Data-driven fusion of turnaround sub-processes to predict aircraft ground time

Mingchuan Luo[1], Michael Schultz[2], Hartmut Fricke[1], Bruno Desart[3]
[1] Institute of Logistics and Aviation, Dresden University of Technology, Dresden, Germany
[2] Institute of Flight Systems, Bundeswehr University Munich, Munich, Germany
[3] EUROCONTROL, Brussels, Belgium

***Abstract*—The robust air traffic network relies on safe and punctual arrivals to and departures from airports. Efficient aircraft ground operations and maintenance thus significantly contribute to stable traffic in- and outbound flows. Any improved prediction of the aircraft turnaround time can help reduce local delays and their propagation through the network. Key is forecasting the related operational states to allow for adjusted planning and delay mitigation strategies. In this paper, we target to predict incrementally the turnaround time by means of machine learning classification algorithms based on real-life data collected at the airport. A turnaround sub-processes fusion model for improving the forecast precision is developed to integrate the sequential information from the turnaround pattern, which mainly considers the duration of the various turnaround sub-processes and their overlapping conditions. Results indicate that the data-driven fusion model enhances the robustness and reliability of the aircraft turnaround time prediction. It so can efficiently support airport management. We show, that the presented methodology holds universal character, can be applied to any airport holding a significant demand/capacity ratio.**

***Keywords*—aircraft ground operations, machine learning, data-driven, turnaround sub-processes fusion model, turnaround time prediction**

I. Introduction

Efficient aircraft ground handling at airports is important to ensure performance operations in the overall air traffic network. Close cooperation between all involved stakeholders (e.g., airport operators, airlines, ground handlers, air traffic service providers) positively impacts the punctuality and predictability of the aircraft turnaround process. In the aviation industry, it is consensus that each aircraft earns revenue only when they are en-route [1]. The airlines thereby are always looking for maximum air time and minimum airport dwell time. Meanwhile, airports have the same expectation of keeping aircraft ground operations efficient, which is a win-win situation from almost all perspectives of the aviation industry.

Airport collaborative decision making (A-CDM) is a process that consists of sharing information between stakeholders of the complex airport system to provide a common situational awareness and to enable mutual strategies to solve operational challenges. It was developed in establishing operational milestones method for every joint arrival and departure aircraft activity to improve the efficiency of airports and the air traffic network [2]. By giving airport stakeholders access to the shared data from different sources, airports are able to make more accurate predictions about their operational progress in the next planning horizon [3]. Integrated management is embodied in an airport operations center (APOC), where all stakeholder operators coordinate tasks to monitor and maintain the agreed performance targets in their respective areas of responsibility [4].

The aircraft turnaround, as part of the aircraft trajectory over the day of operations, has to be part of optimization strategies for minimizing flight delays and ensuring flight connection [5]. In this context, aircraft ground handling depended on buffer time can absorb inbound delays and enhance slot adherence at airports [6]–[8]. Some previous research focused on the critical path of the aircraft turnaround and exhibited that both land- and airside processes can be bottlenecks [9]–[11]. Whenever these processes are part of the critical turnaround path, the effects could also propagate an accumulating delay through the ATM network [12]–[14]. Another approach to model the aircraft turnaround is the Resource-Constrained Project Scheduling Problem (RCPSP), which is the most typical airport ground operations and proposed as a basic tool for real-time decision support of the aircraft turnaround disruption management that provides the rescheduling possibilities [15]–[17]. However, the disadvantages of both models are that it does not capture the operational uncertainties of the concrete aircraft turnaround operations. In our research, we will propose a novel data-driven fusion model to integrate the sequential information of the aircraft ground operations, which can to some extent describe the interdependencies between the turnaround sub-processes and improve the off-block time predictability.

Investigations on turnaround reliability show significant improvement potentials in standardization, data quality and availability, process design, integrated planning, and optimization. The speed and extent, with which data is shared, have massively increased over the last years as well as the need for implementation of new methods to evaluate this data that will further guarantee the sustainability of the air traffic network, e.g., based on the Automatic Dependent Surveillance–Broadcast (ADS-B) data, a simplified A-CDM process was created for small airports in [18], and from a real airport turnaround dataset to confirm the target off-block time (TOBT) adherence, the feature importance of available data was analyzed in [19].

This article comprises five further sections. Section II provides insight into the research objective and technical back-

ground. Section III introduces the data analysis and exploration of the available turnaround data set. In Section IV, we describe the research design and methodology concretely. Section V summarizes the simulation results. Finally, Section VI gives the main conclusions and discusses potential extensions on the subject, as well as the utility of the study.

II. Background

The turnaround time refers to the sum of all activities running on its critical path, subject to ground handling activities. These activities are called turnaround sub-processes [20]. Turnaround is considered to finish when all doors of the aircraft are closed, all ground support equipment (GSE) is disconnected, the aircraft is ready to leave and the chocks are removed [21].

A. Objectives

The main objective of this research is to predict the turnaround time of an aircraft with high reliability and to make predictions as early as possible in the aircraft ground period. Actual A-CDM milestones introduced in [2] do not provide distinctive information about those turnaround sub-processes during in-block and off-block. We consequently restore the complicated aircraft ground operation processes to provide precise off-block time prediction for the APOC, representing all stakeholders in the air traffic network. Empirical turnaround sub-process data and available domain knowledge are the model inputs.

Many computer vision techniques have been used at airports to monitor and collect information about the ongoing turnarounds, which form the foundation of today's situational awareness to all airport stakeholders for real-time manual surveillance and collaborative decision making in the APOC. The available data sets describe aircraft ground operations of each turnaround. We extract the duration of all turnaround sub-processes to build the prediction model. To be able to make earlier prediction in real application, the duration information can be collected by the historical records firstly, even we can estimate them regarding the airlines, ground handlers and aircraft types, and, more concretely, by inferred from the aircraft payloads, such as the number of passengers for deplaning/boarding, amount of cargo for unloading/loading and flight distance for fueling (cf. [22]). As the turnaround process proceeds, the durations of the already finished sub-processes can be replaced easily by the actual values collected, and more actual values will gradually improve the forecast results of the aircraft turnaround time. Here the start and end time stamps of each turnaround sub-processes are not chosen regarding the prediction always relying on the end of the last sub-process the most, inversely decreasing. In this research, we will design the classification-based prediction method. Moreover, the turnaround sub-processes fusion model is developed to optimize the prediction further, which mainly considers the duration of the various turnaround sub-processes and their overlapping conditions.

B. Technical approaches and evaluation metrics

In the research, principal component analysis (PCA) is used to reduce the variables of the available data sets for the data exploration and visualization, which can be used to characterize the data in a smaller dimension and keep the maximal information from the original variables [23]. PCA is a mathematical dimensionality reduction method that uses an orthogonal transformation to transform a set of potentially linearly correlated variables into a new set of linearly uncorrelated variables, also called principal components, so that the new variables can be used to characterize the data in a smaller dimension.

To train the turnaround process stochastic, three supervised learning techniques have been adopted and implemented in Python using the machine learning library *scikit-learn* [24], which are linear regression, decision tree and random forest. Other supervised learning methods, such as support vector machine (SVM) or neural network, etc., can also realize the same functions. But they are not in our options because considering their required computational power and time consumption while not leading to more accurate predictions.

Linear regression models the relationship between one or more independent and dependent variables using the least square function called a linear regression equation [25]. A decision tree tests each attribute under a series of given conditions, shunts to different branches, and consecutively to the leaf nodes of the decision tree to get the final result, whose basic process follows the divide and conquer strategy [26]. Random forest is an ensemble algorithm, which belongs to the bagging type. By combining multiple weak classifiers, the final result is voted or averaged, so that the result of the overall model has higher accuracy and generalization performance. It can achieve good results mainly due to "random" and "forest", the first making it resistant to over-fitting, the latter leading to more accuracy [27].

In this context, accuracy expresses the ratio of the number of samples correctly predicted to the total number of samples for a given test data set. Besides, one used classic performance measure is the Root Mean Squared Error (RMSE).

In machine learning, the confusion matrix [28] is an error matrix, which is used to visually evaluate the performance of the supervised learning algorithms. Each row of the confusion matrix means the true class, while each column of it indicates the predicted class. Thereby the samples that are predicted correctly locate in its diagonal, and the distribution information of the wrong predicted samples is also expressed in detail by the confusion matrix.

III. Data analysis and exploration

Nowadays, many airports have equipped the aircraft stand monitoring system so that all the operations around the aircraft can be supervised and recorded in APOC, which includes the total ground handling activities. The actual dataset utilizing for our research is collected between a busy time span of 2019. It provides the operations data at a common European airport without the COVID-19 negative effects at a

high demand/capacity ratio. We only focus on the turnaround process, therefore, the data that presents the stand occupancy overnight is omitted. The data structure of the turnaround in this real airport will be presented firstly. In the following, the inspirations of the data analysis are shown.

A. Data Structure

The recorded aircraft ground operations data reflect well a typical turnaround process containing seven activities, which are deplaning (dep.), unloading (unl.), catering (cat.), cleaning (cle.), fueling (fue.), loading (loa.) and boarding (boa.). We define two conditions to identify "valid" turnaround records: The first one limits the aircraft turnaround time. As such, e.g. overnight stand occupancy would be omitted and possible very short turnarounds are also not considered, as they are sporadic and do not require the complete ground services. The second is that one valid turnaround process record must contain sub-process data.

The first condition can be described in (1):

$$C_1 = \begin{cases} 1 & if\ t_T \in (t_{min}, t_{max}],\ \exists\ t_T \in \mathcal{T}, \\ 0 & otherwise, \end{cases} \tag{1}$$

where t_T notes the turnaround duration and $\mathcal{T}$ is the whole data collection. Thus C_1 restricts that the duration of each valid turnaround record must belong to $(t_{min}, t_{max}]$ mins.

Furthermore, the second condition C_2 will demonstrate the validity of the turnaround data according to Eqs. (2) and (3):

$$c_{dep.} = \begin{cases} 1 & if\ \exists\ t_{dep.} \in \mathcal{T}_{dep.}, \\ 0 & otherwise, \end{cases} \tag{2}$$

$$\vdots$$

$$C_2 = c_{dep.} \vee c_{unl.} \vee c_{cat.} \vee c_{cle.} \vee c_{fue.} \vee c_{loa.} \vee c_{boa.}. \tag{3}$$

$c_{dep.}$ verifies the special sub-process deplaning in a certain turnaround process. When the operation time of deplaning $t_{dep.}$ was recorded in the deplaning data collection $\mathcal{T}_{dep.}$, we consider this sub-process valid. In (3), the existence conditions of other sub-processes $c_{unl.}$, $c_{cat.}$, $c_{cle.}$, $c_{fue.}$, $c_{loa.}$ and $c_{boa.}$ have the same definition with $c_{dep.}$. Therefore, C_2 means that one valid turnaround process data contains at least one sub-process.

The total determination function C unites these two predefined conditions in (4):

$$C = C_1 \wedge C_2. \tag{4}$$

By limiting the turnaround duration to $(20, 60]$ mins, the data set from the real airport owns 22,080 valid aircraft turnaround records. Each record also includes its corresponding domain knowledge, such as scheduled in-block and off-block time (SIBT and SOBT), estimated in-block and off-block time (EIBT and EOBT), actual in-block and off-block time (AIBT and AOBT), target off-block time (TOBT), arrival delay, flight properties (e.g., the aircraft type, the airline) and time conditions (e.g. time of the day, month).

The complete turnaround data structure is illustrated in Fig. 1, where the first bar presents the amount of the total turnaround, the rest characterizes the number of the available sub-processes, and the bandwidth between bars describes the count of the co-existed sub-processes, i.e., data for both sub-processes are recorded simultaneously in a common turnaround process.

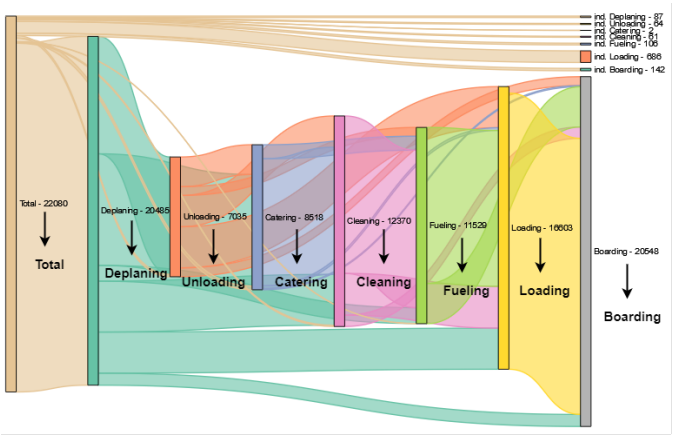

Fig. 1. Turnaround data structure.

We assume the combination of the complete turnaround sub-processes in "deplaning – unloading – catering – cleaning – fueling – loading – boarding". Concerning data deficiencies, few complete turnaround process records exist, so any sub-processes recorded in a concrete turnaround process are seen as being linked together. For example, if a turnaround process only collects data for deplaning, fueling and boarding, then the turnaround sub-processes combination is "deplaning – fueling – boarding". All connection numbers between the available sub-processes are summarized in Table I, which compensate the concrete numbers for the manifold connections. Besides, the direct connections from the "Total" bar to the "individual (ind.) Sub-process" bars in the top right corner of Fig. 1 represent the turnaround collections that only have one sub-process record. The graph of the turnaround data structure roughly shows the availability of the real airport data.

TABLE I
CONNECTION NUMBERS BETWEEN THE AVAILABLE TURNAROUND SUB-PROCESSES

Source	**Destination**					
	Unl.	Cat.	Cle.	Fue.	Loa.	Boa.
Dep.	6895	6552	2989	929	2414	706
Unl.	-	1747	1839	514	2413	522
Cat.	-	-	7467	823	159	69
Cle.	-	-	-	9250	2453	667
Fue.	-	-	-	-	9164	2365
Loa.	-	-	-	-	-	16219

B. Data Exploration

The aircraft turnaround process is, as introduced, stochastic by nature. Small disturbances can invoke significant delay,

since time windows are tightly planned. So we aim at exploring the relationship between the turnaround sub-processes and the total turnaround duration. We concentrate on the data for which all sub-processes were recorded to avoid interference and unconvincing support from the incomplete turnaround sub-processes' collections.

There are 1,143 turnaround records with full sub-processes. We separate these turnaround time values according to Table II. The whole duration interval of $(20, 60]$ min is divided into four segments of 10 minutes each, and sample number in each segment is indicated separately.

TABLE II
TURNAROUND DURATION SEGMENTS IN (20, 60] MIN

Turnaround duration segment	Interval (mins)	Number
0	(20, 30]	14
1	(30, 40]	413
2	(40, 50]	506
3	(50, 60]	210

We consider the segments of the different turnaround durations as the indicators. All data is collected from the real world, and the numbers of the four segments are extremely unbalanced. Most of the turnaround durations are within 30 min to 50 min. In the shortest duration interval from 20 min to 30 min only 14 records exist. Fig. 2 exhibits the frequency distribution histogram plot of the turnaround duration, where we find that most of the data samples of the turnaround time are around 40 min.

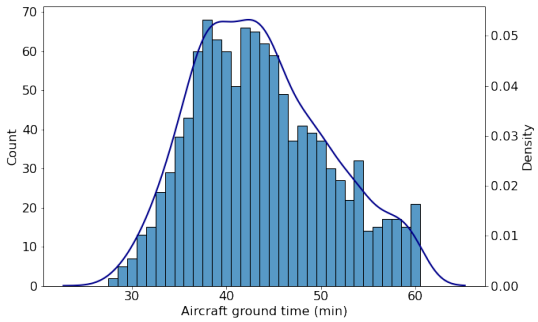

Fig. 2. Frequency distribution histogram plot of the turnaround duration.

As introduced in Section II-B, PCA is a mathematical dimensionality reduction method that uses orthogonal transformation for a set of potentially linearly correlated variables to create a new set of linearly uncorrelated variables, also known as principal components, so that the new synthetic variables can characterize the data at reduced dimensions and without concrete variable units [23]. Selecting the durations of the seven full recorded turnaround sub-processes and their domain knowledge as the variables for PCA processing, where the domain knowledge will be introduced in Section IV-A, we intend to receive the new generated variables in two dimensions. Fig. 3 shows the 2D plot of these variables, and the different turnaround duration segments 0, 1, 2 and 3 are also distinguished into four colors according to the color bar at the right in this picture. The two points lying on the top left and bottom right corner of this figure has some extreme feature values that will be regarded as outlier in the following analysis. We find that, because of the short sample number of segment 0, it's difficult to observe them in this figure and all of them are covered by other dots. However, despite the serious aggregation of the rest data, the dots of segment 1 mainly locate on the right part of the spot pile, the dots of segment 2 stand in the middle, and those of segment 3 distribute slightly to the left. Therefore, some potential rules for turnaround sub-processes and their domain knowledge might exist to affect the total ground time allocation.

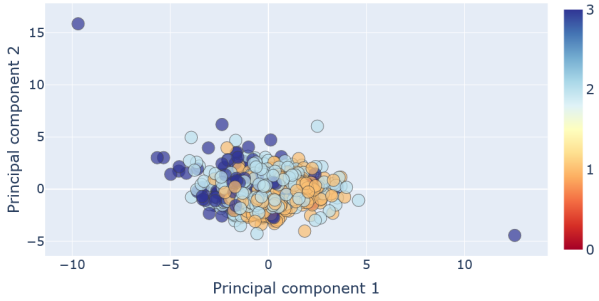

Fig. 3. Data visualization.

IV. RESEARCH DESIGN AND METHODOLOGY

The turnaround process consists of multiple ground operations that can take place independently of each other. In a real airport environment, for one certain turnaround process, the start time of each sub-process can hardly be ensured under many factors, such as the limitation of the ground handling resources, the arrival delays or the air traffic time slots, etc. However, the information about the duration of each sub-process is obtained conveniently that they can be collected directly from the similar turnaround records or estimated empirically the concrete aircraft payloads (the number of passengers for deplaning and boarding, baggage for unloading and loading, flight distance for fueling, etc.). In our research we adopt a heuristic classification-based method to forecast the turnaround time. In particular, a pioneering turnaround sub-processes fusion model is built to optimize the prediction.

A. Classification-based prediction

For improving the prediction reliability, we intend to limit the prediction range in this model, that firstly the turnaround duration is divided into several short intervals, which will be the targets of the classification methods. Besides, the turnaround sub-process durations and their corresponding domain knowledge that can be obtained immediately after the aircraft stands on its position are selected as the model

inputs, which helps the prediction model absorb the maximal accessible information.

Table II has depicted the turnaround duration intervals previously, which will be classification labels. Table III presents the variables for the classification-based prediction, which are the turnaround sub-process durations and the corresponding domain knowledge from the real airport, where the categorical data (airline and aircraft type) has been transferred to the numerical data, the scheduled occupancy is known as SOBT - SIBT, the arrival delay can be obtained by AIBT - SIBT, the estimated arrival difference is calculated by AIBT - EIBT, and the estimated occupancy means EOBT - AIBT. Furthermore, the AIBT has changed into the accumulated value in minutes counting from 0 o'clock, as known as the absolute AIBT feature, the daytime of AIBT represents the aircraft landing in the early morning, morning, afternoon, evening, and night of the day and the month information can tell the seasonal effect during the year. Afterwards, inside each turnaround time segment, we will further forecast the precise turnaround time by regression algorithms only based the sub-process durations.

TABLE III
FEATURE VARIABLES FOR THE CLASSIFICATION-BASED PREDICTION

Feature index	Feature variables
1	Deplaning duration
2	Unloading duration
3	Cleaning duration
4	Catering duration
5	Fueling duration
6	Loading duration
7	Boarding duration
8	Airline
9	Aircraft type
10	Scheduled occupancy
11	Arrival delay
12	Estimated arrival difference
13	Estimated occupancy
14	Daytime of AIBT
15	Absolute AIBT
16	Month

B. Turnaround sub-processes fusion model

As known to all, even though the aircraft ground operations in a turnaround process are stochastic and overlapping with the others, a simple but essential logic limitation is still inevitable, which is that usually the catering, cleaning and fueling go after deplaning, boarding can just start after the catering, cleaning and fueling done, and loading always occurs after unloading. In this case, any sequential indicators will benefit the prediction. The proposed turnaround sub-processes fusion model intends to determine the turnaround pattern which can imply the sequential information of the aircraft ground operations.

Let us define any two turnaround sub-processes firstly, the precedent sub-process is p_i with start timestamp s_i, end timestamp e_i and duration t_i, similarly the subsequent one p_j with s_j, e_j and t_j, which have been shown in Fig. 4. If these two sub-processes start at the same time, we always consider the one with longer duration as p_i. Then the relationship $R_{i,j}$ between these two sub-processes can be described as:

$$R_{i,j} = \begin{cases} p_i, p_j & if\ e_i \leqslant s_j, \\ p_i & if\ (s_i \leqslant s_j)\, and\, (e_i \geqslant e_j), \\ F_{i,j} & if\ (s_i < s_j)\, and\, (e_i > s_j)\, and\, (e_i < e_j), \end{cases} \tag{5}$$

where $F_{i,j}$ is the fusion function that will be discussed at details in the following. When p_i, p_j appear at the same time in (5), it describes that these two sub-processes are independent with the other, so we nominate this situation as "independence". The unique p_i shows that the longer sub-process covers the shorter one on the timeline that the remained duration can represent both turnaround sub-processes directly, which can be seen as "coverage". In Fig. 4, when two sub-processes have overlapping part $\lambda_{i,j}$, which is calculated by $e_i - s_j$, in this case the fusion function $F_{i,j}$ will determine if these two sub-processes merge or not.

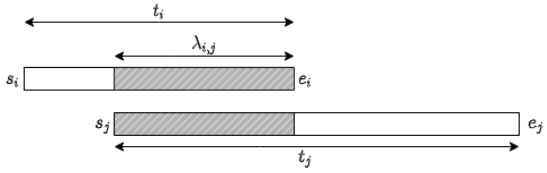

Fig. 4. General turnaround sub-processes fusion instance.

Before the definition of $F_{i,j}$, we adopt ε_i and ε_j to present the percentage of the overlapping part over the precedent turnaround sub-process and the subsequent one, which are called as "fusion factors" and defined in Eqs. (6) and (7), respectively:

$$\varepsilon_i = \lambda_{i,j} \,/\, t_i, \tag{6}$$

$$\varepsilon_j = \lambda_{i,j} \,/\, t_j. \tag{7}$$

The fusion function $F_{i,j}$ takes the form based on the magnitudes of ε_i and ε_j that are percentages. They will be divided into four ranges, each of which is 25% long. Thus we note σ_1, σ_2 and σ_3 representing 25%, 50% and 75%, respectively. Here we can increase the number of ranges for more precise division if necessary. In the initial stage, $F_{i,j}$ is defined as:

$$F_{i,j} = \begin{cases} p_i & if\ (\varepsilon_i \in (0,\, \sigma_1],\ \varepsilon_j \in (\sigma_2,\, 1))\ or \\ & \quad (\varepsilon_i \in (\sigma_1,\, \sigma_3],\ \varepsilon_j \in (\sigma_3,\, 1)), \\ p_j & if\ (\varepsilon_i \in (\sigma_2,\, \sigma_3],\ \varepsilon_j \in (0,\, \sigma_1])\ or \\ & \quad \varepsilon_i \in (\sigma_3,\, 1), \\ p_i, p_j & if\ (\varepsilon_i \in (0,\, \sigma_1],\ \varepsilon_j \in (0,\, \sigma_2])\ or \\ & \quad (\varepsilon_i \in (\sigma_1,\, \sigma_2],\ \varepsilon_j \in (0,\, \sigma_1]), \\ (p_i, p_j) & if\ (\varepsilon_i \in (\sigma_1,\, \sigma_3],\ \varepsilon_j \in (\sigma_1,\, \sigma_3]). \end{cases} \tag{8}$$

In (8), according to the ranges of the two variables ε_i and ε_j, the fusion function $F_{i,j}$ shows three forms, the "coverage" p_i or p_j, the "independence" p_i, p_j and the "fusion" (p_i, p_j),

which refine the relationship function $R_{i,j}$. In order to elaborate the determination of the fusion function $F_{i,j}$, the detailed range combination of ε_i and ε_j is summarized in Table IV.

TABLE IV
DETERMINATION OF THE FUSION FUNCTION $F_{i,j}$

ε_j \ ε_i	(0, 25%]	(25%, 50%]	(50%, 75%]	(75%, 100%)
(0, 25%]	p_i, p_j	p_i, p_j	p_j	p_j
(25%, 50%]	p_i, p_j	(p_i, p_j)	(p_i, p_j)	p_j
(50%, 75%]	p_i	(p_i, p_j)	(p_i, p_j)	p_j
(75%, 100%)	p_i	p_i	p_i	p_j

Next, we intend to integrate the processed information onto the regression prediction, which will be realized by the corresponding turnaround labelling systems. Firstly the concrete relationship between any two possible related sub-processes ("A" and "B") named as "State A-B" are described by specific labels to exhibit their sequential information. For an instantiated introduction, Table V notes the possible fusion states of deplaning and unloading, where the positive and negative label values "1", "−1", "2", "−2", "3" and "−3" are chosen to express the remained sub-process on "coverage" and the specific sub-processes sequence on "fusion" and "independence".

TABLE V
EXAMPLE OF THE FUSION STATES OF DEPLANING AND UNLOADING

State dep.-unl.	labels
dep.	1
unl.	−1
(dep., unl.)	2
(unl., dep.)	−2
dep., unl.	3
unl., dep.	−3

Secondly, we create the basic flags of all the seven sub-processes in chronological order. To distinguish with the last mentioned labelling, we use "original (o)", "fused (f)" and "concealed (c)" to indicate the individual flag of each sub-process here. When two sub-processes are independent, both flags are labelled as "o". In the "coverage" type, flag "o" can only be assigned to the remained sub-process and besides, flag "c" to the covered one. The "fusion" state will transform the flag of the precedent sub-process to "o" and the subsequent one to "f". Fig. 5 exhibits the comparison of the designed basic flags and critical path flags based on a turnaround example from the used dataset. Each bar records the corresponding sub-process start time and length. The definition of the critical path is that "It is the sequence of activities which add up to the longest overall duration. It is the shortest time possible to complete the project. Any delay of an activity on the critical path directly impacts the planned project completion date [29]." The research conducted in [19] describes the critical turnaround path in 4 lanes that were "dep. – cle. – boa.", "dep. – cat. – boa.", "dep. – fue. – boa." and "unl. – loa.". In this case, we note that the "dep. – fue. – boa." lane is the critical path and has been marked already. In addition, the designed basic flags are tagged on the right hand side of each bar according to the rules defined previously, expressing the turnaround time as closely as possible. The labelled unloading and loading fill in the gaps between sub-processes on the critical path to some extent.

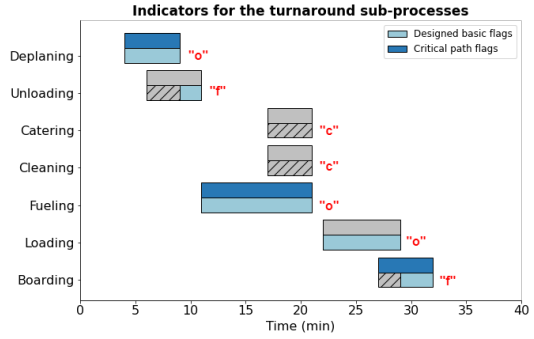

Fig. 5. Comparison of the designed basic flags and critical path flags.

In any case, each turnaround process can be labeled by its own representative sub-processes. By applying the fusion model to all sub-processes inside one aircraft turnaround, its specific turnaround pattern will be generated. Then that the durations of the turnaround sub-processes, the basic flags of the sub-processes and the labels for every possible related sub-processes pair, these three sections jointly construct the inputs to the regression model, where the categorical flags "o", "f" and "c" of the second section will be transferred to the numerical data in the prediction. The Table VI shows the developed input template. In the section of the labels for every possible related sub-processes pair, we don't consider the fusion pairs of "dep.-loa.", "dep.-boa.", "unl.-loa." and "unl.-boa." according to the reality. Normally regarding the security reasons fueling is not allowed to work with the passenger at presence, unless this operation with secured fuel pipes or with fire brigades during the deplaning or boarding. Therefore, "dep.-fue." and "fue.-boa." can still be kept in our turnaround pattern.

V. RESULTS

The classification-based prediction results of the real airport turnaround time is discussed in this section. What's more, we conduct an experiment of the turnaround fusion model after the classification to further optimize the aircraft ground time forecast. We always split 30% of the total dataset as the test data in every prediction model, therefore, the remaining 70% is the train data. To ensure the forecast robustness, we will run 100 times with different random numbers to separate the test and training datasets.

A. Classification-based prediction

Here the chosen classification methods are the decision tree and random forest. After the prediction, we notice that the

TABLE VI
INPUT TEMPLATE INVOLVING TURNAROUND PATTERN

Sec.	Inputs	Sec.	Inputs	Sec.	Inputs
1	Duration dep.	2	Flag dep.	3	State dep.-unl.
1	Duration unl.	2	Flag unl.	3	State dep.-cat.
1	Duration cat.	2	Flag cat.	3	State dep.-cle.
1	Duration cle.	2	Flag cle.	3	State dep.-fue.
1	Duration fue.	2	Flag fue.	3	State unl.-cat.
1	Duration loa.	2	Flag loa.	3	State unl.-cle.
1	Duration boa.	2	Flag boa.	3	State unl.-fue.
				3	State cat.-cle.
				3	State cat.-fue.
				3	State cat.-loa.
				3	State cat.-boa.
				3	State cle.-fue.
				3	State cle.-loa.
				3	State cle.-boa.
				3	State fue.-loa.
				3	State fue.-boa.
				3	State loa.-boa.

classification accuracy of the decision tree is 76.32%, and the random forest own a better accuracy, which is 83.22%.

Fig. 6 shows the normalized confusion matrix of the random forest, where we see the predicted classes are basically consistent with the true classes. Due to the likeness of the feature values from the turnaround intervals $(30, 40]$ min and $(40, 50]$ min, their prediction results merge slightly. However, this consequence is acceptable because the majority of the data set is located in these two intervals. Many of the data samples lay close to 40 min as shown in Fig. 2 so that the prediction model would have difficulty on classifying these data samples.

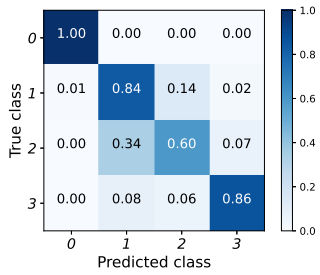

Fig. 6. Normalized confusion matrix of random forest classification.

The RMSEs of the regression predicted results inside each turnaround interval are summarized in Fig. 7. In the short turnaround interval, the RMSE value of linear regression is only 0.5 min. With the turnaround duration increasing, the RMSE values become correspondingly larger. In the interval about $(30, 40]$ min, this value is around 2.5 min. It is close to 3 min in the $(40, 50]$ min and $(50, 60]$ min turnaround interval.

B. Experiment of the turnaround sub-processes fusion model

For improving the classification-based prediction, an experiment of the turnaround sub-processes fusion model is done to validate if it can optimize the regression prediction model within the major turnaround intervals. Therefore, we choose the duration of $(30, 50]$ min with total 919 data records.

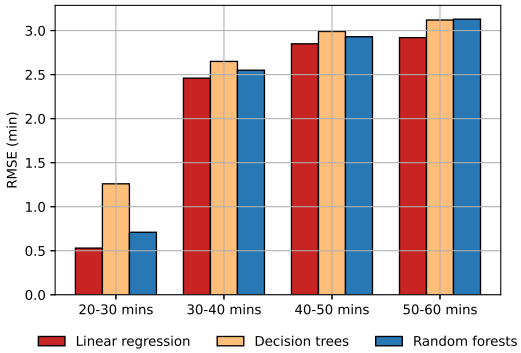

Fig. 7. Regression prediction results inside the turnaround intervals.

The prediction results with and without the turnaround pattern are illustrated in Fig. 8. After adding the turnaround pattern we discover that the linear regression method doesn't converge due to the large number of input variables, the RMSE of decision tree almost remains the same, but the value of the random forest drops the most from 4.64 min to 4.36 min. In each running time the test and training datasets are always reshuffled so that we can see the prediction as stable and convictive. Based on its ensemble learning characteristic, random forest is able to capture more information from our constructed turnaround pattern than only from the turnaround sub-process durations. Even though the prediction enhancement is not that significant, considering that all the data is derived from real-life data set and concentrates on the time point of around 40 min inside the busiest intervals, it still demonstrates that the turnaround sub-processes fusion model can further improve the performance of the regression prediction.

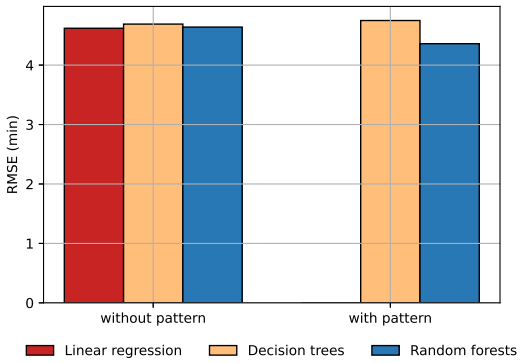

Fig. 8. Comparison of the prediction results with and without turnaround pattern.

VI. CONCLUSION

In this paper, we firstly explored the available turnaround data. Then the classification-based model was developed to predict the exact turnaround time at the A-CDM milestone of the aircraft in-block, which can provide the notable time horizon for the airport management on the operational level. More importantly, we proposed the data-driven turnaround

sub-processes fusion model that used the turnaround pattern to integrate the sequential information into the regression prediction. The advantages of this model are that we can set the initial pattern according to the particular turnaround critical path based on the empirically statistic, and then with the turnaround processing, this pattern can be updated dynamically at the end timestamp of each sub-process. With a sufficiently populated accurate dataset, the reliable turnaround pattern even can be accessed in advance that will build the robust aircraft turnaround time regression model. Because the model is data-driven, it can be easily embedded into any given airport, allowing airport stakeholders to benefit from the reliable turnaround time forecasts. Moreover, the turnaround pattern as the prediction model inputs contains all the detailed information that a sophisticated and advanced machine learning algorithm (e.g. neural network) would not over-perform. Agile and efficient machine learning algorithms are sufficient to bring accurate predictions.

Next, we intend to study the impacts of data loss of the turnaround sub-processes for the fusion model so as to guarantee the robustness and compatibility of the turnaround time prediction.

ACKNOWLEDGMENT

This publication was supported and funded by EUROCONTROL with project ABM4APOC via contract NO.19-220468-C. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of EUROCONTROL.

REFERENCES

[1] Haiqiang Wang, Min Wang, and Yang Wu. Development of an aircraft turnaround time estimation model based on discrete time simulation. In *ITITS*, pages 29–36. IOS Press, 2017.

[2] EUROCONTROL. *Airport CDM implementation manual v5.0*, 2017.

[3] Michael Schultz, Stefan Reitmann, and Sameer Alam. Predictive classification and understanding of weather impact on airport performance through machine learning. *Transp. Res. Part C Emerg.*, 131:103119, 2021.

[4] EUROCONTROL. *Airport network integration v1.1*, 2018.

[5] Judith Rosenow and Michael Schultz. Coupling of turnaround and trajectory optimization based on delay cost. In *2018 Winter Simulation Conference (WSC)*, pages 2273–2284, 2018.

[6] Hartmut Fricke and Michael Schultz. Delay impacts onto turnaround performance. In *8th USA/Europe ATM Seminar*, 2009.

[7] Michael Schultz, Thomas Kunze, B Oreschko, and H Fricke. Dynamic turnaround management in a highly automated airport environment. In *Proceedings of the 28th ICAS*, pages 4362–4371, 2012.

[8] Hartmut Fricke and Michael Schultz. Improving aircraft turn around reliability. In *3rd ICRAT*, pages 335–343, 2008.

[9] John P Braaksma and John H Shortreed. Improving airport gate usage with critical path. *Transportation Engineering Journal of ASCE*, 97(2):187–203, 1971.

[10] Hyunjee Jin, Elena Garcia, and Dimitri N Mavris. Simulation of integrated approach for aircraft turnaround process. In *AIAA MSTC*, page 0418, 2018.

[11] Hasnain Ali, Yash Guleria, Sameer Alam, and Michael Schultz. A passenger-centric model for reducing missed connections at low cost airports with gates reassignment. *IEEE Access*, 7:179429–179444, 2019.

[12] Jinn-Tsai Wong and Shy-Chang Tsai. A survival model for flight delay propagation. *Journal of Air Transport Management*, 23:5–11, 2012.

[13] Caterina Malandri, Luca Mantecchini, and Vasco Reis. Aircraft turnaround and industrial actions: How ground handlers' strikes affect airport airside operational efficiency. *J. Air Transp. Manag.*, 78:23–32, 2019.

[14] Jan Evler, Michael Schultz, Hartmut Fricke, and Andrew Cook. Stochastic delay cost functions to estimate delay propagation under uncertainty. *IEEE Access*, 10:21424–21442, 2022.

[15] Cheng-Lung Wu and Robert E Caves. Modelling and optimization of aircraft turnaround time at an airport. *Transportation Planning and Technology*, 27(1):47–66, 2004.

[16] J Evler, M Lindner, H Fricke, and M Schultz. Integration of turnaround and aircraft recovery to mitigate delay propagation in airline networks. *Computers & Operations Research*, 138:105602, 2022.

[17] Ehsan Asadi, Michael Schultz, and Hartmut Fricke. Optimal schedule recovery for the aircraft gate assignment with constrained resources. *Computers & Industrial Engineering*, 162:107682, 2021.

[18] Michael Schultz, Judith Rosenow, and Xavier Olive. Data-driven airport management enabled by operational milestones derived from ADS-B messages. *J. Air Transp. Manag.*, 99:102164, 2022.

[19] Mingchuan Luo, Michael Schultz, Hartmut Fricke, Bruno Desart, Floris Herrema, and Rocío Barragán Montes. Agent-based simulation for aircraft stand operations to predict ground time using machine learning. In *DASC*, 2021.

[20] M Schultz, T Kunze, B Oreschko, and H Fricke. Microscopic process modelling for efficient aircraft turnaround management. In *Air Transport and Operations Symposium*, 2013.

[21] Jenaro Nosedal Sanchez and Miquel A Piera Eroles. Causal analysis of aircraft turnaround time for process reliability evaluation and disruptions' identification. *Transportmetrica B: Transport Dynamics*, 6(2):115–128, 2018.

[22] Michael Schultz. The seat interference potential as an indicator for the aircraft boarding progress. In *SAE Technical Paper 2017-01-2113*, 2017.

[23] Jian Yang, David Zhang, Alejandro F Frangi, and Jing-yu Yang. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE TPAMI*, 26(1):131–137, 2004.

[24] Scikit-learn. see "https://scikit-learn.org/stable/index.html" for details.

[25] George AF Seber and Alan J Lee. *Linear regression analysis*, volume 329. John Wiley & Sons, 2012.

[26] Anthony J Myles, Robert N Feudale, Yang Liu, Nathaniel A Woody, and Steven D Brown. An introduction to decision tree modeling. *Journal of Chemometrics*, 18(6):275–285, 2004.

[27] Gérard Biau and Erwan Scornet. A random forest guided tour. *TEST*, 25(2):197–227, 2016.

[28] Ying-hui Kong and Mei-li Jing. Research of the classification method based on confusion matrixes and ensemble learning. *Comput Eng Sci*, 34(6):111, 2012.

[29] Jesse Santiago and Desirae Magallon. Critical path method. *CEE320, Winter 2013*, 2009.