

Towards Conformance Criteria for Ensuring Fairness Among Airspace Users in Collaborative Optimization of Flight Lists in Air Traffic Flow Management

Tobias Harzfeld*, Sebastian Gruber†, Christoph G. Schuetz‡, Marie Carré§, Michael Brügger¶, Anaïs Lacroix||, Christoph Fabianek**, Christoph Rihacek††, and Eduard Gringinger‡‡

*Johannes Kepler University Linz, Linz, Austria

ORCID: 0009-0003-0275-6207

†Johannes Kepler University Linz, Linz, Austria

ORCID: 0000-0002-5714-1870

‡Johannes Kepler University Linz, Linz, Austria

ORCID: 0000-0002-0955-8647

§Swiss International Air Lines, Zurich, Switzerland

marie.carre@swiss.com

¶Zurich Airport, Zurich, Switzerland

michael.bruegger@zurich-airport.com

||Skyguide, Geneva, Switzerland

anaïs.lacroix@skyguide.ch

**Frequentis AG, Vienna, Austria

ORCID: 0009-0002-4410-8796

††Frequentis AG, Vienna, Austria

christoph.rihacek@frequentis.com

‡‡Frequentis AG, Vienna, Austria

ORCID: 0000-0003-3897-3003

Abstract—In air traffic flow management (ATFM), unexpected events such as bad weather conditions regularly reduce airport capacity. In such cases of temporarily reduced capacity, the EUROCONTROL Network Manager issues a regulation, which results in arriving flights being assigned a new target time of arrival (TTA), usually on a first-planned, first-served (FPFS) basis. These TTA assignments, however, may be inefficient due to the varying delay costs of individual flights. Therefore, one of the solutions of the SESAR project HARMONIC aims to develop a platform for collaborative optimization of arrival flight lists – a Target Time Management System (TTMS). The TTMS allows airspace users (AUs) and the airport to provide their preferences as input for an optimization of the arrival flight list. In this regard, one challenge is to define conformance criteria applying to the preferences submitted by the AUs in order to reduce the possibility that the preferences of individual AUs intentionally manipulate the optimization process, thus ensuring fairness among AUs. These conformance criteria should be easy to understand and have little or no negative impact on the optimization results. In this paper, we investigate the impact of different configurations of easy-to-understand conformance criteria on the quality of the optimization results. The experimental evaluation shows that the application of conformance criteria has only a minor impact on the quality of the flight lists. Our results will inform the choice of conformance criteria for the live trials of the TTMS developed in the course of the HARMONIC project.

Index Terms—air traffic control, flight prioritization

I. INTRODUCTION

In air traffic flow management (ATFM), which aims to ensure an optimal flow of aircraft [1], each flight is assigned a specific arrival, departure, or en route slot. Unexpected events, such as adverse weather conditions, can lead to reduced capacity and thus an imbalance between capacity and demand, where a slot assigned to a flight may no longer be available [1]. In situations of reduced capacity, EUROCONTROL's Network Manager issues a 'regulation', triggering a reallocation of slots to flights [2] based on the Computer-Assisted Slot Allocation (CASA) algorithm, which follows the first-planned, first-served (FPFS) principle [3].

The FPFS principle for the allocation of ATFM slots in situations of reduced capacity can be considered fair from an ATFM perspective, as no airspace user (AU) is favored and all flights are treated in the same way [2]. Following the FPFS approach, the original flight order is largely maintained, although flights are postponed, resulting in generally evenly distributed flight delays [2], [4]. Different flights, however, are affected differently by the delay, with each flight incurring different delay costs, e.g., depending on the number of passengers missing a connection [5].

An underlying assumption of the User-Driven Prioritization Process (UDPP) developed by EUROCONTROL [2] is the following: AUs best know the cost structures and thus priorities of their flights, and therefore allowing AUs to rearrange flight lists accordingly will result in reduced delay costs. UDPP, however, only allows AUs to reschedule their flights internally, limiting the potential cost savings that could be achieved when allowing prioritization of flights across different AUs.

The SlotMachine exploratory research project [6], which was conducted under the umbrella of the Single European Sky Air Traffic Management Research (SESAR) initiative, shared the UDPP's underlying assumption that AUs know best the priorities of their flights and developed a platform for collaborative optimization of flight lists across different airspace users. The SlotMachine platform employs secure multi-party computation through secret-sharing to keep airspace users' preferences confidential [7]. An auction-based market mechanism using delay credits ensures equity over time and fairness [5].

In the industrial research project HARMONIC [8], which is currently being conducted under the SESAR umbrella, the SlotMachine platform serves as the basis for the development of a Target Time Management System (TTMS), where airspace users express their preferences for the optimization of the flight list in form of a *weight map*, assigning weights for each combination of flight and target time of arrival (TTA); each airspace user only assigns weights for its own flights. A higher weight indicates a higher preference of the AU for a specific flight receiving a specific TTA. The development of the TTMS is a joint effort between Zurich Airport, Swiss International Air Lines (SWISS), the AIT Austrian Institute of Technology (AIT), Frequentis AG, Skyguide, and the Institute of Business Informatics – Data & Knowledge Engineering of Johannes Kepler University (JKU) Linz.

Unlike the SlotMachine platform, the HARMONIC project's TTMS does not employ an auction-based market mechanism, but rather employs inequity weights to ensure equity over time [9] and *conformance criteria* to ensure fairness. Given the competitive nature of the aviation market, with the presence of potentially adverse AUs, individual AUs can intentionally provide weights that do not reflect true preferences but allow the AU to gain an unfair advantage in the optimization process. Even without a market mechanism in place, AUs should be incentivized to report their preferences as truthfully as possible and be hindered from manipulating the system to gain an unfair advantage. To address this issue, in the course of the TTMS development, experts from Zurich Airport, SWISS, Skyguide, and the Institute of Business Informatics - Data & Knowledge Engineering of JKU Linz developed a set of conformance criteria that the weight maps submitted by AUs must meet.

While conformance criteria for the weight maps submitted by AUs to the TTMS may improve fairness, the conformance criteria nevertheless inevitably distort the true preferences of AUs. This paper, therefore, investigates the impact of proposed conformance criteria on the quality of the solutions found by

the TTMS under the assumption that AUs aim to honestly submit preferences for their flights as truthfully as possible given different configurations of the conformance criteria. The underlying research question is the following:

How do different configurations of the conformance criteria for the preferences of the airspace users affect the quality of the flight list found by the TTMS?

To investigate the impact of conformance criteria, in this paper, we conduct a series of optimizations using synthetic but realistic datasets. The basis for the generation of those synthetic datasets were summary data of *criticality maps* obtained from SWISS, where criticality maps are already employed in operations for prioritization of flights. A criticality map indicates for each flight the criticality of multiple available TTAs, where a lower value indicates a higher preference for a specific TTA being assigned to a certain flight. This specific method of calculating criticality maps and using those criticality maps for prioritization of flights within the airline is specific to SWISS. For more information on these criticality maps, we refer to [10]. The summary data of real-world criticality maps from SWISS serve as the basis for obtaining realistic profiles of AU preferences for TTAs.

Based on the summary data of real-world criticality maps obtained from SWISS, different archetypes of flights were identified—namely, high-priority, medium-priority, flexible, and special-case flights—with different criticality curves. For different scenarios with varying characteristics regarding the number of flights per AU and the number of flights per archetype, realistic synthetic criticality maps were generated. These criticality maps serve as the *ground truth* for evaluating the impact of the conformance criteria on the optimization result. The criticality maps are first converted into weight maps for the TTMS, with the weight maps following different configurations of the conformance criteria. The quality of the flight lists found by the TTMS using the weight maps following the conformance criteria is then compared with the quality of the flight lists found by an optimization based on the original criticality maps. Based on the experimental results reported in this paper, promising configuration parameters for the conformance criteria are identified, which can be used in validation exercises conducted in the further course of the HARMONIC project.

The remainder of this paper is structured as follows. Section II provides background information on the TTMS developed in the HARMONIC project, the concepts of equity and fairness in flight prioritization, and the methods used for data analysis. Section III introduces the conformance criteria for weight maps submitted to the TTMS. Section IV describes the generation of the synthetic test data required for the experiments. Section V presents the experimental setup. Section VI describes the experimental results. Section VII concludes the paper with a summary of the main findings and an outlook on future work.

II. BACKGROUND

This section first presents the TTMS of the HARMONIC project, briefly discusses the notions of equity and fairness, and gives an overview on time series clustering used as the basis for synthetic data generation.

A. Target Time Management System

One objective of the HARMONIC research project is the development of a *Target Time Management System* (TTMS). Fig. 1 shows a simplified architecture of the TTMS. The core components of the TTMS are the Orchestrator and the Optimizer. The airport and AUs provide the available TTAs, the flights, and the initial flight list to the Orchestrator. In addition, the airport submits preferences for the available TTAs for all flights, and the AUs provide preferences for the available TTAs, each AU for its own flights. The preferences are expressed with weight maps that contain numerical values for each combination of TTA and flight. The higher the numerical value, the higher the priority of a particular flight for a particular TTA. Table I shows an example of such a weight map.

The preferences of the AUs may be specified based on the individual cost structure of their flights. These costs are characterized by a non-linear function [7] and depend on, for example, the number of passengers missing a connection or required crew rotation [5]. Within a delay target of a flight, the costs increase moderately. However, if the delay target is exceeded, the costs increase considerably. The non-linear cost function can thus also be regarded as a step function, whose shape is determined by the delay targets. The weights for the preferences of a flight may be derived from its inverted cost function.

In general, the TTMS is either hosted by a trusted platform provider, e.g., the airport, or by a provider that is not completely trusted, i.e., an honest-but-curious third-party service provider [11]. With an honest-but-curious provider, since the preferences of the AUs are confidential information, methods for privacy-preserving computation are required to keep the platform provider from snooping [11]. In that case, the Optimizer uses an evolutionary algorithm to find a near-optimal assignment of flights to TTAs within the time available for optimization [11].

If the TTMS is hosted by a trusted platform provider, there is no need for privacy-preserving computation. In that case, the weights of the AUs are visible to the platform provider [11]. The Optimizer simply combines the weight maps of the AUs and the airport, and employs a deterministic algorithm, e.g., the Hungarian method, to find an optimal assignment of flights to TTAs [11].

The Orchestrator of the TTMS returns an optimized flight list to the stakeholders after optimization for further processing. The stakeholders or the Orchestrator may evaluate the optimized flight list against internal key performance indicators (KPIs) [11]. If the optimized flight list satisfies the predefined KPIs, a request with the optimized TTA assignments may be sent to the EUROCONTROL Network Manager [11].

TABLE I
EXAMPLE WEIGHT MAP

	TTA 1	TTA2	TTA 3	TTA 4	TTA 5
Flight 1	100 000	50 000	10 000	4 000	1 000
Flight 2	-	-	5 000	2 000	100
Flight 3	-	10 000	10 000	-	-

B. Equity and Fairness in Flight Prioritization

In case of a demand–capacity imbalance in ATFM, limited resources should be allocated in the most optimal manner [12]. A well-established and widely accepted paradigm for this allocation of resources is *first-planned, first-served* [2]. In this process, flights are assigned ATFM slots such that the original order of the flights is (generally) preserved, meaning that they are pushed back in time [2], [4]. The FPFS paradigm preserves equity since all flights are treated equally [2], [4], and fairness is guaranteed as AUs do not have a way to influence the outcome in their favor—no preferences by AUs are considered.

An approach that enhances the flexibility of AUs regarding the prioritization of flights is the *User-Driven Prioritization Process* (UDPP) [2]. This process allows AUs to swap slots to prioritize their own flights. However, this flexibility is limited, as swapping is only allowed internally, i.e., an AU cannot swap the ATFM slots with other AUs. Thus, UDPP ensures that ATFM swaps do not negatively impact other AUs [2], guaranteeing equity and fairness.

In the SlotMachine platform, which allows for ATFM slot swapping between AUs, an auction-based market mechanism [5] aims to ensure equity and fairness. In the proposed market mechanism, each AU can offer ATFM slots of its own flights and bid on ATFM slots that they are interested in, using delay credits to bid for ATFM slots. In the absence of such a market mechanism, different means of ensuring equity over time—i.e., no AUs are systematically favored—and fairness—i.e., individual AUs cannot “trick” the system into being systematically favored. In this paper, we investigate conformance criteria for the weight maps that express AU preferences as a means of ensuring fairness.

C. Time Series Clustering

Cluster analysis is an unsupervised learning task that aims to identify groups of similar data objects in unlabeled data [13]. For each of these classes, also known as clusters, the data objects within the same cluster are typically more similar to each other than to data objects in other clusters [14]. Different clustering algorithms exist to identify patterns in data. A popular, widely used algorithm is *k-means*, which aims to minimize the variance within the clusters [14].

Time series data can be described as a set of numerical values arranged in chronological order and characterized by their continuous and dynamic nature [15], [16]. For the purposes of cluster analysis, each time series is considered as an object [16], [17]. The objective of time series clustering is then to find clusters of similar time series. Similarity of the

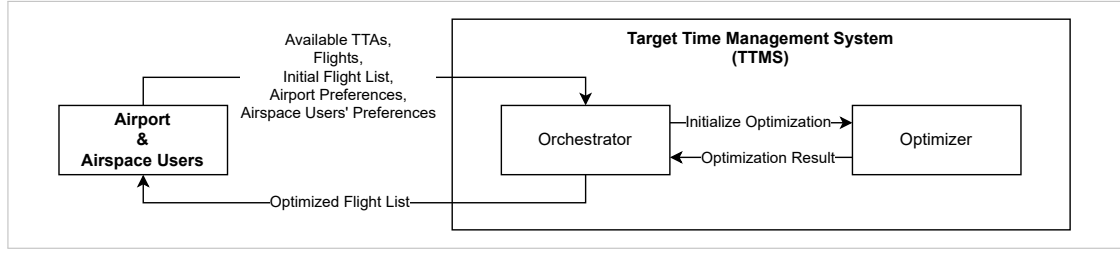


Fig. 1. Simplified architecture of the Target Time Management System (TTMS) developed in the HARMONIC project

different time series refers to the similarity of the time, shape, and change [18].

The k-means clustering algorithm may be adapted for clustering time series data. To measure the similarities in time series data, *dynamic time warping* (DTW) may be used as a distance metric for clustering [16]. In this paper, time series k-means clustering is applied in combination with DTW to identify different *archetypes of flights* based on the summary data of real-world criticality maps obtained from SWISS.

III. CONFORMANCE CRITERIA

The conformance criteria are applied to the weight maps of AUs to reduce the possibility of AUs intentionally manipulating the optimization process with the weights the AUs provide for optimization, thus aiming to ensure fairness among AUs. The conformance criteria shall be easy to understand and have little or no negative impact on the quality of the optimization result. The following conformance criteria were defined in workshops with domain experts from SWISS, Zurich Airport, and Skyguide as well as researchers from JKU Linz.

A. Minimum and Maximum Weight

This conformance criterion limits the weights to a minimum and maximum value to prevent AUs from specifying arbitrarily high or low weights, thus improving the comparability of weight maps among AUs. Let w_{min} be the minimum weight, w_{max} be the maximum weight and W_k be the weight map of an AU, each weight $w \in W_k$ must satisfy the following condition:

$$w_{min} \leq w \leq w_{max} \quad (1)$$

If $w_{min} = 0$ and $w_{max} = 10\,000$, for example, the weight map W_k in Table I violates the maximum value with the weights for Flight 1 and TTA 1 and TTA 2.

B. Weight Budget

This conformance criterion limits the sum of the weights an AU can spend for prioritizing flights to prevent AUs from specifying any number of high weights even for lower-priority flights. Let b_k be the individual weight budget for an AU, W_k be the weight map of the AU, $|W_k|$ be the number of weights in W_k , and w_i be a weight of W_k for $i \in [1, |W_k|]$, the weight map W_k must satisfy the condition:

$$\sum_{i=1}^{|W_k|} w_i \leq b_k \quad (2)$$

The individual weight budget b_k for an AU is calculated based on the number of the AU's flights affected by the regulation $|F_k|$ and a budget multiplicator m that is constant for all AUs:

$$b_k = |F_k| \cdot m \quad (3)$$

If $m = 10\,000$, for example, the weight budget b_k for the weight map W_k in Table I with $|F_k| = 3$ is 30 000. The weight map in Table I violates the weight budget $b_k = 30\,000$ because a total of 192 100 weights are spent.

C. Minimum Average Weight

This conformance criterion establishes a minimum average weight per flight to prevent AUs from completely de-prioritizing lower-priority flights and forces a minimum number of weights for each flight to provide flexibility for optimization. Let w_{avg} be the minimum average weight, $|S|$ be the number of available TTAs, $W_f \subseteq W_k$ be the weight map for a particular flight, $|W_f|$ be the number of weights in W_f , w_j be a weight of W_f for $j \in [1, |W_f|]$, the weight map W_f must satisfy the condition:

$$\frac{1}{|S|} \cdot \sum_{j=1}^{|W_f|} w_j > w_{avg} \quad (4)$$

The minimum average weight w_{avg} is calculated based on the number of available TTAs $|S|$ and a threshold dividend t that is constant for all AUs:

$$w_{avg} = t \div |S| \quad (5)$$

If $t = 10\,000$, for example, the minimum average weight w_{avg} for each flight of the weight map W_k in Table I with $|S| = 5$ is 2 000. The weight map for Flight 2 in Table I violates the minimum average weight $w_{avg} = 2\,000$ because the average weight is 1 420.

D. Combination of Conformance Criteria

The conformance criteria can also be combined to further promote fairness. For example, using a minimum and maximum weight cannot prevent an AU from specifying the maximum weight for any preferred combination of flight and TTA. Combining minimum and maximum weights with a weight budget can prevent AUs from providing weight maps that contain any number of maximum weights. However, AUs operating multiple flights could still de-prioritize some flights to gain an advantage in prioritizing TTAs that are important

TABLE II
EXAMPLE WEIGHT MAP BASED ON THE WEIGHT MAP FROM TABLE I THAT
COMPLIES WITH A PARTICULAR CONFIGURATION OF THE CONFORMANCE
CRITERIA

	TTA 1	TTA2	TTA 3	TTA 4	TTA 5
Flight 1	10 000	8 430	2 182	1 246	777
Flight 2	-	-	1 412	942	647
Flight 3	-	2 182	2 182	-	-

for competitors, or the AUs only provide a few maximum weights for important TTAs. In this case, combining minimum and maximum weights and a weight budget with a minimum average weight improves fairness compared to using only a single conformance criterion.

Table II provides an example weight map based on the weight map from Table I that complies with the minimum weight $w_{min} = 0$, the maximum weight $w_{max} = 10\,000$, the budget multiplier $m = 10\,000$, and the threshold dividend $t = 3\,000$.

E. Discussion of the Conformance Criteria

The conformance criteria shall be easy to understand and have little or no negative impact on the quality of the optimization results. The conformance criteria defined above were agreed to be easy to understand among domain experts. While additional conformance criteria may be even more beneficial with regard to promoting fairness, they may also increase the difficulty for AUs to create weight maps that represent the true preferences and at the same time comply with the conformance criteria. In order to investigate the impact of different configurations of the conformance criteria on the quality of the optimization results, an experimental evaluation is required.

IV. GENERATION OF SYNTHETIC DATA

This section presents our approach to generating synthetic data based on summary data of real-world flight preferences in order to evaluate the conformance criteria.

A. Summary Data

SWISS calculates criticality values for combinations of flight and TTA to prioritize flights in the event of an ATFM regulation [10]. These criticality values range from -4, i.e., the most important TTA overall, to 4, i.e., the least important TTA overall. For each flight, a criticality value is specified for at least one TTA, which indicates the priority of a TTA for the flight. We refer to a set of criticality values for one or multiple flights as a criticality map. Over time, SWISS stored criticality maps in a database from which we obtained summary data.

The summary data includes the number of the criticality values and the number of physically and operationally infeasible TTAs as well as the minimum, median, and maximum criticality values for each flight of a regulation. We distinguish between feasible TTAs and physically and operationally infeasible TTAs. Feasible TTAs are those where both physical and operational conditions permit landing and the AU indicates

a preference. Physically infeasible TTAs violate fundamental physical principles, e.g., exceeding maximum speed. Operationally infeasible TTAs are physically possible but impractical due to excessive cost or operational constraints.

The summary data also includes a classification and additional characteristics of the criticality values for each flight of a regulation. A flight is classified as a normal case flight or a special case flight based on the first criticality value and the minimum criticality value, with a flight being a normal case if both values are equal and a special case otherwise. The criticality values of a flight are each assigned to one of four buckets: values below the 25th percentile, values between the 25th percentile and the median, values between the median and the 75th percentile, and values at or above the 75th percentile. For a special case flight, however, the procedure differs in one aspect—criticality values from the first criticality value up to the minimum criticality value are assigned to the first bucket. The additional characteristics of the criticality values of a flight are the number of criticality values and the mean slope between the criticality values of each bucket.

The summary data contains information about 180 criticality maps and 2 296 flights. A total of 32 criticality maps were identified for removal because they are either duplicates or contain only a single TTA, leaving 148 criticality maps and 2 092 flights for further processing.

B. Approximated Criticality Profiles

The summary data obtained from SWISS is used to approximate criticality profiles. We refer to a criticality profile as the ordered sequence of criticality values for a flight.

The criticality values of a flight are approximated based on the characteristics of its criticality value buckets. For each bucket, exactly as many criticality values are added to an approximated criticality profile as are originally in the bucket. The approximated criticality values of a bucket are calculated by interpolation based on the bucket's mean slope between criticality values. While the approximated criticality values for normal case flights start with the minimum criticality value and tend to increase over time, the approximated criticality values for special case flights of the first bucket decrease until the minimum criticality value is reached and then start to increase. The physically and operationally infeasible TTAs are added to the approximated criticality profile of a flight without having a criticality value.

In total 2 092 criticality profiles are approximated from the summary data. From the 2 092 approximated criticality profiles, 236 criticality profiles that contain only a single criticality value are removed, leaving 1 856 approximated criticality profiles for further processing.

C. Cluster Analysis

The approximated criticality profiles are clustered to identify similar criticality profiles. We assume a flight is similar to another flight if the criticality values behave similarly over time. Therefore, the *tslearn* library [19] is used for k-means clustering with dynamic time warping as the distance

metric. The hyperparameter search identified using $k = 4$ as promising with an average silhouette score of 0.807.

The clustering results with four clusters are shown in Fig. 2. The 86 approximated criticality profiles in Cluster 1 have the highest priority for a TTA—indicated by the lowest criticality value—at the beginning of their progression. This suggests that these AUs have a strong preference for these flights arriving close to their scheduled time. The criticality values for TTAs after the first criticality value tend to decline. Such criticality profiles are referred to as *high-priority flights*. The 1636 criticality profiles in Cluster 2 are rather variable in terms of TTA preferences, which indicates that arrivals are not particularly time-dependent. Such criticality profiles are referred to as *flexible flights*. The 105 criticality profiles in Cluster 3 are similar to those in Cluster 1, with the difference that these flights have a considerably lower preference for a TTA at the beginning. The criticality values for further TTAs decline by about half after the first criticality value. Such criticality profiles are referred to as *medium-priority flights*. The 29 criticality profiles assigned to Cluster 4 represent flights classified as special cases, characterized by having their lowest criticality value at a later point in their progression. Such criticality profiles are referred to as *special-case flights*. The identified archetypes of flights are used for generating synthetic criticality profiles.

D. Synthetic Criticality Profiles

Synthetic criticality profiles are generated based on the clustering results. Four archetypes of flights are identified in the cluster analysis—namely, high-priority, flexible, medium-priority, and special-case flights. For each archetype, a value range for the number of physically infeasible TTAs and a value range for the number of operationally infeasible TTAs are determined based on the approximated criticality profiles of the archetype. Furthermore, value ranges for six different points in time are calculated for each archetype. More precisely, value ranges for the criticality values at the start, 20th, 40th, 60th, and 80th percentile, and the end are determined to be the median $\pm$ standard deviation of the criticality values at these points in time of the flights assigned to the archetype.

When generating a synthetic criticality profile for a flight, the number of available TTAs and the archetype are required. The number of physically and operationally infeasible TTAs are drawn randomly from the archetype’s corresponding value ranges while ensuring that the criticality profile contains at least two criticality values. For each of the six points in time of an archetype, a criticality value is drawn randomly from the corresponding value range, and the remaining criticality values between these positions are interpolated. If a criticality profile contains less than six criticality values, criticality values are drawn from a subset of the six value ranges.

Fig. 3 shows an example of ten synthetic criticality profiles for each archetype. These criticality profiles reflect the characteristics of their archetype but still differ despite using only six different value ranges for each archetype. To obtain a synthetic criticality map, the synthetic criticality profiles are

combined into a matrix, which in this example has a shape of 40 flights and 40 TTAs.

V. EXPERIMENTAL SETUP

This section describes the datasets for the experimental evaluation, the evaluation process, the different experiments, and the transformation of criticality maps into weight maps.

A. Datasets

The datasets for the experimental evaluation of the conformance criteria are generated with different AUs that operate flights of different archetypes. In this regard, the following types of AUs are distinguished, loosely based on the actual numbers of flights operated by different airlines in typical mid-day and afternoon phases at Zurich Airport. A *hub airline* operates a large number of flights (possibly the majority) at an airport and typically designates the airport as the operational center. A *single-flight airline* operates only a single flight in a given time window, corresponding to airlines based elsewhere but serving the airport. A *multiple-flight airline* operates only a few flights, e.g., five flights, in a given time window, corresponding to smaller airlines based at the airport or airlines based elsewhere but serving the airport frequently, possibly from multiple hubs.

In the experiments, eleven datasets with different distributions for the different types of AUs are used, with each distribution comprising ten samples (or datasets). Table III gives an overview of the different distributions, including the number of flights per archetype for each airline type. The archetypes of the flights for multiple-flight airlines in Distributions 1 to 10 were randomly selected, with each archetype occurring at least once. In Distribution 11, a smaller hub airline is represented by a multiple-flight airline that operates more flights than the others. The combined criticality map for each sample includes one hundred flights, with each flight having criticality values for up to one hundred TTAs. The generated samples were also tested for having a valid assignment solution.

B. Evaluation Process

The evaluation of different configurations of the conformance criteria is based on comparing the results from optimization using the true preferences of AUs and those using the weight maps that comply with a particular configuration of the conformance criteria. For this purpose, synthetic criticality maps serve as the ground truth. The synthetic criticality maps of different AUs are transformed into weight maps that comply with a particular configuration of the conformance criteria and are combined into a weight map. The linear sum assignment is used to find optimal assignments of flights to TTAs based on the ground truth on the one hand and based on the weight maps on the other. These assignments incur costs in terms of criticality values that are compared.

The evaluation process is shown in Fig. 4. For each assignment, an aggregated criticality score is calculated. Considering the assignment as a permutation matrix $X = (x_{ij})$ where $x_{ij} = 1$ indicates that the i^{th} flight is assigned to the j^{th}

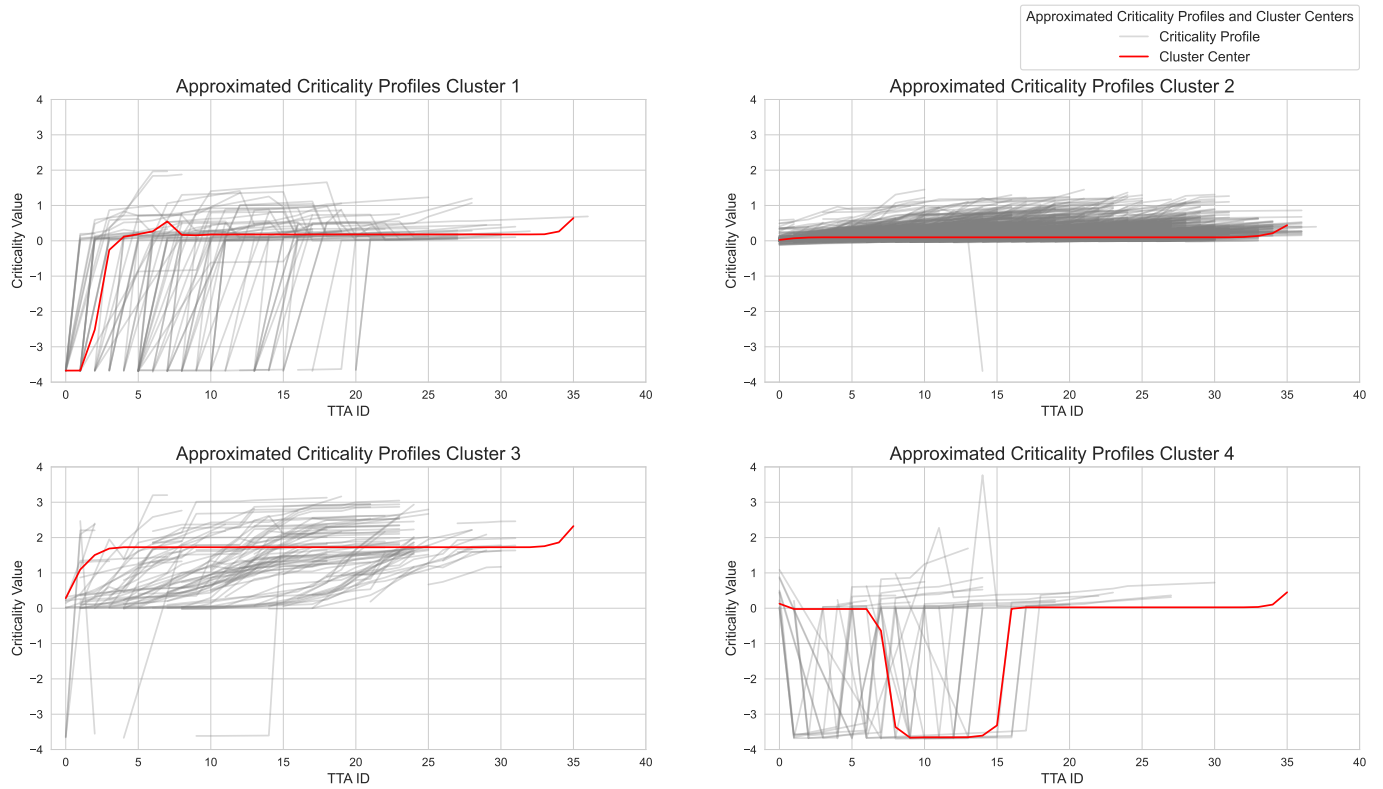


Fig. 2. Visualization of the results of the cluster analysis of the approximated criticality profiles

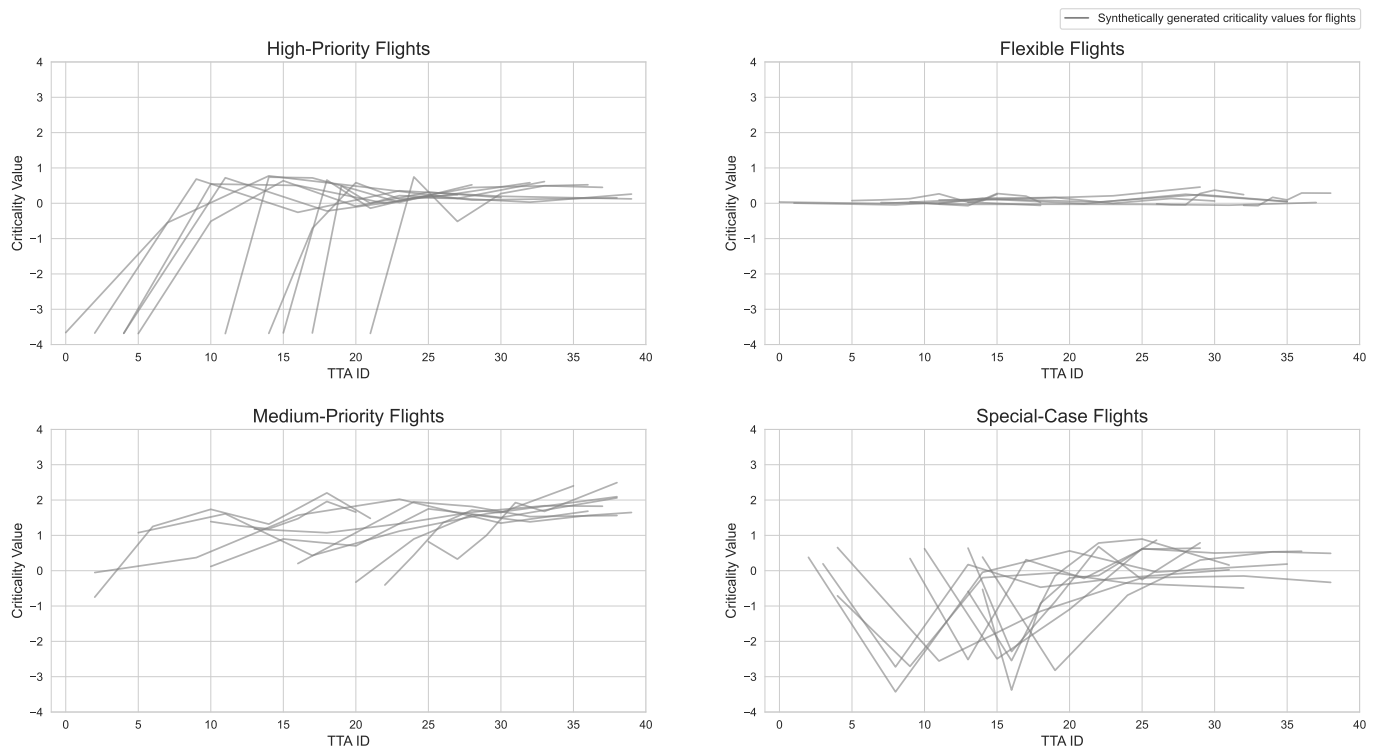


Fig. 3. Example of ten synthetic criticality profiles generated for each archetype

TABLE III
DISTRIBUTIONS OF FLIGHTS PER ARCHETYPE AND AIRLINE TYPE FOR THE GENERATION OF SYNTHETIC DATASETS

ID	Hub Airline					Single-Flight Airline				Multiple-Flight Airline				
	Archetype				Σ Flights	Σ Airlines	Archetype		Σ Flights	Σ Airlines	Archetype			Σ Flights
	1	2	3	4			1				1	2	3	
1	5	40	5	0	50	10	1 per Airline	10	8	5 per Airline			40	
2	10	35	5	0	50	10	1 per Airline	10	8	5 per Airline			40	
3	10	25	10	5	50	10	1 per Airline	10	8	5 per Airline			40	
4	15	15	15	5	50	10	1 per Airline	10	8	5 per Airline			40	
5	15	45	15	0	75	10	1 per Airline	10	3	5 per Airline			15	
6	20	40	15	0	75	10	1 per Airline	10	3	5 per Airline			15	
7	20	35	15	5	75	10	1 per Airline	10	3	5 per Airline			15	
8	15	45	15	0	75	15	1 per Airline	15	2	5 per Airline			10	
9	20	40	15	0	75	15	1 per Airline	15	2	5 per Airline			10	
10	20	35	15	5	75	15	1 per Airline	15	2	5 per Airline			10	
11	5	40	5	0	50	10	1 per Airline	10	1	5	15	5	25	
									5	1	1	1	15	

Archetype 1 refers to high-priority flights, Archetype 2 to flexible flights, Archetype 3 to medium-priority flights, and Archetype 4 to special-case flights.

TTA and $x_{ij} = 0$ indicates no assignment, and the combined synthetic criticality map $C = (c_{ij})$ where c_{ij} indicates the criticality value for assigning the i^{th} flight to the j^{th} TTA (c.f., [20]), the aggregated criticality score σ is calculated as follows:

$$\sigma = \sum_{i=1}^{|F|} \sum_{j=1}^{|S|} x_{ij} \cdot c_{ij} \quad (6)$$

where $|F|$ is the number of flights and $|S|$ is the number of TTAs. The aggregated criticality scores of the assignments are compared to evaluate the impact of the conformance criteria on the quality of the optimization results. If the aggregated criticality scores are equal, the conformance criteria did not affect the quality of the flight lists. If the aggregated criticality scores differ, the increase of the criticality score is calculated, and for each AU and archetype, the number of flights that received a changed TTA assignment compared to the ground truth is counted.

C. Experiments

The experiments investigate the impact of the conformance criteria on the quality of the flight lists and identify suitable parameters for the configuration of the conformance criteria. Choosing parameters for the configuration of conformance criteria is a trade-off between providing enough flexibility for AUs to specify their true preferences while restricting AUs as much as possible to reduce the possibility of manipulating the optimization process. The following three experiments are conducted.

1) *Experiment 1*: The first experiment investigates applying a minimum weight $w_{\min} = 0$ and a maximum weight $w_{\max} = 10\,000$. These parameters were specified together with domain experts and are also used in the following experiments.

2) *Experiment 2*: The second experiment investigates the combination of a minimum and maximum weight and varying weight budgets. Different budget multipliers are tested

under the assumption that a higher weight budget is superior when there is a minimum and maximum weight, as it allows AUs to express their preferences more precisely.

3) *Experiment 3*: The third experiment investigates the combination of a minimum and maximum weight, a weight budget, and varying minimum average weights. The most promising parameter from the second experiment is used for calculating the weight budget, i.e., the budget multiplier $m = 120\,000$. Different threshold dividends are tested under the assumption that a higher minimum average weight is superior, as it forces AUs to specify a minimum amount of weights for each flight.

D. Weight Map Transformation

The synthetic criticality maps are transformed into weight maps that comply with a particular configuration of the conformance criteria. The general transformation process is shown in Fig. 5. The criticality map of an AU is first converted into an initial weight map, which is then modified to comply with a particular configuration of the conformance criteria. If the modified weight map violates the conformance criteria, e.g., due to rounding to integer values, a repair mechanism is applied before the weight map is modified again.

1) *Scaling Function*: Transforming criticality maps into weight maps requires a suitable scaling function that normalizes the original criticality values into a desired weight range. Different possibilities exist for such scaling functions. The choice of scaling function must be informed by the assumptions made regarding the future decision-making processes followed by AUs.

We use a scaling function adapted from *min-absolute scaling* but tailored to the characteristics of the generated criticality maps in compliance with certain observations made in our interactions with airline and airport representatives in the course of the TTMS development. Intuitively, for criticality maps where the minimum value is a negative number, our scaling

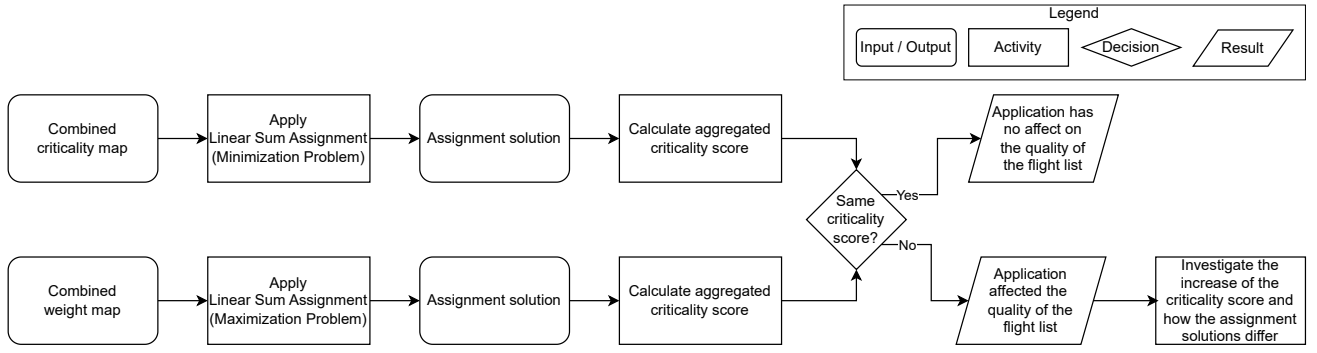


Fig. 4. Evaluation process for assessing the impact of conformance criteria on the quality of the optimization results

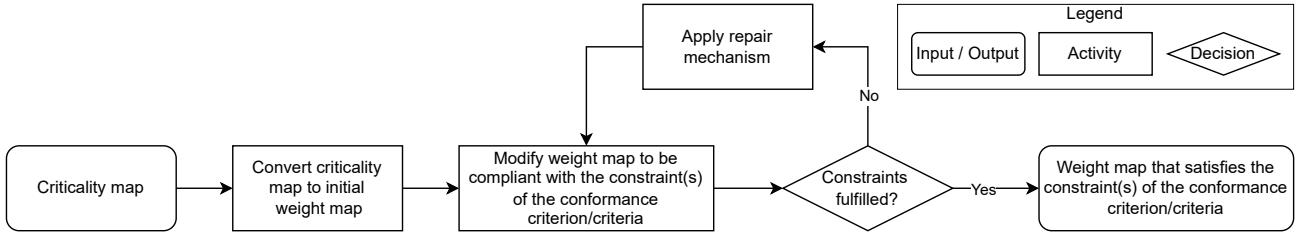


Fig. 5. Process of transforming a criticality map into a weight map that satisfies the constraint(s) of a particular configuration of the conformance criterion/criteria

function transforms the minimum value from a criticality map into 1 and preserves the proportions of the distances between the other values from the criticality map, without a lower bound. However, we shift the values into a positive range for all criticality maps where the absolute value of the minimum of the criticality map is greater than or equal to the maximum, which is the case for the generated datasets. Thus, our scaling function can be defined as follows:

$$c'_i = -1 \cdot \frac{c_i + c_{min}}{2 \cdot |c_{min}|} \quad (7)$$

where c'_i is the normalized value of an original value c_i from the criticality map C and c_{min} is the minimum value in C .

The normalized values are then multiplied with the maximum weight, e.g., a weight of 10 000, to obtain a weight map in the desired range, assuming a minimum allowed weight of 0, which will typically not be assigned by AUs. Note that for criticality maps where the maximum is greater than the absolute value of the minimum, i.e., $c_{max} > |c_{min}|$, negative weights are possible and a different scaling function would be needed if the minimum allowed weight is 0. Furthermore, the minimum of the criticality map must be a negative number in order for the scaling to work properly. These restrictions are derived from the characteristics of the real-world criticality maps obtained from SWISS, which were the basis for the generation of the synthetic datasets employed in the experiments.

The use of the particular scaling function models a certain assumed behavior that AUs exhibit when generating their weight maps for optimization. First, an AU is expected to assign the highest weight to the most important TTA for the

AU's most important flight. The weights for other TTAs and other flights are distributed in reference to the highest weight. The maximum weight being capped, AUs will typically make use of that maximum weight to ensure that their most important flight has a reasonable chance of being assigned a desirable TTA. The minimum weight may then vary between AUs and also between different optimization runs, depending on the specific operational situation of the respective AU. The minimum weight, however, will typically be distinctively greater than zero because the weight maps are an expression of the utility of TTAs for flights rather than costs.

As an alternative, we may employ *min-max scaling* for the normalization of criticality maps into a value range between 0 and 1, where the minimum value in a criticality map is assigned a value of 1 and the maximum value is assigned a value of 0. To avoid values of 0 in weight maps—a value of 0 may be reserved for operationally infeasible but physically possible flights in the future—*min-max scaling with a custom value range* could be used, where the maximum in the original criticality map is transformed into a certain minimum weight, e.g., 2 500, for the weight map, which is set to a constant value. We believe, however, that a fixed minimum weight across all AUs as well as a fixed minimum weight per AU across time is unrealistic. Therefore, in the current series of experiments, which are reported in this paper, we opted for the custom scaling function.

2) *Transformation for Experiment 1:* This transformation scales a criticality map into a weight map that complies with a minimum and maximum weight using our scaling function.

The scaled weights of the initial weight map are multiplied by the maximum weight $w_{max} = 10\,000$.

3) *Transformation for Experiment 2*: This transformation requires first performing the transformation for Experiment 1. The weights of the weight map are then again scaled based on the available weight budget b_k and the current sum of the weights. The weight budget may be violated due to, e.g., rounding to integer values, which requires a repair mechanism. If more weights are spent than the weight budget allows, the weights of the least favored TTAs are iteratively reduced by one. If fewer weights are spent than the weight budget allows, the weights of the most favored TTAs below the maximum weight are iteratively increased by one if possible. If iteratively increasing the weights of the most favored TTAs is not possible because the weights for all feasible TTAs of flights of an AU have already been assigned the maximum weight, the transformation terminates successfully because the weight budget does not have to be spent entirely.

4) *Transformation for Experiment 3*: This transformation requires first performing the transformations for Experiment 2. If the weights of a flight do not comply with the minimum average weight, the weights are increased to exceed the minimum average threshold. If the weights of a flight still do not comply with the minimum average weight despite having the maximum weight for the feasible TTAs, additional weights are assigned to operationally infeasible TTAs until the minimum average weight is satisfied. In this case, for the evaluation, the affected operationally infeasible TTAs of such a flight are assigned the least favored criticality value, i.e., a value of 4, in the criticality map. If the weights of a flight do not satisfy the minimum average weight and no additional operationally infeasible TTAs for assigning weights exist, the flight is excluded from the evaluation.

VI. EXPERIMENTAL RESULTS

This section presents the results of the experimental evaluation of the impact of different configurations of the conformance criteria on the optimization results.

A. Experiment 1

The first experiment investigates the application of the conformance criterion that requires each weight to be greater than or equal to a minimum weight $w_{min} = 0$ and less than or equal to a maximum weight $w_{max} = 10\,000$. This configuration of the conformance criterion had no impact on the quality of the flight lists for the following number of samples for each distribution.

- Distribution 1: 6 / 10 samples
- Distribution 2: 7 / 10 samples
- Distribution 3: 8 / 10 samples
- Distribution 4: 6 / 10 samples
- Distribution 5: 10 / 10 samples
- Distribution 6: 7 / 10 samples
- Distribution 7: 7 / 10 samples
- Distribution 8: 7 / 10 samples
- Distribution 9: 6 / 10 samples

- Distribution 10: 10 / 10 samples
- Distribution 11: 7 / 10 samples

The other samples of the distributions show an increase of the criticality score and changed TTA assignments compared to the ground truth. Sample 4 of Distribution 9 had the overall highest relative increase of the criticality score compared to the ground truth—namely, 0.01547 %. In this sample, 11 of the 75 hub airline's flights received a changed TTA assignment compared to the ground truth, and one of the 15 single-flight airlines' flights received a changed TTA assignment compared to the ground truth. Sample 3 of Distribution 2 had the overall lowest relative increase of the criticality score compared to the ground truth—namely, 0.00002 %. In this sample, one of the 10 single-flight airlines' flights received a changed TTA assignment compared to the ground truth, and one of the 40 multiple-flight airlines' flights received a changed TTA assignment compared to the ground truth.

In total, 29 of the 110 samples across the different distributions show an increase of the criticality score and changed TTA assignments compared to the ground truth. These differences arise from rounding the weights during the transformation process and can be considered negligible. Therefore, applying the conformance criterion—with the constraint of minimum and maximum weights—to the preferences of the AUs does not considerably impact the quality of the flight lists.

B. Experiment 2

The second experiment investigates applying the minimum weight $w_{min} = 0$, the maximum weight $w_{max} = 10\,000$, and different budget multipliers for calculating weight budgets. The budget multiplier started at 10 000 and is increased in steps of 10 000 up to 200 000. The relative increase of the criticality score compared to the ground truth is calculated for each sample and budget multiplier. Fig. 6 shows the progression of the mean relative increase of the criticality score for each distribution.

Different budget multipliers have varying effects on the aggregated criticality scores and, consequently, on the mean relative increases of the criticality scores. The progressions of the mean relative increases of the criticality scores of the distributions remain quite stable until using a budget multiplier $m > 120\,000$. This suggests that using a budget multiplier $m > 120\,000$ likely causes the criticality maps to be transformed into weight maps that no longer reflect the true preferences of the AUs.

We assume that the best budget multiplier is the highest budget multiplier that causes a minimal relative increase of the criticality score compared to the ground truth. The best individual budget multiplier for four of the 110 samples is $m = 120\,000$. For an additional 35 samples, using the budget multiplier $m = 120\,000$ instead of the best individual one, i.e., the best individual budget multiplier $m > 120\,000$, had no further impact on the relative increase of the criticality score compared to the ground truth. Therefore, the budget multiplier $m = 120\,000$ for calculating the weight budget of

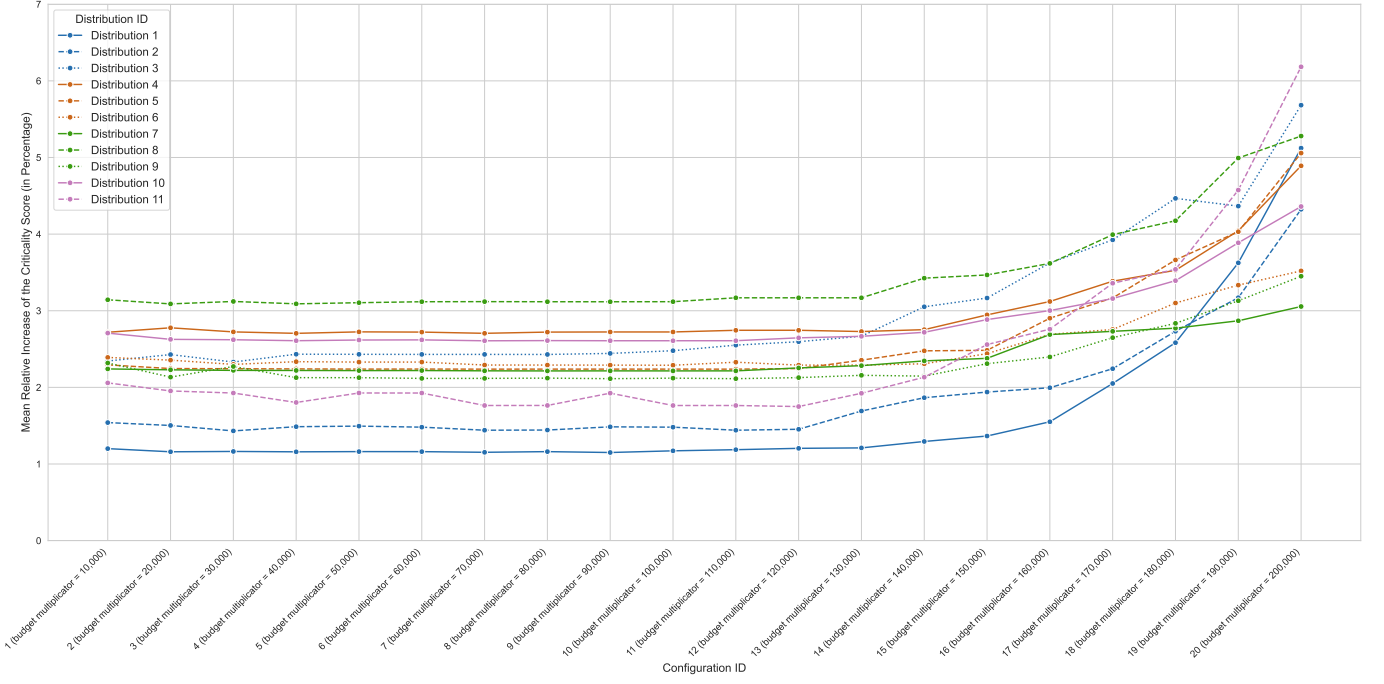


Fig. 6. Progressions of the mean relative increases of the criticality scores of the distributions compared to the ground truth when using different budget multipliers

an AU appears to be a suitable parameter for the configuration of the conformance criteria.

The differences between using each sample's best budget multiplier and using the budget multiplier $m = 120\,000$ are rather small. Using the best individual budget multiplier of each sample leads to an overall average relative increase of the criticality score of 2.08 % compared to the ground truth. Using the budget multiplier $m = 120\,000$ leads to an overall average relative increase of the criticality score of 2.23 % compared to the ground truth, which is 0.15 percentage points higher than the overall average relative increase of the criticality score when using the best budget multipliers of each sample.

The differences between using the best individual budget multiplier and using the budget multiplier $m = 120\,000$ are also rather small with regard to the changed TTA assignments. Using the best individual budget multipliers results in an average of 19.51 % changed TTA assignments compared to the ground truth for the hub airline's flights, an average of 31.36 % changed TTA assignments compared to the ground truth for the single-flight airlines' flights, and an average of 23.69 % changed TTA assignments compared to the ground truth for the multiple-flight airlines' flights. Using the budget multiplier $m = 120\,000$ results in an average of 20.08 % changed TTA assignments compared to the ground truth for the hub airline's flights, an average of 31.73 % changed TTA assignments compared to the ground truth for the single-flight airline's flights, and an average of 25.11 % changed TTA assignments compared to the ground truth for multiple-flight airlines' flights.

C. Experiment 3

The third experiment investigates applying the minimum weight $w_{min} = 0$, the maximum weight $w_{max} = 10\,000$, the budget multiplier $m = 120\,000$ for calculating the weight budgets, and different threshold dividends for calculating the minimum average weights. The threshold dividend started at 10000 and is increased in steps of 2500 up to 60000. The relative increase of the criticality score compared to the ground truth was calculated for each sample and threshold dividend. Fig. 7 shows the progression of the mean relative increase of the criticality score for each distribution.

Different threshold dividends have varying effects on the aggregated criticality scores and, consequently, on the mean relative increases of the criticality scores. The progressions of the mean relative increases of the criticality scores of the distributions remain quite stable until using a threshold dividend $t > 27\,500$. The progressions remain stable again between the threshold dividend $t = 27\,500$ and $t = 37\,500$ and increase considerable with threshold dividends $t > 37\,500$, especially for Distribution 11. This suggests that using a threshold dividend $t > 27\,500$ or $t > 37\,500$ causes less important flights to become more important than they actually are for the AUs.

We assume that the best threshold dividend is the highest threshold dividend that causes a minimal relative increase of the criticality score compared to the ground truth. The best individual threshold dividend for 18 of the 110 samples is $t = 27\,500$. For an additional 15 samples, using the threshold dividend $t = 27\,500$ instead of the best individual one, i.e., the best individual threshold dividend $t > 27\,500$, had no further

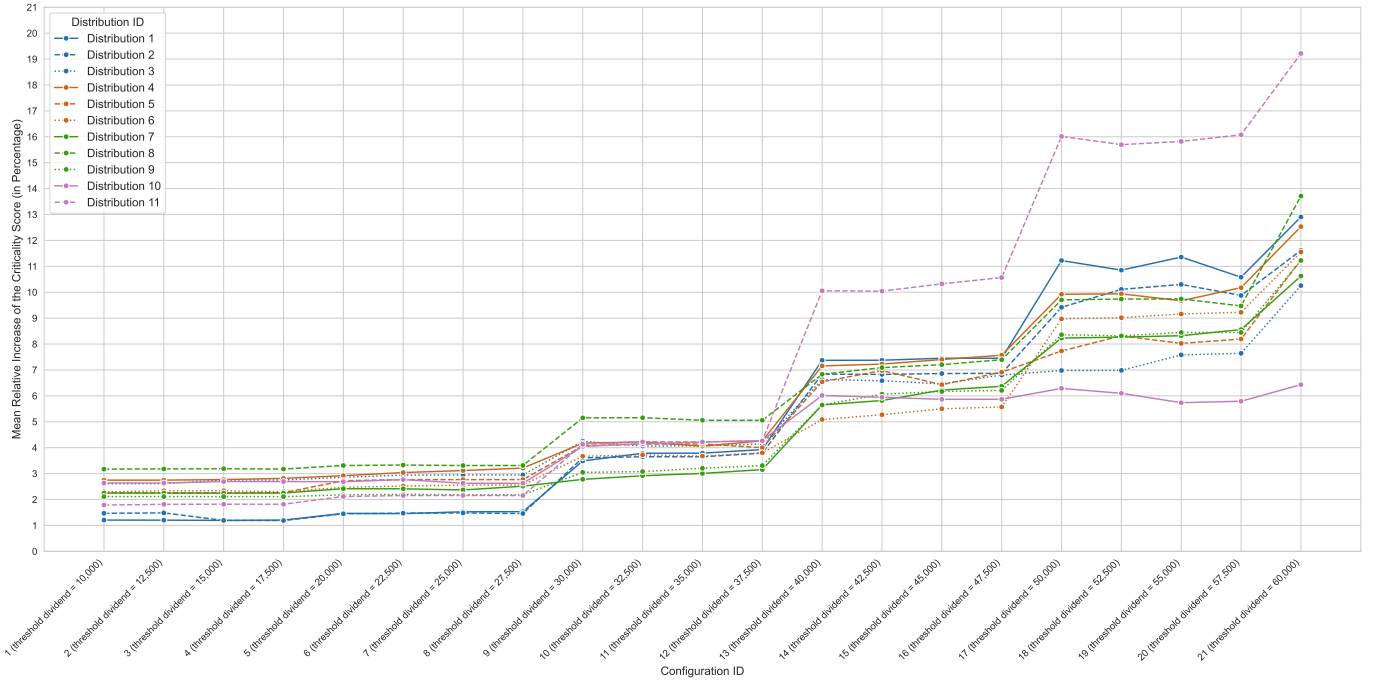


Fig. 7. Progressions of the mean relative increases of the criticality scores of the distributions compared to the ground truth when using different threshold dividends

impact on the relative increase of the criticality score compared to the ground truth. Therefore, the threshold dividend $t = 27\,500$ for calculating the minimum average weight for a flight appears to be a suitable parameter for the configuration of the conformance criteria.

The differences between using each sample's best threshold dividend and using the threshold dividend $t = 27\,500$ are rather small. Using the best individual threshold dividend of each sample leads to an overall average relative increase of the criticality score of 2.08 % compared to the ground truth. Using the threshold dividend $t = 27\,500$ leads to an overall average relative increase of the criticality score of 2.48 % compared to the ground truth, which is 0.40 percentage points higher than the overall average relative increase of the criticality score when using the best threshold dividend of each sample.

The differences between using the best individual threshold dividend and using the threshold dividend $t = 27\,500$ are also rather small with regard to the changed TTA assignments. Using the best individual threshold dividend results in an average of 19.53 % changed TTA assignments compared to the ground truth for the hub airline's flights, an average of 30.42 % changed TTA assignments compared to the ground truth for the single-flight airlines' flights, and an average of 23.95 % changed TTA assignments compared to the ground truth for the multiple-flight airlines' flights. Using the threshold dividend $t = 27\,500$ results in an average of 21.68 % changed TTA assignments compared to the ground truth for the hub airline's flights, an average of 31.88 % changed TTA assignments compared to the ground truth for the single-flight airline's flights, and an average of 25.77 % changed TTA assignments

compared to the ground truth for multiple-flight airlines' flights.

VII. CONCLUSION

In this paper, we described conformance criteria for weight maps submitted by AUs to the TTMS developed in the HARMONIC industrial research project. Furthermore, we investigated the impact of those conformance criteria on the quality of the flight lists found by the TTMS. The conformance criteria were specified in a collaborative effort of experts from HARMONIC project partners Zurich Airport, SWISS, Skyguide, and the Institute of Business Informatics – Data & Knowledge Engineering of JKU Linz. In particular, three conformance criteria were specified—namely, minimum and maximum values for weights, a weight budget per AU, and a minimum average weight per flight—with various configuration options. Collectively, these conformance criteria aim to promote fairness in collaborative optimization of flight lists. However, the evaluation of the proposed conformance criteria in terms of their ability to actually promote fairness was not the focus of this paper and is left to future work. Future work may also investigate additional conformance criteria for weight maps.

We used summary data from real-world criticality maps obtained from SWISS to identify four archetypes of flights—namely, high-priority, medium-priority, flexible, and special-case flights. We then defined eleven distributions with varying shares of flights from each archetype. For each distribution, ten sample sets of flights were randomly generated. Flights were also grouped by airlines, with airlines varying in their

characteristics, i.e., hub airlines, multiple-flight airlines, and single-flight airlines were considered. The thus generated synthetic criticality maps served as the ground truth in the experiments to determine the impact of conformance criteria on the quality of the flight lists found by the TTMS.

The generated synthetic criticality maps were transformed into weight maps for the TTMS, following varying configurations of the conformance criteria, i.e., different budget multipliers and threshold dividends to determine the weight budget per airline and the minimum average weight per flight, respectively. The assumption behind the transformations was that AUs strive to transform criticality maps as truthfully as possible into weight maps for the TTMS. The quality of the solutions found by the TTMS, evaluated with respect to the original criticality maps, was then compared to the quality of the best solutions according to the criticality maps. The results show, on the one hand, that increasing the budget beyond a certain point, while at the same time setting minimum and maximum limits for weights, leads to arbitrary weight distributions that do not reflect the true priorities of TTAs for flights from the original criticality maps; that point was generally reached when using a budget multiplier beyond 120 000 to calculate the budget per AU. On the other hand, given the presence of a weight budget per AU, setting the minimum average weight too high leads to the weights not reflecting the different priorities of flights for AUs, which was generally the case when using a threshold dividend beyond 27 500 to calculate the minimum average weight per flight. For future validation exercises in the course of the HARMONIC project, the following configuration parameters for the conformance criteria, therefore, seem promising as a starting point:

- a minimum weight of 0 and a maximum weight of 10 000,
- a budget multiplier of 120 000 as the basis for calculating the budget per AU, and
- a threshold dividend of 27 500 for calculating the minimum average weight per flight.

Future work will conduct experiments with different underlying assumptions regarding the behavior of AUs when compiling their weight maps.

ACKNOWLEDGMENTS

This paper is a condensed version of the thesis submitted by Tobias Harzfeld [21] at Johannes Kepler University Linz. Parts of the thesis were reused in this paper. We refer to the thesis for a more extensive description of the research results presented in this paper.

This work was conducted as part of the HARMONIC project. This project has received funding from the SESAR Joint Undertaking under grant agreement No 101114675 under the European Union's Horizon Europe research and innovation program. UK participant NATS in Horizon Europe Project HARMONIC receives funding from UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant number 10091990]. This work has received funding from the Swiss State Secretariat

for Education, Research and Innovation (SERI). The views expressed in this paper are those of the authors.



Co-funded by
the European Union



Project funded by



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs,
Education and Research EAER
State Secretariat for Education,
Research and Innovation SERI

REFERENCES

- [1] L. Budd and S. Ison, Eds., *Air transport management: an international perspective*. London ; New York: Routledge, Taylor & Francis Group, 2017.
- [2] N. Pilon, L. Guichard, Z. Bazso, G. Murgese, and M. Carré, "User-driven prioritisation process (UDPP) from advanced experimental to pre-operational validation environment," *Journal of Air Transport Management*, vol. 97, p. 102124, 2021.
- [3] EUROCONTROL, "ATFCM Users Manual," May 2024. [Online]. Available: <https://www.eurocontrol.int/publication/atfcm-users-manual>
- [4] S. Ruiz, H. Kadour, and P. Choroba, "An Innovative Safety-neutral Slot Overloading Technique to Improve Airspace Capacity Utilisation," in *Proceedings of the 9th SESAR Innovation Days*, Dec. 2019.
- [5] C. G. Schuetz, Sergio Ruiz, E. Gringinger, C. Fabianek, and T. Lorünser, "An auction-based mechanism for a privacy-preserving marketplace for ATFM slots: ICAS conference 2022," in *33rd Congress of the International Council of the Aeronautical Sciences*, C. G. Schütz, Ed., 2022, pp. 1–14.
- [6] European Union, "A Privacy-Preserving Marketplace for Slot Management," <https://doi.org/10.3030/890456>, 2020.
- [7] C. G. Schuetz, E. Gringinger, N. Pilon, and T. Lorünser, "A privacy-preserving marketplace for air traffic flow management slot configuration," in *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)*. IEEE, 2021, pp. 1–9.
- [8] European Union, "HARMONISED network through smart technology and Collaboration," <https://doi.org/10.3030/101114675>, 2023.
- [9] M. Carré, "Multi-Airline Operations Optimization under major disruptions," Ph.D. dissertation, Université Clermont Auvergne, Jan. 2024. [Online]. Available: <https://theses.hal.science/tel-04547756>
- [10] L. Caranti, M. Carré, L. Weber, D. Heilmann, and R. Stevens, "Optimization of regulated airline arrival flows via target time management an enhanced slot swapping process," in *Proceedings of the 13th SESAR Innovation Days*, 11 2023.
- [11] S. Gruber, P. Feichtenschlager, C. Fabianek, E. Gringinger, and C. G. Schuetz, "Towards a heuristic optimizer for a target time management system in air traffic flow management," in *2024 AIAA DATC/IEEE 43rd Digital Avionics Systems Conference (DASC)*, 2024, pp. 1–10.
- [12] T. Vossen, M. Ball, R. Hoffman, and M. Wambsganss, "A general approach to equity in traffic flow management and its application to mitigating exemption bias in ground delay programs," *Air Traffic Control Quarterly*, vol. 11, no. 4, pp. 277–292, 2003.
- [13] Z. Ghahramani, *Unsupervised Learning*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 72–112.
- [14] A. Jain, P. Duin, and Jianchang Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.
- [15] T. chung Fu, "A review on time series data mining," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164–181, 2011.
- [16] S. Aghabozorgi, A. Seyed Shirkhorshidi, and T. Ying Wah, "Time-series clustering – a decade review," *Information Systems*, vol. 53, pp. 16–38, 2015.
- [17] R. P. Kumar and P. Nagabhushan, "Time series as a point - a novel approach for time series cluster visualization," in *Proceedings of the 2006 International Conference on Data Mining, DMIN 2006, Las Vegas, Nevada, USA, June 26-29, 2006*, S. F. Crone, S. Lessmann, and R. Stahlbock, Eds. CSREA Press, 2006, pp. 24–29.

- [18] X. Zhang, J. Liu, Y. Du, and T. Lv, "A novel clustering method on time series data," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11 891–11 900, 2011.
- [19] R. Tavenard, J. Faouzi, G. Vandewiele, F. Divo, G. Androz, C. Holtz, M. Payne, R. Yurchak, M. Rußwurm, K. Kolar, and E. Woods, "Tslearn, a machine learning toolkit for time series data," *J. Mach. Learn. Res.*, vol. 21, no. 1, Jan. 2020.
- [20] K. Schuetz, C. G. Schuetz, and S. Jaburek, "Privacy-preserving implementation of local search algorithms for collaboratively solving assignment problems in time-critical contexts," in *2023 IEEE Congress on Evolutionary Computation (CEC)*, 2023, pp. 1–10.
- [21] T. Harzfeld, "Evaluation of conformance criteria for preference indication in collaborative flight prioritization in air traffic flow management," Master's thesis, Johannes Kepler University Linz, 2025.