Journal of Air Transport Management 95 (2021) 102109

Contents lists available at ScienceDirect

Journal of Air Transport Management

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

Estimating entry counts and ATFM regulations during adverse weather conditions using machine learning

Aniel Jardines *, Manuel Soler, Javier García-Heras

Department of Bioengineering and Aerospace Engineering, Universidad Carlos III de Madrid, Leganés, Madrid, Spain

ARTICLE INFO

Keywords:
Thunderstorms
Air traffic flow management
Data science
Machine learning
ATM data

ABSTRACT

In recent years, convective weather has been the cause of significant delays in the European airspace. With climate experts anticipating the frequency and intensity of convective weather to increase in the future, it is necessary to find solutions that mitigate the impact of convective weather events on the airspace system. Analysis of historical air traffic and weather data will provide valuable insight on how to deal with disruptive convective events in the future. We propose a methodology for processing and integrating historic traffic and weather data to enable the use of machine learning algorithms to predict network performance during adverse weather. In this paper we develop regression and classification supervised learning algorithms to predict airspace performance characteristics such as entry count, number of flights impacted by weather regulations, and if a weather regulation is active. Examples using data from the Maastricht Upper Area Control Centre are presented with varying levels of predictive performance by the machine learning algorithms. Data sources include Demand Data Repository from EUROCONTROL and the Rapid Developing Thunderstorm product from EUMETSAT.

1. Introduction

Convection is a well known aviation hazard; turbulence, wind shear, lighting and hail are elements within thunderstorms that can be catastrophic for aircraft. Aviation research related to thunderstorms and convection has typically focused on flight specific solutions, e.g.: the development of trajectory optimisation algorithms in the presence of thunderstorms (Sauer et al., 2014; Zhang et al., 2017; Hentzen et al., 2018; González-Arribas et al., 2019; Seenivasan et al., 2020); or the development of decision support tools to analyse active flights in the en route environment and find simple and efficient route corrections around convective weather (McNally et al., 2015). In the United States, the Federal Aviation Administration regularly publishes a National Severe Weather Playbook. The National Playbook is a collection of Severe Weather Avoidance Plan (SWAP) routes that have been prevalidated and coordinated with the impacted air traffic control centres. However, in Europe, where many ANSPs each control a portion of the airspace, a network wide perspective focused on reducing the impact of adverse weather on the Air Traffic Flow Management (ATFM) process is lacking.

Proper execution ATFM requires well organised collaboration between stakeholders, including aircraft operators, Air Navigation Service Providers (ANSPs), Meteorological (MET) service providers and the Network Manager (NM) (EUROCONTROL, 2019a). On days with strong thunderstorm activity, the airspace system conditions can be highly volatile making it difficult to balance aircraft operator demand with airspace capacity. Thunderstorms can move quickly and exhibit lifecycles that can develop and dissipate within 1–2 h, making them difficult to anticipate over longer time horizons. As a result, convective weather has been typically managed during the execution of a flight (at a more tactical level), having significant impact on the efficiency of the Air Traffic Management (ATM) system. In 2018, 4.8 million minutes of en route ATFM delay were due to adverse weather in the European airspace, a 124 percent increase vs 2017 (EUROCONTROL, 2019b). In the top 10 days of convective activity over Europe in 2018, more than 1 million minutes of en route delay were accumulated due to adverse weather, with the cost of ATFM delay estimated at €100 per minute (Cook and Tanner, 2015), weather has a significant financial impact on the ATM system.

Performing ATFM operations in a convective weather environment is particularly complex due to the dynamic nature of thunderstorms and their effect on air traffic demand and airspace capacity. At different time horizons, the weather information available widely varies. For shorter time horizons, the weather information relies heavily on observations and extrapolation of radar and satellite images; data can

* Correspondence to: Department of Bioengineering and Aerospace Engineering, Universidad Carlos III de Madrid, Avenida de la Universidad 30, Leganés, Madrid 28911, Spain.

E-mail addresses: ajardine@ing.uc3m.es (A. Jardines), masolera@ing.uc3m.es (M. Soler), gcarrete@ing.uc3m.es (J. García-Heras).

https://doi.org/10.1016/j.jairtraman.2021.102109
Received 2 September 2020; Received in revised form 8 March 2021; Accepted 25 June 2021
Available online 7 July 2021

be provided every 15 min proving to be fairly accurate but limited by the forecast time horizon, typically less than one hour. For longer time horizons (3 to 48 h ahead), where ATFM processes are put in place, weather information relies on numerical weather prediction (NWP) products with varying accuracy. However, the large computational effort required to run NWP tools results in limitations in spatiotemporal granularity and the refresh rate of the forecast, typically around 12 h (6 h in Limited Area Models), all in all resulting in forecasts that are not capable of accurately capturing convective phenomena. Detecting capacity-demand imbalances in the airspace network due to weather should be done on a continuous basis over varying time horizons and, when possible, up to 2 days ahead of operation (covering, thus, both the pre-tactical and tactical ATFM phases).

The motivation behind this research is to obtain better understanding of the ATFM process around convective weather events. While it is well understood that convective weather is disruptive to ATFM operations, being able to characterise the network imbalances due to weather can be quite complex. We strongly believe that in order to improve the ATFM process during convective weather, it is first necessary to quantify key aspects such as the changes in demand, the changes in capacity, and the effectiveness of regulations. The research presented in this paper is a first attempt at applying data-driven methods to better understand the problem. To do so, we will rely on machine learning algorithms. By collecting historical data related to traffic demand, ATFM regulations, traffic trajectories, weather forecast, and storm observations we will build machine learning algorithms to better predict the demand and the regulations due to weather in a given sector.

Applying machine learning and machine learning techniques to ATM is an extremely active area of research, and has proved to useful in applications such as, just to mention a few: predicting controller workload (Gianazza, 2017), trajectory prediction (Marcos et al., 2017; Fernández et al., 2017; Gallego et al., 2019), identifying anomalous events leading to unsafe arrivals (Jarry et al., 2020), predicting arrival times (Wang et al., 2018), and the characterisation of trajectory adherence to standard routes in complex airspace volumes (Carmona et al., 2020). Similarly, the use of machine learning to enhance weather prediction is also an active area of research. Convolution LSTM networks and support vector machines have proven successful for nowcasting applications (Xingjian et al., 2015; Han et al., 2017), as well as forecasting techniques that combine NWP and observations as input (Mecikalski et al., 2016). Applying machine learning techniques on numerical weather models has proved to enhance thunderstorm prediction beyond the nowcasting range up to 15 h in advance (Kamangir et al., 2020). The use of ensemble forecasting techniques has also made it possible to provide a probabilistic thunderstorm forecast up to multiple days in advance (Bouttier and Marchal, 2020).

Despite the vast research activity on machine learning applications to ATM or convective weather (and, of course, other disciplines) in the last years, the tackling of problems combining both fields exhibits a significant gap. In particular, and to the best of authors' knowledge, such combined application (ATM and convective weather) of machine learning techniques are non-existing today in the literature.

The main contribution of this paper is the integration of data sources capturing traffic and weather conditions, as an initial building block to improve air traffic flow management processes during convective weather events. We analyse the historical data (aerial traffic, ATFM regulations, and thunderstorm activity), quantify the impact and associated performance, and try to understand and discuss the ATFM operational complexities. We build and compare various supervised regression and classification learning algorithms, namely logistic regression, decision tree, random forest, and a neural network, to find functional relationships that can predict three features that are relevant to ATFM processes, i.e., sector entry count, weather regulations in a particular sector, and regulated traffic entry counts. Results are presented for the Maastrict Upper Area Control Centre airspace using data covering a period of 68 days (some of them with significant storm activity) in May–June–July 2018.

The paper is structured as follows: The data sources used, the data processing methods, and the final data set we have considered are described in , Section 3, and Section 4. The machine learning algorithms are presented in Section 5. The results obtained in the MUAC use case are presented and discussed in Section 6. Finally, some conclusion are drawn in Section 7.

2. Data sources

In building the machine learning algorithms we utilised a dataset composed of historical weather observations and historical traffic information. In using actual weather observations rather than forecasts to build our models, we are assuming we have access to "perfect weather forecast". It is acknowledged that this is an unrealistic assumption that is not compatible in an operational context. However, we believe that any relationships that exist between weather and traffic will be more easily captured by using actual weather observations.

2.1. Weather observations

Geostationary satellites with orbital periods that match the earths rotation provide continuous observation of specific regions of the Earth. Visual and infrared satellite imagery captures vital information regarding cloud cover, water vapour and temperature that allow for monitoring and tracking of weather.

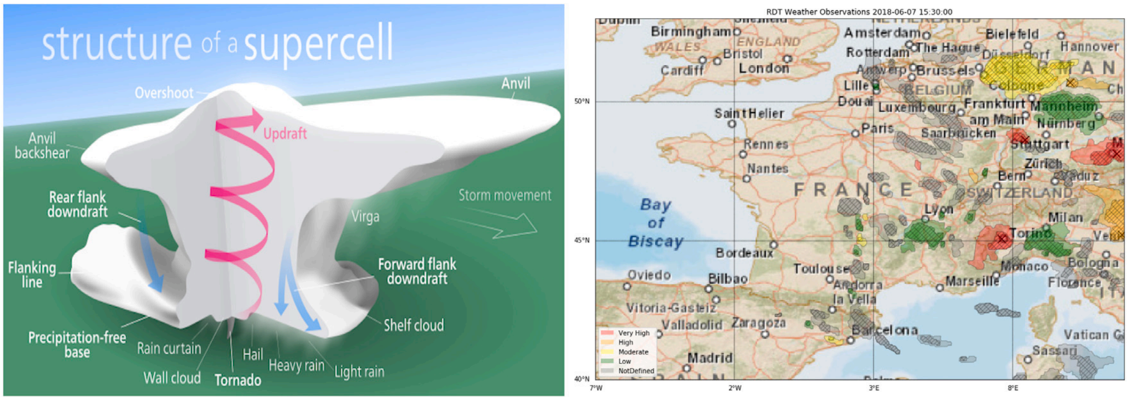
The Rapid-Development Thunderstorm (RDT) algorithm was developed by Meteo-France within the EUMETSAT NWC-SAF framework. The RDT algorithm employs primarily geostationary satellite data to provide information about clouds related to significant convective systems from the mesoscale (200–2000 km) down to hundreds of metres (Lee et al., 2020).

The RDT product covers the geographical region of Europe, and outputs storm data on a 15 min interval. For each cloud cell, the RDT product defines a series of parameters capturing the location, shape, movement, severity, and life cycle phase.

In our research we focus on parameters defining the altitude of the cloud top, the contour coordinates of the top cloud and shelf cloud, the severity of the storm, and the location of the overshoot if present. Fig. 1 shows a thunderstorm schematic identifying the various storm features and a sample of the RDT product.

2.2. ATM data

The Demand Data Repository (DDR) from EUROCONTROL was used to analyse historical demand and airspace environment. DDR provides the most accurate picture of past pan-European air traffic demand. The ALLFT files from DDR include an abundance of information for each flight by modelling of the filed, regulated and current flight plans and trajectories. For our research we use data from the Current Tactical Flight Model (CTFM), the CTFM uses a combination of the last filed flight plan and available radar data to compute the closest estimate of the flight trajectories handled by controllers on the date of operations. More specifically, we make use of the CTFM airspace profile, this profile provides information on the ATC unit airspaces and elementary sectors the aircraft encountered throughout the flight; the data provides the time, location and altitude of the entry and exit points for a flight as it traverses the various elements that make up the European airspace. Fig. 2 shows a graphical representation of the data extracted from DDR. Additionally, we also utilise DDR data relating to the regulations that occurred. From the ALLFT files we are able to know if a flight was regulated and which regulation was most penalising. From the environment files we are able to see regulation specific data, including the location, time and reason for the regulation.



(a) Convective cell schematic. (b) Rapidly Developing Thunderstorm (RDT) product visualisation.

Fig. 1. Convective cell schematic 1(a) and RDT visualisation 1(b). Fig. 1(b) shows data from June 7th, 2018 at 15:30. Hatch pattern contours indicate shelf clouds, colours represent storm severity, while an "X" indicates the location of overshoots.
Source: https://commons.wikimedia.org/wiki/File:Supercell.svg.

(a) Filed and Current trajectory for flight from LFBO to EHAM. (b) Map of Flight Information Regions in western Europe .

Fig. 2. Flight trajectory data 2(a) and airspace data 2(b) extracted from EUROCONTROL's Demand Data Repository.

3. Data processing

A challenge in performing any type of met-ATM analysis is integrating the various types of weather and traffic data. In this particular instance, the weather data is provided in the form of polygons defined by the RDT product. As mentioned in the previous section, each RDT polygon contains information regarding characteristics of the storm. In the case of the DDR traffic data, there is data specific to flights and data specific to airspace sectors. In order to build a machine learning algorithm it is necessary to integrate the various data types onto a common domain.

Our idea is to integrate the various data sources onto a 4 dimensional grid. One can imagine splitting the airspace into discrete volumes of airspace defined by longitude, latitude, altitude and time window (see Fig. 3).

Storm polygons and cloud top altitude provided by the RDT product can be projected on the 4D grid, enabling us to identify grid elements which contain a storm, along with the associated storm characteristics such as severity or overshoots. Airspace geometry data from the DDR environment files can also be projected onto the grid to define sectors as a set of grid elements. 4D flight trajectories can also be represented as the set of grid elements intersect.

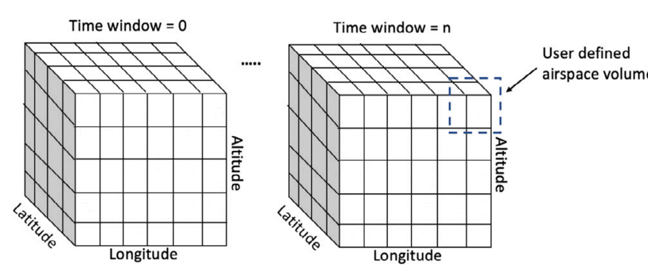


Fig. 3. User defined grid will determine the scale at which traffic and weather features are extracted.

In processing each data type it is important to consider the granular scale of the grid, as the choice in spatial temporal granularity will impact the data and what the machine learning algorithms are capable of learning. For our analysis we used a spatial definition of 0.1 × 0.1 degrees in latitudinal–longitudinal, and a non-uniform altitude distribution that would give us more granularity in the upper airspace. Our grid ended up with 6 altitude levels, these are defined in Table 1. The table indicates the altitude ranges in pressure level, used in processing the

Table 1
Altitude level definition.

Altitude Level	Pressure Range (hPa)	Altitude Range FL
5	≤200	≥390
4	200–250	390–340
3	250–300	340–300
2	300–400	300–240
1	400–700	240–100
0	700-Surface	100 - Surface

storm data, and flight level, used in processing the traffic and sector information. In processing the data, we kept the spatial granularity fixed, however we experimented with different granularities in the temporal scale. Given the RDT product provides data in 15 min intervals, this was a logical time window to use. Additionally, we also processed the data using a 1 h time window.

Figs. 4 and 5 show examples of the RDT and DDR data projected onto the 4D grid.

4. Model architecture

Our dataset consisted of 68 days of historical traffic and weather data from May 4th through July 10th 2018. Although weather and traffic data was available for the entire pan-European airspace, we narrowed our focus of our models to the Maastricht Upper Area Control Centre (MUAC). Using the 4D grid we constructed time series representations of the features that will become inputs to the models as well as the outputs. The weather and traffic information in grid form covering the 68 day period for the entire pan-European airspace consisted in a data volume of roughly 35 GB, however a time-series representation of only the MUAC airspace scaled down to 200 MB.

4.1. Model features

The weather input to our models were based on the storm characteristics from the RDT product. Using the 4D grid we were able to define the storm characteristics that occurred with MUAC. Considering that MUAC extends vertically from FL 245 to FL660, we estimated the volume by only considering the air-blocks in altitude levels 2–5, as described in Table 1. Based on the 4D grid, 14,596 airblocks were used to represent the MUAC airspace volume, the weather features were formulated as the fraction of airblocks where the feature was present. The complete list of weather features used for our models is provided below:

- Overshoot - Number of overshoots in MUAC
- Storm Cell - Fraction of MUAC airblocks containing a storm cell
- Shelf Cloud - Fraction of MUAC airblocks containing the shelf cloud of storm cell
- Low - Fraction of MUAC airblocks with low storm severity
- Moderate - Fraction of MUAC airblocks with moderate storm severity
- High - Fraction of MUAC airblocks with high storm severity
- Very High - Fraction of MUAC airblocks with very high storm severity
- Not Defined - Fraction of MUAC airblocks with storm severity "not defined"

Besides the eight weather features described above, we also included features related to the day of the week and hour of the day. The time related input features were encoded using a binary categorical representation. For example, the day of the week Tuesday is represented as [0, 1, 0, 0, 0, 0, 0]. Similarly, hour of the day was encoded using a vector with 24 elements. Given the highly cyclic behaviour of air traffic, these time related features would allow our models to capture the weekly and daily patterns in the data, and in a sense, the demand on the sector.

Given that 68 days of data were available, using a 15 min time window interval resulted in a dataset containing 6528 samples. A second dataset was also formulated using a one hour time window temporal scale, this 1 h dataset contained 1632 samples. In the case of the 1 h dataset, the values in model features related to weather were averaged hourly. For example, in the case of the "Storm Cell" feature, if the 15 min values for fraction of MUAC volume containing a storm cell were $[.40, .30, .10, 0]$, the hourly value would equal $.20$, an average of the four values.

4.2. Learning tasks

Using the supervised learning technique implies that we have data representative of the model output that can be used to train the model. Traffic data from the DDR ALLFT and environment files were used to formulate the model output data. In this paper we propose three separate supervised learning tasks; two regression and one classification.

4.2.1. Entry count

The first learning task is to use the weather features to predict the entry count for the MUAC airspace. Entry count is defined as the number of flights entering in a sector during a selected time period. Using the DDR airspace profiles, a list of the flights entering in the MUAC was obtained. We would expect that during convective weather events the entry count of the sector would be reduced. This learning task can be categorised as a supervised learning regression problem since we want the algorithm to provide us with numerical estimate of the entry count.

4.2.2. Regulated entry count

The second learning task is to predict the regulated entry count for the MUAC airspace within the time window. Because a flight can be regulated for various reasons and various locations throughout the European network, we limit the definition of "regulated entry count" to aircraft whose most penalising regulation was declared by MUAC and for weather reasons. From the DDR ALLFT files, it is possible to see the most penalising regulation for each flight entering MUAC. We would expect that convective weather events in the MUAC would be correlated with the number of aircraft that are penalised by weather regulations. One could also think of this value of "regulated entry count" as a proxy for the rate applied to the weather regulation. Given that we want the algorithm to provide us with a numerical value, this learning task can be also categorised as a supervised learning regression task.

4.2.3. Active weather regulation

Our last learning task is to predict if a weather regulation will be active in the MUAC airspace within the time window given the weather conditions. From the DDR environment files we were able to obtain a list of all the regulations that were active in the MUAC airspace. MUAC is a large and complex airspace that can simultaneously activate various regulations to mitigate its various traffic flows. In an attempt to simplify the problem, we treat this task as a binary and try to predict only if a weather regulation is active. Given our algorithm has to chose from one of two options or "classes", this learning task can be categorised as a supervised learning binary classification problem.

To get a better understanding of the data, Fig. 6 provides a visual representation of the model features and learning tasks for four consecutive Tuesdays with varying weather conditions in the MUAC. In each graph, the colourful shaded areas along the bottom represent the severity and percentage of airspace volume containing storms as indicated by the left vertical axis. The black line and dashed red line indicate the MUAC entry count and regulated entry count respectively. Entry count values can be read using the right vertical axis. The light

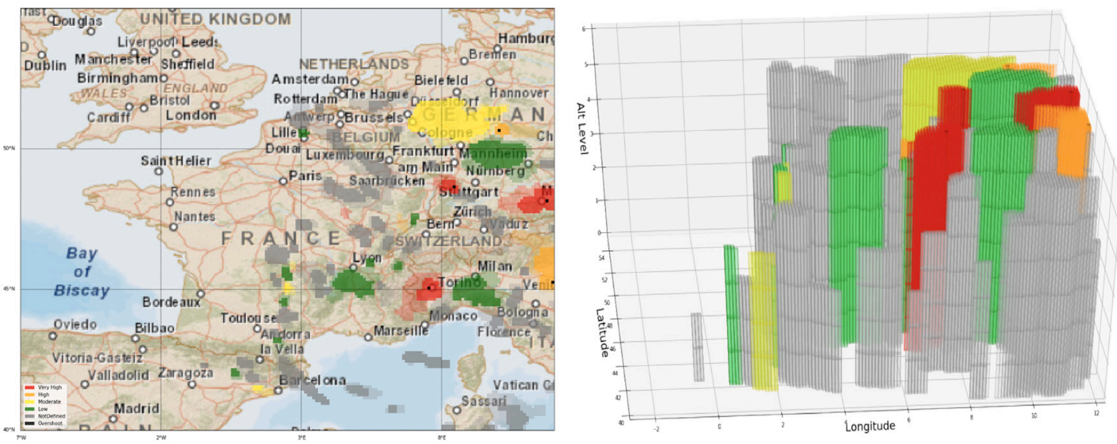


Fig. 4. Storm data from RDT product projected onto 4D grid.

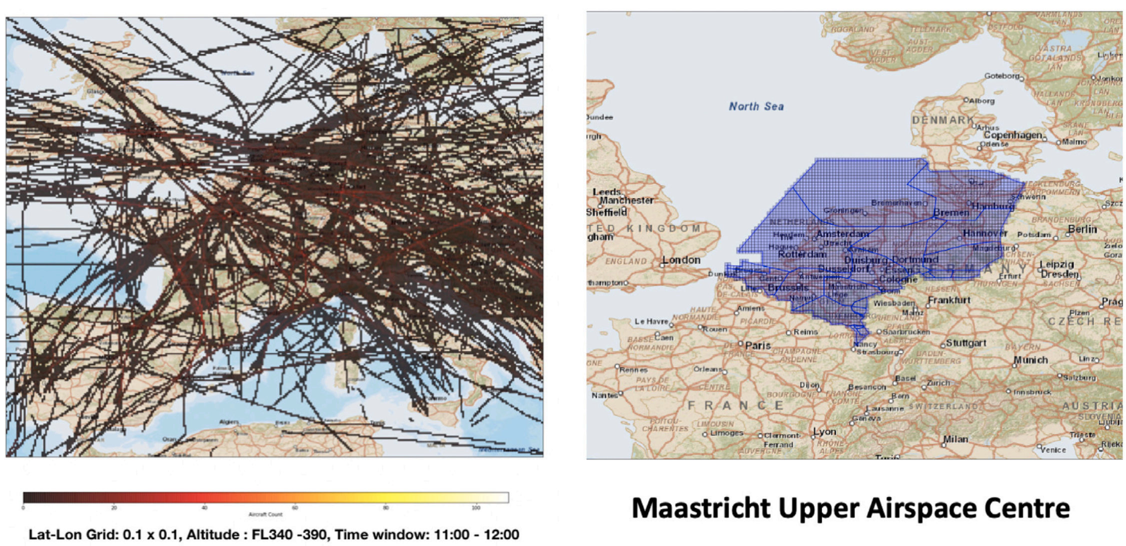


Fig. 5. 4D flight trajectory and sector geometry data from DDR projected onto 4D grid.

blue vertical shaded regions on the graphs, indicate that a weather regulation is active within the MUAC. From the figures for May 22nd and May 29th, we can observe that both days experienced afternoon storms, causing weather regulations to become active, and an increase in regulated traffic. The graph for May 29th shows more severe weather covering a larger portion of the airspace, the graph shows a reduced entry count into the airspace, with the majority of traffic being regulated. The image for June 5th, shows a weather regulation active during the afternoon resulting in an increase in the regulated entry count. While not visible on the graph, the data shows a very small percentage of airspace volume contained weather. We can assume a storm on the edge of the airspace caused the regulation. The image corresponding to June 12 shows a day without any storms or weather regulations.

Fig. 7 shows a graphical representation of the entire data set. The weather related features are shown in blue, while the dependent variables representative of the learning tasks are shown in green. From the figure we can notice several peaks for the various weather features, these peaks correspond with storm activity in the MUAC airspace. The figure represents the data using a 15 min time window. Fig. 8 provides a diagram showing the flow of input data, machine learning models, and outputs.

5. Machine learning algorithms

The first step in developing the algorithms was to split the data into training and testing subsets, an 80/20 training to testing split ratio was chosen. Given the data set consisted of a time series covering the 68 day period from May 4th, 2018 to June 10th, 2018, a sequential split of the data would have been preferred; using the first portion of the time series to train the models and the remaining portion for testing. However, because the majority of the convective activity in our dataset occurs in the two week period between May 28th and June 11th (depicted in Fig. 7), splitting up the data sequentially would not have provided us with a similar number of convective events between training and testing. Due to having a limited amount of data, we chose to use a random split of our data, implying that every time interval observation is treated as independent. It is acknowledged that treating each time interval as independent is an unrealistic assumption, however we believed the models would still capture the trends due to weather. Given the two types of supervised learning problems, a set of regression algorithms was chosen to predict the Entry Count and the Regulated Entry Count in MUAC, and a different set of algorithms was chosen for the classification task of predicting if a weather regulation was active.

In our analysis we experimented with four types of machine learning techniques suitable for regression and classification; Multiple Linear/

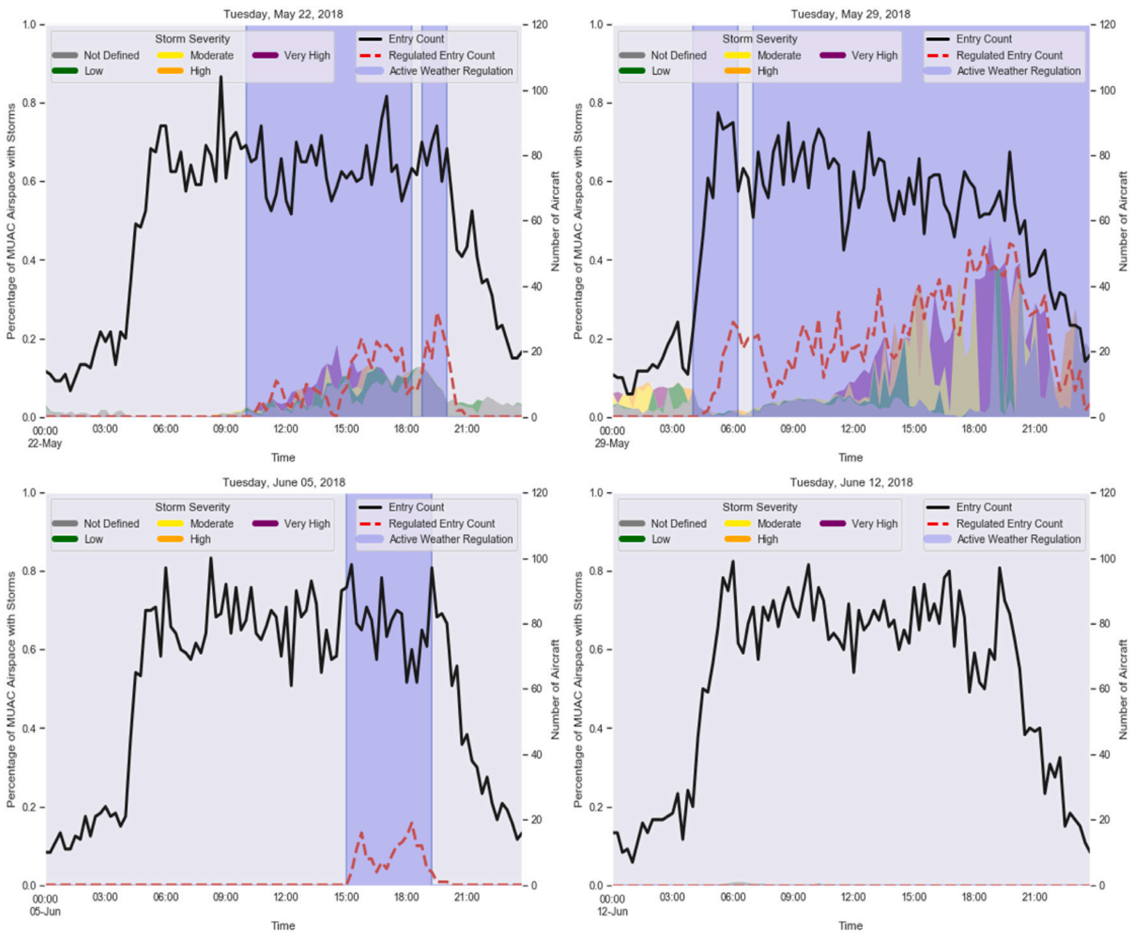


Fig. 6. Traffic and Weather situation in the MUAC airspace for four consecutive Tuesdays. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Logistic Regression, Decision Trees, Random Forests, and Artificial Neural Networks.

(a) Multiple Linear Regression: is used to determine the mathematical relationship between two or more explanatory variables and one dependent variable. It is an extension of ordinary least-squares regression but with more than one explanatory variable. The multiple linear regression model for dependent variable y and k explanatory variables $x_1, x_2, ..x_k$ can be expressed as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon,$$

for $i = 1,2, \ldots n$, given n observations where β_0 is a constant, $\beta_1, \beta_2, ..\beta_k$ are slope coefficients and ε is the model deviation or error term. For the classification case, logistic regression uses a similar formulation and enables the modelling binary outcomes. In logistic regression, we can find a linear relationship between the explanatory variables and the log-odds of an event, defining the probability of success as p and the probability of failure as $(1 - p)$. The model can be formulated using the equation below:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}.$$

(b) Decision Trees: are a popular supervised learning technique that can be used for both regression and classification problems. A decision tree predicts the target function based on several independent variables. The decision tree works by splitting the training dataset into subsets that contain similar values of the dependent variable. The decision tree is created starting with a single root node, at each node the data is split into two or more child nodes based on the data feature values. The data splits are performed with the goal of maximising the information gain using the concept of entropy from information theory. Where the entropy (H) and the information gain (IG) after splitting set S on attribute A are defined as

$$H(S) = \sum_{c \epsilon C} -p(c)log_2 p(c);$$

$$IG(A, S) = H(S) - \sum_{t \epsilon T} p(t)H(t).$$

In the entropy equation, S is the data for which entropy is being calculated, C represents the set of classes in S, and p(c) is the proportion of the number of elements in class c to the number of elements in set S. In the information gain equation T represents the subsets created from splitting set S and $p(t)$ is the proportion of the number of elements in subset t to the number of elements in set S.

(c) Random Forest: is an ensemble machine learning technique also suitable for regression and classification problems based on the Decision Tree theory. In a Random Forest, multiple decision trees are created using a random subsamples of the entire dataset with replacement. The final algorithm prediction comes from the average prediction from the multiple trees, or "forest".

(d) Neural Networks (NN): are machine learning techniques that attempt to mimic the structure of the human brain with an interconnected network of neurons or nodes. The nodes in the network are organised using multiple layers. NN architecture will have one input layer consisting of the model features, one or more hidden layers, and an output layer. The nodes in the model are connected with edges, each with an associated

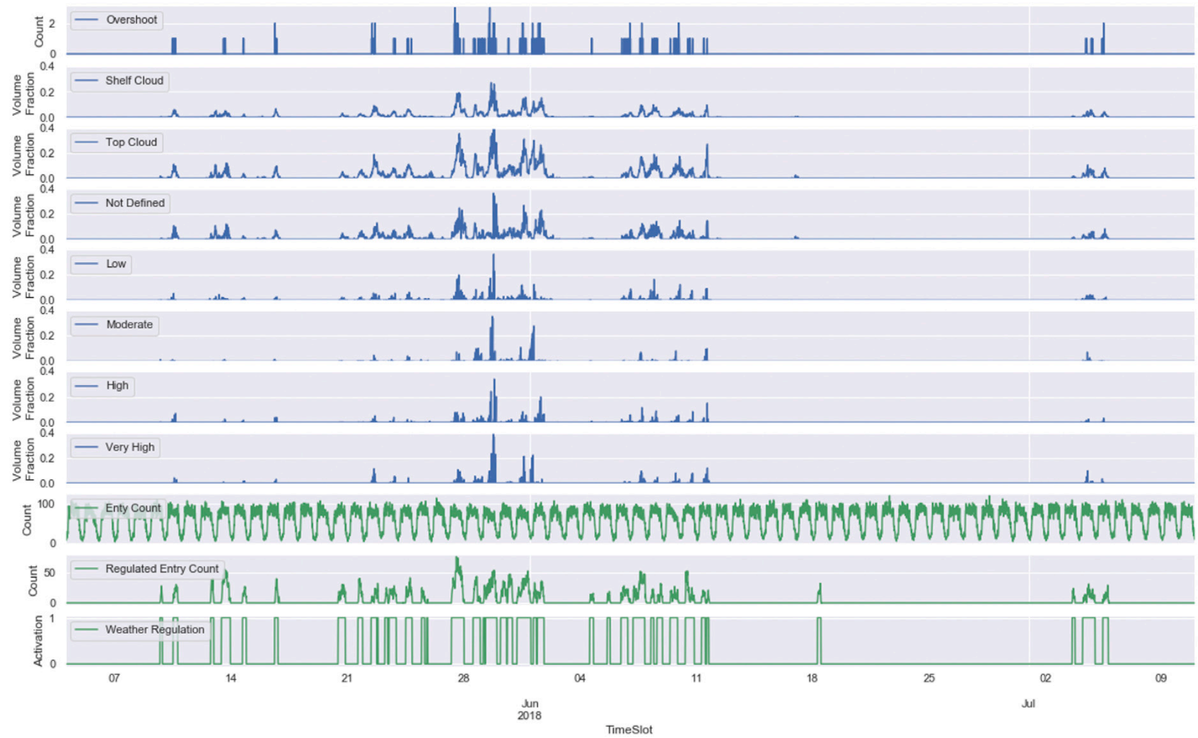


Fig. 7. Graphical representation of entire data using 15 min time window. Weather features are shown in blue and dependent variables representative of learning tasks in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

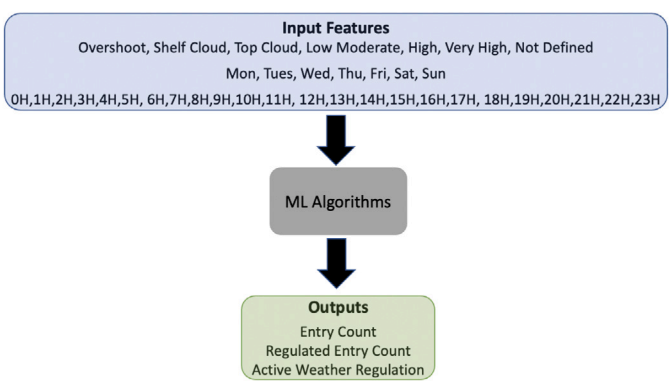


Fig. 8. Machine Learning Process.

weight representative of strength of the connection. The output of each node can be defined using the following equation:

$$Output = f(w_1 \cdot X_1 + w_2 \cdot X_2 + \cdots + w_k \cdot X_k + b),$$

where w represents the weight associated with the connection, and X is the input value, k is the number of inputs and b is a biased term. The function f is a non-linear activation function based on the sum of its inputs. In our models we use the rectified linear unit activation function:

$$f(x) = max(0, x).$$

During the training phase of the NN model, the connection weights are adjusted to minimise a loss function between the final output and the target function.

All of the models used in the analysis were developed using existing data science libraries in python. Below we provide a brief description on the architecture of each algorithm.

5.1. *Supervised regression learning*

Four algorithms were chosen to perform the learning task of predicting the Entry Count and Regulated Entry Count into the MUAC. The same algorithm architectures were used to fit the 1 h and 15 min time interval datasets. The input data for each learning task was the same and consisted of weather features and time features relating to day of the week and hour of the day. The time related data was treated as a categorical features, resulting in 7 features for day of the week and 24 features for hour of the day. The final input to our models consisted of 39 features, 8 weather features and 31 time features. In the case of the 15 min time interval data set, an extra feature was added to account for the minutes.

5.1.1. *Multiple linear regression*

The python library *statsmodels* was used to create a multiple linear regression model using ordinary least squares.

5.1.2. *Decision tree regressor*

The *DecisionTreeRegressor* function from python library *Scikit-learn* was used to create the decision tree regression model. The max depth parameter, which defines the maximum path length from the tree root to a leaf, was set to 20.

5.1.3. *Random forest regressor*

The *RandomForestRegressor* function from python library *Scikit-learn* was used to create random forest model. Similar to the decision tree model the max depth was set to 20. The random forest model uses an ensemble of decision trees to make its prediction. The number of estimator parameter, which defines the number of trees in the ensemble was set to the default value of 100.

5.1.4. Neural network regressor

The neural network model was built using the python library *keras*. The model contained one input layer, four hidden layers with sizes 30, 15, 9, 3 and one output layer. All nodes in the model used a rectified linear activation function. The model was trained with the *Adam* (Adaptive Moment Estimation) optimiser and a mean absolute error loss function. Prior to training the input data was scaled using a *StandardScaler* function. During training 20 percent of the training data was used to validate.

5.2. Supervised classification learning

The following four algorithms were chosen to perform the learning task of predicting if a weather regulation was active in the MUAC airspace within a time window given the weather conditions.

5.2.1. Logistic regression

The *LogisticRegression* function from the python library *Scikit-learn* was used to build the classification model.

5.2.2. Decision tree classifier

The *DecisionTreeClassifier* function from python library Scikit-learn was used to create the decision tree classification model, similar to the regression case, a max tree depth of 15 was selected.

5.2.3. Random forest classifier

The *RandomForestRegressor* function from python library *Scikit-learn* was used to create the random forest classification model.

5.2.4. Neural network classifier

The *MLPClassifier* function from the *Scikit-learn* was used to build the neural network model. The model contained four hidden layer of sizes 30, 15, 9, 5 and one output layer. All nodes in the model used a rectified linear activation function. The model was trained with the *LBFGS* (Limited-Memory Broyden–Fletcher–Goldfarb–Shanno) optimiser using a log-loss function. Prior to training the input data was scaled using a *StandardScaler* function. During training 20 percent of the training data was used for validation.

6. Results

We present results for each of the three learning tasks described in Section 5 and the two time interval data sets. The same training/testing data split was used to fit each of the algorithms, for each learning task. In the case of the supervised regression type learning tasks (entry count and regulated entry count), the actual vs predicted plots are shown along with the corresponding values of R-squared (R^2) and Root Mean Squared Error (RMSE). For the classification task of determining if a weather regulation was active, a confusion matrix is shown for each algorithm tested.

6.1. Entry count results

Fig. 9 shows results for the learning task of predicting the entry count for the MUAC airspace given the weather conditions. Results indicate that all machine learning algorithms do a fairly good job as estimating how many aircraft will enter the airspace within the time window given the weather conditions, however the Random Forest and Neural Network algorithms work best for this problem. Given the highly cyclical traffic pattern in MUAC airspace, it is possible that the strong correlation in the results may be due to the features reflecting the day of the week and hour of the day. Furthermore, the data shows the algorithms are better at predicting entry count using an hourly time interval. This improved prediction in the hourly dataset is likely due to the reduced variability in entry count for in hourly time intervals when compared to 15 min intervals.

6.2. Regulated entry count results

Results for the learning task of predicting the number of aircraft entering the MUAC under a weather regulation is presented in Fig. 10. From the figure, it is evident that a correlation exists between the weather condition and the number of regulated aircraft. For this task, we can notice the Random Forest algorithm performs best, followed by the Decision Tree and Neural Network. The results in this task, are not as accurate as in the previous task of predicting total entry count, however, given the correlation values in the data we can conclude the weather features are influencing the model predictions. Furthermore, the results suggest that the models perform better when using the more precise 15 min time window interval weather data.

6.3. Weather regulation activation

Fig. 11 shows the prediction of an active weather regulation in the MUAC airspace. The confusion matrices gives insight into the classification skill of the four algorithms. Ideal values for a matrix are 1 in the upper left (True Negatives) and lower right (True Positives) corners, and 0 in the upper right (False Negatives) and lower left (False Positives). In our data, a label of 1 indicates a weather regulation was active. From the figure we can see that again the Random Forest Algorithm seems to perform best, giving the right percentage of True Positives. Again, the data suggest that the algorithms show better skill with the more precise weather information of 15 min time interval dataset.

6.4. Feature importance

In an attempt to understand the influence of the weather features on the model predictions, the impurity-based feature importance metric for the Random Forest models is presented. In this section we focus only on the Random Forest algorithm since it outperformed the others in each of the three learning tasks. The feature importance is a built-in model attribute that indicates how well the selected feature can divide the data into separate sets with similar responses. While the feature importance values may be biased towards high cardinality features, and the statistics are specific only to the training data set, they still provide some insight into the models. From Fig. 12 we can see that in the case of predicting Entry Counts, the most relevant features relate with hour of the day, specifically 0H, 23H, 1H, 2H and 3H. These features are times where the MUAC entry counts drop drastically, thus it is not surprising that the model ranks these features as high when predicting entry count. For this specific learning task, Top Cloud is the most important weather feature ranking 12th overall. It is interesting to note that this feature ranks higher than those dealing with the day of the week. In the case of the predicting Regulated Entry Count, the weather related features Shelf Cloud and Top Cloud stand out from the rest. This seems to indicate that the model prediction for this learning task is based primarily on percentage of airspace volume being taken up be convective weather. Lastly, the task of predicting if a Weather Regulation is Active is perhaps the most sensitive to weather, with the top seven ranking features being weather related. As in the case of predicting Regulated Entry Count, the Shelf Cloud and Top Cloud features seem to be most important. The high ranking feature importance of the Shelf Cloud and Top Cloud features seem to indicate that the volume of airspace being occupied be convection is perhaps more important than the severity. In the future, if we intend to build similar models using weather forecasts, it will be important to correctly predict the location and altitude of the convective weather.

Overall, the results indicate that there is a relationship between the weather data and the traffic patterns in the MUAC. While the degree of skill and correlation varies by learning objective, being able to predict how the current system will respond given a weather input is an important first step in determining the adjustments that need

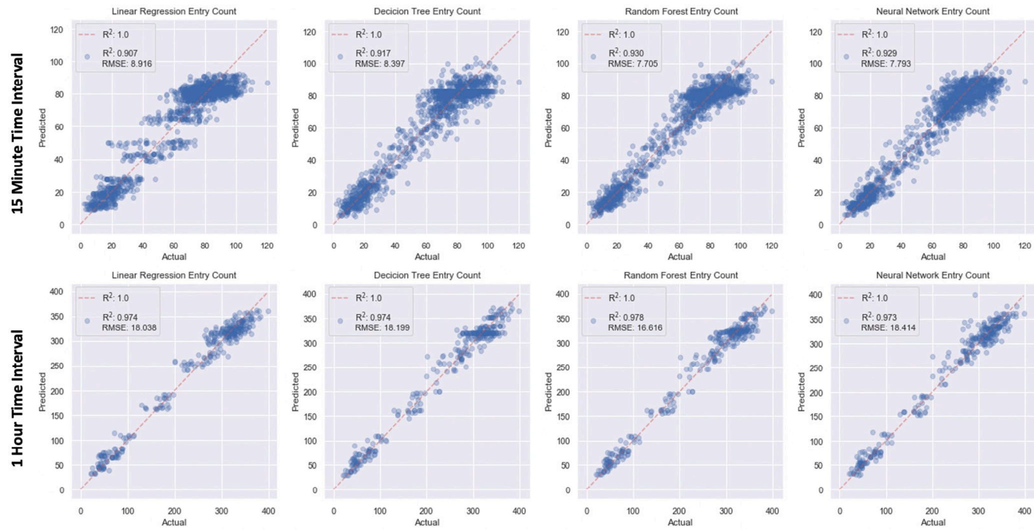


Fig. 9. Results for the learning task of predicting entry count for the MUAC airspace.

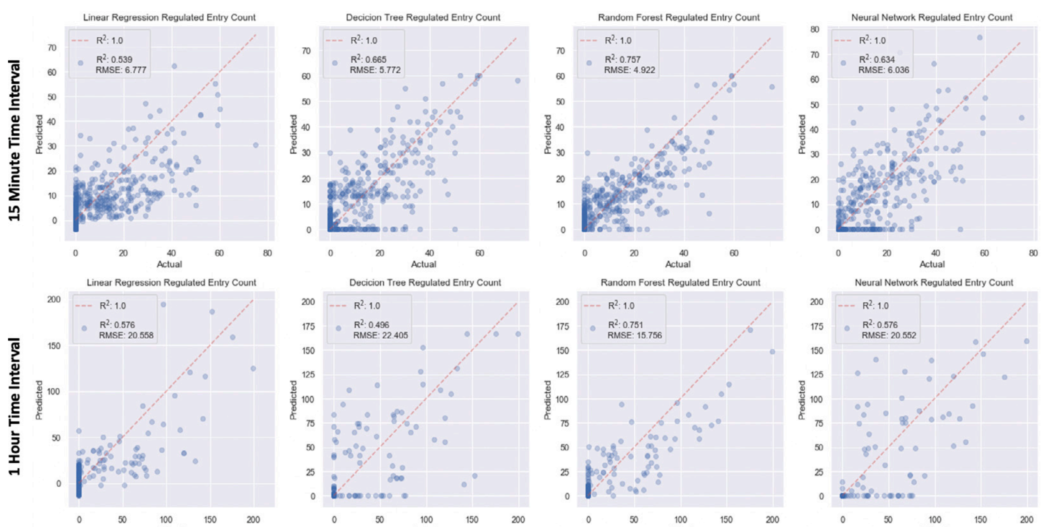


Fig. 10. Results for the learning task of predicting the regulated entry count for the MUAC airspace.

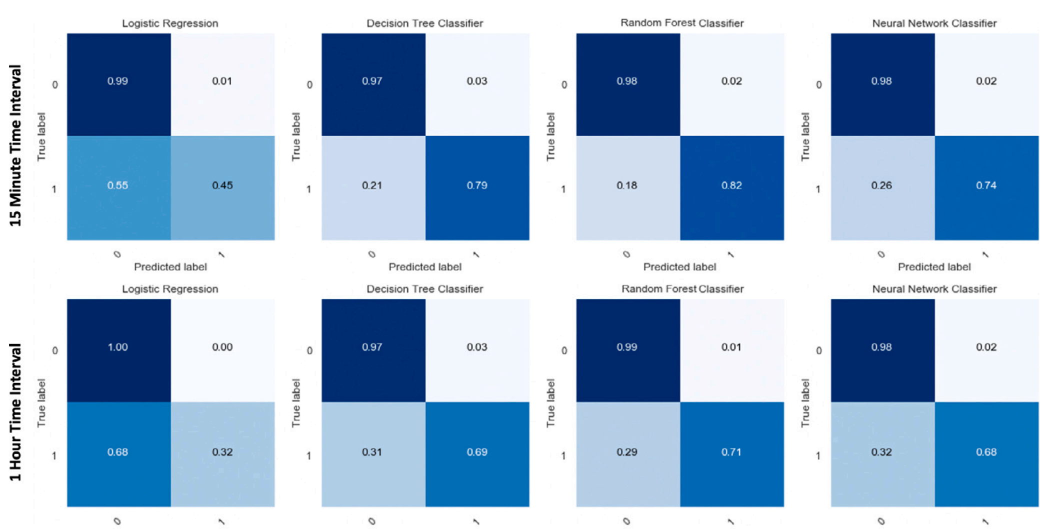


Fig. 11. Results for the learning task of predicting if a weather regulation is active in the MUAC airspace.

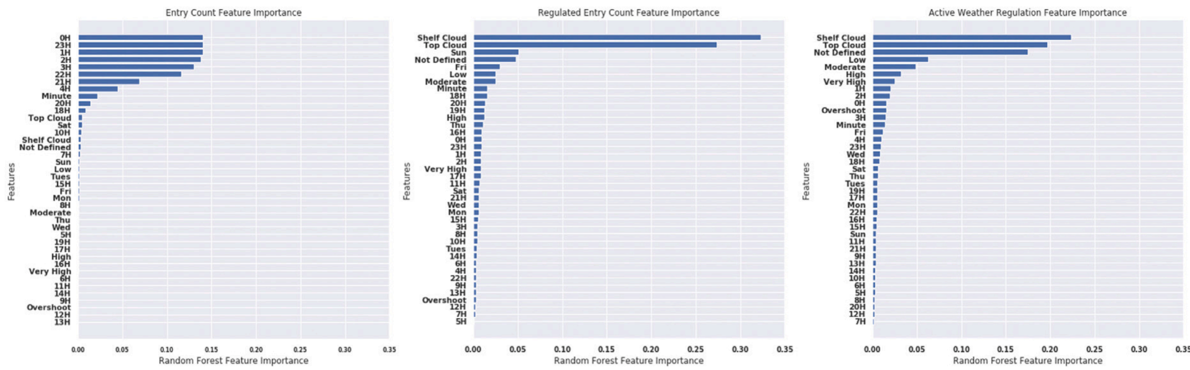


Fig. 12. Impurity-based feature importance of Random Forest models.

to be made to improve ATM performance. While we may be able to estimate the system response in terms of entry counts and whether a weather regulation is active, we still lack information on whether the estimated system response is an appropriate response. However, supposing critical situations in the system exhibit similar responses, the models may be useful in detecting these critical situations.

7. Conclusion and future work

In this paper we set out to demonstrate a methodology for integrating historical weather and air traffic data. A time series representation of the traffic and weather situation over 68 days was obtained. Preliminary implementation of multiple machine learning algorithms show correlation exists between weather conditions and characteristics in the MUAC airspace. Initial results varied for depending on learning task and the Random Forest algorithm proved to perform better than Linear Regression, Decision Tree and Neural Network architectures, however models need to be refined and tested with additional datasets to validate their accuracy. Results also provide some insight into how the temporal granularity of the data can impact the results.

The next steps in our research will focus on refining several of the assumptions that are currently made. With the use of actual forecasts, we will move away from the assumption of having a "perfect weather forecast". Rather than using storm observations, we envision having probabilistic weather forecast as inputs to our machine learning models. These probabilistic storm forecasts will be the output of ongoing parallel work, in which we integrate data from numerical weather prediction tools and storm observations to develop machine learning algorithms for weather prediction. Machine learning has already started making advances in the area of meteorology to improve the accuracy of weather forecasts, we aim to leverage these knowledge gains to improve the ATFM process.

Second, we want to improve our assumption of treating each interval in our time series as independent samples, it is obvious that events such as storms and regulations occurring at a given time instance will impact what happens at future instances. In the future we aim to build model architectures that are able to capture the time dependencies relating to how far in the future a prediction is being made, while incorporating all the traffic and weather information available at the time of prediction.

Lastly, we hope to improve the representation of traffic and regulation data. Air traffic patterns in the MUAC, or any other airspace are complex. Being able to discern between the traffic flows and traffic volumes within a sector would greatly enrich the quality of the data, enabling improved model performance and prediction of specific traffic characteristics.

The ultimate goal is to accurately predict the system response given the weather information. By having improved weather forecast and better understanding how that weather will impact the traffic, we hope to improve the air traffic flow management operations.

CRediT authorship contribution statement

Aniel Jardines: Conceptualization, Methodology, Data curation, Writing - original draft. **Manuel Soler:** Conceptualization, Supervision, Writing - review & editing, Funding acquisition. **Javier García-Heras:** Supervision, Writing - review & editing.

Acknowledgements

This work is partially supported through the Spanish Government initiative *Proyectos de I+D+I "RETOS INVESTIGACIN" Ministerio de Ciencia, Innovación y Universidades* by the project entitled "Management of Meteorological Uncertainty for More Efficient Air Traffic: Meteorological Data Provision and Thunderstorm Avoidance" (RTI2018-098471-B-C32).

References

Bouttier, F., Marchal, H., 2020. Probabilistic thunderstorm forecasting by blending multiple ensembles. Tellus A 72 (1), 1–19.

Carmona, M.A.A., Nieto, F.S., Gallego, C.E.V., 2020. A data-driven methodology for characterization of a terminal manoeuvring area in multi-airport systems. Transp. Res. C 111, 185–209.

Cook, A., Tanner, G., 2015. European airline delay cost reference values: Updated and extended values.

EUROCONTROL, 2019a. ATFCM Operations Manual. EUROCONTROL, https://www.eurocontrol.int/publication/atfcm-operations-manual.

EUROCONTROL, 2019b. An Assessment of Air Traffic Management in Europe during the Calendar Year 2018. 2018 Performance Review Report, EUROCONTROL, https://www.eurocontrol.int/air-navigation-services-performance-review.

Fernández, E.C., Cordero, J., Vouros, G., Pelekis, N., Kravaris, T., Georgiou, H., Fuchs, G., Andrienko, N., Andrienko, G., Casado, E., 2017. DART: A machine-learning approach to trajectory prediction and demand-capacity balancing. In: Seventh SESAR Innovation Days, Belgrade, Serbia, 28th–30th November.

Gallego, C.E.V., Comendador, V.F.G., Carmona, M.A.A., Valdés, R.M.A., Nieto, F.J.S., Martínez, M.G., 2019. A machine learning approach to air traffic interdependency modelling and its application to trajectory prediction. Transp. Res. C 107, 356–386.

Gianazza, D., 2017. Analysis of a workload model learned from past sector operations.

González-Arribas, D., Soler, M., Sanjurjo-Rivo, M., Kamgarpour, M., Simarro, J., 2019. Robust aircraft trajectory planning under uncertain convective environments with optimal control and rapidly developing thunderstorms. Aerosp. Sci. Technol. 89, 445–459,

Han, L., Sun, J., Zhang, W., Xiu, Y., Feng, H., Lin, Y., 2017. A machine learning nowcasting method based on real-time reanalysis data. J. Geophys. Res.: Atmos. 122 (7), 4038–4051,

Hentzen, D., Kamgarpour, M., Soler, M., González-Arribas, D., 2018. On maximizing safety in stochastic aircraft trajectory planning with uncertain thunderstorm development. Aerosp. Sci. Technol. 79, 543–553.

Jarry, G., Delahaye, D., Nicol, F., Feron, E., 2020. Aircraft atypical approach detection using functional principal component analysis. J. Air Transp. Manag. 84, 101787.

Kamangir, H., Collins, W., Tissot, P., King, S.A., 2020. Deep-learning model used to predict thunderstorms within 400 km2 of south Texas domains. Meteorol. Appl. 27 (2), e1905.

Lee, J.-G., Min, K.-H., Park, H., Kim, Y., Chung, C.-Y., Chang, E.-C., 2020. Improvement of the rapid-development thunderstorm (RDT) algorithm for use with the GK2A satellite. Asia-Pacific J. Atmospheric Sci. 56 (2), 307–319, [Online]. Available: https://doi.org/10.1007/s13143-020-00182-6, Marcos, R., Ros, O.C., Herranz, R., 2017. Combining visual analytics and machine learning for route choice prediction. Application to pre-tactical traffic forecast. In: Seventh SESAR Innovation Days, Belgrade, Serbia, 28th–30th November.

McNally, D., Sheth, K., Gong, C., Sterenchuk, M., Sahlman, S., Hinton, S., Lee, C.H., Shih, F.-T., 2015. Dynamic weather routes: two years of operational testing at American airlines. Air Traffic Control Q. 23 (1), 55–81,

Mecikalski, J.R., Rosenfeld, D., Manzato, A., 2016. Evaluation of geostationary satellite observations and the development of a 1–2 h prediction model for future storm intensity. J. Geophys. Res.: Atmos. 121 (11), 6374–6392,

Sauer, M., Hauf, T., Forster, C., 2014. Uncertainty analysis of thunderstorm nowcasts for utilization in aircraft routing. In: 4th SESAR Innovation Days (SID2014), pp. 1–8.

Seenivasan, D.B., Olivares, A., Staffetti, E., 2020. Multi-aircraft optimal 4D online trajectory planning in the presence of a multi-cell storm in development. Transp. Res. C 110, 123–142.

Wang, Z., Liang, M., Delahaye, D., 2018. A hybrid machine learning model for short-term estimated time of arrival prediction in terminal manoeuvring area. Transp. Res. C 95, 280–294.

Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-c., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Adv. Neural Inf. Process. Syst. 802–810.

Zhang, B., Tang, L., Roemer, M., 2017. Probabilistic planning and risk evaluation based on ensemble weather forecasting. IEEE Trans. Autom. Sci. Eng. 15 (2), 556–566.