*Proceedings of the 2019 Winter Simulation Conference*
*N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.*

# CLASSIFICATION OF WEATHER IMPACTS ON AIRPORT OPERATIONS

Michael Schultz

Institute of Logistics and Aviation
Dresden University of Technology
Hettnerstr. 1-3
Dresden, 01069, GERMANY

Stefan Reitmann

Institute of Flight Guidance
German Aerospace Center
Lilienthalplatz 7
Braunschweig, 38108, GERMANY

Sameer Alam

School of Mechanical & Aerospace Engineering
Air Traffic Management Research Institute
Nanyang Technological University
639798, SINGAPORE

## ABSTRACT

Classification of weather impacts on airport operations will allow efficient consideration of local weather events in network-wide analysis. We use machine learning approaches in our contribution to correlate weather data from meteorological reports and airport performance data, which contains flight plan data with scheduled, actual movements and delays. In particular, we applied unsupervised learning to cluster performance impacts at the airport and classify the respective weather data with recurrent and convolutional neural networks. Thus, our machine learning approach allows for both an appropriate matching of decreased airport performance during the occurrence of local weather events as well as a delay estimation based on weather forecast and flight plans. This paper serves to illustrate the potential of classifications with machine learning methods and is the basis for further investigations on this topic. We update current expert-based weather classifications and provide a better understanding of local and network-wide interdependencies in air transportation.

## 1 INTRODUCTION

Weather has a significant impact on airport operations and the performance of the whole aviation network. Delayed operations may caused by airport capacity constraints due to severe weather conditions. The prediction of aircraft processes along their whole trajectories is required to achieve punctual operations. Uncertainties during the airborne phase of flights represent only a minor impact on the overall punctuality. In the current operational environment, ground tasks gain more relevance. The focus on ground operations will allow the different stakeholders to define and maintain a comprehensive 4D aircraft trajectory over the day of operations. Using a reliable and predictable departure time is one of the main tasks of the ground activities. Mutual interdependencies between airports, as departing delays propagate thought the network, result in system-wide far reaching effects. In 2016, reactionary delays continued to be the main delay cause, followed by turn around delays, accounting for 46% of departure delays (Eurocontrol 2017).

Flight deviations are important for the air traffic management and induced by weather and traffic situations as well as controller actions (e.g. directs (Bongiorno et al. 2017)). Typical average standard deviations for airborne flights are 30 s at 20 min before arrival (Bronsvoort et al. 2009), but could increase to 15 min when the aircraft is still on the ground (Mueller and Chatterji 2002). The average time variability (measured as standard deviation) during the flight phase (5.3 min) is higher than in the taxi-out (3.8 min) and in the taxi-in (2.0 min) phases, but it is still significantly lower than the variability of both the departure (16.6 min) and arrival (18.6 min) phases (Eurocontrol 2017). The changes experienced during the gate-to-gate phase are comparatively small, leading to a translation of departure variability into arrival one (Tielrooij et al. 2015). Thus, the arrival punctuality is driven by the departure punctuality and all stakeholders play a significant role on the system-wide punctuality performance. Weather related delays are reported as the second most common cause of en-route air traffic flow management (ATFM) delays (18%) (Eurocontrol 2017). For airports, the closer they operate to their maximum capacity, the more severe is the impact of a capacity loss due to external events such as weather (cf. Cook et al. (2010), Sohoni and Erat (2015)).

Current research in the field of flight and airport operations addresses economic, operational and ecological efficiency (Rosenow et al. 2018; Gerdes et al. 2018; Standfuß et al. 2018; Rosenow et al. 2017; Niklaß et al. 2017; Santos et al. 2017; Gerdes et al. 2016; Kaiser et al. 2012; Carlier et al. 2007). The propagation of delay in the network is paramount when assessing the impact of congestion (Campanelli et al. 2016; Ivanov et al. 2017). This is particularly critical when estimating the resilience of the Air Traffic Management (ATM) system and the impact of different mechanisms on the expected performances' variations (Cook et al. 2016; Proag and Proag 2014; Cook et al. 2009). Dynamic traffic situations emerge from traffic flow patterns across Europe and to/from intercontinental flows, military operations (Islami et al. 2017), volcanic ash eruptions (Luchkova et al. 2015), zones of convective weather (Kreuz et al. 2016), prevention of contrails (Rosenow et al. 2017), consideration of commercial space operations (Kaltenhäuser et al. 2017) and integration of new entrants (Sunil et al. 2015). Current research also address passengers metrics to evaluate flight performance (Montlaur and Delgado 2017), which can be particularly relevant when optimizing arrival flows at airports under uncertainty (Delgado and Prats 2014; Buxi and Hansen 2013). Thus, delay generation due to weather impacts including location and time of the primary delay generation and its evolution are relevant to capture the complexity of the system dynamics.

With a focus on airport operations, the weather phenomena could be categorized by the ATM Airport Performance (ATMAP) weather algorithm (Eurocontrol 2011) provided by the Eurocontrols Performance Review Unit (PRU), which aims to quantify the weather conditions at European airports (measure of the intensity and duration of weather phenomena). Thus, a group of experts identifies relevant aviation weather factors and considers that these factors are additionally coupled with the availability of local airport technologies (such as precision approaches in poor visibility conditions) and aircraft characteristics (such as defined tolerances for crosswind and tailwind). Furthermore, the ATMAP algorithm weight the different weather factors, that similar ATMAP weather scores will result in comparable impacts on airport operations, although they are based on different weather events (such as high wind speeds or low visibility conditions). The ATMAP algorithm considers five weather classes (ceiling and visibility, wind, precipitations, freezing conditions, dangerous phenomena) and also considers different degrees of severity per weather class. In Figure 1 the daily ATMAP weather score (diamond) is displayed against the airport performance at Frankfurt airport using on-time performance (delay $<$ 15 min) and cancellations. Since airports are located at different places, the specific characteristics and impacts of classified weather events are different.

The following definitions are used in the ATMAP algorithm: *weather phenomenon* is a single meteorological element which impacts the safety of aircraft during air and ground operations; *weather class* is a group of one or more weather phenomena affecting the airport performance; *severity code* is a ranking number of the weather class status (from best to worst); *coefficient* represents the assignment of a score to a given severity code in order to describe the nonlinear behavior of various weather phenomena. The PRU proposes a multi-step procedure to determine the ATMAP weather score: in a first step, a given METAR (Meteorological Aviation Routine Weather Report) observation at the airport will be assessed by specifying the severity code and its associated coefficient for each weather class. This METAR message is parsed, filtered, and transformed to a quantified measures (coefficients). In a second step, these weather class

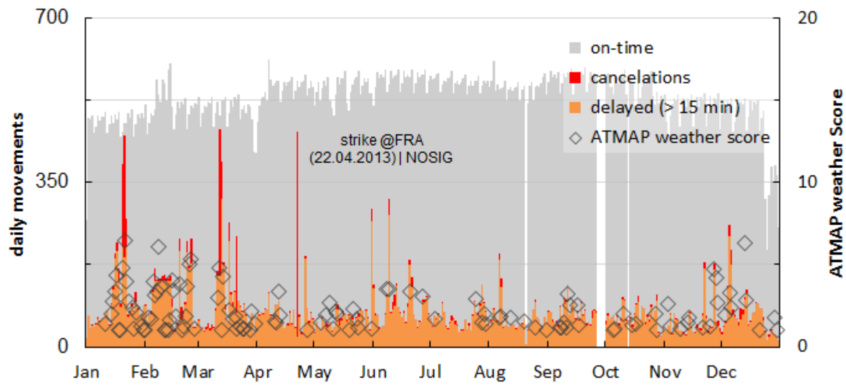


Figure 1: Airport performance data and ATMAP score at Frankfurt Airport (2013) (Schultz et al. 2018).

coefficients are summed up to the corresponding ATMAP weather score (per METAR message). Finally, for a given time interval (hours of operations), the sum of all ATMAP weather scores are divided by the number of METAR observations to calculate an average ATMAP weather score per time interval (e.g., per hour, per day).

With our contribution we provide a machine learning approach to quantify the impact of local (severe) weather conditions to the airport performance. Therefore we use local weather data (METAR) and airport performance data (scheduled and actual flight plan) to correlate the complex dependencies in the airport systems. We categorize the airport performance and backtrack/ evaluate possible causes from the observed weather phenomenon. Finally, we will show a first approach to map individual weather phenomenon to airport performance impacts, which will be a basis for a new method to overcome the limitations of the current ATMAP algorithm (based on expert judgments).

## 2 WEATHER AND AIRPORT PERFORMANCE

Our dataset the analysis consists of flight plans and weather data of major European Airports with more than 60 million flights from the years 2014 and 2015. These data sets include scheduled, actual times of specific aircraft movements and air traffic relevant weather data (airport specific METAR data). We used a subset from this comprehensive dataset with a focus on London Gatwick airport (EGKK), as one of the most busiest single runway airport. Due to the high dense runway operations at EGKK with nearly no operational buffers, we expect that weather phenomena will have a direct impact to the airport operations.

### 2.1 Weather Data

Current weather conditions are usually recorded at each airport in the form of METARs (Meteorological Aviation Routine Weather Report (Federal Aviation Administration 2016)). METARs are reported in combination with a Terminal Area/Aerodrome Forecast (TAF). While TAF provides forecast values (cover a period ranging from 3 hours to 30 hours), METAR data are measured values and accepted as valid for the next 30 min. The unscheduled special weather report (SPECI) is another format representing significant changes in airport weather conditions. The time of update and the update interval of a METAR weather report are not harmonized and implemented differently worldwide. In addition to information about the location, the day of the month and the UTC-time ('EGKK 190850Z'), the METAR contains information about wind, visibility, precipitation, clouding, temperature, and pressure that are relevant for the air traffic, especially for the airport operations.

An exemplary weather information derived from a METAR dataset (average values per day) is shown in Figure 2 and exhibits: temperature, dew point, wind direction and speed, humidity, and pressure. Besides this general weather information, some additional measurements were available related to adverse weather situations, such as information about wind gusts, runway conditions (e.g., ice layer) and thunderstorm related clouds, as well as calculated values of the Runway Visual Range (RVR). The use of METAR weather records for data analysis demands for a detailed analysis, since specific characteristics exist and the data integrity is not assured by the data provider. Typically, data lacks (partial) loss of significant information, such as wind data, dew-point data, or runway condition information (e.g., depth of deposit), variable units of measure, or incomplete information about airport runway conditions. To allow for an appropriate analysis of the weather phenomena, the METAR is decoded step-wise. The information has to be parsed, filtered and transformed to a usable measure in the context of the comparison to the airport performance.

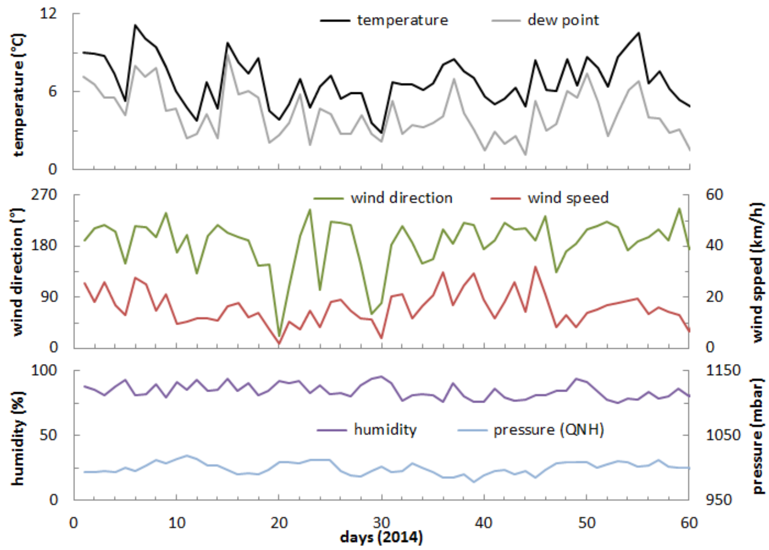


Figure 2: Weather data from the first 60 days in 2014 at Gatwick airport.

## 2.2 Standard ATMAP Approach

The current ATMAP algorithm quantifies and aggregate major weather conditions at airports, which have significant impact on the airport operations. Five different weather classes with a significant influence on aircraft and airport operations are included: (1) ceiling and visibility; (2) wind; (3) precipitation; (4) freezing conditions; and (5) dangerous phenomena. These five different weather classes are linked to the associated maximum coefficient defined by the ATMAP algorithm. Compared to the other weather classes (coefficient max 3-5), dangerous phenomenon have a high particular impact on airport operations which results in the highest coefficients. For both cumulonimbus (CB) and towering cumulus clouds (TCU), the ATMAP coefficients are ranging from 3 to 10 (TCU) or from 4 to 12 (CB) depending on the cloud coverage. Showery precipitation and intensive precipitation can lead to a further increase of the coefficient values up to 18 or 24 for TCU as well as CB. Other dangerous phenomena with impact on the safety of aircraft operations can be divided into three groups: 30 points (heavy thunderstorm), 24 points (e.g., sandstorm, volcanic ash), and 18 points (small hail and/or snow pellets). According to the time period used at Figure 2 (first 60 days of 2014, daily average), the ATMAP algorithm could quantify these measurements into a ATMAP weather score. The aggregated scores for the particularly observed weather classes indicate the severity of a weather phenomenon with an increasing value (see Figure 3).

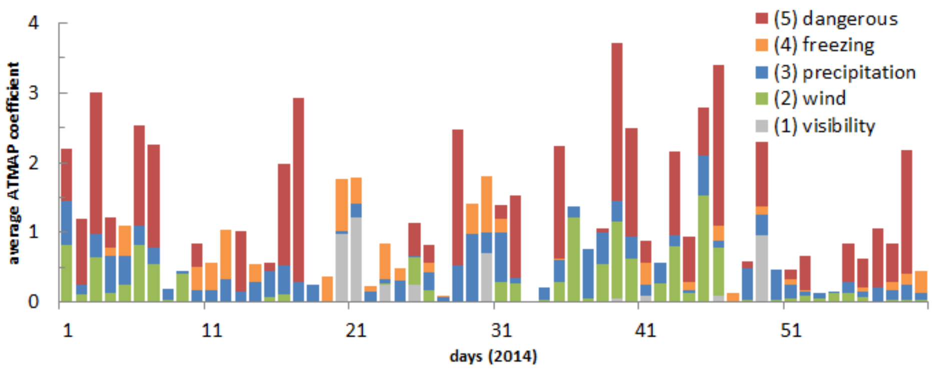


Figure 3: Corresponding ATMAP weather score (expert judgment) for weather data from the first 60 days in 2014 at Gatwick airport.

### 2.3 Airport Performance

The performance of an airport is mainly related to the number of aircraft movements handled (airport capacity). In this case, the term capacity generally refers to the ability of a given transportation facility to accommodate a traffic volume (e.g., movements) in a given time period (e.g., on hourly, daily, or yearly basis). If the air traffic demand approaches or exceeds the given airport capacity, the congestion of provided infrastructure increases which results in delays and cancellations. This demandcapacity imbalance is a key cause of unpunctual operations and affects different components of the whole airport system on airside (e.g., runways, taxiways, aprons) and landside (e.g., passenger handling (O'Flynn 2016; Schultz 2010; Schultz 2018; Ali et al. 2019)). Results of a data analysis from Frankfurt airport show that more than 45% of the variability in daily punctuality are related to local weather impacts (Röhner 2009).

Flight delays are defined as the difference between the scheduled and actual times of arrivals and departures. Reference points for flights are usually their on- and off-block times at gate or apron positions but sometimes are defined by the time when the aircraft is passing the runway threshold during start or landing. In this context, punctuality of an airport/airline is determined as the proportion of flights delayed less than 15 min according to all flights occurred. Time buffers are often strategic elements to improve punctuality and mitigate delay costs (Eurocontrol 2017; Cook et al. 2016). The definition of delay can vary according to the stakeholder so that a lot of terms and definitions have been established in the aviation domain, such as acceptable delay, on-time performance, reactionary delays, or passenger delay minutes.

### 2.4 Flight Plan and Weather Data

A higher detailed picture of airport operations and their weather dependencies will be arise, if airport performance and flight plan data (schedule, actual arrival/ departure times, delays) are combined with the aggregated weather data. Figure 4 exhibits the rapid delay increase to 795 minutes at the heavily utilized Gatwick (accumulated delay minutes from all flights in a one hour period) at the beginning of the day of operations due to a two hour period of fog (associated with an ATMAP weather score of 5).

## 3 MACHINE LEARNING APPROACH

Our approach differs significantly from the basic considerations made at the ATMAP algorithm (Eurocontrol 2011), where degradation of airport performance is linked to five major weather classes. One example is wind categorization: coefficient 1 stands for wind speed small than 15 kt, coefficient 4 for greater than 30 kt. These categories are taken as generally valid for all European airports and are not linked to specific airports or regions. We take a critical view of this, since the location of an airport and the local meteorological conditions will significantly impact its performance. These points are addressed with our

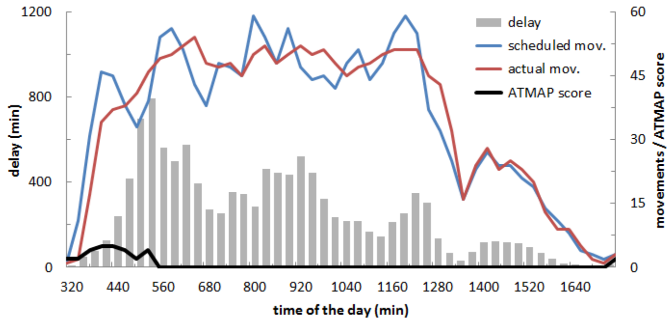


Figure 4: Airport performance data and ATMAP weather score at Gatwick airport (EGKK).

proposed machine learning approach, where we focus on two core features in the model development: (1) the model must be impact-based (i.e. we link effects to their causes), (2) the model must be adaptive (i.e. enable an airport-specific assessment). In order to evaluate the performance of an airport, it is essential to use the performance (delay in scheduled operations) itself as a benchmark. Therefore, weather categorization are converted into an *inverse problem* - deduce from an output (airport performance) its causes (weather). The method proposed includes the following five steps to cope with non-linear time series and interdependencies: (1) *data preparation* of flight schedules/ weather data, (2) *clustering*, class creation of impact data, (3) *model creation*, parameterization and setup, (4) *model training*, application of model to data, and (5) *evaluation*, error measurement.

## 3.1 Classification with Neural Networks

It is essential to distinguish between classification and regression. Classification is about predicting a label and regression is about predicting a quantity. Classification predictive modeling is the task of approximating a mapping function $f$ from input variables $X$ to discrete output variables $Y$. The output variables are often called *labels* or *categories*. The mapping function predicts the class or category for a given observation. Classification problems can be solved by a variety of methods within and outside machine learning. They all have advantages and disadvantages and their applicability depends on the particular application. Examples are Logistic Regression, k-Nearest Neighbors and Support Vector Machines (SVM). However, neural networks also offer the possibility to classify non-linear, multivariate data. Thus it is possible to build an adaptive decision support that delivers complex but fast outputs to specific input sets. This is the reason why we focus on neural networks in our application. There are two main approaches to neural networks that are suitable for time series classification and that have proven successful several applications: (1) Convolutional Neural Network Models (CNN) and (2) Recurrent Neural Network Models (RNN), especially their sub-type Long Short-Term Memory (LSTM). RNNs and LSTMs are recommended to detect short-term correlations that have a natural order, while CNN is better able at deriving long-term repeated interdependencies. The reason for this is that RNN could take advantage of the time correlation between measurements, and CNN is better at learning deep traits contained in recursive patterns (Wang et al. 2018). The benefit of using LSTMs for sequence classification is that they can learn from the raw time series data directly (see Schultz and Reitmann (2019)).

## 3.2 Impact Clustering

In order to enable the process of classification, the target data streams must be labeled (see Figure 5). This label creation on the effects (not causes) can be done algorithmically or added manually based-on expert knowledge. The labels of the target variables should represent effect categories, which reflects the severity of the respective effect on the airport performance. In this context, Eurocontrol 2007 labeled a delay between $-15$ min to 5 min with '*on-time*', with no significant influence on airport performance.

Further categories can be found to capture the severity of a growing delay, where the specific airport and demand/capacity conditions have to be taken into account as well. In particular, this form of label creation requires local expertise. The exclusive consideration of delay is call a 1D target label, where the creation of intervals is a recommended technique (see Section 4). For the algorithmic, multidimensional label creation different methods can be used, which allow a categorization without expert knowledge. In the field of machine learning, this is referred to *unsupervised learning*. Typically, unsupervised algorithms make inferences from datasets using only input vectors without referring to known (or labeled) outputs. A basic method from this area is *k-Mean clustering*. This method searches for a defined number of $k$ groups in a dataset which are similar to each other and takes into account underlying patterns. The k-Mean algorithm starts with a first set of randomly selected centroids used as a starting point for each cluster, and then performs iterative calculations to optimize the centroid positions (Hartigan and Wong 1979).

# 4 APPLICATION

The neural network serves as an adaptive intermediate model and processes weather and airport performance data (ATMAP 2.0 approach). The selection of these data can be done algorithmically as well as by expert knowledge. The same applies to the classification of the impact (performance impact at the airport). The neural network itself is determined by its parameters, its structure and the range of data available (see Figure 5). We implemented the given neural networks in *Python 3.6.5* using the open-source deep learning library *Keras 2.2.4* (frontend) with the open-source framework *TensorFlow 1.12.0* (backend) and *Scipy 1.0.0* (routines for numerical integration and optimization). Training and testing were performed on GPU (NVIDIA Geforce 980 TI) using CUDA as parallel computing platform and application programming interface. Similar experiments have also been carried out for regression, in which LSTM models were used to predict delays (cf. Reitmann and Schultz (2018)).

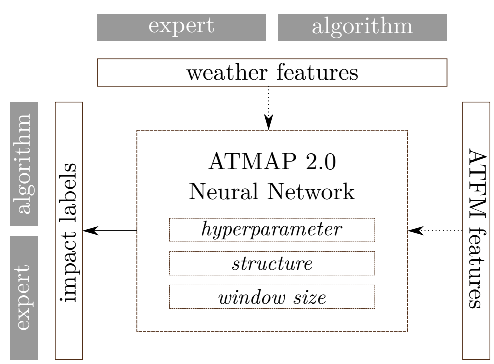


Figure 5: Classification of time-discrete data streams for the ATMAP 2.0 approach.

## 4.1 Data Preparation

The raw data must be prepared for the use of machine learning, which includes the selection of features for input and output and the classification of output. Similar to clustering, the choice of features also offers the possibility of solving this algorithmically or of drawing on expert knowledge. The feature selection for the following applications is made with expert knowledge. Features are all numerically accessible factors of METAR dataset (see Input A in Table 1). The airport performance is decisively determined by the relationship between demand and capacity, where capacity significantly influenced by weather events. An imbalance between the demand (scheduled movements) and capacity (actual movements) results in delays, which are added as a supplementary airport performance indicator.

Both delay and the deviation of number of scheduled flights ($n_{flights\_scheduled}$) from number of actual flights ($n_{flights\_actual}$) can be used as outputs. Since a uniform database is indispensable for the learning

Table 1: Input feature selection of weather and airport performance data.

| Input | Features |
|---|---|
| Input A | wind speed, visibility, temperature, humidity, pressure, wind direction, heat index, wind gust speed, number of actual aircraft movements (arrivals and departures) |
| Input B | Input A + number of scheduled aircraft movements (arrivals and departures) |

process and METAR data is provided in 30 min intervals the output values need to refer to these intervals. Thus, either the METAR data is mapped to single flight events or the flight data is aggregated to the 30 min interval of the weather data. For the following analysis we see the constant 30 min time slots as an advantage for the learning process and implement the second approach. As aggregated value, the average, absolute delay (difference between scheduled and actual time at the gate) is used. Additionally, the deviation in numbers of flights ($\Delta flights = n_{flights\_scheduled} - n_{flights\_actual}$) is calculated per 30 min slot. The outputs are either mapped to intervals for the delay (1D case) or clustered by delay and Δflights (2D case). For both cases we create $k = 5$ impact labels. The result of the k-Mean clustering is shown in Figure 6. The combination of two indicators is intended to reflect the effect of meteorological conditions on the dynamics of an airport in greater detail. The limits of the intervals for 1D are inspired by Eurocontrol (2007).

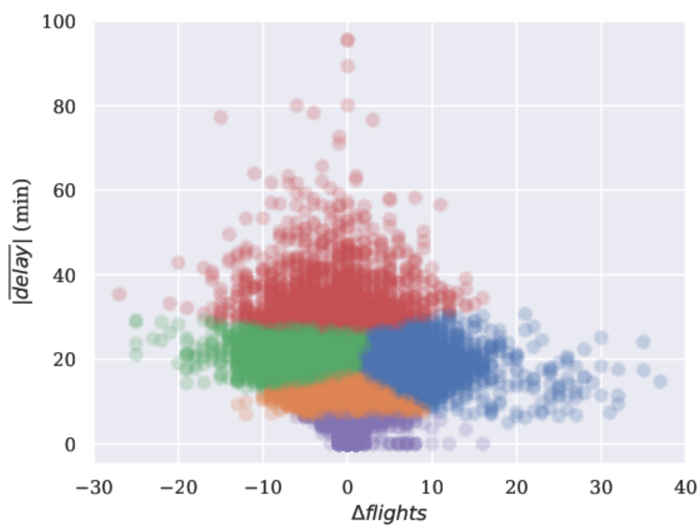


Figure 6: k-Mean-clustered 2D-data set with $k = 5$ clusters.

### 4.2 Neural Network Setup

The result of a learning process of neural networks depends essentially on their structure and parameterization. A certain set of parameters determines this process (see Table 2). For the following experiments, parameters should remain uniform across applications, others should be changed. The optimizer determines the learning rate ($\eta = 10^{-3}$) of the network, the number of epochs the repetitions. This increases to a total of 300, since it is a stochastic process that is performed 15 times each for evaluation. The number of layers determined the complexity of the network, the window size includes the inclusion of past time steps in the calculation. A higher value would have no relation to reality with 48 half-hour sequences per day. A total of 24,733 sequences are available, of which 66% are used for training and 34% for testing.

Table 2: Hyperparameter setup of the networks.

| Parameter | Config I | Config I | Config III |
|---|---|---|---|
| Individual | $n_{layer}$ = 100<br>window size = 50 | $n_{layer}$ = 100<br>window size = 6 | $n_{layer}$ = 10<br>window size = 6 |
| Constant | optimizer = Adam, $n_{epochs}$ = 20, batch size = 64 | | |

### 4.3 Model Fitting and Evaluation

As already described, each application is executed 15 times. The reason for this is that neural networks are stochastic, which means that a different specific model is created when training the same model configuration with the same data. This is an advantage of the network, because it gives the model its adaptability, but requires a more complex assessment of it. Table 3 shows the final accuracy after 15 applications of the respective model - for a better overview the corresponding $\sigma$ have been omitted. These values include the correctly mapped input-output ([weather,demand]-delay) combinations and the quality of the trained net.

Table 3: Accuracy of LSTM and CNN with different data sets and hyperparameters.

| | | **Config I** | | **Config II** | | **Config III** | |
|---|---|---|---|---|---|---|---|
| | | *1D* | *2D* | *1D* | *2D* | *1D* | *2D* |
| **LSTM** | Input A | 82.1% | **89,6%** | 72.5% | 66.0% | 70.5% | 72.0% |
| | Input B | 64.3% | 55.7% | 63.8% | 52.6% | 42.8% | 47.9% |
| **CNN** | Input A | 90.1% | **96.9%** | 74.1% | 68.8% | 72.1% | 62.8% |
| | Input B | 36.3% | 61.3% | 42.8% | 37.7% | 42.8% | 43.0% |

The best results are achieved uniformly with a 2D target label using the most elaborate learning configuration. In both cases, this refers to Input A, considering weather phenomena and demand. The computing times dropped considerably from configuration I to III. It should also be noted that the use of hybrid paradigms (e.g. CNN-LSTM, ConvLSTM) should be focused in future analyses.

### 4.4 Model Summary

The results in Table 3 indicate that there are some satisfactory solutions provided by the LSTM and CNN considering different prerequisites, data and configurations. Since this is the first time, to our knowledge, that machine learning is used to derive a model to predict local airport performance from weather data. At this stage, it is still challenging to evaluate the impact of individual meteorological components on the overall delay, as in the initial ATMAP approach (Eurocontrol 2011). Consequently, we see the observed delay as a combinatorial product of several dependent inputs. While the given traffic demand from the flight plan (scheduled arrivals and departures) offers a wide forecast horizon, the weather forecast is limited by the time range of the Terminal Area/Aerodrome Forecast (e.g. 6-9 hours weather forecast with an update of every 3 hours). A short, representative example is given in the following (Table 4 and Figure 7) and refers to flight plan data at EGKK on $3^{rd}$ September 2014. The values of Table 4 and Figure 7 represent the delay labels (cluster numbers) of the k-Mean clustering.

In Table 4 the actual weather label is shown in comparison to the prediction of our LSTM and CNN approach. The prediction quality is nearly 100%, only the LSTM provides an underestimated delay situation at the airport at timeslot $t+4$. The advantage of neural networks, especially recurrent paradigms,

Table 4: Label prediction for $3^{rd}$ September 2014, 30 min timeslots.

| | timeslot | | | | | |
|---|---|---|---|---|---|---|
| | $t+1$ | $t+2$ | $t+3$ | $t+4$ | $t+5$ | $t+6$ |
| | label | | | | | |
| **EGKK** | 4 | 4 | 1 | 2 | 2 | 2 |
| **LSTM** | 4 | 4 | 1 | ***1*** | 2 | 2 |
| **CNN** | 4 | 4 | 1 | 2 | 2 | 2 |

is the integration of parallel or past knowledge. The LSTM and CNN approaches were used in their best configuration (for input A, configuration I, 2D). The start pulse thus comprises a sequence of 50 time steps of preceding time slots. The data refer to the presented example day from Figure 4, but are already labelled and assigned to the slots of the weather data. Figure 7 depicts the underlying dataset.


Figure 7: $3^{rd}$ September 2014, labelled and slotted.

The application shown is only intended as an example of a possible use of the model. In further investigation we want to find out, how a decision support for the user can be derived from the trained knowledge. An example would be an optimal adaptation of the demand to meteorological conditions to minimise the overall delay. Furthermore, a complex quantification of the individual components of METAR makes sense, which, however, should be addressed separately due to complex internal interactions of the weather components.

## 5 CONCLUSION

We investigated a quantification of the influence of meteorological conditions on the individual airport performance at London Gatwick airport (EGKK) using machine learning. Two different paradigms of neural networks were used and combined in order to process the corresponding data foundation in a target-oriented way. The accuracy of the trained networks was compared and finally exemplarily applied. The developments in this paper contrast with the mechanism used by the ATMAP algorithm from Eurocontrol (2011), since our approach follows a effect-to-cause relation. From the grouped performance effects at the airport (e.g. delayed operations), machine learning methods were used to draw conclusions about the underlying weather data. The grouping took place by means of unsupervised learning, the classification by supervised learning. LSTM and CNN approaches show satisfactory results and allow a weather-related decision support for future airport operations. A more detailed splitting of the data set makes sense in the future in order to gain model complexity. In particular, a separation of arrivals and departures and the associated delays makes sense in the future, as arrival delays also represent reactionary delays of the feeder airports. In this respect, a more specific splitting of the data set would also be useful. The weather will only have a significant influence on the capacity when the airport is working at its capacity limits as far as possible and can no longer serve the demand. We aim to continue our experiments in this direction. Therefore, we want to investigate further airports and airport clusters (Schultz et al. 2018) with regard to their specific weather impacts.

## REFERENCES

Ali, H., Y. Guleria, S. Alam, V. Duong, and M. Schultz. 2019. "Impact of Stochastic Delays, Turnaround Time and Connection Time on Missed Connections at Low Cost Airports". In *USA/Europe ATM R&D Seminar (13th ATM Seminar)*.

Bongiorno, C., G. Gurtner, F. Lillo, R. N. Mantegna, and S. Miccich. 2017. "Statistical Characterization of Deviations from Planned Flight Trajectories in Air Traffic Management". *Journal of Air Transport Management* 58:152–163.

Bronsvoort, J., G. McDonald, R. Porteous, and E. Gutt. 2009. "Study of Aircraft Derived Temporal Prediction Accuracy using FANS". In *Proceedings of the* $13^{th}$ *ATRS World Conference*.

Buxi, G., and M. Hansen. 2013. "Generating Day-of-Operation Probabilistic Capacity Scenarios from Weather Forecasts". Volume 33, 153–166.

Campanelli, B., P. Fleurquin, A. Arranz, I. Etxebarria, C. Ciruelos, V. M. Eguluz, and J. J. Ramasco. 2016. "Comparing the Modeling of Delay Propagation in the US and European Air Traffic Networks". *Journal of Air Transport Management* 56:12–18.

Carlier, S., I. de Lpinay, J. Hustache, and F. Jelinek. 2007. "Environmental Impact of Air Traffic Flow Management Delays". In $7^{th}$ *USA/Europe Air Traffic Management Research and Development Seminar (ATM2007)*.

Cook, A., L. Delgado, G. Tanner, and S. Cristbal. 2016. "Measuring the Cost of Resilience". *Journal of Air Transport Management* 56:38–47.

Cook, A., G. Tanner, and P. Enaud. 2010. "Quantifying Airline Delay Costs - The Balance between Strategic and Tactical Costs". In $14^{th}$ *Air Transport Research Society (ATRS) World Conference*.

Cook, A., G. Tanner, V. Williams, and G. Meise. 2009. "Dynamic Cost Indexing  Managing Airline Delay Costs". *Journal of Air Transport Management* 15(1):26–35.

Delgado, L., and X. Prats. 2014. "Operating Cost based Cruise Speed Reduction for Ground Delay Programs: Effect of Scope Length". *Transportation Research Part C: Emerging Technologies* 48:437–452.

Eurocontrol 2007. "A Matter of Time: Air Traffic Delay in Europe". Technical report.

Eurocontrol 2011. "Algorithm to Describe Weather Conditions at European Airports". Technical report, Performance Review Unit.

Eurocontrol 2017. "Performance Review Report  An Assessment of Air Traffic Management in Europe During the Calendar Year 2014, 2015, 2016". Technical report, Performance Review Commission.

Federal Aviation Administration 2016. "Advisory Circular 00-45 - Aviation Weather Services". Technical report, FAA.

Gerdes, I., A. Temme, and M. Schultz. 2016. "Dynamic Airspace Sectorization using Controller Task Load". Delft, Netherland.

Gerdes, I., A. Temme, and M. Schultz. 2018. "Dynamic Airspace Sectorisation for Flight-Centric Operations". *Transportation Resarch Part C: Emerging Technologies* 95:460–480.

Hartigan, J. A., and M. A. Wong. 1979. "Algorithm AS 136: A K-Means Clustering Algorithm". *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28(1):100–108.

Islami, A., M. Sun, S. Chaimatanan, and D. Delahaye. 2017. "Optimization of Military Missions Impact on Civilian 4D Trajectories". In *ENRI International Workshop on ATM/CNS (EIWAC 2017)*.

Ivanov, N., F. Netjasov, R. Jovanovi, S. Starita, and A. Strauss. 2017. "Air Traffic Flow Management Slot Allocation to Minimize Propagated Delay and Improve Airport Slot Adherence". *Transportation Research Part A: Policy and Practice* 95:183–197.

Kaiser, M., J. Rosenow, H. Fricke, and M. Schultz. 2012. "Tradeoff Between Optimum Altitude and Contrail Layer to Ensure Maximum Ecological En-route Performance using the Enhanced Trajectory Prediction Model". In *2nd International Conference on Application and Theory of Automation in Command and Control Systems (ATACCS)*.

Kaltenhäuser, S., F. Morlang, T. Luchkova, J. Hampe, and M. Sippel. 2017. "Facilitating Sustainable Commercial Space Transportation Through an Efficient Integration into Air Traffic Management". *New Space* 5 (4):244–256.

Kreuz, M., T. Luchkova, and M. Schultz. 2016. "Effect Of Restricted Airspace On The ATM System". In *WCTR Conference*.

Luchkova, T., R. Vujasinovic, A. Lau, and M. Schultz. 2015. "Analysis of Impacts an Eruption of Volcano Stromboli Could Have on European Air Traffic". In *USA/Europe ATM R&D Seminar (11th ATM Seminar)*.

Montlaur, A., and L. Delgado. 2017. "Flight and Passenger Delay Assignment Optimization Strategies". *Transportation Research Part C: Emerging Technologies* 81:99–117.

Mueller, E., and G. Chatterji. 2002. "Analysis of Aircraft Arrival and Departure Delay". In *Proceedings of the AIAA ATIO Conference*.

Niklaß, M., B. Lührs, V. Grewe, K. Dahlmann, T. Luchkova, F. Linke, and V. Gollnick. 2017. "Potential to Reduce the Climate Impact of Aviation by Climate Restricted Airspaces". *Transport Policy*:In Press.

O'Flynn, S. 2016. "Airport Capacity Assessment Methodology - ACAM Manual". Technical Report 1.1, Eurocontrol.

Proag, S.-L., and V. Proag. 2014. "The Cost Benefit Analysis of Providing Resilience". *Proc. Economics and Finance* 18:361–368.

Reitmann, S., and M. Schultz. 2018. "Computation of Air Traffic Flow Management Performance with Long Short-Term Memories Considering Weather Impact". *Artificial Neural Networks and Machine Learning – ICANN 2018* 11140:532–541.

Röhner, P. 2009. *Modelling of Punctuality at Frankfurt Airport*. Ph. D. thesis, Gottfried Wilhelm. Leibniz Universität Hannover.

Rosenow, J., H. Fricke, T. Luchkova, and M. Schultz. 2018. "Minimizing Contrail Formation by Rerouting around Dynamic Ice-supersaturated Regions". *Aeronautics and Aerospace Open Access Journal* 2 (3):105–111.

Rosenow, J., H. Fricke, and M. Schultz. 2017. "Air Traffic Simulation with 4D Multi-criteria Optimized Trajectories". In *Winter Simulation Conference*, edited by W. K. V. Chan, A. DAmbrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 2589–2600. IEEE.

Rosenow, J., M. Lindner, and H. Fricke. 2017. "Impact of Climate Costs on Airline Network and Trajectory Optimization: A Parametric Study". *CEAS Aeronautical Journal* 8 (2):371–384.

Santos, B. F., M. M. E. C. Wormer, T. A. O. Achola, and R. Curran. 2017. "Airline Delay Management Problem with Airport Capacity Constraints and Priority Decisions". *Journal of Air Transport Management* 63:34–44.

Schultz, M. 2010. *Entwicklung eines Individuenbasierten Modells zur Abbildung des Bewegungsverhaltens von Passagieren im Flughafenterminal*. Ph. D. thesis, Technische Universität Dresden.

Schultz, M. 2018. "Implementation and Application of a Stochastic Aircraft Boarding Model". *Transportation Research Part C: Emerging Technologies* 90:334–349.

Schultz, M., S. Lorenz, R. Schmitz, and L. Delgado. 2018. "Weather Impact on Airport Performance". *Aerospace* 5.

Schultz, M., and S. Reitmann. 2019. "Machine Learning Approach to Predict Aircraft Boarding". *Transportation Research Part C: Emerging Technologies* 98:391–408.

Sohoni, M. G., and S. Erat. 2015. "Can Time Buffers Lead to Delays? The Role of Operational Flexibility". *SSRN*.

Standfuß, T., I. Gerdes, A. Temme, and M. Schultz. 2018. "Dynamic Airspace Optimisation". *CEAS Aeronatuical* 9 (3):517–531.

Sunil, E., J. Hoekstra, J. Ellerbroek, F. Bussink, D. Nieuwenhuisen, A. Vidosavljevic, and S. Kern. 2015. "Metropolis: Relating Airspace Structure and Capacity for Extreme Traffic Densities". In *USA/Europe ATM R&D Seminar (11th ATM Seminar)*.

Tielrooij, M., M. C. Borst, M. van Paassen, and M. Mulder. 2015. "Predicting Arrival Time Uncertainty from Actual Flight Information". In 11$^{th}$ *USA/Europe Air Traffic Management Research and Development Seminar (ATM2015)*.

Wang, J., Y. Chen, S. Hao, X. Peng, and L. Hu. 2018. "Deep Learning for Sensor-based Activity Recognition: A Survey". *Pattern Recognition Letters*. arXiv: 1707.03502.

## AUTHOR BIOGRAPHIES

**MICHAEL SCHULTZ** Michael Schultz is principle investigator and Privatdozent at TU Dresden, Institute of Logistics and Aviation. He was Head of the Air Transportation Department at the German Aerospace Center (DLR, 2014-2019). He holds a habilitation degree on Aviation/Aerospace (2019) and a PhD in Aviation Engineering (2010). His research focus on data-driven (machine learning) and model-based approaches to improve air traffic management. In particular, he researches on dynamic, flow-centric management of airspaces, inter-airport coordination, performance-based airport operations, and advanced concepts for the future air (urban) traffic management. His email address is michael.schultz@tu-dresden.de.

**STEFAN REITMANN** is PhD student at the German Aerospace Center (DLR). In 2015 he received his diploma in Traffic Engineering at the Dresden University of Technology with focus on traffic flow sciences. In 2016 and 2017 he was visiting scientist at the State University St. Petersburg (SPbU). His research focus is on machine learning and big data analysis, especially for the usage of neural networks in traffic sciences. His email address is stefan.reitmann@dlr.de.

**SAMEER ALAM** is an Associate Professor and Program Director of Artificial Intelligence at the Air Traffic Management Research Institute, Nanyang Technological University, Singapore. He obtained his PhD in Computer Sc. from University of New South Wales, Australia in 2008. His research interests are in complex system modelling, machine learning and optimization algorithms applied to air traffic management problems. His email address is sameeralam@ntu.edu.sg.