by the model, *e.g.* size and congestion, and therefore their explicit consideration is not necessarily required to achieve high-performance models (as shown in this article).

Information about the expected congestion of the network is also used. In this case, we show the models the normalised Occupancy Count (OC) and Entry Count (EC) between all the elementary crossed sectors in the planned routes. OC and EC are normalised with respect to the average number of flights inside the sector in the previous thirty

Table 2
Input features grouped by topic.

Topic	Feature	Definition	Type
Operational time	Hour departure	Hour from SOBT. Value from 0 to 23	Static
	Day week departure	Day week form SOBT. Value from 0 to 6	
	Month departure	Month form SOBT. Value from 0 to 11	
Airport static information	Size departure airport	Three size {small, medium, large}	Static
	Size arrival airport	Three size {small, medium, large}	
	Hub departure	Used as a hub by the airline {no, yes}	
	Hub arrival	Used as a hub by the airline {no, yes}	
Airport demand	Normalised departures hour	Departures respect the same hour	Dynamic
	Normalised departures day	Departures respect the same day of week	
	Normalised arrival hour	Arrivals respect the same hour	
	Normalised arrival day	Arrivals respect the same day of week	
Network demand	Normalised OC average	OC/avg. OC. Most crowded crossed sector	Dynamic
	Normalised OC max	OC/max. OC. Most crowded crossed sector	
	Normalised EC average	EC/avg. EC. Most crowded crossed sector	
	Normalised EC max	EC/max. EC. Most crowed crossed sector	
Weather	Visibility depart/arrival	Directly from NOAA divided by 12000	Dynamic
	Wind depart/arrival	Directly from NOAA (Knots) divided by 30	
	u-wind depart/arrival	Directly from NOAA (Knots) divided by 30	
	Temperature depart/arrival	Directly from NOAA (F) divided by 125	
	Rel. humidity depart/arrival	Directly from NOAA divided by 0.0015	
	Geopotential depart/arrival	Directly from NOAA divided by 25000	
	Ventilation rate depart/arrival	Directly from NOAA divided by 40000	

days in the same hour. They are also normalised using the maximum values considering the previous thirty days, in the same hour of the day. Finally, the weather conditions at the departure and destination airports are modelled based on the NOAA weather forecast.

4.3. Output targets — Labelling

The expected outcome of the models depends on the selected ML algorithm, the labelling used, and the type of prediction (binary, regression or classification). For the binary classifiers that predict the *probability of regulated flight* regulation, a label equal to zero indicates a non-regulated flight; otherwise, the label is one. This differentiates between the subsets R and N as indicated in Section 2. The *protected location* would use a label equal to zero if the flights were regulated due to aerodrome congestion, while one corresponds to congestion in the airspace, generating subsets A and S. The model that predicts whether the ATFM delay will be *zero* would utilise a label equal to one if the flight had zero minutes of delay, for the definition of subsets Z and D.

Note that while the output from a random forest model can provide a measure of confidence or certainty about a prediction, it is not a probability in the strict sense. However, you can indirectly estimate class probabilities with a random forest by looking at the proportion of trees that vote for a particular class (Liaw, Wiener, et al., 2002).

For the *probability distribution of ATFM delay*, as the name indicates, we want to predict a probability distribution, combining a regressor and a classifier, as explained in Section 3.3.2. For the regressor, the labelling used is the actual issued minutes of ATFM delay to each of the flights. For the probabilistic classifier, once the prediction error from the regressor has been computed on the training set, the next step is to discretise and bin the error of this first model to train a probabilistic classifier model to predict these errors.

4.4. Feature selection analysis

Fig. 4 shows the sum of the scores (F-statistics) by the different features grouped per *topic* obtained applying an ANOVA analysis as explained in Section 3.1. This provides a measure of each feature's relative importance in the problem's context. Note that different *topics* have different amounts of features. For instance, the topic of operational time has three features, while the topic of weather contains 14. Therefore, weather may report a much larger overall relative importance. However, no normalisation according to the number of features per topic has been applied to show their overall contribution.

The previous figure (Fig. 4) provides an overall view of the results obtained during the future selection analysis. However, as Section 3.1 explained, an individual analysis has been performed per model.

For the *probability of regulated flight*, weather information shows high significance when explaining the probability of the event. Especially the wind components and speed, geopotential,[4] and temperature. It is also highly correlated with the size of the arrival airport, the congestion at both airports and network demand features. On the other hand, the operational information and the congestion at the airports are less correlated with the target features, but their correlation cannot be ignored. Finally, the feature that indicates whether the arrival airport is used as a hub by the airline does not add information; thus, it is removed from the final training dataset.

When estimating the *protected location*, the ANOVA analysis shows that the static information about the airports has high variability with the labelling, providing meaningful information. Especially the size of the airports and whether the departure airport is used as a hub by the airline. Similar to the previous case study, the most correlated weather features related to the weather are the wind, followed by the geopotential, and the ventilation rate[5] of the departure airport. The rest of the input features do not present a clear pattern. However, the contribution is not negligible, except for the congestion in the same hour and the ventilation rate of the arrival airport, which are removed from the training dataset as the score is smaller than one.

Focusing on whether the ATFM delay will be *zero*, the static information about the origin/destination airports is the most correlated information with the target, and weather information plays a minor role than in the previous case studies. Furthermore, something important to note is that the overall score of the features is significantly smaller. Previously, the observed scores for individual features were around three, while now they are around 0.5. This indicates that the overall significance of the selected features is very low when predicting

[4] Geopotential is a way of measuring height in the atmosphere that considers the Earth's gravity, but instead of using physical meters or feet, we use a hypothetical unit of energy called gravitational potential energy. By measuring the geopotential height of a pressure surface, we can better understand how the atmosphere is structured and changes over time.

[5] Ventilation rate is based on multiplying transport wind to mixing height. Mixing height represents the height of the mixed layer (or parcel of air) that would rise in the atmosphere due to atmospheric mixing or turbulence. Transport wind is the average wind speed through the mixed layer.

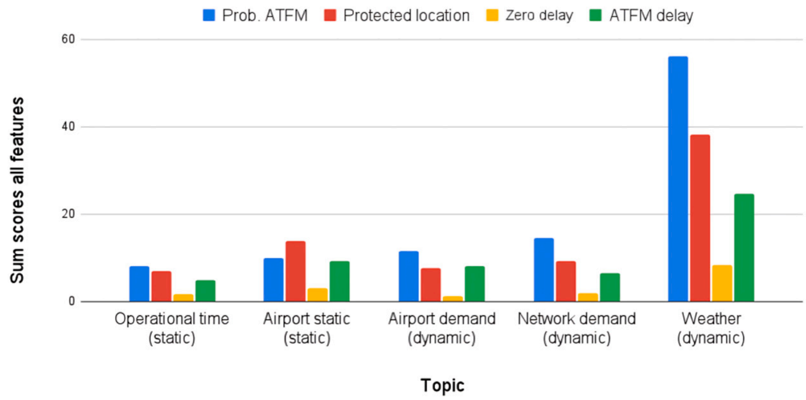


Fig. 4. Feature correlation ATFM characteristics.

whether the identified ATFM regulation will issue a delay equal to zero minutes for specific flights. The results make sense as the final imposed ATFM delay comes from the CASA algorithm based on the first-in-first-served principle. For instance, the condition of the network does not determine the issued ground delay. The congestion at the airports, the month of the year, and the visibility are removed from the final training dataset due to their low correlation.

Finally, the feature selection analysis between the selected features and the *actual ATFM delay* shows a medium level of variance compared to the previous three case studies. Interestingly, although the selected features exhibit a lower correlation for the labelling used to predict whether the delay is zero, this is not the case for the actual ATFM delay with an average score of around two. The most correlated features are the departing hour and the characteristics of the departure airport. Then, the wind, the geopotential, the network congestion normalised by the maximum historical values, and the temperature. All features with a score lower than one are removed from the training dataset.

4.5. Model training

Model training includes selecting the model and hyper-parameters that best fit each problem. The approach followed in this article is based on a *GridSearch analysis* to find the best models and their hyper-parameter (Hastie, Tibshirani, Friedman, & Friedman, 2009) using the training/validations datasets.

4.5.1. Binary classifiers training

The balanced accuracy score has been used for the binary classifiers as the scoring function to ensure that the selected algorithm and hyper-parameters are as optimal as possible for both classes for the final model. Based on the literature, the analysed algorithms are the MLPs (*i.e.*, NN), decision tree, random forest, AdaBoost, and Linear SVCs.

After the analysis, the supervised algorithm that stands out is the **random forest classifier**, with different configurations for the target variable. Table 3 shows the hyper-parameters of the different binary models and a reminder of the subset of flights used to train the models as described in Section 2. For the *probability of regulated flight* and *protected location*, the GridSearch analysis reports that the ML algorithm is a random forest classifier using a criterion equal to 'gine', a maximum depth of 50, and 200 estimators. For the binary classifier that aims to predict whether the *ATFM delay is going to be zero*, the results indicate that also a random forest classifier is the best candidate, but using a criterion equal to 'gine', a maximum depth of 50, and 25 estimators.

The results obtained from the GridSearch analysis match what can be found in the literature, showing that simpler models tend to perform better than complex ones for flight-centric approaches. For instance, Gui et al. (2019) show that random forest-based models report higher performance than complex artificial NN (*e.g.* LSTMs, which tend to under-perform and overfit) when predicting flight delays.

Another important benefit of using this type of classifier is the low training and prediction complexity. A random forest presents a prediction complexity of approximately $O(N log(N))$ in the worst-case scenario. N is the number of trees or estimators. This simplifies to approximately $O(1100)$ operations with 200 estimators, which is relatively small and indicates that making predictions with tree-based models should be computationally efficient, even for large datasets and many estimators.

4.5.2. Probability distribution training

Similar to the binary classifiers, different supervised algorithms have been evaluated using a *GridSearch* analysis to discover the best possible candidate to predict the ATFM delay of regulated flights. To do so, MLPs for regression, decision tree regressors, random forest regressors, AdaBoost regressors, and Ridge have been studied. In each case, all possible combinations of a pre-defined search space have been used to train models and discover the best possible candidate. Table 4 collects the hyper-parameters of the best model to predict the *minutes of ATFM delay*, which is a **random forest regressor** using a 'squared_error' as the criterion, a maximum depth of 100, a maximum number of features equal to 'auto' which means that all the features are used, and 25 estimators. Note that this model is trained on the subset of flights that are regulated and have a positive delay (D).

Once trained, the regressor and the prediction error are computed on the training set, and the next step is to discretise and bin the error from the regressor so that a classification model can be trained to predict these errors. After the error distribution analysis, an error range of [−20, 20] minutes is selected, corresponding to the 90th percentile of the probabilities in the distribution error with a bin size equal to 2 min to ensure a good resolution.

Similar to the previous tree-based classifiers, the tree-based regressor and the selected NN present a low training and prediction complexity (e.g., the NN uses one hidden layer with 25 units). Focusing on the prediction complexity, the models will be very efficient in evaluating new scenarios and predicting the expected ATFM delay.[6] Therefore, as in the previous cases, there is no need to use advanced techniques such as the ones based on knowledge distillation to reduce the complexity of the trained models (Xiao et al., 2023; Xiao, Xing, Zhao, et al., 2024).

Table 5 collects the hyper-parameters for the multi-output classifier required to predict the probability distribution of the error of the regressor. The main specification of the classifier has been extracted from (De Falco & Delgado, 2021), and different combinations of MLPs have been tested to find the best hyper-parameters. The analysis reports that the best performance is obtained using the optimiser *Adam*, one

[6] A prediction can be generated in approximately 0.2 s with a 9th generation Intel Core i7.

Table 3
Binary classifiers models parameters.

Target problem	Training subset	ML algorithm	Hyper-parameters	Value
Probability regulated flight (yes VS no)	F	Random forest classifier	Num. estimators Max. depth Criterion	200 50 gini
Protected location (aerodrome VS airspace)	R	Random forest classifier	Num. estimators Max. depth Criterion	200 50 gini
Zero ATFM delay (zero VS non-zero)	R	Random forest classifier	Num. estimators Max. depth Criterion	25 50 gini

Table 4
Regressor — Probability distribution ATFM delay models parameters.

Target problem	Training subset	ML algorithm	Hyper-parameters	Value
ATFM delay (minutes)	D	Random forest regressor	Num. estimators Max. depth Max. features Criterion	25 100 auto squared_error

Table 5
Classifier — Probability distribution ATFM delay.

Target problem	Training subset	ML algorithm	Hyper-parameters	Value
ATFM delay (prob. distribution)	D	MLP	Learning rate Num. hidden layers Num. units Batch size Optimiser Activation functions Weights initialisation Learning rate	0.0001 1 25 64 Adam Relu Glorot uniform 0.0001

hidden layer with 35 neurons, a 'relu' activation function, and a learning rate equal to 0.0001. The GridSearch analysis is used to define the hidden layers.

It is worth mentioning that using cross-entropy as a loss function for the classification model enables us to consider the probabilities of each class (bin) in the prediction as a probability for the real value being on that bin. Therefore, the possible values on all bins (shifted by the expected value of the regression model) can be understood as a probability distribution of the target variable (De Falco & Delgado, 2021).

5. Results

The performance of the optimised models was evaluated on the test set, using the typical 80%–20% train–test split. Section 5.1 shows the performance of the different models. Section 5.2 presents the feature analysis based on SHAP values to interpret the models' predictions. Section 5.3 provides some examples of the advice capabilities of the system, integrating the outcome of the different models as integrated into an expert system. Finally, Section 5.4 compares the results from this article with the ones obtained in a previous publication (Mas-Pujol, Salamí, & Pastor, 2021) where post-operational data were used.

To better show the contribution of the *dynamic* datasets (and features), the performance of the models will be compared to a *static model* trained using only the *static* data (and features). For each of the models, it is presented the proposed metrics using only static data (*i.e.*, operational time and static airport information) and all the proposed features (*i.e.*, *static* and *dynamic* features).

5.1. Performance metrics

Table 6 shows the results from the binary classifier. The *probability of regulated flight* exhibits a balanced performance with all the metrics around 82%. It is interesting to see how the dynamic features significantly increase the performance of the models, especially the recall. This indicates that the additional features positively contribute to the detection of possible samples (*i.e.*, regulated flights), also increasing the overall accuracy. The *protected location* model for regulated flights shows the best performance, taking advantage of the dynamic features. For this model, we can see that the precision is the metrics with a larger improvement, indicating that the model can distinguish positive predictions better. The better recall and precision, indicating a more accurate detection of airspace regulations, improved around 10% of the overall performance with a balanced accuracy of around 87%. Finally, the model that predicts whether the ATFM delay will be zero reported accuracy and F1-Score close to 70%. The drop in the performance is expected due to the low correlation of the input features (see Fig. 4, compared to the previous models). Probably, the model will benefit from features related to the information used by the CASA algorithm, such as the expected moment in which the flight will enter the congested region. However, this approach will require a tool that identifies what and when the crossed traffic volumes will be regulated in the selected prediction horizon, information unavailable at D-1.

Note that the results for the *protected location* are conditioned on the probability of a flight being regulated, as the model is trained on the subset of flights that are regulated (R). The protected location will be only estimated for flights expected to be regulated. A single model which directly predicts if a flight will be regulated and the protected location has been tried, *i.e.*, trained using all the flights (F) and differentiating between three classes: non-regulated, regulated due to airspace or regulated due to airports. This model shows a slightly worse performance, with an accuracy of 0.69, while the current approach shows a combined accuracy of 0.71. Moreover, the Dispatcher3 project's advisory board validated the current approach, stating that the proposed approach will add more value to the process as it will allow the decision-maker to consider the area where the flight would be regulated if that is the case, following a Bayesian approach.

Table 6
Evaluation metrics binary classifiers.

Model	Features	Accuracy	Recall	Precision	F1-Score
Prob. flight regulated	Static	0.71	0.66	0.85	0.74
	Static and Dynamic	**0.82** (+0.11)	**0.81** (+0.15)	**0.82** (−0.03)	**0.82** (+0.08)
Protected location	Static	0.78	0.76	0.75	0.75
	Static and Dynamic	**0.87** (+0.09)	**0.84** (+0.08)	**0.89** (+0.14)	**0.86** (+0.11)
Zero delay	Static	0.59	0.68	0.58	0.62
	Static and Dynamic	**0.68** (+0.09)	**0.67** (−0.01)	**0.69** (+0.11)	**0.69** (+0.07)

Table 7
Evaluation probability distribution.

Features	MAE (mins)	Mean 90% probability (mins)	PICP (%)
Static	12.38	20.13	0.83
Static and Dynamic	**9.58** (−2.8)	**12.87** (−7.26)	**0.88** (+0.05)

Table 7 presents the results when predicting the probability distribution of the expected ATFM delay for regulated flight with a non-zero delay. As can be seen, the addition of dynamic features improves performance considerably. The value of the MAE, which estimates the error between actual and expected delay, is reduced to 9.6 min (an improvement of 2.8 min). Furthermore, the results show a reduction of around 7 min in the uncertainty: from 20 to 13 min. The predicted probability distribution is narrower, requiring a smaller time range to cover 90% of the overall probability. Using only static features, the PICP (actual ATFM delay inside the predicted distribution) is 83%, while it improves by 5%, to 88%, when adding the dynamic features. Therefore, the dynamic data not only reduces the uncertainty of the predictions (narrower probability distribution) but increases the PICP (ATFM delay falls within the predicted distribution). Similarly to the zero delay model, the model will probably benefit from additional features linked to the information used by the CASA algorithm.

5.2. Feature importance — SHAP values

The graphs below aggregate the features and observations for the top 10 most relevant features, showing on the y-axis the name of the features and the x-axis the average contribution to the final prediction. A colour schema has been used to easily identify topics of the different input features (*e.g.* light blue for operational information), and those topics that only contain static features are distinguished using a diagonal line pattern.

Fig. 5 depicts the SHAP analysis for the *probability of regulated flight, purple for airport-related static information*. The model prioritises the size of the arrival airport, the expected wind, the network demand, the geopotential, and operational information such as the day of the week or the hour. Aerodrome ATFM regulations are issued due to high demand at the destination airport; thus, the importance of the arrival airport's size as a bigger airport implies more traffic and higher probabilities of regulations. The relevance of wind at the arrival airport can drive similar conclusions. The expected network demand is directly related to airspace ATFM regulations, and the geopotential is linked to the altitude of the airports, *e.g.* to the likelihood of adverse weather conditions. Finally, the hour and day of the week are highly relevant because most regulations are issued in the morning to avoid downstream effects or on the weekends due to the larger traffic volume.

Fig. 6 shows the SHAP analysis of the *protected ATFM location (aerodrome or airspace)*. In this case, *static* features play an important role. The size of the arrival airport is the most relevant input feature, followed by whether the arrival airport is used as a hub by the airline and some operational information, such as the day of the week or the departing hour. The model mainly considers wind-related information as the *dynamic* features.

Fig. 7 presents the SHAP analysis of the trained model to predict *whether the expected ATFM delay of a regulated will be zero or non-zero*. In this case, the most relevant features are the flight operational information, the geopotential, and the wind. However, if we focus our attention on the values on the x-axis, the range of reported SHAP values is much smaller than in the previous models (previously from 0 to 0.1, and now from 0 to 0.05). This indicates that the model cannot extract much information from the features directly related to the observed low performance.

For the *probability distribution of ATFM delay*, Fig. 8 presents the SHAP analysis of the trained random forest regressor, which estimates the minutes of ATFM delay. The analysis shows that the hour of departure is the most relevant feature. The next most relevant features are the wind, the network demand, the congestion at the airports, and the geopotential. Blue dots represent low values of the features while red high values. Thus, the SHAP values of the hour of departure indicate that most non-zero ATFM regulations are implemented in the morning, while high wind, network demand, and geopotential might be responsible for the regulation.

Finally, Fig. 9 shows the results from the SHAP analysis for the classifier used to characterise the error of the regressor. The *static* hour, day, and size of the arrival airport are, in this case, the most important features. The wind, the airport's congestion, and the departure aerodrome's size follow them. The selected error range is [−20, 20] minutes with a bin size equal to 2 min (20 classes), as explained in Section 4.5.2. Therefore, class 0 corresponds to −20 min correction, class 10 to 0 min, and class 20 to 20 min. If we focus our attention on the legend in Fig. 9, we can see that the most frequent classes are between 0 and 10, indicating the classifier is trying to compensate for an overestimation from the regressor. As seen before, predicting the minutes of ATFM delay is the most challenging part.

5.3. Operational usage

The previously presented models can be integrated into an expert system that provides the tactical planners and duty managers with the relevant information when defining the operational plan on D-1 and when monitoring and managing the fleet on the day of operations, respectively. The proposed system, therefore, combines the outcome of the different models, ensuring that only the necessary information is displayed and avoiding an overload of information. Furthermore, the proposed colour schema indicates the uncertainty of the models. Uncertain predictions are shown in red; otherwise, green is used.

The simplest scenario is the one where the flight is not regulated. The closer to zero the predicted value by the *probability of regulated flight* model, the less likely it is for the flight to be regulated. For example, Fig. 10(a) shows that, in this case, the flight has a 12% probability of being regulated. The outcome is displayed in green as the model is certain about the prediction.

On the other hand, if the model that predicts the probability of regulation is uncertain, *i.e.*, probability $\geq 0.25 \leq 0.75$, the outcome is displayed in red, and the analysis is extended by evaluating the protected location and whether the issued delay would be zero (see

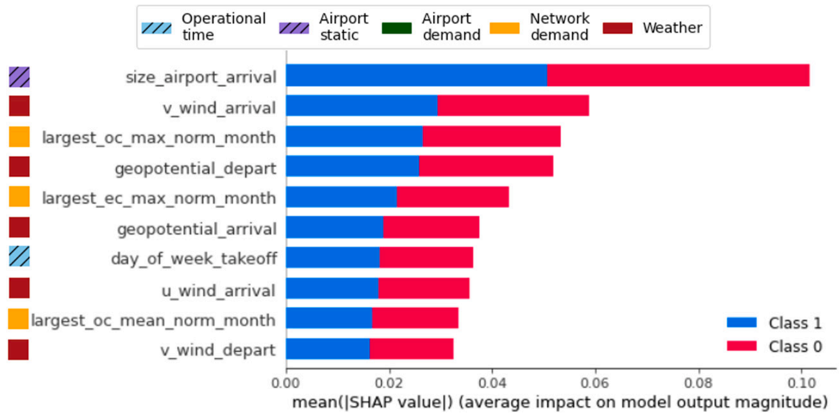


Fig. 5. Top-10 SHAP analysis probability of regulated flight.

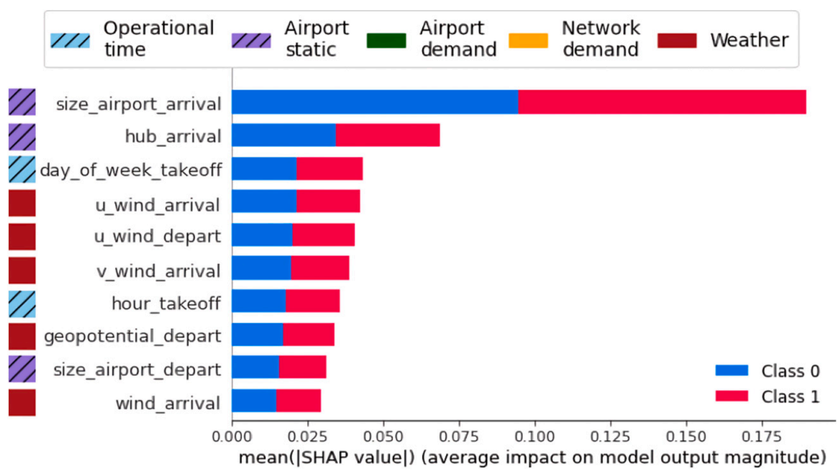


Fig. 6. Top-10 SHAP analysis protected ATFM location.

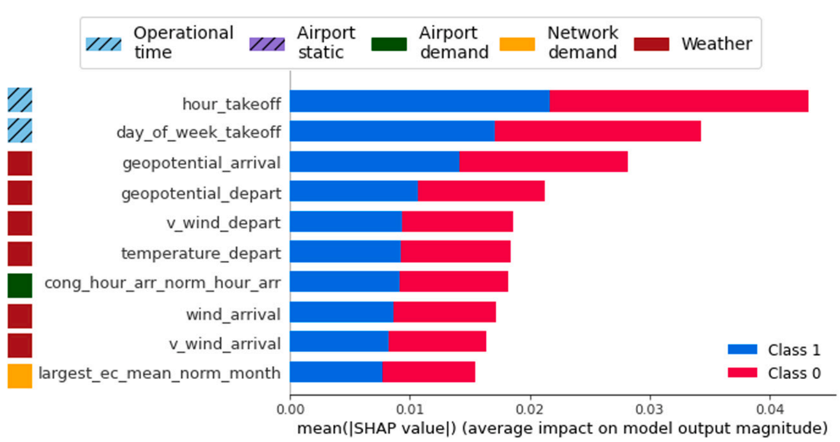


Fig. 7. Top-10 SHAP analysis zero ATFM delay.

Fig. 10(b)). If the flight is expected to be regulated, *i.e.*, probability of being regulated >0.75, the predictions of location and amount of delay will also be carried out and displayed (as shown in Figs. 11 and 12). The protected location is represented on the integration view by changing the top level between *aerodrome* or *airspace* as a function of the highest predicted value.

For flights with high uncertainty or likely to be regulated with a delay different than zero, the system triggers the models required to show the amount of delay assigned to the flight. Fig. 11 shows the predictions at D-1 obtained for a flight from *LIQR* to *LFPO* with a SOBT at 8:50. The framework predicts with a high probability the flight will be regulated due to airspace congestion, but it is uncertain about the amount of delay (probability zero delay around 0.5). If the flight is regulated with a positive delay, the expected ground delay is 5 min with an uncertainty (90th percentile) of 14 min.

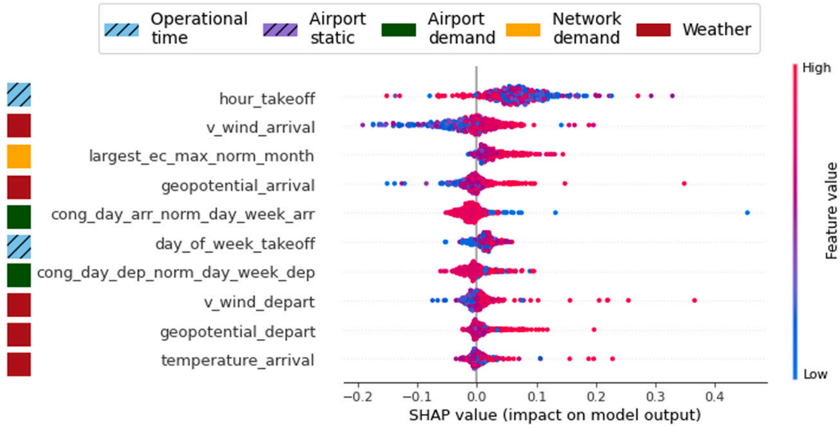


Fig. 8. Top-10 SHAP analysis regressor ATFM delay.

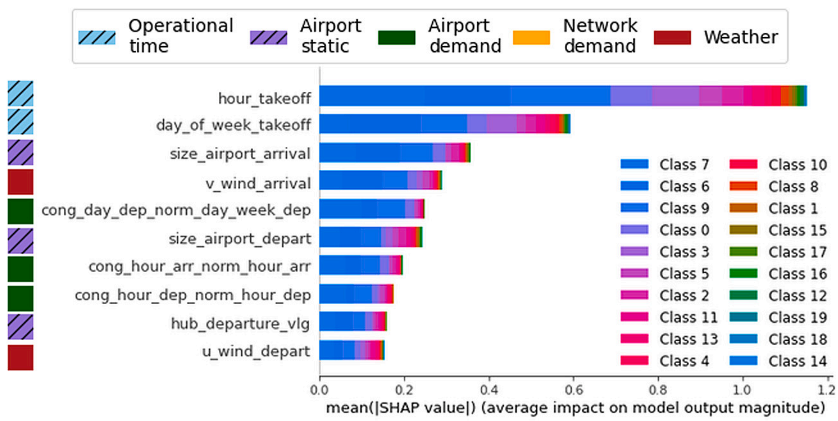
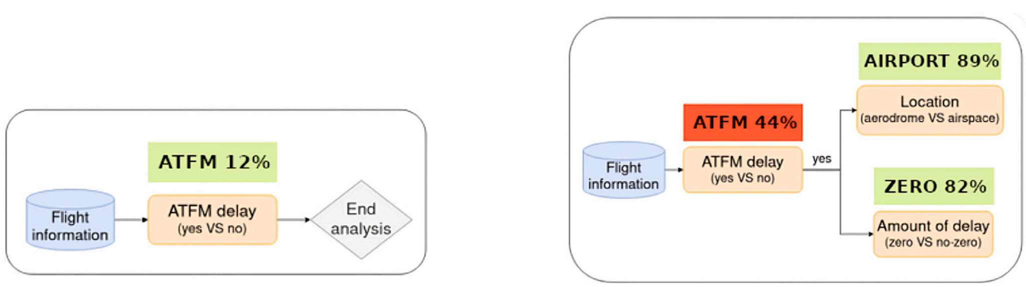


Fig. 9. Top-10 SHAP analysis multi-output classifier ATFM delay.



(a) Certain probability no regulated flight

(b) Uncertain probability regulated flight & Certain zero delay

Fig. 10. Certain and uncertain probability of ATFM regulation.

Airlines need to monitor flights that are highly likely to be regulated closely. Therefore, the tactical planner could benefit from identifying those flights during the pre-tactical phase and actively producing new flight plans and preparing solutions (*e.g.* aircraft swaps) to reduce the impact of ATFM delays on their fleet. Nonetheless, airlines, through the duty manager, can also benefit from uncertain predictions as red flags linked to flights that should be monitored as their status could change during the day of operations. The proposed framework could also be used to evaluate the newly produced flight plans, using a what-if scenario, to determine the implications of changes in the initial flight plan. For example, Fig. 12 evaluates the same previous flight but with a different horizontal route, crossing less congested sectors. The new proposed route has been extracted from historical data, using a similar principle to the one proposed in the PREDICT implementation used in the article, *i.e.*, using historical information. In this new scenario, the models are still certain about the ATFM regulation, indicating the flight is still regulated but with a high probability of a delay equal to zero. Other combinations could be tested if desired.

5.4. Available at prediction-horizon vs post-operational data sources

This section compares the performance of the models presented in this article and the one obtained in (Mas-Pujol et al., 2021). In (Mas-Pujol et al., 2021), the models were trained using data available post-operationally, *e.g.* demand estimated using the last filled flight plan, instead of from PREDICT, and actual weather information (METARs)

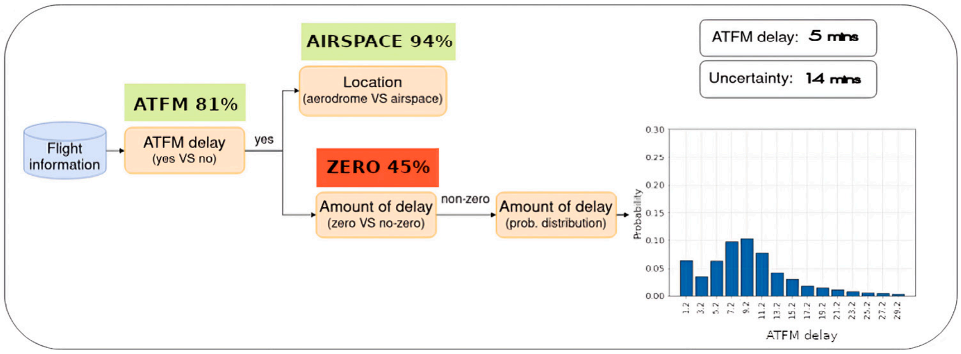


Fig. 11. Certain probability regulated flight & Uncertain probability zero delay.

Table 8
Comparison of performance binary classifiers using forecasts and perfect data.

Model	Experiment	Accuracy	Recall	Precision	F1-Score
Prob. regulated flight	Post-operational	0.88	0.90	0.86	0.89
	Forecast	**0.82** (−0.06)	**0.81** (−0.05)	**0.82** (−0.04)	**0.82** (−0.07)
Protected location	Post-operational	0.84	0.80	0.83	0.82
	Forecast	**0.87** (+0.03)	**0.84** (+0.04)	**0.89** (+0.09)	**0.86** (+0.04)
Zero delay	Post-operational	0.73	0.75	0.71	0.76
	Forecast	**0.68** (−0.05)	**0.67** (−0.08)	**0.69** (−0.03)	**0.69** (−0.07)

Table 9
Comparison of performance probability distribution ATFM delay using forecasts and perfect data.

Experiment	MAE (mins)	Mean 90% probability (mins)	PICP (%)
Post-operational	9.37	13.56	0.87
Forecast	**9.58** (+0.21)	**12.57** (−1.01)	**0.88** (+0.01)

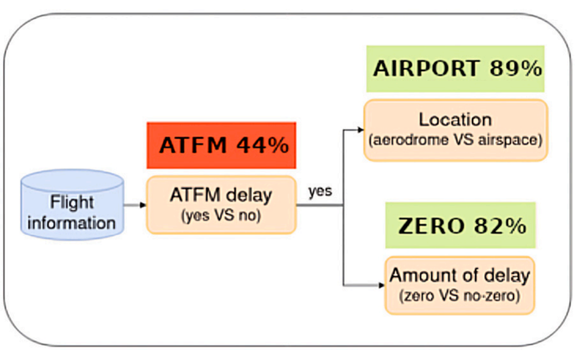


Fig. 12. Certain probability regulated flight & Certain probability zero delay.

at the moment of operation instead of a weather forecast. These data are not available at the prediction horizon D-1; however, it provided a baseline to compare the performance of the models when using information with less uncertainty.

Table 8 presents the results on the accuracy, the recall, the precision, and the F1-Score for the models of (Mas-Pujol et al., 2021) (post-operational) and the ones of this article (forecast). As shown, for the *probability of regulated flight* and *zero delay* models, using forecast data, there is a drop in performance of around 5%. However, the model for *protected location*, which identifies if the regulation impacting the flight is located in the airspace or the arrival airport, shows an improvement of 5% when using the forecast data.

Similarly to the *protected location* model, using data sources based on the forecast (available at the prediction horizon) also improves the performance of *probability distributions of ATFM delay*. Table 9 compares the performance obtained by the model trained in this article and the one using post-operational data from (Mas-Pujol et al., 2021). Using the information available in the prediction horizon reduces the prediction's uncertainty with minimal accuracy reduction and an equivalent PICP.

Overall, using post-operational data does not improve all the models and provides a minimal enhanced performance while complexifying the deployment of the models.

6. Conclusions

The European ATM Master Plan (SESAR, 2020) envisions earlier information sharing between stakeholders, aiming to improve collaborative planning related to demand-capacity imbalances. The available data and the use of ML techniques could be key to incorporating new tools into the current pipeline when facing ATFM regulations. Tools based on supervised techniques could be the answer due to their fast response to new scenarios, creating a human-machine methodology where the human is still in charge of critical situations.

In this article, we propose using ML models trained on data available at D-1 to infer the impact of ATFM regulations on individual flights. This critical information would support the airlines' tactical planners in improving their operational plans, anticipating disruptions and highlighting which flights should be closely monitored by duty managers during the day of operations. Not only if the flight is regulated, but the main characteristics of the regulations could improve the planning and implementation of preventive actions. Concretely, we have trained models able to predict the probability of regulations for individual flights, the protected location, whether the issued ATFM delay will be zero, and the probability of the minutes of delay for regulated flights with information available at D-1. These particular models have been selected as they provide actionable operational information to the user. Furthermore, the low prediction complexity of the proposed models overcomes any time constraint issue that can arise when deploying the models (Paleyes, Urma, & Lawrence, 2022).

The features analysis highlights the most relevant features for each of the four selected ATFM models. The main conclusion from this analysis is the need for specialised features when predicting information related to the minutes of ATFM delay. Moreover, the models will always benefit from extra information about the less frequent regulations (*e.g.* military actions). However, ensuring that the airline's data sources are available is essential.

Results show that, even though the models are imperfect, they can extract patterns in data to accurately identify ATFM-related information. The models can predict which flight will be regulated and the protected location (aerodrome or airspace) with an accuracy of around 85%, whether the issued delay will be zero with an accuracy of 70%, and reported a MAE of 9 min between the actual as estimated minutes of ATFM delay. Despite the challenges found predicting the minutes of ATFM delay, the approach selected based on (De Falco & Delgado, 2021) to predict a probability distribution of values, rather than a single value, enriches the advice capabilities of the system clearly showing to the end user the uncertainty of the predictions.

Nowadays, some tools are available for the Flow Manager Positions (FMPs) to assess the impact of ATFM regulations before activating them (*e.g.* SIMulation and EXperiment (SIMEX) (EUROCONTROL, 2022a)). The models presented in this article enable airspace users to develop equivalent tools identifying possible downstream effects on their fleet. The integrated view proposed in this article could be used as a *what-if* function to evaluate new flight plans before submitting them. Moreover, machine learning tools can be seamlessly integrated with the current highly automated methodology, as the SHAP analysis proved a realistic behaviour.

The results obtained in this article using the forecast available at the prediction horizon D-1 have been compared with the ones reported using only *static* (*i.e.*, not using data which evolves) and post-operational models (Mas-Pujol et al., 2021). When comparing the performance against *static* models, the accuracy reported improves around 10% for the binary classifiers (*i.e.*, probability of regulated flight, the protected location, and whether the delay will be zero) and by 3 min for the delay of non-zero regulated flights while reducing the uncertainty by 7 min and increasing the PICP by 5% (actual ATFM delay is within the predicted probability distribution). Note, however, that *static* models can be used with a much longer look ahead prediction horizon as they do not depend on *dynamic* factors, which might evolve as the day of operations approaches. It is worth noticing how post-operational data does not improve all the models. This indicates that using data similar to the one available when the ATFM regulations were implemented (forecast available at D-1) is more suitable. When the regulations were issued, post-operational data was not available; therefore, using this information to predict the regulations is inefficient. Moreover, using data available at D-1 ensures all the models can be triggered as all the features can be computed at that time horizon.

As the delay used for the labelling is the final experienced ATFM delay by the flights, the impact of mitigation actions, *e.g.* reroutings, is already captured by the model. After applying any possible mitigation actions, the actual ATFM is the required delay to solve the demand-capacity imbalance. The models will, therefore, assist the duty managers (and dispatchers) in identifying the actual delays that the flights would experience.

The proposed method and system have some limitations that could be overcome in future work. The accuracy of downstream models is conditioned on the previous ones. Additional features, such as en-route weather information, could improve the accuracy of models. The PICP of the probability distribution of ATFM delay model could also be improved, as around 10% of the actual ATFM delays do not fall within the distribution. This might require fine-tuning the error range and resolution to discretise the regressor error. Furthermore, even though the novel selected method has potential, future research could benefit from comparing this approach with more conventional techniques to predict probability distributions. Also related to the selected models, it could be interesting to study whether models based on attention mechanisms could improve the overall performance, such as (Xiao, Xing, Qu, et al., 2024) where the authors present a densely knowledge-aware network for classification tasks.

Another future improvement could be to consider when the flights will enter the congested region. This could impact the model which estimates the amount of ATFM delay because this is the main source of information used by the CASA algorithm when issuing the actual ATFM delay. However, this requires a prior estimator to identify the congested traffic volumes, which is difficult with a large prediction horizon; however, Mas-Pujol, Salamí, and Pastor (2022) showed this might be possible.

As the models presented are trained on subsets of data which consider the materialisation of some conditions, *e.g.* the location of the regulation is subject to the flight being regulated, some difficulties might be experienced by the end-user who needs to consider the outcome of the models in this context. However, this keeps the human in the decision-making process and provides all the information that aligns with the training and practice of the tactical and duty managers.

Related to the capabilities of the presented models, they could be integrated into the current decision support tools used by airlines to estimate and minimise the impact of disrupted operations (*e.g.* reactionary delay including the probability of missing ATFM slots). When developing the system presented in this article, we ensured that the data used for the training and the execution of the models was available to the airline operating centre at D-1. Therefore, as information becomes available, the models can easily be re-trained over time to minimise the impact of any potential data drift. The models only use features available and stored by the airlines, such as the departing time or information about the origin and destination airports; information about expected network demand (or the input required to the PREDICT model) can be obtained from the NM. Therefore, airlines should be able to deploy, maintain and update the models within their facilities.

CRediT authorship contribution statement

Sergi Mas-Pujol: Conceptualization, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Luis Delgado:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing.

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 authors do not have permission to share data.

Acknowledgements

This work has been performed as part of Dispatcher3 innovation action, which has received funding from the H2020 – Clean Sky 2 Joint Undertaking (JU) under grant agreement No 886461. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the Clean Sky 2 JU members other than the Union. The opinions expressed herein reflect the authors' views only. Under no circumstances shall the Clean Sky 2 Joint Undertaking be responsible for any use that may be made of the information contained herein.

References

Andribet, P., Baumgartner, M., Garot, J.-M., & of the Air, M. (2022). Reinventing European air traffic control based on the covid-19 pandemic experience. *Utilities Policy*, *75*, Article 101343.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*: *vol. 4*, Springer.

Bolić, T., Castelli, L., Corolli, L., & Rigonat, D. (2017). Reducing ATFM delays through strategic flight planning. *Transportation Research Part E: Logistics and Transportation Review*, *98*, 42–59.

Cook, A., Delgado, L., Tanner, G., & Cristóbal, S. (2016). Measuring the cost of resilience. *Journal of Air Transport Management*, *56*, 38–47, Long-term and Innovative Research in ATM.

Cook, A. J., & Tanner, G. (2015). *European airline delay cost reference values, updated and extended values*: *Tech report*, London, UK: University of Westminster, v4.1.

Dalmau, R. (2022). Predicting the likelihood of airspace user rerouting to mitigate air traffic flow management delay. *Transportation Research Part C: Emerging Technologies*, *144*, Article 103869.

Dalmau, R., & Gawinowski, G. (2024). The effectiveness of supervised clustering for characterising flight diversions due to weather. *Expert Systems with Applications*, *237*, Article 121652.

Dalmau, R., Genestier, B., Anoraud, C., Choroba, P., & Smith, D. (2021). A machine learning approach to predict the evolution of air traffic flow management delay. In *14th USA/Europe air traffic management research and development*. New Orleans, LA, USA.

Dalmau, R., Zerrouki, L., Anouraud, C., Smith, D., & Cramet, B. (2021). Are all the requested air traffic flow management regulations actually indispensable? In *11th SESAR innovation days*. Virtual Event.

de Arruda Junior, A. C., Weigang, L., & Milea, V. (2015). A new airport collaborative decision making algorithm based on deferred acceptance in a two-sided market. *Expert Systems with Applications*, *42*(7), 3539–3550.

De Falco, P., & Delgado, L. (2021). Prediction of reactionary delay and cost using machine learning. In *Airline group of the international federation of operational research society*. Atlanta, GA, USA: AGIFORS.

De Falco, P., Kubat, J., Kuran, V., Rodriguez Varela, J., Plutino, S., & Leonardi, A. (2023). Probabilistic prediction of aircraft turnaround time and target off-block time. In *13th SESAR innovation days*. Seville, Spain.

Delgado, L., Gurtner, G., Cook, A., Martín, J., & Cristóbal, S. (2020). A multi-layer model for long-term KPI alignment forecasts for the air transportation system. *Journal of Air Transport Management*, *89*, Article 101905.

Delgado, L., Mas-Pujol, S., Skorobogatov, G., Argerich, C., & Gregori, E. (2022). Dispatcher3 – Machine learning for efficient flight planning – Approach and challenges for data-driven prototypes in air transport. In *Towards sustainable aviation summit*. Toulouse, France: Association Aéronautique et Astronautique de France.

Dispatcher3 Consortium (2020). *D1.1 - Technical resources and problem definition*: *Tech report*, Dispatcher3 Consortium, v4.1.

EUROCONTROL (2019). *Performance review report: An Assessment of air traffic management in europe during the calendar year 2018*: *Tech report*, EUROCONTROL, Performance Review Commission, Brussels, Belgium.

EUROCONTROL (2022a). *ATFCM operations manual*: *Tech report*, EUROCONTROL, Brussels, Belgium.

EUROCONTROL (2022b). *ATFCM user manual*: *Tech report*, EUROCONTROL, Network Management Directorate, Brussels, Belgium.

EUROCONTROL (2022c). *EUROCONTROL Forecast 2022–2024*: *Tech report*, EUROCONTROL, Brussels, Belgium.

EUROCONTROL (2023). *EUROCONTROL data snapshot #38 on the path to recovery for intra-European and intercontinental flights*. EUROCONTROL, Brussels, Belgium.

García-Heredia, D., Molina, E., Laguna, M., & Alonso-Ayuso, A. (2021). A solution method for the shared resource-constrained multi-shortest path problem. *Expert Systems with Applications*, *182*, Article 115193.

Garrigó, L., Alsina, N., Adrienko, N., Andrienko, G., Piovano, L., & Blondiau, T. (2016). Visual analytics and machine learning for air traffic management performance modelling. In *6th SESAR innovation days*. Delf, Netherlands.

Gopalakrishnan, K., & Balakrishnan, H. (2017). A comparative analysis of models for predicting delays in air traffic networks. In *12th USA/Europe air traffic management research and development seminar*. Seattle, Washington, USA.

Gudmundsson, S. V., Cattaneo, M., & Redondi, R. (2021). Forecasting temporal world recovery in air transport markets in the presence of large economic shocks: The case of COVID-19. *Journal of Air Transport Management*, *91*, Article 102007.

Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z., & Zhao, D. (2019). Flight delay prediction based on aviation big data and machine learning. *IEEE Transactions on Vehicular Technology*, *69*(1), 140–150.

Gurtner, G., Delgado, L., & Valput, D. (2021). An agent-based model for air transportation to capture network effects in assessing delay management mechanisms. *Transportation Research Part C: Emerging Technologies*, *133*, 103–358.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction*: *vol. 2*, Springer.

Jardines, A., Eivazi, H., Zea, E., García-Heras, J., Simarro, J., Otero, E., et al. (2024). Thunderstorm prediction during pre-tactical air-traffic-flow management using convolutional neural networks. *Expert Systems with Applications*, *241*, 122–466.

Jardines, A., Soler, M., & García-Heras, J. (2021). Estimating entry counts and ATFM regulations during adverse weather conditions using machine learning. *Journal of Air Transport Management*, *95*, Article 102109.

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2017). *Data analysis: A model comparison approach to regression, ANOVA, and beyond*. Routledge.

Khosravi, A., Nahavandi, S., & Creighton, D. (2010). A prediction interval-based approach to determine optimal structures of neural network metamodels. *Expert Systems with Applications*, *37*(3), 2377–2387.

Kistan, T., Gardi, A., Sabatini, R., Ramasamy, S., & Batuwangala, E. (2017). An evolutionary outlook of air traffic flow management techniques. *Progress in Aerospace Sciences*, *88*, 15–42.

Lambelho, M., Mitici, M., Pickup, S., & Marsden, A. (2020). Assessing strategic flight schedules at an airport using machine learning-based flight delay and cancellation predictions. *Journal of Air Transport Management*, *82*, Article 101737.

Liaw, A., Wiener, M., et al. (2002). Classification and regression by randomforest. *R News*, *2*(3), 18–22.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems*: *vol. 30*.

Luo, X., Wu, H., Wang, Z., Wang, J., & Meng, D. (2021). A novel approach to large-scale dynamically weighted directed network representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(12), 9756–9773.

Mas-Pujol, S., De Falco, P., & Delgado, L. (2022). Pre-tactical advice using machine learning for air traffic flow management delay estimation. In *Airline group of the international federation of operational research society*. Virtual: AGIFORS.

Mas-Pujol, S., De Falco, P., Salamí, E., & Delgado, L. (2022). Pre-tactical prediction of ATFM delay for individual flights. In *41st digital avionics systems conference*. Portsmouth, Virginia, USA: IEEE.

Mas-Pujol, S., Salamí, E., & Pastor, E. (2021). Predict ATFCM weather regulations using a time-distributed recurrent neural network. In *40th digital avionics systems conference*. San Antonio, Texas, USA: IEEE.

Mas-Pujol, S., Salamí, E., & Pastor, E. (2022). RNN-CNN hybrid model to predict C-ATC CAPACITY Regulations for en-route traffic. *Aerospace*, *9*(2), 93.

National Oceanic and Atmospherics Administration (NOAA) (2024). Global forecast system (GFS). (Accessed 7 May 2024).

Paleyes, A., Urma, R.-G., & Lawrence, N. D. (2022). Challenges in deploying machine learning: A survey of case studies. *ACM Computing Surveys*, *55*(6), 1–29.

Pamplona, D. A., & Alves, C. J. P. (2019). Testing the air delay variability. *Australian Journal of Basic and Applied Sciences*, *13*(10), 16–24.

Rebollo, J. J., & Balakrishnan, H. (2014). Characterization and prediction of air traffic delays. *Transportation Research Part C: Emerging Technologies*, *44*, 231–241.

Sanaei, R., Pinto, B. A., & Gollnick, V. (2021). Toward atm resiliency: A deep cnn to predict number of delayed flights and atfm delay. *Aerospace*, *8*(2), 28.

Schultz, M., Reitmann, S., & Alam, S. (2021). Predictive classification and understanding of weather impact on airport performance through machine learning. *Transportation Research Part C: Emerging Technologies*, *131*, Article 103119.

SESAR (2020). *European ATM master plan: Digitalizing Europe's aviation airspace*: *Tech report*, Brussels, Belgium: SESAR Joint Undertaking.

Sridhar, B., Chatterji, G. B., & Evans, A. D. (2020). Lessons learned in the application of machine learning techniques to air traffic management. In *AIAA AVIATIoN 2020 FORUM*. London, United Kingdom: AIAA.

Vossen, T. W., Hoffman, R., & Mukherjee, A. (2012). Air traffic flow management. *Quantitative Problem Solving Methods in the Airline Industry: a Modeling Methodology Handbook*, 385–453.

Wang, Z., Liao, C., Hang, X., Li, L., Delahaye, D., & Hansen, M. (2022). Distribution prediction of strategic flight delays via machine learning methods. *Sustainability*, *14*(22), 15180.

Xiao, Z., Tong, H., Qu, R., Xing, H., Luo, S., Zhu, Z., et al. (2023). CapMatch: Semi-supervised contrastive transformer capsule with feature-based knowledge distillation for human activity recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15.

Xiao, Z., Xing, H., Qu, R., Feng, L., Luo, S., Dai, P., et al. (2024). Densely knowledge-aware network for multivariate time series classification. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.

Xiao, Z., Xing, H., Zhao, B., Qu, R., Luo, S., Dai, P., et al. (2024). Deep contrastive representation learning with self-distillation. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *8*(1), 3–15.

Xie, Y., Pongsakornsathien, N., Gardi, A., & Sabatini, R. (2021). Explanation of machine-learning solutions in air-traffic management. *Aerospace*, *8*(8), 224.

Yousefzadeh Aghdam, M., Kamel Tabbakh, S. R., Mahdavi Chabok, S. J., & Kheyrabadi, M. (2021). Optimization of air traffic management efficiency based on deep learning enriched by the long short-term memory (LSTM) and extreme learning machine (ELM). *Journal of Big Data*, *8*(1), 1–26.