Fifteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2023)

Machine Learning to Predict Convective Weather and its Impact on En-Route Capacity

Marta Sánchez, Danlin Zheng, Pablo Gil
CRIDA, A.I.E.
Madrid, Spain
mscidoncha@e-crida.enaire.es

Gilles Gawinowski, Ramon Dalmau
EGSD/INO/SHO&ENG
EUROCONTROL Innovation Hub (EIH)
Brétigny-Sur-Orge, France

Manuel Soler, Javier García-Heras, Iván Martínez, Aniel Jardines
Department of Aerospace Engineering, Universidad Carlos III de Madrid
Madrid, Spain. masolera@ing.uc3m.es

***Abstract*—Network performance is very sensitive to weather and uncertainty in its prediction. We address these challenges through the contribution to an Artificial Intelligence (AI)-based Network Operations Plan. This plan is enhanced by including a probabilistic weather prediction tailored to ATFM, ATM and weather data integration and demand and capacity imbalance characterization at the pre-tactical and tactical phases of ATFM. We integrate all these modules into a visualization tool aimed at supporting human's decision-making. An operational assessment has been conducted. The improvements for FMPs and NM in pre-tactical and tactical ATFM are: 1) Ability to understand the weather convective prediction and to identify critical weather areas; 2) Ability to understand and manage the capacity reduction prediction; 3) Ability to manage the situation at the network level; 4) Ability to improve the efficiency and productivity of human performance (workload, usability).**

***Keywords*—ATFM; Artificial Intelligence; Weather**

I. Introduction

Despite the COVID crisis and the war in Ukraine, aerial traffic is close to recovering 2019 levels (85% in Europe). Traffic forecasts expect air traffic to increase a 10-20% over the next 7 years[1] and the analysis of actual conditions suggests that all-causes Air Traffic Flow Management (ATFM) delays may increase exponentially. It becomes critical for network management to address demand and capacity balancing in an efficient way so that the associated measures actually contribute to minimise delays at network and local levels.

Network prediction and performance is very sensitive to weather and the uncertainty in its prediction. In addition, current ATFM operations are not evaluated from a systematic perspective. These two factors together lead to a strong dependency on the experience of human operators.

In this paper, linked to the EU project ISOBAR[2], we address these challenges through the contribution to an Artificial Intelligence (AI)-based Network Operations Plan. This plan is enhanced by including in its scope a probabilistic weather prediction tailored to ATFM, ATM, weather data integration and demand and capacity (DC) imbalance characterisation.

Indeed, the possibility of using Artificial Intelligence (AI) techniques is emerging with great force in various fields. In a recent Nature communications survey [1] the authors claim that AI can enable the accomplishment of 134 targets across all the goals established in 2030 Agenda for Sustainable Development, yet it may also inhibit 59 targets. This includes Earth sciences (and meteorological prediction) [2] and the aviation domain [3]. Nevertheless, the use of AI in problems related to ATFM operations is still scarce.

Early work aimed at assessing the impact of convective weather on ATFM began in the United States building a model to estimate aircraft deviation probabilities in the vicinity of convective weather for terminal areas based on historical analysis of traffic flows and weather information [4]. Then, this model was further developed to deal with en-route traffic [5] which later gave rise to the convective weather avoidance model (CWAM) [6]. Later on, this model became a cornerstone of the Route Availability Planning Tool (RAPT), an automated decision support tool of the Federal Aviation that improves management of flight departures at airports during thunderstorms [7]. In addition, nowadays CWAM is at the core of the NextGen Weather Processor (NWP). On the contrary, the implementation of such decision support tools in operation is still a pending aspect in Europe. Nonetheless, initiatives like the Cross Border Weather Forecast promoted by EUROCONTROL Network Manager show a strong determination in this regard [8].

Instead of addressing the problem from a pure trajectory perspective, other works have attempted to measure the impact of convective weather into ATFM in terms of Air Traffic Control (ATC) sector capacities drops. By applying graph theory, [9] measures sector capacity drops as a function of traffic flow complexity and blockage. In [10], the impact of convection on sectors in the US network is quantified based on the fraction of the sector's area covered by convective weather forecasts. Finally, in [11], sector capacity drop were estimated by estimating the maximum workload that the respective Flow Management Position (FMP) team assigned to a sector can manage safely [11].

If we resort to the usage of machine learning to shape models that can leverage historical traffic and weather data to

[1]EUROCONTROL - STATFOR Team, "EUROCONTROL Seven-Year Forecast February 2022," Brussels, 2022.

[2]https://isobar-project.eu/

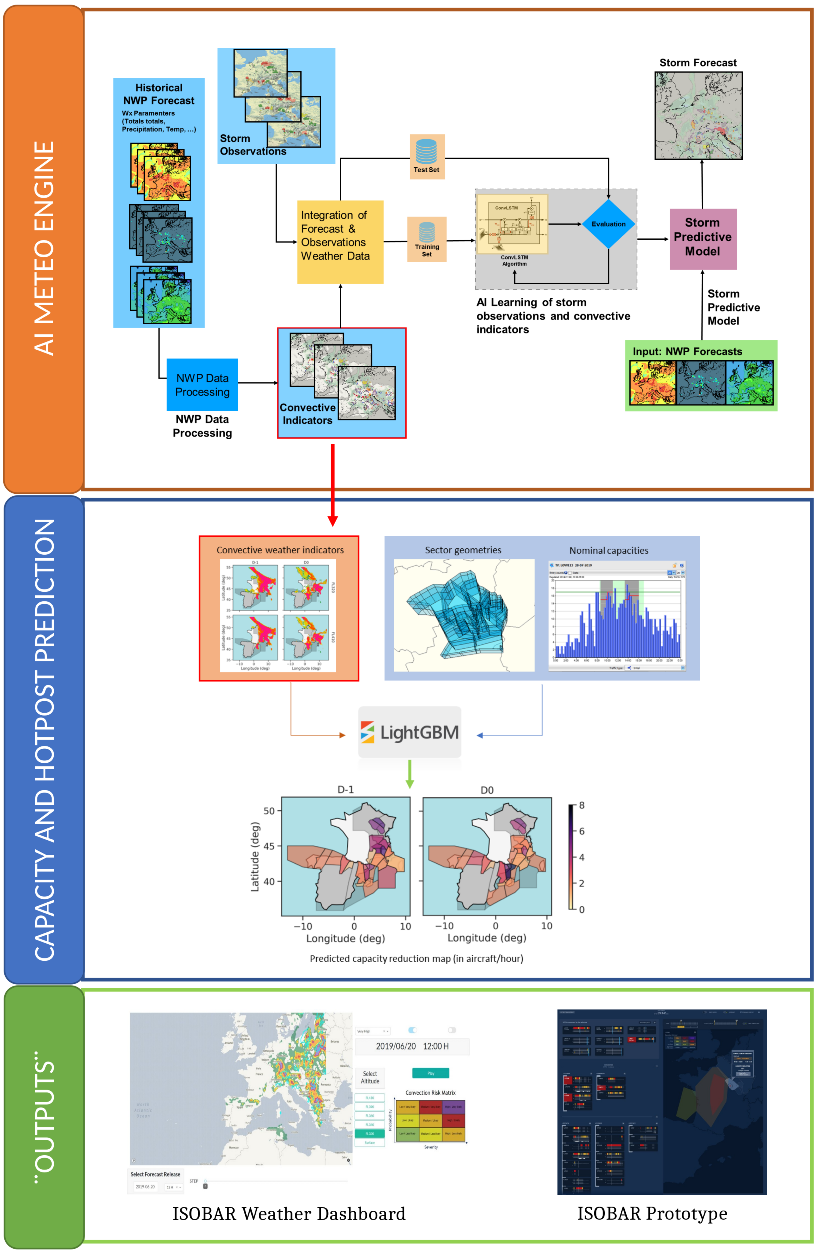

Figure 1: Concept Figure.

enhance ATFM processes during convective weather events, previous work is limited. In [12], the authors compare different supervised machine learning architectures for predicting sector entry counts, sector weather regulation activation and regulated entry counts in the Masstric Upper Area Control Centre (MUAC) airspace. The same sector-based approach was used by [13] to develop a decision support tool for FMP units aimed at predicting the intervals of time at which an specific traffic volume must be regulated. Results show that proposed methodologies can predict ATFM measures due to convective weather accurately at a sector level which could eventually help traffic managers to detect overloads and sector clousures at a tactical level. However, from a pre-tactical level, a network-centric approach capable of predicting system-wide implications is still lacking. Moreover, the integration of the Dynamic Airspace Configuration (DAC) concept will require to assess the impact of convective weather on ATFM from a holistic view, as the airspace sectorisation will be dynamically adjusted to match airspace capacity with traffic demand.

All in all, we present for the first time an AI-based, network-centric prototype to assess capacity-demand imbalances due to convective weather at the pre-tactical and tactical phases of ATFM. The ISOBAR prototype (as we coin it) leverages an AI-based Meteorological prediction module (see Section II), a capacity and hotspot characterization module (see Section III), and a visualization tool aimed at aiding human's decision-making (Section IV). The operational assessment of the ISOBAR prototype is provided in Section V. See Figure 1 for a concept overview.

II. AI-BASED WEATHER FORECASTING

An AI-based approach was used to accurately predict and improve situational awareness around convective weather. Several AI-based models were developed utilizing a combination of weather forecast and weather observation data to train machine learning algorithms to predict the occurrence of convective weather including characteristics such as severity, cloud top altitude, overshoots, and lightning up to 36 hours in advance. The enhanced digital weather information produced by the AI-based weather forecast has the potential to improve the planning processes and mitigation strategies of air traffic management when dealing with disruptive weather events.

A dataset of historical weather data was provided by the organizations Agencia Estatal de Meteorologia (AEMET), Météo-France and Earth-networks to develop the AI-weather forecasting models. The AI model learning tasks were formulated to utilize data from the numerical weather products as input and predict the observational data as output. A separate model was created for each NWP product. In this way, the models are capable of interpreting the raw data from the NWPs to provide a representation of the expected convective weather.

Three AI algorithms were trained, validated, and tested using three sources of numerical weather prediction (NWP) products. Each of the forecasts exhibited differences with respect to the geographical region of coverage, set of available parameters, and spatial resolution, providing several use cases to trial the AI-based methodology for convective weather prediction.

- The European Centre for Medium-Range Forecast Ensemble Prediction System, which covers the pan-European area;
- The high-resolution regional forecast gammaSREPS product, produced by the Spanish meteorological agency AEMET;
- The high-resolution regional forecast AROME product produced by Météo-France.

From each NWP product, parameters were selected based on their ability to provide information related to atmospheric factors such as instability, moisture, and triggering mechanisms that could lead to convection.

To create the target function, or "labeled data" for each of the models, two sources of thunderstorm observation data were utilized; satellite observations and lightning detection data. Satellite observations were obtained from the Rapid-Development Thunderstorm (RDT) developed by Météo-France within the EUMETSAT NWC-SAF framework. This product employs geostationary satellite data to provide information about clouds related to significant convective systems from the mesoscale (100–1000 km) down to tenths of kilometers [14]. The lightning detection data was provided by Earth-networks' Total Lightning Network product. Characteristic of the observational data utilized from these sources is provided below:

- The RDT product provided convective cell polygons at 15-minute frequency. Each polygon provided characteristics related with the storm severity rating (Low, Medium, High, and Very High) and the cloud top altitude.
- Earth-Networks Total Lightning Network product provided the timing, location, and peak current measurements of both intra-cloud and cloud-to-ground flashes.

In order to train the models, it was necessary to blend the forecast and observational data. This blending process allowed for a spatial-temporal resolution harmonization between the NWPs and observations for each model. Furthermore, the observational thunderstorm characteristics such as severity, cloud top altitude, and overshooting tops were transformed into binary code representation. The final training, validation, and test data sets covered convective periods from the summers 2018 and 2019.

During the development, several machine-learning algorithms were explored. In all instances, the learning tasks were formulated as supervised binary classification problems using cross-entropy loss functions. during the training process. Early versions of the model were conducted using simple neural networks (NN) architectures, these architectures only considered the values of a set of parameters provided at a grid point to provide a convection prediction at that location. However, later versions explored more complex model architectures such as long-short-term memory (LSTMs), to account for the temporal behavior of the input parameters, and Convolutional Neural Networks (CNNs) to exploit the spatial characteristics within the NWP data. While some sophisticated model architectures performed better at predicting certain features such as thunderstorm severity and cloud top altitude, overall all models showed positive results indicating that an AI-based approach can be utilized to predict the occurrence of convection. An operational assessment of the results found that forecasts covering a large area such as the ECMWF EPS could be best employed when dealing with weather problems on the network-wide scale, while high-resolution forecasts such as the gammaSREPS and AROME are better suited for dealing with weather issues at the local level. Results from the various NWP forecasts can be seen in Figure 2.

The use of machine learning for convection prediction allows for the implementation of an automatic digital warning system to provide ATM Flow Managers and other decision-makers with accurate hourly predictions of convective weather at longer lead-times. The current research also tackled integration challenges of translating the results from the neural network models into an ATFM operational environment. Proper visualization of model results is required to provide users with a format that is ready for uptake within ATM operations. Model results are transformed into polygons using the defined convection risk matrix, shown in Figure 3, providing decision-makers with a visualization that they are already familiar with. A visualization of final model results in the final format is provided in Figure 4.

The digital solution can be easily integrated with existing ATM tools and systems using an API. The tool also allows for the customization of convection risk tolerance depending on the user or region in question. Based on the enhanced weather information provided by the AI algortihms, weather

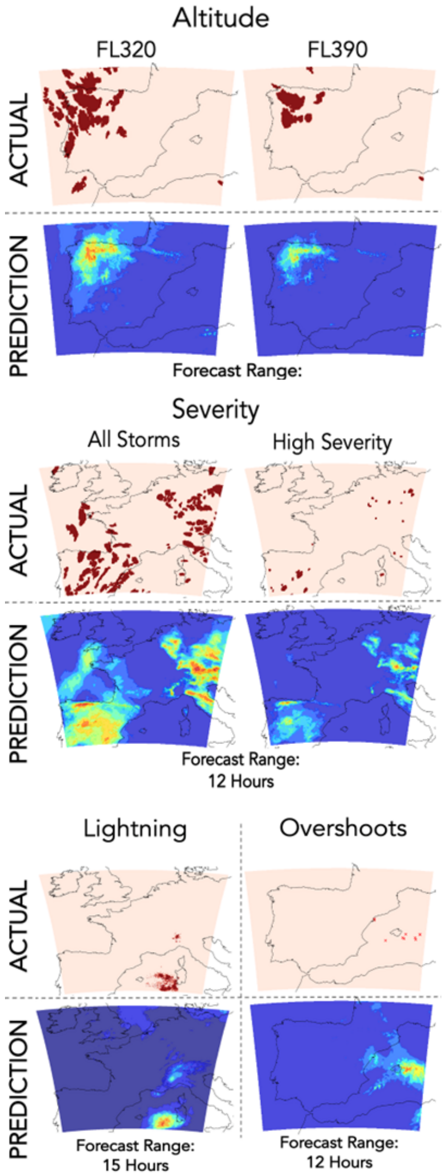

Figure 2: Results showcasing AI-based models skill at predicting thunderstorm, severity, altitude, lightning and overshoots.

decisions impacting re-route and sector capacity adjustments can start being addressed at the pre-tactical planning phase, allowing for more efficient mitigation measures.

III. Capacity and Hotspot Prediction

We present a model that, given a convective weather forecast, predicts capacity reduction (in entries per hour) at individual traffic volumes (TVs). As a result, by providing the expected capacity reduction ahead of the weather event, this component supports the preparation and negotiation of mitigation measures. The model was built on the assumption that the greater the airspace volume affected by convective weather, the greater the capacity loss. The precise relationship between these two quantities, however, is not evident. Accordingly, developing a model with rules established by experts would be a time-consuming and error-prone effort.

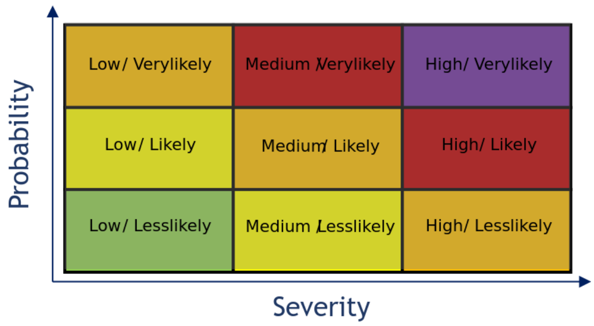

Figure 3: Convective weather risk levels used in operation

Instead, the capacity prediction model was trained on historical capacity and weather data using cutting-edge machine learning techniques to learn this relationship without the need for human intervention.

Specifically, the model used in this work is an ensemble of gradient-boosted decision trees (GBDTs) [15]. This type of model was selected because GBDTs have been demonstrated to outperform NNs in various applications, particularly on tabular datasets like the one presented in this section [16]. Additionally, GBDTs are more interpretable than NNs, making them more likely to be trusted by human operators.

The GBDTs model uses as input (1) the total volume of the airspace sector associated with the TV; (2) the proportion influenced by the various weather risk levels shown in Fig. 3 (green, yellow, orange, red and violet) at a specific time, which are provided by the AI-based weather forecasting model presented in Section II; (3) the nominal capacity when the weather is clear; and (4) the identifier of the ACC to which the airspace sector belongs. As such, the model is fed with seven numerical and one categorical feature as observation vector $\boldsymbol{x}$ per TV and time, and outputs a scalar y representing the predicted capacity reduction. Figure 5 illustrates the computation of features for a simple example, which is in 2D merely for visualisation purposes. The reader must keep in mind that, in reality, the features of the observation vector are computed in terms of volume rather than area.

It is worth noting that convective weather is typically reported at specific flight levels. To determine the volume intersections, we first compute the overlapping area in two dimensions using the Shapely Python package, and then consider the height of the airspace sector, as well as the vertical separation with the next and previous flight levels where weather was reported. This process is highly efficient and does not require a significant computational burden.

The model was trained and tested using the convective weather forecasts from the AI-based weather forecasting model. In particular, the experiment focused on the months with the highest number of en-route air traffic flow management (ATFM) regulations caused by weather all over the European Civil Aviation Conference (ECAC) area during 2018 and 2019: June, July and August (i.e., summer).

Each observation $(\boldsymbol{x}, y)$ of the dataset corresponds to one TV at a specific time. It should be noted that, despite the fact that the operational assessment described in Section V only

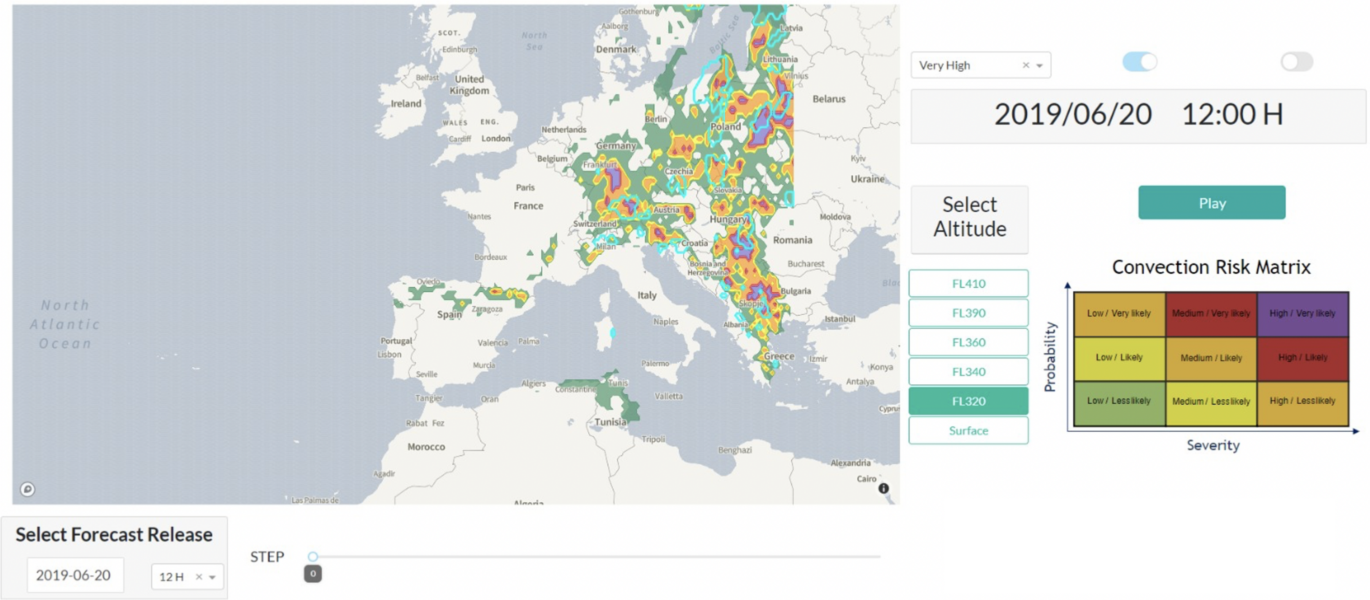

Figure 4: AI-based weather forecasting dashboard provides digital weather information

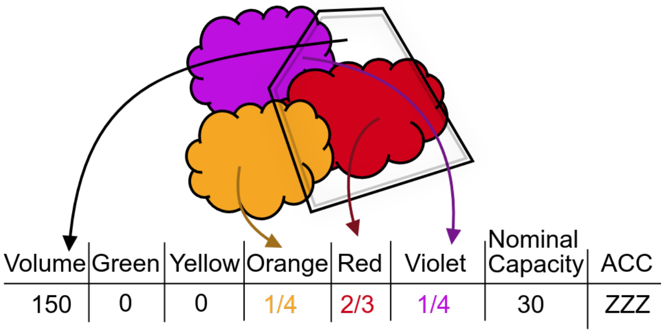

Figure 5: Example of observation vector computation

focused on Spanish and French airspace, all TVs regulated over the ECAC area during the period under consideration were taken into account when creating the dataset, with the goal of training a model with the ability to generalise as much as possible. For each observation, the eight features that compose $\boldsymbol{x}$ were determined as explained above, whereas y is the entry rate of the ATFM regulation applied to the TV.

The observations corresponding to the last month of the period under consideration (i.e., August 2019), as well as the dates selected for the operational assessment presented in Section V, were used for testing (i.e., to evaluate the performance of the model on unseen data), whereas the remaining 13K observations were used for training purposes.

Table I compares the regression metrics for the AI-based capacity reduction model predictions on the test set to a dummy baseline that assumes no capacity reduction.

According to Table I, the model predicted the actual capacity of the regulated TVs during the times considered in the test set with a mean absolute error (MAE) of 2.8 entries per hour, reducing the prediction error of the baseline by 3.7 entries per hour (57%). Table I reveals similar and even more

TABLE I. Regression metrics on the test set. Units are entries per hour except for the R2, which is dimensionless

Model	MAE	MSE	R2	MAXAE
Baseline (zero capacity reduction)	6.5	64.0	0.09	23
AI-based capacity reduction model	2.8	13.0	0.81	16

remarkable findings for the remaining regression metrics: the mean squared error (MSE), the coefficient of determination (R2), as well as the maximum absolute error (MAXAE).

It should be noted that the GBDTs model was not the only one designed during the project to predict capacity reduction using AI techniques. Rather than constructing a structured dataset[3] suited for tree-based models by determining the intersection of convective weather risk polygons and the airspace sectors associated with the TVs, a more pragmatic approach consists of processing the convective weather forecast as if it were an image using CNNs, then *pool* the result into a latent vector, and finally use a standard feed-forward neural network with as many outputs as possible TVs to determine the predicted capacity reduction of each one.

Like with many machine learning tasks, however, training a neural network requires far more data than training a tree-based model. A neural network trained on *only* 13K observations is unlikely to reach the same performance as the GBDTs model (see Table I). Furthermore, another advantage of tree-based models is that they are relatively easy to interpret. A well-known method for interpreting the predictions of a tree-based model is the Shapley method [16], which is build on principles from game theory. Mathematically speaking, the Shapley value a specific feature i for a given observation represents the average marginal contribution of i on the output

[3]In the machine learning terminology, a structured dataset consists of a table with as many rows as observations and as many columns as features.

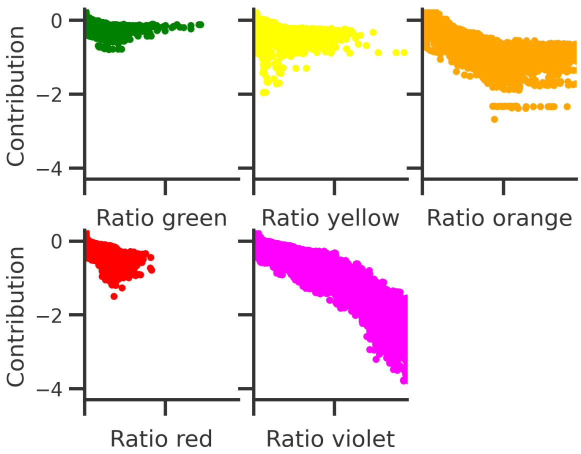

Figure 6: Learned contribution of the risk ratio features (in entries per hour, as the model output)

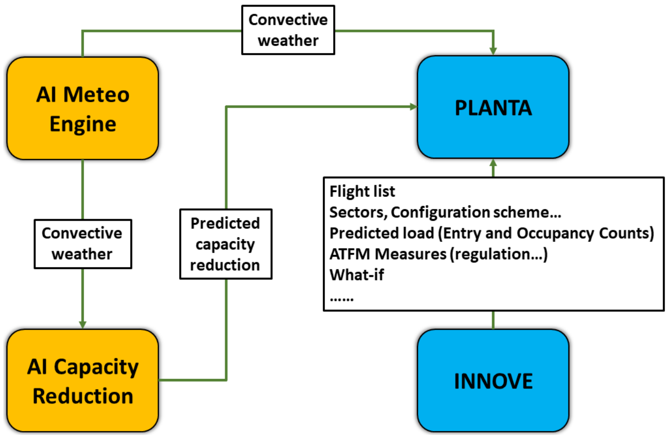

Figure 7: ISOBAR Prototype Architecture

of the model across all possible combinations of features. This method is becoming more popular in the literature as machine learning applications need model interpretability, particularly in high-risk applications or those where errors could result in considerable financial losses. The details about the Shapley method are beyond the scope of this paper, so the reader is directed to the original article for more information.

Figure 6 shows the dependence plot for the proportion of airspace sector (0 means no intersection, and 1 means fully covered) influenced by the various convective weather risk levels. In this kind of graph, the horizontal axis represents the value of the input feature, and the vertical axis is the contribution on the model output (i.e., the Shapley value). Each dot represents one observation, and the Shapley values are expressed in the same units as the model output.

The hypothesis stated at the beginning of this section is supported by Figure 6: the greater the ratio of airspace volume affected by convective weather, the greater the capacity loss. Furthermore, and as expected, the higher the risk, the greater the capacity loss. For instance, when a TV is fully covered of high severity and probability convective weather (violet), the capacity drops by 2 to 4 entries per hour only due to this feature. It should be noted that other features may contribute to a further drop in capacity. The vertical dispersion observed for a given value in the horizontal axis results from the fact that the Shapley value of a feature for a specific observation is dependent on the value of the other features (e.g., the total volume of the TV or the ACC identifier).

IV. ISOBAR PROTOTYPE

A prototype has been built to run a human-in-the loop validation exercise to assess the operational acceptability and performance of AI components. The exercise covers the use case of detection of a severe convective weather and detection of France/Spain cross-border imbalances in pre-tactical phase (D-1) and re-assessed in the tactical phase (D-0). The objective was to validate the local FMP (ANSP) and Network Manager operators interactions with AI components to manage convective weather situations. The prototype is structured on two main components (Convective Cells and Storm Prediction, and Demand and Capacity Characterization).

The prototype was built-up on two ATFM simulators developed by EUROCONTROL:

- INNOVE is a ATFM real-time simulator emulating the full capabilities of the Network Management functions with B2B web service connectivity and REST services. The simulations are based on data recorded by the operational NM system. Traffic data is based on allFT+ data files and airspace data on DDR2 data files.
- PLANTA connected to the INNOVE B2B back-end simulator offers the full HMI look and feel capabilities emulating the FMP and NM working positions. Additional HMI capabilities have been developed to visualize the convective weather prediction and the predicted capacity reduction.

The two components, AI Met Engine and AI Capacity Reduction, have been integrated into the architecture shown in Figure 7 with information flow.

Two HMI capabilities have been developed to display convective information and capacity reduction. In the figure below is an illustration of the convective geographic map on which end users can select the level of probability and severity to display.

Four prototype positions have been set-up to run French ACCs (Reims, Marseille), Spanich ACC (Barcelona) and the Network Manager to perform the evaluation assessment. See the visualization of one of this prototypes in Figure 8.

V. OPERATIONAL ASSESSMENT

The Met Engine was subject to a formal evaluation of its operational impact in pre-tactical Air Traffic Flow Management (ATFM) process. The evaluation was part of a more ample exercise focused on validating human interactions with Artificial Intelligent (AI) components supporting the management of convective weather situations. The Met Engine was one of this components, which was also previously tested outside the operational context loop, in order to assess the level of accuracy of the resulting forecasts.

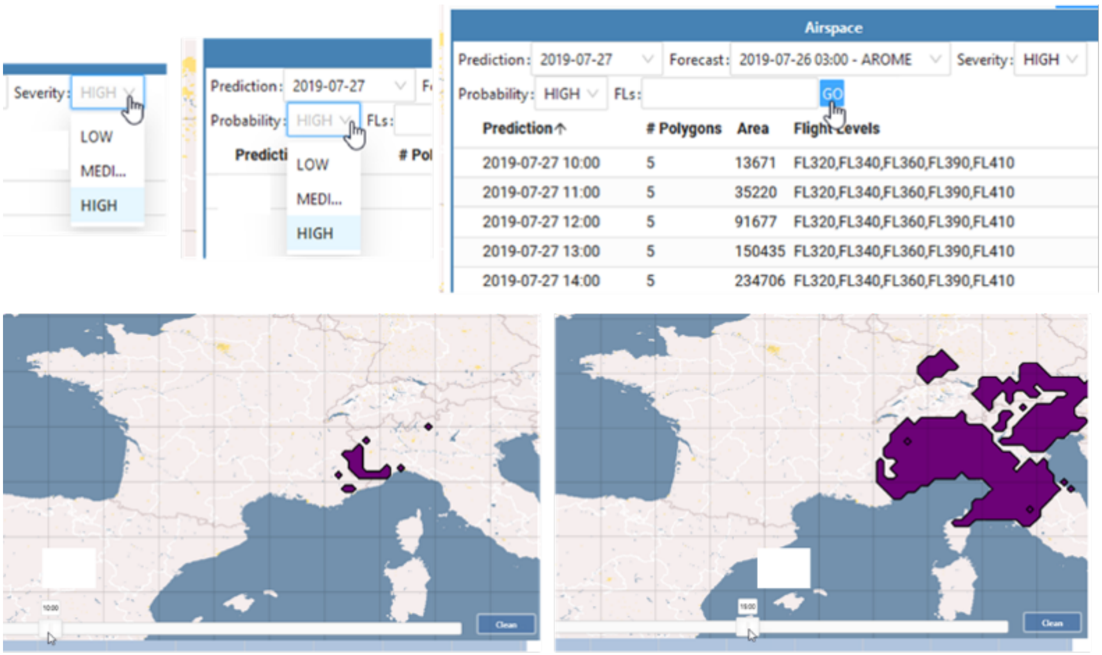

Figure 8: Prototype – Convective Weather Visualization (Left map : high probability and high severity, right map : low probability and high severity).

The evaluation exercise was performed including a forecast of a weather-induced critical situation with negative impact on airspace capacity, requiring pre-tactical ATFCM intervention. The cross-border scenario encompassed various capacity hotspots in en-route sectors of Spain and France forming a network spot or netspot. Therefore the participation of EUROCONTROL, as Network Manager, and Flow Management Positions (FMP) of ENAIRE and DSNA, as Air Navigation Service Providers (ANSP) of both states, was required.

In the following Table II, more details of the operational assessment are included.

TABLE II. Operational Assessment details

Mode	Human-In-The-Loop Real-Time experimentation
Platform	PLANTA/INNOVE
Participants	3 FMP from DSNA 1 FMP from ENAIRE 1 NM ops from EUROCONTROL 1 meteorologist from MeteoFrance
Simulation dates	26^{th} July, 27^{th} July and 27^{th} August (2019)
Simulation area	Reims and Marseille ACCs for DSNA Madrid and Barcelona ACCs for ENAIRE
Number of trials	Three, one per simulated date

The exercise covered the use case *Detection and Resolution of a Netspot in Pre-Tactical Phase (D-1) and re-assessed in the Tactical Phase (D-0)*, exploring how human actors (NM and FMP controllers) can manage:

- the processing of new information dealing with convective weather prediction and weather capacity reduction.
- the enhanced cross-border collaborative process involving NM and FMPs.

The method for the evaluation of the Use Case was to confront the air traffic controllers with a critical weather induced situation affecting the capacity of various en-route sectors within the geographical scope of the exercise. The Real Time simulation platform displayed different information in each position, with the relevant geographical granularity, capacity and traffic data for each actor. The information at D-1 focused on meteorological forecast and on estimation of impact on the sectors' capacities, according to the calculations of the module for capacity and hotspot prediction. The platform allowed the controllers to select the capacity reduction of each sector, the traffic impacted and the mitigation measures needed. An updated forecast was displayed to simulate the situation at D-0, in order to evaluate stability of the solutions selected at D-1 in view of the new, more accurate, data. The assessment of the **operational usability and acceptability** of the **enhanced ATFM procedure** was based on observations of researchers, general de-briefings and questionnaires after each simulation session.

The questionnaires were designed around three evaluation objectives relevant for the operational evaluation of the Met Engine [17]. These objectives have been established in reference to the gradual introduction of higher levels of automation enclosed in the European ATM Master Plan [18]. Each objective has been addressed to each one of the levels defined: level 1 (support in information acquisition and analysis), level 2 (task support) and level 3 (actions selection).

The three objectives, along with the aimed research questions and the hypotheses proposed for their evaluation, are presented in the subsection below, followed by another subsection presenting the key results.

A. Objectives and Hypothesis

OBJECTIVE 1 - Integration of convective weather information.

This objective aimed at the following research questions:

- How to represent the severity of the convective phenomenon and the probability of occurrence to support the awareness and understanding of operational actors?
- Which is the most digestible way of visualization of convective information for operational actors?

Three hypothesis were formulated to verify this objective:.

– *H1.1*: a convective risk matrix mixing the severity of the convective phenomenon and the probability of occurrence supports the understanding of the weather situation by operational actors.
– *H1.2*: the convective information displayed as coloured polygons (according to risk matrix) on a geographical map allows a very good understanding and analysis of weather phenomenon by the operational actors.
– *H1.3*: the convective information displayed superimposed to sector information allows the operator to assess the impact on each traffic volume.

OBJECTIVE 2 - Enhancement of ATFCM process facing weather critical situations at pre-tactical and tactical levels.

This objective aimed at the following research question:

- Which procedure is best from pre-tactical D-1 to tactical D0 to improve the weather-related problems and solutions, integrating weather information and DCB detection and resolution in a collaborative process?

Two hypothesis were formulated to verify this objective:

– *H2.1*: a workflow from pre-tactical D-1 to tactical D0 supports a collaborative process to better manage in an anticipated manner the weather-related problems and solutions.
– *H2.2*: the workflow is split in two parts: to prepare a weather scenario and implement some measures in the pre-tactical phase and to reassess the situation and to manage residual imbalances in the tactical phase.

OBJECTIVE 3 - Precise characterisation of demand and capacity imbalances due to convective weather

This objective aimed at the following research questions:

- Which is the most digestible way of visualization of the predicted DCB impact for operational actors?
- Which object can support the management of cross-border convective weather?

Three hypothesis were formulated to verify this objective:

– *H3.1*: the predicted DCB impact displayed as alerts in the configuration dashboard allows a very good detection of weather-related DCB impact by the operational actors.
– *H3.2*: the predicted DCB impact displayed as a threshold in the traffic volume monitoring allows a very good understanding and analysis of weather-related DCB impact by the operational actors.
– *H3.3*: the Netspot object is the proper reference to manage the cross-border weather operations and is easily manageable by operational actors for its creation/ modification/ cancellation and to classify its coordination status as proposed/ coordinated/ implemented.

Figure 9 summarizes the objectives, research questions, and hypotheses for the Operational Assessment.

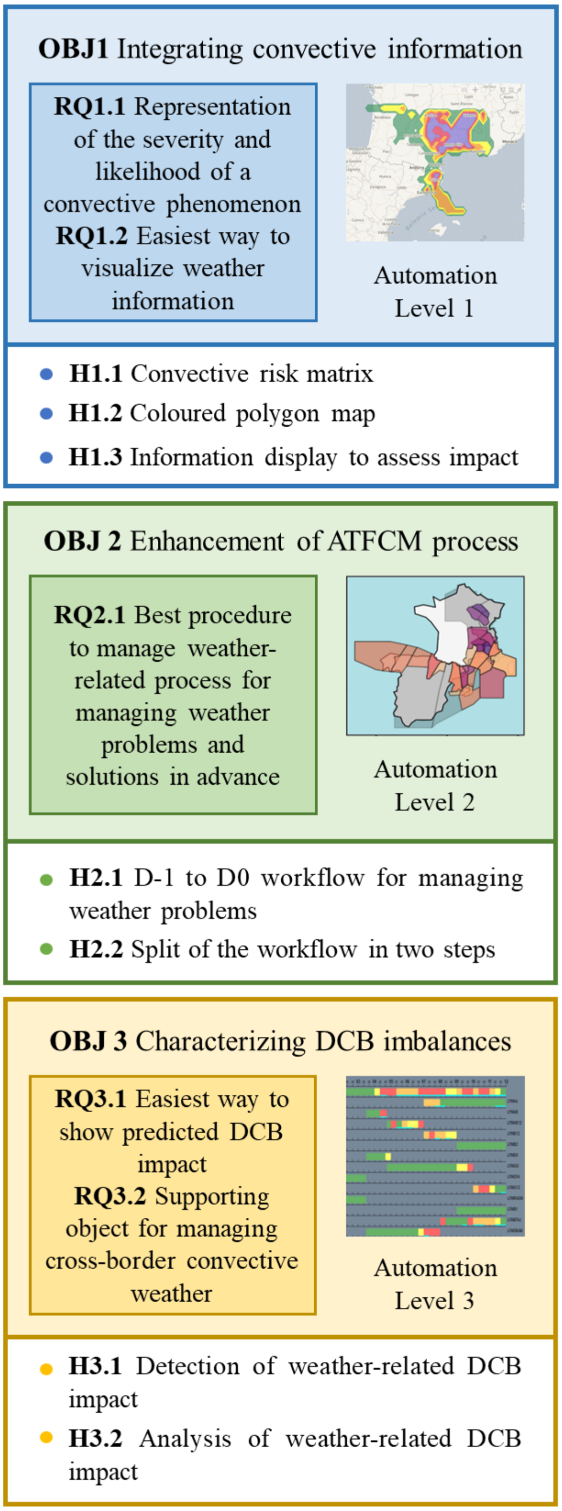

Figure 9: Objectives, research questions and hypotheses for the operational assessment

B. Results

Regarding OBJ. 1, both FMP and NM reported that **the convective risk matrix allowed to assess the severity and probability of weather phenomena** as low, medium and high helping them to easily identify the areas requiring pre-tactical actions. They added the proposal of using this matrix as a selection tool for more granular or combined information selecting the particular combination of severity and probability to display.

Both actors also reported that **the colour code for risk representation was understandable and helpful to assess the weather situation** in terms of severity and probability of occurrence. The granularity of convective information displayed in simplified polygons was usable and understandable, so all actors could easily identify the convective areas and their depicted forecasted evolution through time.

The **weather information was found adequate and very useful] to determine the set of traffic volumes (TVs) with potential capacity reduction**. From here, it was intuitive to cross-check that information with the predictions of the capacity and hotspots prediction model.

- As a future improvement for maturing the Met Engine tool, all actors agreed that the best option will be to present the operator either a single aggregated output fed by the different providers (GSREPS, ECMWF, AROME), or a combination of the different weather forecasts incorporating a dispersion metric representing their degree of similarity.

The workflow for the enhancement of ATFCM process facing weather critical situations stated in the OBJ. 2 was very positively rated by the actors involved. It resulted **very helpful for anticipating critical situations at D-1** providing a better decision-making and a better network stability to avoid last-minute snow-ball effect and smooth their workload.

In addition, **the splitting of the workflow in two parts between pre-tactical and tactical phases was also very positively assessed**. It served for a better anticipation of problems and solutions starting at D-1 and managing residual problems at D0. The workflow led to a reliable performance compared to the use of meteo observations, although a meteorologist was considered necessary for refining the analysis.

It was observed a **very good stability along time of the capacity reduction prediction caused by convective weather situations between D-1 and D0**, making possible to start planning solutions at D-1. In the Figure 10, this high stability of predictions at D-1 and D0 is illustrated, showing the impacted polygons in similar severity colors.

- Although the workflow was well defined, the implementation strategy for the mitigation solutions at pre-tactical phase was not very clear for the actors involved. Further investigations should be conducted to propose new working methods for a better design of the strategy. At the same time, this investigation will help to confirm the ATFM solvers stability performance.

In respect of OBJ. 3, the operational staff agreed on the **relevance, realism and precision of the weather-related capacity reduction values**. They also reported positively the possibility of comparing the value of the maximum demand expected with the predicted capacity value. This, together with the possibility of easily adjusting the monitoring values in the case of convective weather, supported the operators situational awareness while detecting weather imbalances.

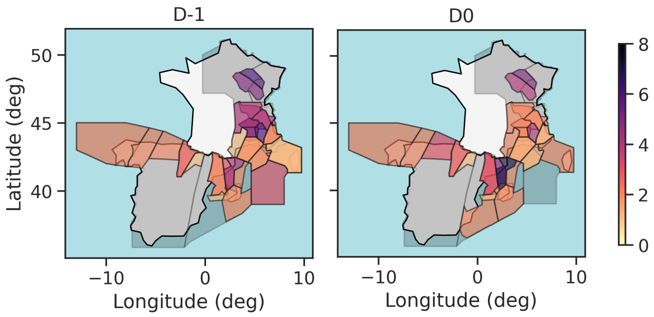

Figure 10: Predicted Capacity Reduction, July 27th at 08:00

The **configuration dashboard helped to reach to a common understanding of the cross-border weather problems**, allowing to visualize in detail the weather-related capacity drop in the entry count schemes. The display interface (see Figure 11) also allowed the operators to compare the demand with the initial monitored capacity value and the weather-related capacity reduction values.

Finally, both FMP and NM reported that **the Netspot was the proper object to identify the cross-border weather problems**. Thanks to it they were able to identify easily the cross-border convective events with the weather-related capacity reduction prediction and to understand the imbalance propagation at the network level. In the same way, they declared the Netspot an easily manageable concept when creating, modifying or canceling it and coordinating its status.

- Further characterization of demand and capacity imbalances due to convective weather should incorporate filtering options to the capacity reduction predictor, as well as the possibility of crossing the information of the impacted TV with its resulting reduction on the visual map. In this way, operational staff would more easily detect and, consequently, analyse weather-related DCB impact helping to maintain their situational awareness.

VI. Conclusions and Recommendations

We have shown how Artificial Intelligence enhances weather prediction up to 36 hours ahead and predicts capacity drops at individual traffic volumes, resulting in a promising methodology to assess capacity-demand imbalances due to convective weather at pre-tactical and tactical level.

To illustrate the benefits of the methodology, a prototype has been built on two ATFM EUROCONTROL simulators (INNOVE and PLANTA) to run human-in-the-loop evaluation exercises. Results show that decisions can be better anticipated at D-1 because predictions on capacity reduction caused by convective weather exhibit reasonable stability over time. Besides, decisions on colour code on risk representation and dashboard configuration are helpful in flow manager tasks.

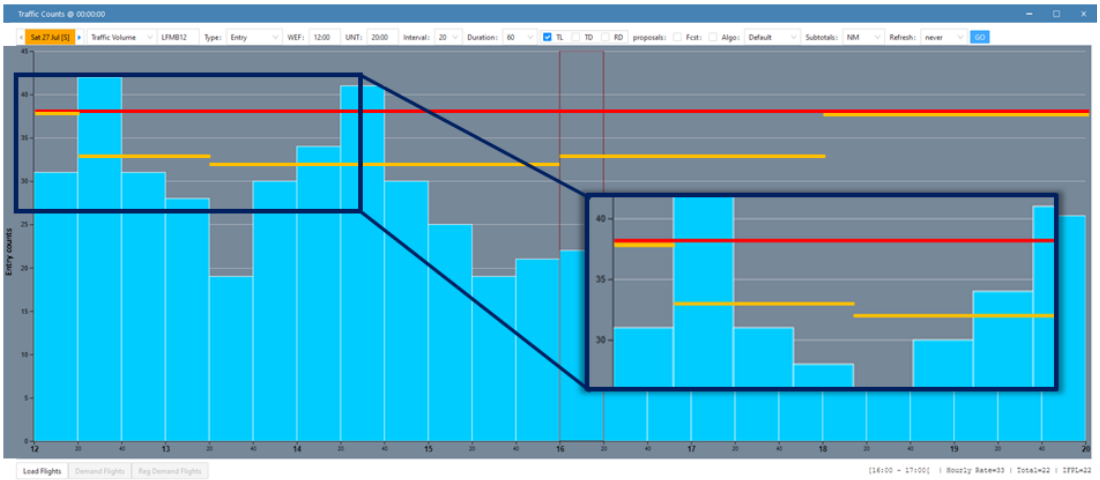

Figure 11: Snapshot of the configuration dashboard – Entry counts along with the Monitoring Value Threshold (red) and the Weather Capacity Reduction (orange)

To sum up, the operational and performance improvements for FMPs and NM in pre-tactical and tactical ATFM are:

- Ability to understand the weather convective prediction and to identify critical weather areas.
- Ability to understand and manage the capacity reduction prediction.
- Ability to manage the situation at the network level.
- Ability to improve the efficiency and productivity of human performance (workload, usability, . . .).

Future research steps will focus on three main paths. First, AI-Based weather forecasting should present a unified result from a single weather product or a combination of them. At the same time, the tool should be able to self-evaluate the level of accuracy of the forecast and show it to the user. Second, a study on the mitigation solutions devoted to the workflow for the enhancement of ATFCM process should be done for a better design of the strategy. Finally, the characterization of demand and capacity should incorporate filtering options and enhancement in the visualization containing the impacted TV with the prediction of capacity reduction value.

ACKNOWLEDGMENT

ISOBAR project has received funding from the SESAR Joint Undertaking (JU) under grant agreement No 891965. The SESAR JU receives support from the European Union's Horizon 2020 research and innovation programme and the SESAR JU members other than the Union.

REFERENCES

[1] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S. D. Langhans, M. Tegmark, and F. Fuso Nerini, "The role of artificial intelligence in achieving the sustainable development goals," *Nature communications*, vol. 11, no. 1, p. 233, 2020.

[2] R. Aggarwal and R. Kumar, "A comprehensive review of numerical weather prediction models," *International Journal of Computer Applications*, vol. 74, no. 18, 2013.

[3] S. L. Brunton, J. Nathan Kutz, K. Manohar, A. Y. Aravkin, K. Morgansen, J. Klemisch, N. Goebel, J. Buttrick, J. Poskin, A. W. Blom-Schieber, *et al.*, "Data-driven aerospace engineering: reframing the industry with machine learning," *AIAA Journal*, vol. 59, no. 8, pp. 2820–2847, 2021.

[4] D. A. Rhoda and M. Pawlak, "An assessment of thunderstorm penetrations and deviations by commercial aircraft in the terminal area," *Massachusetts Institute of Technology, Lincoln Laboratory, Project Report NASA/A-2*, vol. 3, 1999.

[5] D. Rhoda, E. Kocab, and M. Pawlak, "Aircraft encounters with convective weather in en route vs. terminal airspace above memphis, tn," in *10 Conference on Aviation, Range, and Aerospace Meteorology*, 2002.

[6] R. DeLaura, M. Robinson, M. Pawlak, and J. Evans, "Modeling convective weather avoidance in enroute airspace," in *13th Conference on Aviation, Range, and Aerospace Meteorology, AMS, New Orleans, LA*, Citeseer, 2008.

[7] R. DeLaura, "Route availability planning tool," tech. rep., Massachusetts Institute Of Technology Lexington Lincoln Lab, 2012.

[8] EUROCONTROL, *Operational Instruction. Network Cross Border Weather Collaboration Procedure*, 2022.

[9] L. Song, C. Wanke, D. Greenbaum, and D. Callner, "Predicting sector capacity under severe weather impact for traffic flow management," in *7th AIAA Aviation Technology, Integration and Operation Conference (ATIO)*, 2007.

[10] C. Russell, "Predicting airspace capacity impacts using the consolidated storm prediction for aviation," in *10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, 2010.

[11] J. D. Welch, J. Y. Cho, N. K. Underhill, and R. A. DeLaura, "Sector workload model for benefits analysis and convective weather capacity prediction," in *Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013), Chicago, IL*, 2013.

[12] A. Jardines, M. Soler, and J. García-Heras, "Estimating entry counts and atfm regulations during adverse weather conditions using machine learning," *Journal of Air Transport Management*, vol. 95, 2021.

[13] S. Mas-Pujol, E. Salamí, and E. Pastor, "Predict atfcm weather regulations using a time-distributed recurrent neural network," in *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)*, pp. 1–8, 2021.

[14] J.-G. Lee, K.-H. Min, H. Park, Y. Kim, C.-Y. Chung, and E.-C. Chang, "Improvement of the rapid-development thunderstorm (rdt) algorithm for use with the gk2a satellite," *Asia-Pacific Journal of Atmospheric Sciences*, vol. 56, no. 2, pp. 307–319, 2020.

[15] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *30th Advances in Neural Information Processing Systems (NIPS)*, vol. 30, pp. 3146–3154, 2017.

[16] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable ai for trees," *Nature machine intelligence*, vol. 2, pp. 56–67, 2020.

[17] ISOBAR Consortium, *Report on ISOBAR Evaluation and roadmap for ISOBAR B2B service. Ed. 02.00*, 2022.

[18] SESAR Joint Undertaking, *ATM Master Plan 2022 - Progress report*, 2022.