EXPLORATORY RESEARCH

sesar JOINT UNDERTAKING

D5.2 Final tactical model and results

Deliverable ID:	D5.2
Dissemination Level:	PU
Project Acronym:	BEACON
Grant:	893100
Call:	H2020-SESAR-2019-2
Topic:	SESAR-ER4-08-2019
Consortium Coordinator:	University of Westminster
Edition date:	09 February 2023
Edition:	01.00.06
Template Edition:	02.00.03

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Authoring & Approval

Authors of the document

Name / Beneficiary	Position / Title	Date
Gerald Gurtner / UoW	Project coordinator	07/11/2022
Tatjana Bolic / UoW	Project member	03/11/2022

Reviewers internal to the project

Name / Beneficiary	Position / Title	Date
Gerald Gurtner / UoW	Project coordinator	07/11/2022
Tatjana Bolic / UoW	Project member	07/11/2022

Reviewers external to the project

Name / Beneficiary	Position / Title	Date

Approved for submission to the SJU By - Representatives of all beneficiaries involved in the project

Name / Beneficiary	Position / Title	Date
Gérald Gurtner / UoW	Project Coordinator	28/11/2022

Rejected By - Representatives of beneficiaries involved in the project

Name and/or Beneficiary	Position / Title	Date

Document History

Edition	Date	Status	Name / Beneficiary	Justification
00.01.00	15/10/2022	Internal	Gerald Gurtner / UoW	Draft for SJU
01.00.01	03/11/2022	Internal	Gerald Gurtner / UoW	Internal draft for review
01.00.02	07/11/2022	Submitted	Gérald Gurtner / UoW	First submission to SJU
01.00.03	28/11/2022	Submitted	Gérald Gurtner / UoW	Resubmission after SJU feedback
01.00.04	17/01/2023	Submitted	Gérald Gurtner / UoW	Resubmission after second round of SJU comments
01.00.05	31/01/2023	Submitted	Gérald Gurtner /UoW	Resubmission after third round of SJU comments

EUROPEAN PARTNERSHIP

Co-funded by the European Union

01.00.06	09/02/2023	Submitted	Gérald Gurtner /UoW	Resubmission after fourth round of SJU comments

BEACON

BEHAVIOURAL ECONOMICS FOR ATM CONCEPTS

This deliverable is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 893100 under European Union's Horizon 2020 research and innovation programme.

Abstract

This deliverable presents the final results obtained with the model developed in BEACON to estimate the impact of new mechanisms for air traffic flow (ATFM) regulation[1] resolutions using in particular behavioural economics as part of the modelling process. Compared to the previous deliverable, it presents similar analyses, obtained on an extended geographical scope and more focused on understanding the multi-dimensional impact of the mechanisms. The results demonstrate that the impact on different airlines and airports is heterogeneous. Furthermore, it explores the relationship existing between regulation features, airline characteristics, and indicators like cost saved per flight. High-level conclusions on the impact of the different mechanisms are given, in terms of absolute and relative savings as well as in terms of equity and fairness.

Finally, the human-in-the-loop simulations performed during the project are described.

[1] Note that we use terms ATFM regulation and hotspot interchangeably in this deliverable.

Table of Contents

List of Tables

List of Figures

1 Introduction

In this deliverable, we explore the mechanisms defined during the course of the BEACON project. Similarly to what was done in D5.1, we use the tactical simulator Mercury to build a realistic dataset of regulations that are used in quick simulated 'games'. These games are played notionally by airlines – represented by humans or AI routines – who try to use different mechanisms in order to find collectively the best resolution of a given hotspot. As a reminder from D5.1, the games have two main ingredients: the central optimiser, that gathers preferences from airlines, and the airlines themselves, that express them.

These games were played in two settings. First, we organised human-in-the-loop simulations where members of the industry played as one airline (the other ones being played by an AI). This allowed us to gather important feedback on the mechanisms. Second, we performed fast-time simulations, with only AIs playing the games.

We have defined In D5.1 different types of behaviours for the AIs in order to explore the consequences of the mechanism resolution under different situations, including an 'honest' , 'rational' , and 'bounded' one. The 'honest' one gives its most truthful estimation of its costs to the central optimiser, the 'rational' one declares costs so that its profit is maximised, and the 'bounded' one does the same with a general utility of prospect function. Indeed, In economy in general, it is to be expected that human and corporate players would not be giving the most truthful costs if they can gain an advantage not doing so. This type of behaviour may affect adversely the resolution of the hotspot.

In D5.1 we focused on one airport (LFPG was used as an example) and on the inner workings of the model. Here, we expand the analyses to include more airports (21), we focus on most realistic efficiency estimations, and we explore the differential impact of the mechanisms on the airlines, depending on their size in the regulation for instance.

The deliverable is structured as followed:

- Section 2 presents the human-in-the-loop simulations performed during the course of the project and the feedback we got from the participants.
- Section 3 presents the methodology for the fast time simulations: the choice of airports, some preliminary descriptive analysis of the regulation dataset, and the experiments to be carried out.
- Section 4 presents the results obtained with the model. It shows the results of the calibration process for the Credit Mechanism (CM), results aggregated per airlines, per airports, and analysis of the impact of different variables on the efficiency of the mechanisms, equity metrics etc.
- Finally, section 5 draws some conclusions.

2 Human-in-the-loop simulations

Within the WP5 human-in-the-loop simulations (HITL), using real-time simulations, were performed. The primary goal of these experiments was to compare the behaviour of real dispatchers against their modelled counterparts. In particular, this task focused on estimating the importance of behaviours deviating from rationality.

To carry out the HITL simulations, the development of human-machine-interface (HMI) interface was necessary, and the adaptation of the Mercury model to enable the interaction between the participants and the simulator. The HMI enabled the participants to play the dispatcher role which would have been otherwise played by artificial intelligence subroutine in the simulator. The developed HMI and communication protocols are first described, followed by the description of the HITL experimental setup and the results obtained.

2.1 HMI and communication between the HMI and the simulator

2.1.1 HMI details

The first interface view displayed when the user opens the interface's URL is the **waiting view**. This particular view was designed to inform the user that the interface is waiting for data from the simulator. When the user opens the interface, the platform sends a request to the simulator and waits for the ATFM regulation data. This view will appear as a loading screen for as long as it takes for the simulator to find a new regulation and send the data.

Essentially, the user should just wait until the next screen appears, which happens when the simulator finds the hotspot. Figure 1 displays the waiting view.

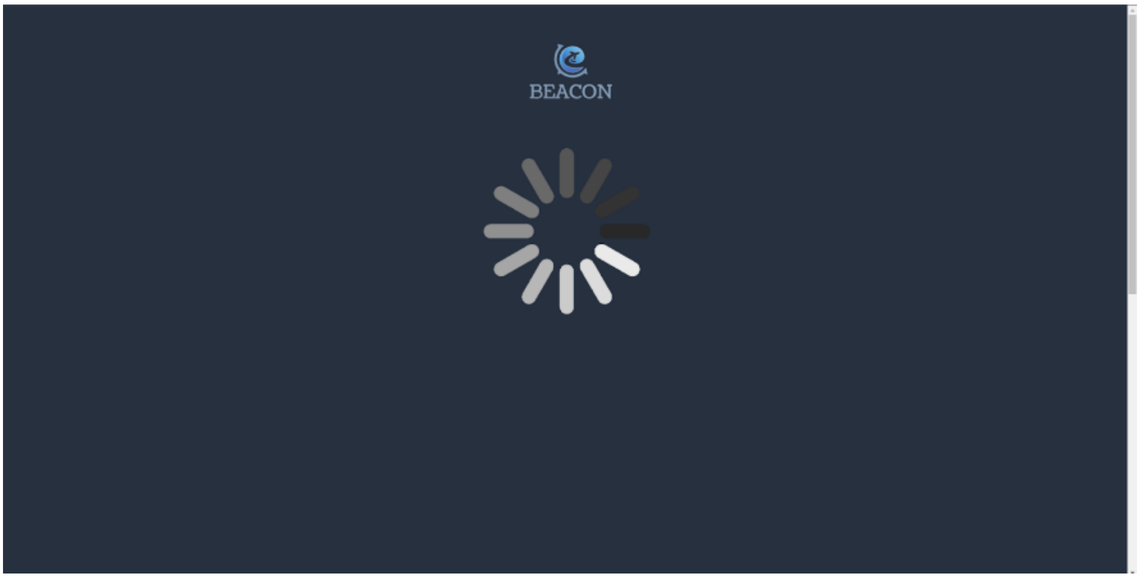


Figure 1: Waiting view.

The waiting view should redirect the user automatically to a **mechanism view** as soon as the interface receives data of a new hotspot. In this view, the user can see a title of the mechanism in use, and the table containing information on the flights within the hotspot.

The HMI acts as an interface between the model and human interactions. It receives information regarding the affected flights each time an ATFM regulation[2] occurs and displays this information on screen so that the participant can make flight prioritization decisions. The essential part of the process was deciding which information to show and how, settling on an interactive table. The table contains a series of data fields (columns) for each affected flight (row) with information related to schedule and regulation. In addition, the HMI contains a series of columns where the participant could input/alter the values for the flight, in order to carry out prioritization for flight in question.

The interface has two different mechanisms available: ***the UDPP+ISTOP mechanism*** and the ***Credits mechanism.*** The HMI view varies slightly with the type of mechanism but overall, the information on the impact of the hotspot on the flights is the same across both mechanisms. The columns meant to be filled by the user to carry out the prioritization of flights are easy to perceive: the cells are empty, and they are presented in a light blue colour indicating that they are editable.

Furthermore, to assists the user in decision-making process, the HMI allows the user to select a particular flight and display the costs across the flight would incur across different available slots (see Figure 2). Costs take into account costs of delay, duty and care for passengers, possible reactionary delay costs, etc, which are computed by the Mercury simulator.

Figure 2 shows the UDPP+ISTOP mechanism view, and the interactive table with information on flights in the hotspot. The first set of columns contains the data on *Schedule information*, followed by the *Regulation information* and the columns for input of decisions by users. For this mechanism, the editable columns are Margin and Jump.

[2] Note that we use terms “ATFM regulation” and “hotspot” interchangeably in this deliverable.

Figure 2: UDPP+ISTOP mechanism dummy view, displaying costs for two flights below the table.

As can be seen in Figure 2, if the user ticks the check-boxes in the very first column of the table, the costs of those flights appear at the bottom of the screen. The costs plot has the name of the flight on the left side of the y-axis to avoid confusion. The costs for flights are displayed in the order of being chosen by the user. The y-axis denotes costs, and the x-axis denotes available slots in the regulation.

Schedule information consists of the flight id, origin, destination, scheduled time of arrival (STA), connection buffer, ground time (GT) buffer, and estimated time of arrival (ETA). Figure 3 shows part one of the table with the synthetic data used in the simulation[3], where more details can be seen in the Regulation information columns. The columns are controlled time of arrival (CTA), slot, delay, passengers, connection buffer, GT buffer, connection details, pax costs, and total flight costs. The CTA gives the time assigned to the flight by the First Planned First Served (FPFS) algorithm. 'Slot' gives the index of the slot within the regulation and 'delay' gives the delay assigned through the assigned slot (given in minutes).

[3] Note that code for Air France (AFR) is used to describe the flights in realistic way, as the data used in the simulation is synthetic, not real one.

BEACON

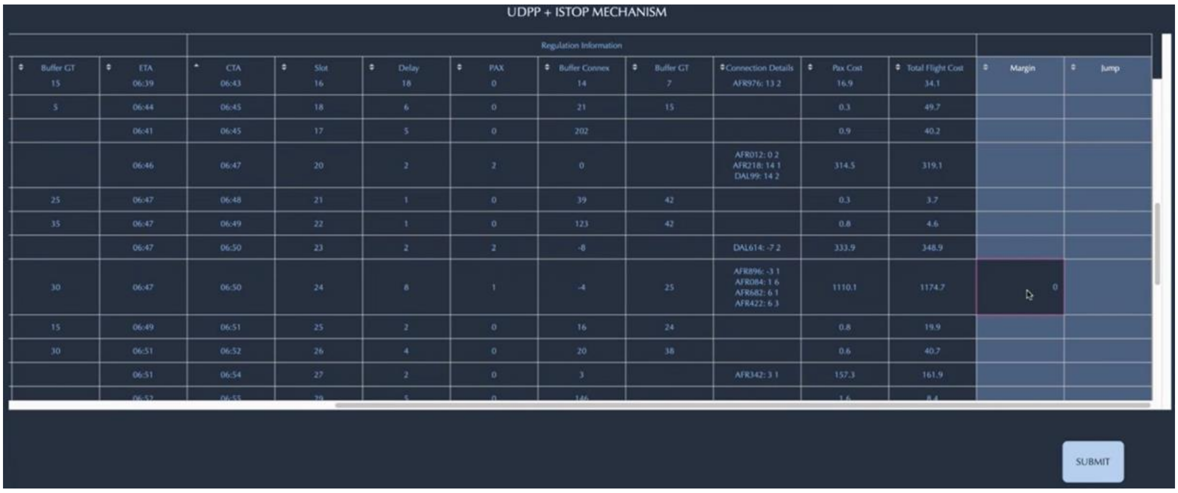
UDPP + ISTOP MECHANISM

	Schedule Information											Regulation Information	
	Flight	Origin	Destination	STA	Buffer Connex	Buffer GT	ETA	CTA	Slot	Delay	PAX	Buffer Connex	
	AFR1377	LEBB	LFPG	06:51	1		06:51	06:54	27	2	0	3	
	AFR1559	EGNT	LFPG	06:47	21	25	06:47	06:48	21	1	0	39	
	AFR1765	EKBI	LFPG	07:02	166		07:02	07:05	38	2	0	177	
	AFR342D	LSGG	LFPG	06:20	1	20	06:20	06:20	1	0	0	11	
	AFR1019	EDDF	LFPG	06:30	1	15	06:30	06:33	8	3	0	14	
	AFR1513	LIML	LFPG	06:49	6	15	06:49	06:51	25	2	0	16	
	AFR1579	EIDW	LFPG	07:03	6	10	07:03	07:09	42	5	0	18	
	AF677YS	LFRN	LFPG	06:47	1		06:47	06:50	23	2	2	-8	
	AFR183P	EGLL	LFPG	06:37	1	30	06:37	06:38	12	0	0	14	
	AFR1415	LSZH	LFPG	06:47	116	35	06:47	06:49	22	1	0	123	
	AF651PQ	LFLL	LFPG	07:04	1	25	07:04	07:10	44	5	0	1	
	AFR801T	EGLL	LFPG	07:07	6	55	07:07	07:15	48	7	0	7	

Figure 3. UDPP+ISTOP view with the synthetic data from simulation, part one of the table.

As can be noted better from Figure 4, the connection details field contains the id of the subsequent connecting flight and the number of passengers (divided by class) that should reach it. Pax cost gives the costs at the assigned CTA/slot that are attributable to passenger duty of care. The total flight costs field gives the total cost of delay at the assigned CTA/slot. For the costs at other slots in regulation, the user should tick the check-box in the first column, which displays the costs across slots in the regulation.

Another functionality is that all columns can be sorted numerically using the arrows at the column title to indicate if the sorting is ascending or descending. Furthermore, when the user selects a cell to edit, the cell will be highlighted and with a different colour as we can see in from Figure 2Figure 4 as well. Regarding the editable columns, it is not supposed to be possible for the user to fill it with negative or decimal values. Consequently, if the user tries to insert a decimal value, it will be automatically rounded. Likewise, if the value inserted is negative the cell will turn red, as shown in Figure 2.



UDPP + ISTOP MECHANISM

		Regulation Information									
Buffer GT	ETA	CTA	Slot	Delay	PAX	Buffer Connex	Buffer GT	Connection Details	Pax Cost	Total Flight Cost	Margin	Jump
15	06:39	06:43	16	18	0	14	7	AFR976: 13 2	16.9	34.1		
5	06:44	06:45	18	6	0	21	15		0.3	49.7		
	06:41	06:45	17	5	0	202			0.9	40.2		
	06:46	06:47	20	2	2	0		AFR012: 0 2 AFR218: 14 1 DAL99: 14 2	314.5	319.1		
25	06:47	06:48	21	1	0	39	42		0.3	3.7		
35	06:47	06:49	22	1	0	123	42		0.8	4.6		
	06:47	06:50	23	2	2	-8		DAL614: -7 2	333.9	348.9		
30	06:47	06:50	24	8	1	-4	25	AFR896: -3 1 AFR084: 1 6 AFR682: 6 1 AFR422: 6 3	1110.1	1174.7	0	
15	06:49	06:51	25	2	0	16	24		0.8	19.9		
30	06:51	06:52	26	4	0	20	38		0.6	40.7		
	06:51	06:54	27	2	0	3		AFR342: 3 1	157.3	161.9		

SUBMIT

Figure 4. UDPP+ISTOP view with the synthetic data from simulation, part two of the table.

When the user is finished editing the cells with the response to solve the hotspot, the information is sent to the simulator by pressing the “Submit” button. After pressing the button, the data is sent, and a message is displayed to the user to let them know that the data is sent (see Figure 5). When the results are elaborated, the user is automatically redirected to the **results view.**

Figure 5. Data submission message.

Credit mechanism view. The information part of the table is the same as for UDPP+ISTOP mechanism. New elements include the name of the mechanism and the amount of credits the user has, which is located at the top right of the table in a placeholder. Also, at the top left of the table, the user may find information indicating the cost of each variable as well as the amount of credits available, see Figure 6.

Another difference lies in the fact that the table loads with the initial/standard values (called default parameters in D5.1 and below in this deliverable) for the specific regulation. The user may change or remove the values of the editable cells without violating the experiment restrictions. When the user is happy with the changes, the Submit button is pressed, which then displays the “data is sent” message (Figure 5) and eventually switches to Results view (after the simulator sends back the elaborated data).

Figure 6. Credit mechanism, dummy view.

Results view. The simulator computes the results of the experiment with the data of the user's response and sends the data to be displayed in the HMI, results view.

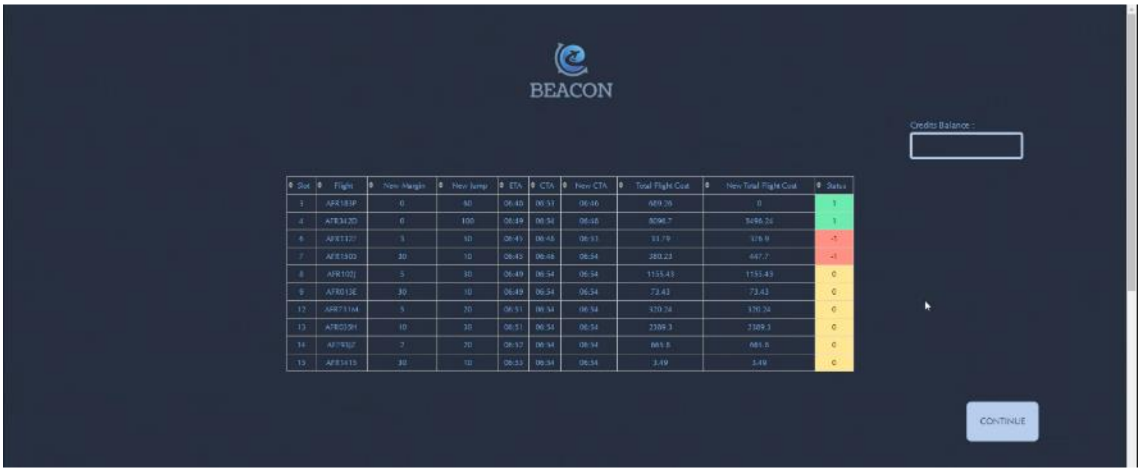


Figure 7. Results view.

There is also a placeholder with the remaining credits from the user's response, however, this concerns only the credits mechanisms. As such, the field is empty for UDPP+ISTOP, and will contain he correct information for the Credits mechanism.

Finally, the experiment ends with this last view, so the user may leave the interface by closing the tab, or start a new experiment by hitting the button "Continue". Such will automatically replace the results view with the waiting view, where the process starts all over again.

This section concerns the communication established between the Human in the Loop Interface and the simulator. Therefore, contains a brief introduction to the communication protocol used to achieve a successful communication, as well as the communication flow and the respective messages exchange during.

2.1.2 Communication between HMI and simulator

2.1.2.1 Architecture

The protocol chosen to establish the communication between the developed interface and the model was the ZeroMQ library.

- **ZeroMQ**

ZeroMQ is a library that allows the implementation of communication and messaging between applications and processes through a fast and asynchronously process. Therefore, it is a high-performance asynchronous messaging library, aimed at use in distributed or concurrent applications.

Likewise, it supports common messaging patterns such as publisher/subscriber, request/reply, client/server, among others. Also, it has a variety of transports, among them, TCP, in-process, inter-process, multicast and WebSocket, making inter-process messaging as simple as inter-thread messaging. This keeps the code clear, modular and extremely easy to scale.

- **Client/Server Pattern**

The pattern settled to implement the communication between the interface and the simulator is the client/server pattern. Since the interface will depend directly on the data provided by the simulator to know which mechanism should present to the user, it was settled that the interface should act as a client and the simulator as a server, see Figure 8.

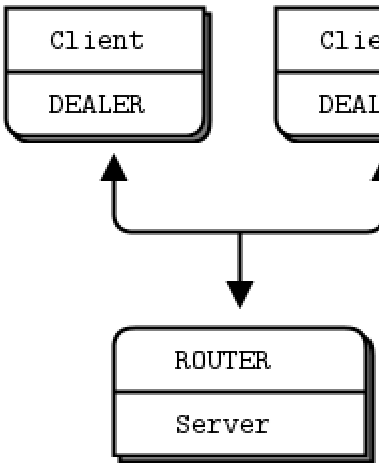


Figure 8. Client/server pattern

To clarify how the pattern operates, here are some steps important to notice:

- The clients (in our case only one, the interface) connect to the server and send requests.
- For each request, the server sends 0 or more replies.
- The clients can send multiple requests without waiting for a reply.
- The server can send multiple replies without waiting for new requests.

2.1.2.2 Communication Flow HMI Simulator

After selecting the pattern and protocol to use, it was established a communication flow to make sure that everything worked as expected.

The communication between the HMI and the simulator takes the following sequence of steps, see Figure 9:

1. The simulator, acting as a server, is initiated, and waits for a request.
2. The HMI sends a request resorting to zmq.REQ to request data to the simulator.
3. The simulator responds to this request using zmq.REP.
    a. The response should be the data from the UDPP+ISTOP mechanism or the credits' mechanism, which should respect a previously established data format.
4. When the simulator sends the response, the HMI will show to the user the table correspondent to that data response.

5. The user performs a series of input actions into the table columns. This information is sent back to the simulator by pressing a button labelled as "Submit". Such will cause the HMI to send to the simulator the data from the user resorting to zmq.REQ.
6. After calculating the results with the user's input, the simulator responds to this request with the finished data resorting to the zmq.REP.
7. The HMI receives as a response the finish data from the simulator and presents to the user the finish data table.
    a. In this final screen the user does not have to fill values, it serves only to see the final order of the flights and some details of the hotspot. Therefore, if the user wants to start another experiment, just has to press a button identified as "Continue".
8. If the user presses "Continue", the HMI will again send a request to the simulator and wait for data, just like in the beginning. Therefore, the process repeats without intermediate messages.

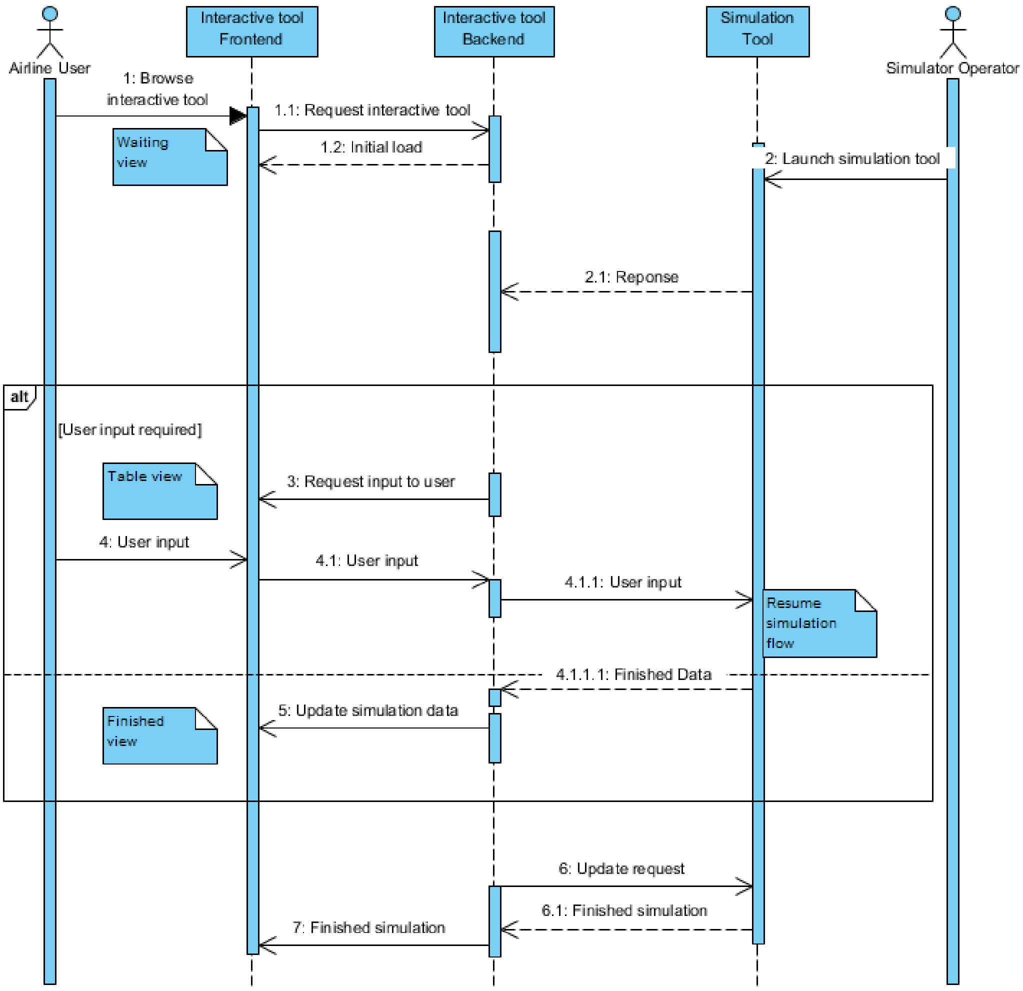


Figure 9: Communication flow between HMI and simulator.

Message Format. To ensure the communication, a message format had to be established in a way that the HMI could validate the messages coming from the simulator.

To agree on the data that will be sent and receive during the communication, JSON SCHEMAS were elaborated.

A JSON SCHEMA is a vocabulary that allows the developer to annotate and validate JSON documents. Essentially, a JSON Schema provides a contract for what JSON data is required for a given application and how to interact with it. JSON Schema is intended to define validation, documentation, hyperlink navigation, and interaction control of JSON data.

It is important to highlight some of the many benefits of applying JSON SCHEMAs in projects. Some of them are the following:

- Description of the existing data formats;
- Provides clear human-machine readable documentation;
- Validates data which is useful for:
    - Automated testing;
    - Ensuring quality of the submitted data;

Hence, in this project we elaborated several JSON SCHEMAS and use them to secure a proper communication. It was established a JSON SCHEMA for each type of message exchanged during the communication:

- From the Client/HMI to the Server/simulator:
    - Request data message;
    - User input data message;
- From the Server/Simulator to the HMI:
    - Message with the data from the UDPP+ISTOP mechanism;
    - Message with the data from the credits' mechanism;
    - Finished data message;

Every time the HMI sent a message to the simulator, it was validated first to ensure that the simulator wouldn't receive incorrect messages that could affect the normal flow of the communication.

Likewise, every time the HMI received a message from the simulator, that message would be validated first to guarantee that all the information corresponded to the expected before presenting in a table format.

2.2 Description of the experiments

Before starting the HITL exercises, the consortium had to complete the following steps:

1. Recruitment of participants, among the employees of the BEACON Advisory Board.
2. Distribution on the Participant information sheet and the consent form to participate in the exercises. The participants needed to sign and email the signed document to the consortium's data officer before starting the exercises, as per procedures described in D8.1, D8.2 and D8.3.
3. Set up of pseudo-anonymisation procedure for the HITL results (so as not to be able to trace specific results to specific participants), following procedures described in D8.2.
4. Conducted a brief on **introduction to the project**, **mechanisms and HITL simulations** to the recruited participants. The objectives of the HITL experiments were presented, highlighting the dimension of behavioural models in prioritization mechanisms and stressing that the participants should act as if in operations, but using the HMI and the information available through it.
5. Before each exercise, the **HMI was described again** to the participants, as some exercises took place days or weeks after the introductory brief. The following was described again:
    a. Description of the HMI: acronyms, mode of use.
    b. Description of how they should interact with the HMI, and what data will be collected (i.e. just numbers inserted and timing of the hotspot resolution, no personal or identifiable data).

Exercise was composed of introduction, and two sessions with BEACON mechanisms. Each session started with ISTOP or Credit mechanism in a random manner. The BEACON team was in contact with the participant throughout the exercise, to troubleshoot any glitches, or answer questions on the HMI or mechanisms' characteristics. During the exercise, the team members were collecting feedback coming from the conversations. At the end of the exercise, the participants were asked to respond (verbally) to a battery of qualitative questions, and to fill in the system usability survey (SUS) for the two mechanisms. The questionnaire can be found in Appendix C.

Qualitative feedback questions:

- Did you feel comfortable exchanging slots with other airlines?
- Was it an issue that you could not see the results for other airlines? If so, in what way?
- Did you make your choices purely on cost minimisation? If 'no', please summarise other factors driving your decisions.
- When making your choices, what were the main objectives you had in mind?
- What did you think about the ISTOP mechanism? Were you expecting these results?
- What did you think about the credits mechanism? Were you expecting these results?

2.3 Analysis of the feedback – verbal and questionnaire

None of the participants had issues with exchanging slots with other airlines. Furthermore, they did not mind not having access to the results for other airlines, as they were focusing on the best solutions they could provide for their own airline.

Some participants made their choices purely using the cost information presented via the HMI, while others took into account other factors, as they would in their daily job. These ranged from trying to obtain the best on-time performance for their flights, over favouring flights with connections, minimising the minutes of delay for passenger, to taking care of higher passenger categories (i.e. business passengers), sometimes assuming that the change of aircraft would be possible for the future rotations. When taking only the costs into account, the participants tried to obtain the slot just before a big jump in the costs, while ensuring that thus obtained ETA would be feasible.

The opinions on the ISTOP mechanism span a spectrum. Some participants declared that it was not intuitive and could not understand the consequences of submitted choices. Other participants declared that the mechanism was understandable, even if it required some learning to get the feeling. The possibility of implementing own scales for margins and jumps was appreciated. These participants noted the improvement in the costs using the ISTOP.

The opinions on Credit mechanism were also divided. For some participants it was easier as this mechanism offered standard values. In case of doubt (in a sense of what to do with a particular flight), the standard values were left in place. Although other participants questioned the meaning of standard values as it was not clear to them how those were set. What was appreciated in this mechanism was that there was a clear trade-off in setting up jumps and margins – for a better position, some credits needed to be spent, and for accepting a worse one, credits were gained. Thus, in a participant's words, "there is a need to give up on some flights' time to build credits, to use it to protect the most valuable flights." The time of creation and roll over of credits was discussed, as something that would be useful to know for the use of the mechanism.

Furthermore, the participants were asked to complete the questionnaire form to collect system usability score ***(SUS)*** scores on both mechanisms. (The questionnaire is in Appendix C).

Average score for Credit mechanism is 65, while it is 70 for ISTOP. Apparently, usability of ISTOP is slightly better. However, with just 6 respondents, it is not possible to test the scores for statistical significance.

Lessons learned regarding the organisation and running of HITL exercises:

- The cost information for flights (in function of the delay) was much appreciated by participants.
- Clear indication of cost savings through the choices is needed by participants. The HMI did show the initial regulated costs and costs after the submission of choices. However, the visualisation of this information could have been better. The information was just displayed, so the participants should have gone through each flight and checked, remembering what they chose for each flight. We appreciate that this is not a very good indication, and that it could be improved.
- When the limited time with participants is available, it is very important to have an efficient introductory session explaining the exercise objective, mechanisms and the HMI details. The consortium provided this briefing. However, we found that the same briefing did not work the same with all the participants, as for some all was rather clear, while for some others the learning curve was much slower.
- One needs to take into account where the participants are connecting from. For example, most were connecting from their work computers, that had varied degrees of connection security in place. In some instances that delayed, or made the exercise impossible, as the HMI was designed

with the goal of collecting the information, but not to be at a level that can be certified as secure connection by the more serious connection security filters.

- One needs to take into account the browsers the participants would use. Most of the participants used Google Chrome or Microsoft Edge, with which HMI worked correctly. However, the setup did not work properly with Safari.

3 Regulation analysis and experiments

Compared to results produced for D5.1, we are interested in this deliverable in expanding the scope of the deliverable by running the 'games' defined in D5.1 on additional airports. However, on one hand, we lack the computational power to run the games on every airport in Europe, and on the other hand, we are not interested in running games on very small regulations. Hence, we start by selecting the airports on which the simulations will be run.

3.1 Choice of airports

One of the goals of this deliverable is to explore in detail the effect of the mechanisms in different situations, for different actors. In particular, we are interested in the case where the number of flights per airlines is well balanced, and cases where it is not.

3.1.1 Herfindahl-Hirschmann Index (HHI)

In order to perform the selection, we use the Herfindahl-Hirschmann Index (HHI), well known in economy to measure the degree of competition existing in a market. In short, it computes the sum of the squares of the market shares for different competitors. Hence, the higher the index, the higher the concentration of the market is. Here, the index is defined using the share in the number of flights of different airlines at a given airport. We use the normalised version:

$$HHI = \frac{(\sum_i^N s_i^2 - 1/N)}{(1 - 1/N)},$$

Where s_i is the share in the number of flights of airline i. The index goes from 0 (perfect balance between airlines) to 1 (one airline has the entire market).

3.1.2 Relationship between HHI and airport size

For the next step, we used the dataset prepared for Mercury (see D2.1, D2.2, and D5.1 for more details). We computed the HHI for each airport in the dataset (restricted to the ECAC area). In Figure 10 we plot the relationship between the HHI and the size of the airport, measured as the total number of flights operating at the airport in a day. In this figure, we also categorised the airports based on their HHI value. For this, the common categories defined by the US Department of Justice [13] are used, defined as:

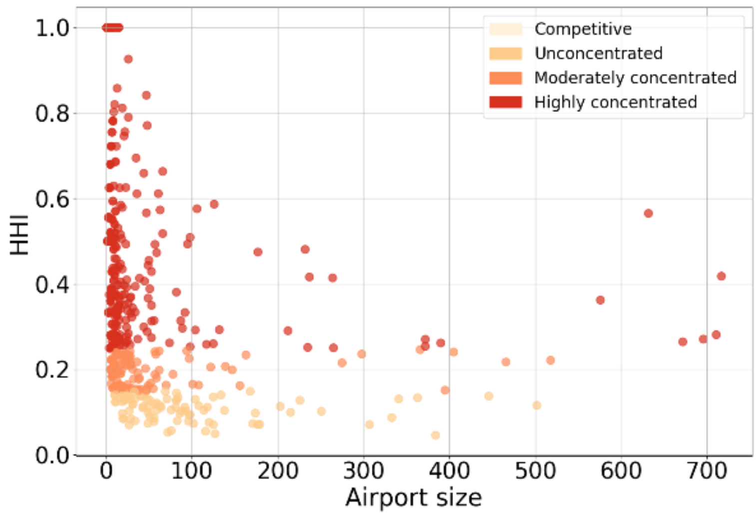


Figure 10: HHI of each airport as a function of the airport size. The points are coloured based on the category of the airport, see text.

- Competitive: HHI below 0.01,
- Unconcentrated: HHI above 0.01 and below 0.15,
- Moderately concentrated: HHI above 0.15 and below 0.25,
- Highly concentrated: HHI above 0.25.

The categories are reflected in the figure with the different colours.

Next, we decided to select the 10 airports in each HHI category with the highest sizes, in order to have sufficiently high number of flights in regulations while testing different levels of competition.

Among these 30 airports, some of them are unusable in our simulations, for various reasons. In particular, airports from UK (Gatwick, Heathrow) cannot be used because of the way flights are regulated there. Indeed, traffic at these airports is regulated through the use of ATFM delays applied to the airspace surrounding the airports. In this study we are focused on arrival regulations applied at the airport itself, and thus we cannot directly include these airports in the analysis. It is left for further research to properly select the airspace regulations that regulate the arrival traffic at these airports. In the end, we are left with a dataset of 21 airports.

3.2 Dataset Analysis

We are interested in exploring the regulations applied that are present in each airport, in terms of sizes (number of airlines in the regulation) and structure (distribution of the different airlines involved in the regulations). In order to do this, we created a dataset with the help of the Mercury simulator. For each airport, we simulate 200 times a scenario where only the flights coming from or going to this airport are simulated. We then extract the regulation information in a similar way as described in D5.1, i.e. gathering information on flights, airlines, cost functions etc.

In this section we use this dataset to explore the regulation structures.

3.2.1 Frequency of regulations at airports

Since the focus of this deliverable is on the computation of realistic KPIs, it is important to have a look at the distribution of regulations across the airports selected for our experiments. Indeed, if for example a given airport displays a high efficiency for one of the mechanisms, it may drive the average over all airports up if regulations hits it very often. On other words, the policymaker should always remember that aggregated values depend in general a lot on underlying distributions.

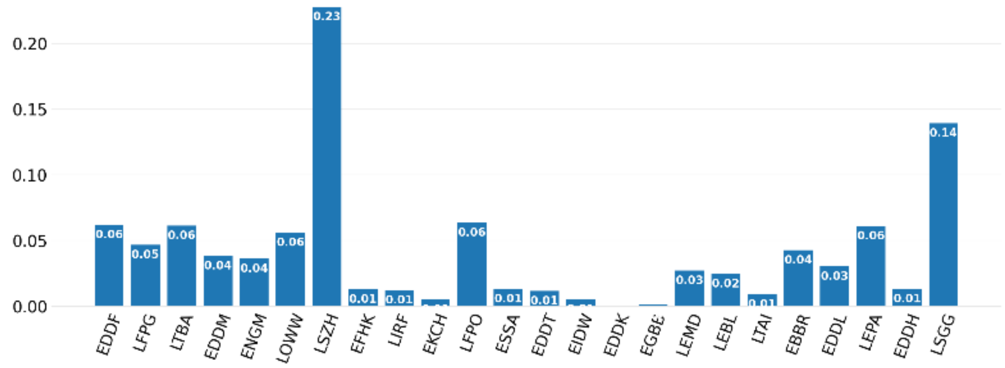


Figure 11: Relative frequencies of regulations happening at the different airports in our dataset.

Figure 11 illustrates the relative frequency in the apparition of regulations for different airports in our dataset. The reason for which this pattern appears is not important for this deliverable. Instead, it is important to remember that these frequencies are unbalanced, with some airports displaying a much higher frequency that others. This is means that some of the KPIs computed in the results section on the efficiency of the mechanisms are very much driven by these airports.

3.2.2 Regulation size

We then move on to the size of the regulation. In Figure 12 we show a violin plot[4] of the distribution of sizes of regulation, in terms of number of flights, for each airport in the dataset (sorted by decreasing average regulation size). It is important to note on this figure that airports have very different distributions. While at some airports the number of flights in regulation gravitates towards a fixed number of flights (LEBL, EDDT), others have a very high heterogeneity (LFPG, EFHK, EIDW). It is likely that the BEACON mechanisms will have different efficiency levels in these cases.

[4] As a reminder, a violin plot is a box plot superimposed with a Kernel Density Estimate (KDE), essentially a smooth representation of the underlying discrete distribution.

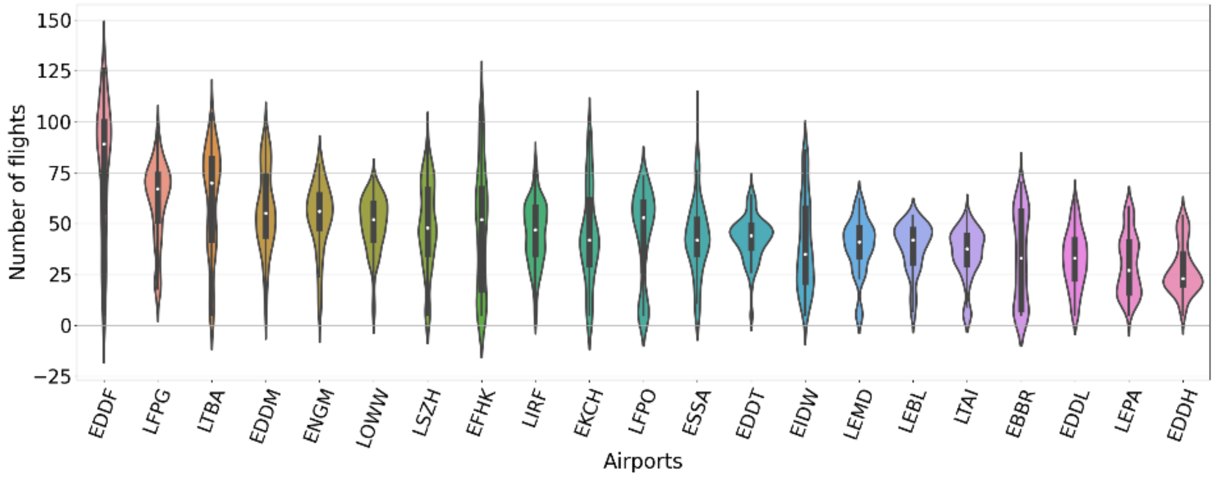


Figure 12: Violin plots (box plot + Kernel Density Estimation) showing the distribution of the number of flights in regulations per airport.

3.2.3 Airlines in regulation

The number of airlines in a regulation is also an important variable. Figure 13 shows a violin plot of the distribution of the number of airlines within regulations. While one can notice the heterogeneity among airports, it is also important to note that there are less differences between airports in terms of the typical number of airlines involved in the regulation.

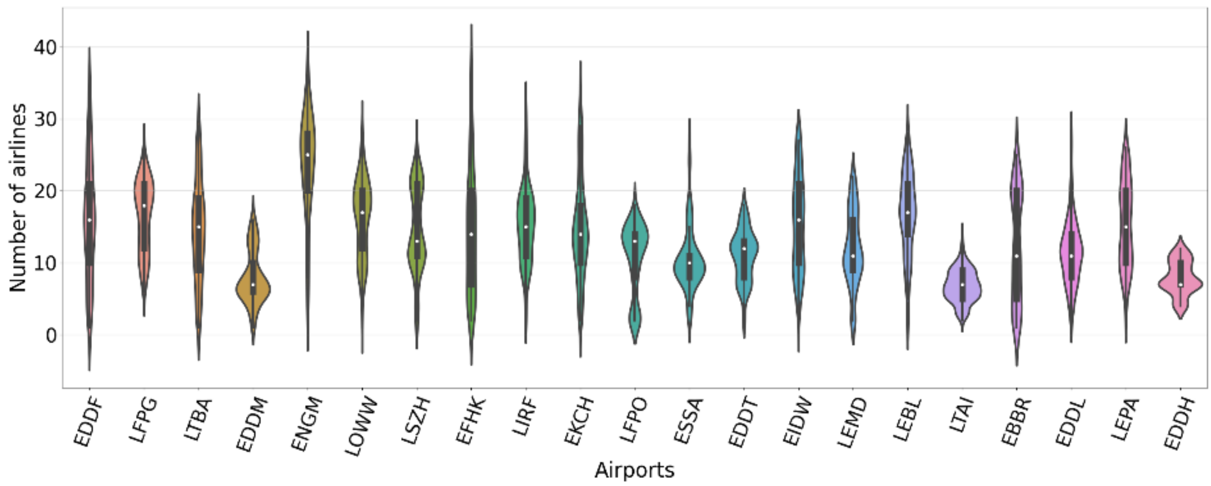


Figure 13: Violin plots showing the distribution of the airline numbers in each airport.

The number of flights and airlines in regulations obviously depends on the size of the airports. Figure 14, explores the relationship between these variables. While it is quite obvious that the average number of flights in regulation would be higher when the airport is bigger (with some high variability nonetheless), it is striking to see that the number of airlines involved does not depend very much on the size of the airport. Hence, we are expecting the *structure* of the regulation, e.g. the flights per airline, to be quite different.

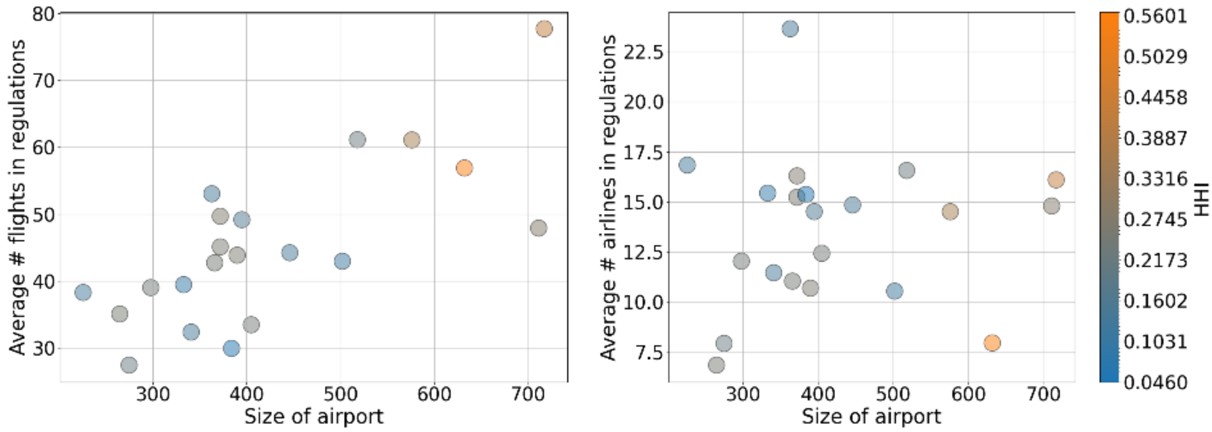


Figure 14: Left: average number of flights in regulations as a function of the size of the airport (number of movements). Right: Average number of airlines in regulations as a function of the size of the airport. Colour of circles: HHI of the corresponding airports.

3.2.4 Flights per airlines

This is exactly what we explore in Figure 15, where we show violin plots for the distribution of the number of flights per airlines (number of flights belonging to each airline in each regulation at a given airport). The heterogeneity is clearly visible: while for instance EDDF, EDDM, or LTAI have a very wide distributions, other airports have a fairly stable number of flights per airline (ENGM). This will be crucial when we compare mechanisms, since the number of flights per airline conditions the efficiency of some mechanisms, in particular UDPP, ISTOP, and NNBOUND (see Table 1 for mechanism description). Figure 16 shows the same relationship in another way. Superimposing all the KDEs in one figure, it is clear that a wide range of variances exist for these airports.

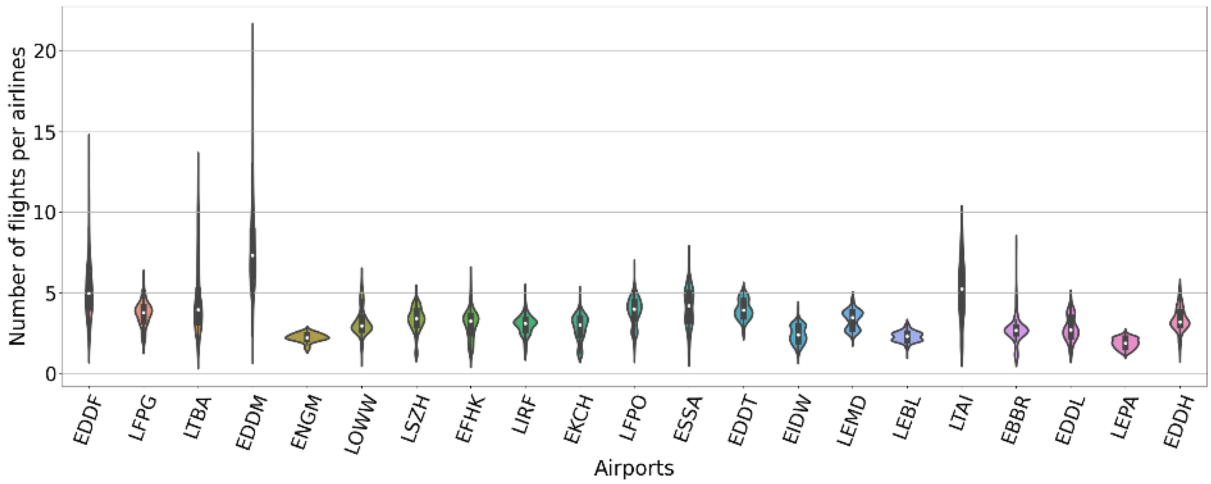


Figure 15: Violin plots showing the distribution of the number of flights per airlines for each airport.

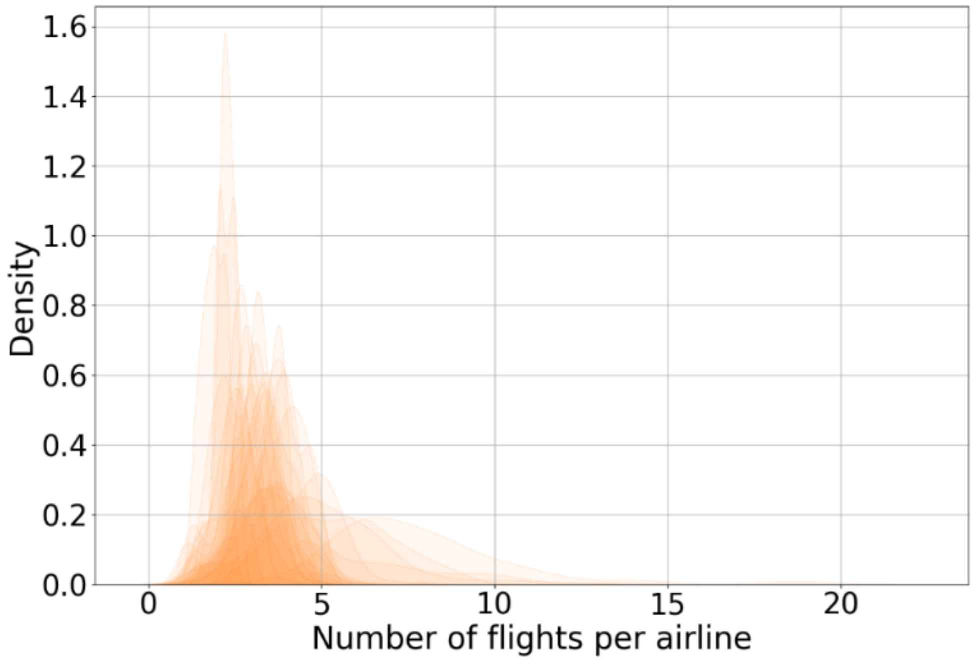


Figure 16: Kernel Density estimation for the number of flights per airline for different airports.

3.2.5 Low- and high-volume users

These considerations lead us to the final analysis we perform on the regulations themselves. In this section, we are interested in the recurring number of flights that each airline has in series of regulations. Indeed, if an airline has only one flight every now and then in regulations at EDDF, then it will be difficult for them to lower their costs, as least with UDPP and ISTOP mechanisms. We thus need to look at how many flights a given airline has a in regulation, on average.

Figure 17 shows examples for three airports, EDDF, LIRF, and LEMD. Each of the airports belongs to a different HHI category. It is clear that EDDF has an overwhelmingly major player, with a high number of very, very small competitors. The situation in LIRF is less drastic, while LEMD can be considered to have at least 4 main players.

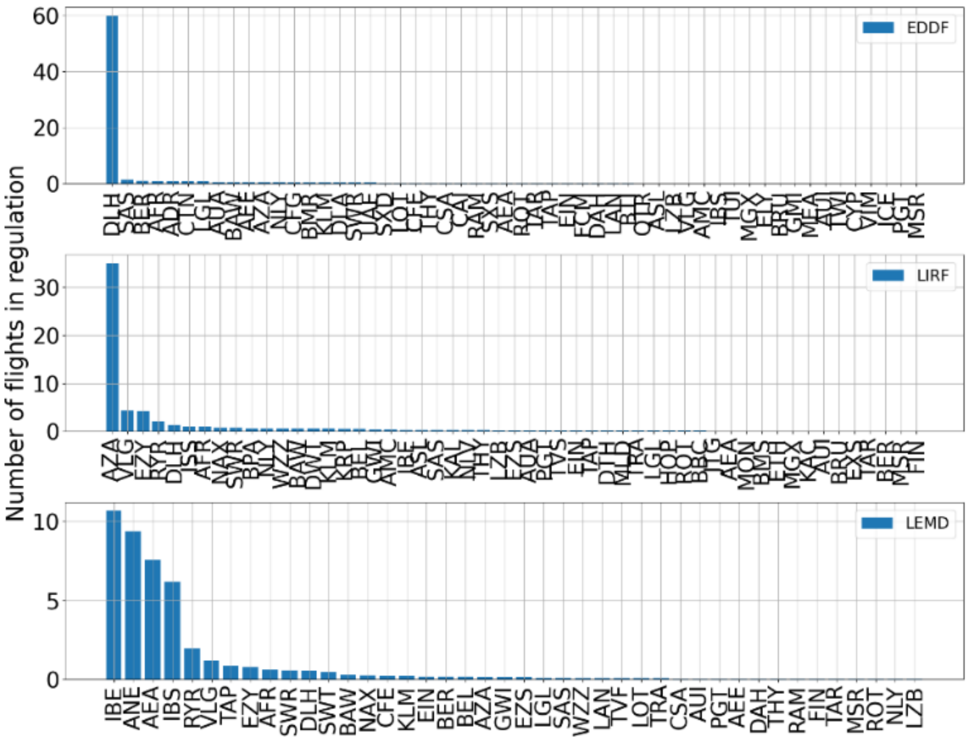


Figure 17: Number of flights in regulations of each airline at three different airports: EDDF, LIRF, and LEMD.

A more systematic approach to this is to plot some the above distributions per category of airport. In Figure 18 we superimposed all these distributions, for each category. It is clear that the airports of the first category display a much higher concentration in their distributions of flights, as expected, but that there is also some heteronegeneity among them.

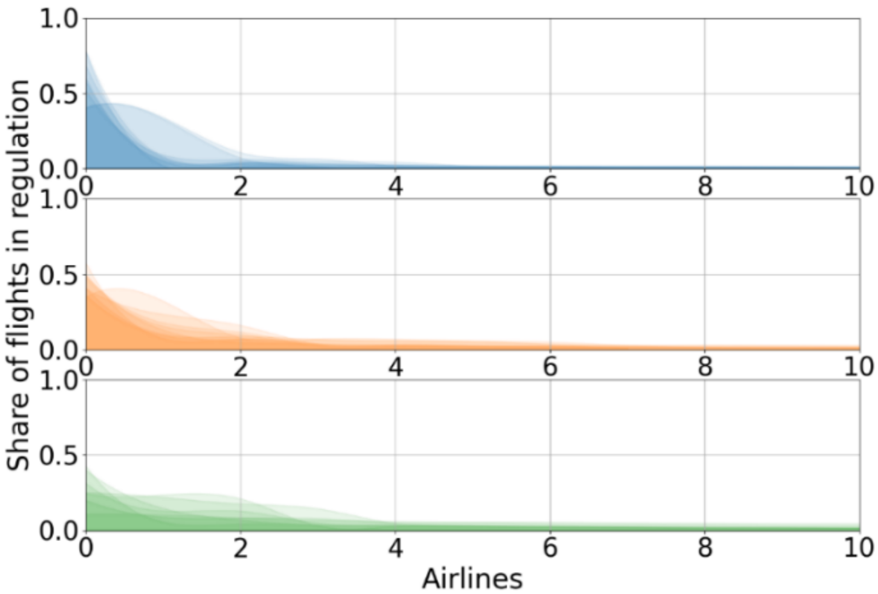


Figure 18: Kernel Density Estimations of the share of flights in regulations at different airports, in three different categories of airport: high HHI (high concentration) in blue, medium HHI, and low concentration.

Based on these distributions, we are able to define an important categorisation for the airlines: low- and high-volume users. One can use several definitions for this, based on quantiles for instance. Here, we used a simple criterion: we say that airline A is a low-volume user (LVU) if its share of flights in regulations is on average smaller than 5%. All the other airlines are high-volume users (HVUs). Figure 19 and Figure 20 show the application of this rule on our dataset by displaying the number (resp. the proportion) of airlines in each category. Note that it might be useful to consider an alternative categorisation of airspace users too, which we present in Appendix A.

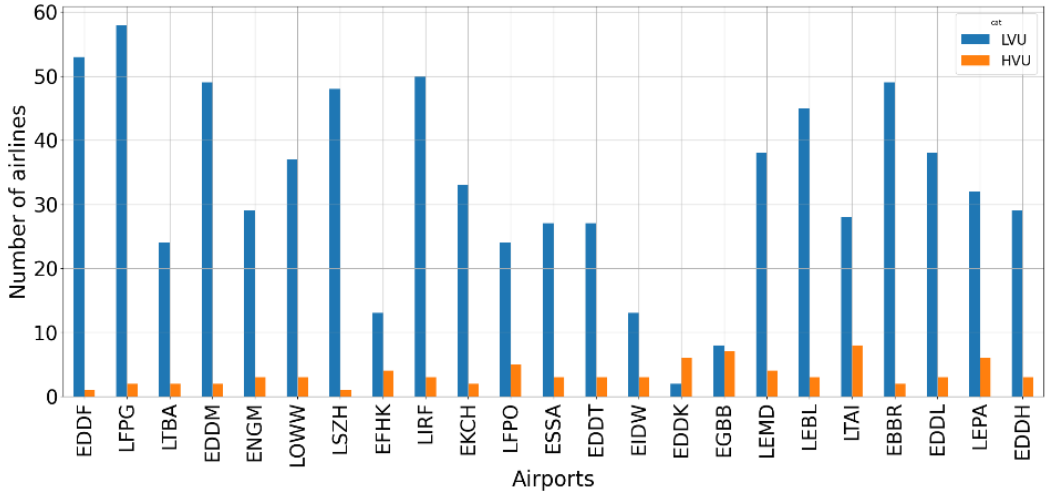


Figure 19: Number of airlines in each category, LVU (Low volume users) and HVU (High volume users) in each airport of the dataset.

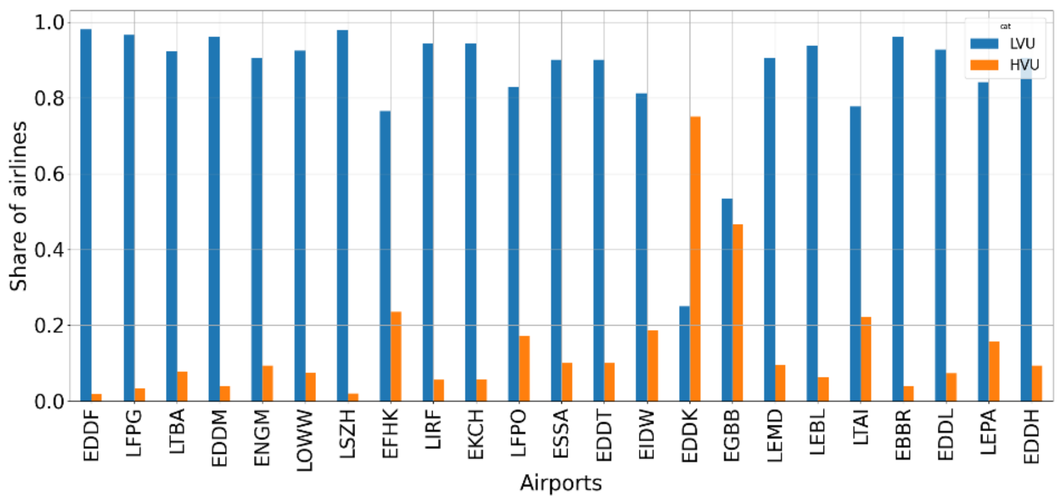


Figure 20: Share of airlines in each category, LVU (Low volume users) and HVU (High volume users) in each airport of the dataset.

3.3 Experiments

The experiments carried out on the dataset presented above are in line with the ones presented D5.1. They consist in a series of 'games' played by agents representing airlines when a regulation hits:

- The NM draws a regulation from the dataset.
- The NM sends information on the regulation to the airlines involved.
- The airlines make decisions for each of their flights, in the form of parameters related to the mechanism used for the resolution (UDPP, ISTOP, etc).
- The airlines' decisions are sent to the NM, which computes the final allocation based on the mechanism selected.
- True costs before (FPFS allocation) and after (final allocation) are computed for the impact assessment.

The same mechanisms are tested as in D5.1 and are briefly described in Table 1. The agents can be of different types, which are described in , again coming from D5.1. We also left the "comment" columns, related to some important characteristics of the respective agents this time. More details about them can be found in D5.1.

Table 2.

3.3.1 Mechanisms

The mechanisms tested in D5.2 are presented in Table 1, which is coming from D5.1. We left the "comment" columns pertaining to some important characteristics of the respective mechanisms. More details about them can be found in D5.1.

Table 1: Mechanisms tested in D5.2.

Mechanism	Description	Comment
UDPP	Based on their own true cost functions, airlines set priorities to their flights. The NM then 'merges' the priorities to have the final allocation.	Use standard UDPP concepts like Selective Priorities.
UDPP+ISTOP	After applying UDPP, the airlines give some parameters to the NM, used to approximate their cost function. The NM suggests two-airline swaps that are beneficial to everybody. The airlines always accept the suggestions.	The approximation is the same used across different mechanisms, see below.
UDPP+ISTOP_TRUE	Idem above, except the airlines provide the real cost function to the NM. Used for benchmarking.	The efficiency of this mechanism should be between UDPP and NNBOUND.

NNBOUND	The airlines give some parameters to the NM, used to approximate their cost function. The NM then performs an optimisation, seeking the minimum total cost across airlines, while keeping changes of cost above zero for each airline (i.e. nobody can lose from the final allocation).	Low volume users will not be affected by this mechanism, since they cannot lose from it and there is no intra-airline suitable swaps for them.
NNBOUND_TRUE	Idem above except that airlines give their true costs to the NM. Used for benchmarking.	
GLOBAL	The airlines give some parameters to the NM, used to approximate their cost function. The NM then performs an optimisation, seeking the minimum total cost across airlines.	Idem NNBOUND but without the "no-negative gain" constraint.
GLOBAL_TRUE	Idem above except that airlines give their true costs to the NM.	Best possible outcome from the social point of view (total cost). Equity likely to be very low.
CM	The airlines give some parameters to the NM, used to approximate their cost function. Each airline pays a certain number of virtual credits based on which parameters they chose. The NM then uses the parameters to rebuild cost functions and finds the global optimum (total minimum cost).	Except for the approximation, this mechanism should be one of the best from the efficiency point of view, while achieving equity on the long run. Can be considered as a simple extension of GLOBAL.

3.3.2 Agent types

, again coming from D5.1. We also left the "comment" columns, related to some important characteristics of the respective agents this time. More details about them can be found in D5.1.

Table 2 shows the different agent types used in D5.2, again coming from D5.1. We also left the "comment" columns, related to some important characteristics of the respective agents this time. More details about them can be found in D5.1.

Table 2: Agent types used in D5.2.

Agent type	Description	Comment
Random	A random agent provides random information to the NM regarding their	This is only used for testing or calibration.

	cost function, usually based on uniform distribution.	
Honest	Agents are said to be 'honest' when they communicate to the NM either their true costs or the best approximation they have of their true costs.	Honest agents are not rational in the sense that they do not try to have the best allocation for themselves.
Rational	Agents are said to be rational when they communicate costs to the NM designed to minimise their expected cost in the mechanism.	The agents in the model are not fully rational from an economic point of view, because we do not have a closed, exact form for their expectation. Instead, we rely on different approximations, see section 4.
Bounded	Short for "agents with bounded rationality", these agents include two "distortions" in their decision-making process, based on prospect theory (PT) and hyperbolic discounting (HD).	Note that bounded-rationality in general is much wider than just a distortion of profit-seeking optimiser's decision-making processes, for instance using heuristics, rules of thumbs etc.

3.3.3 Calibration plan and game definition

In this deliverable, we have access to a wider range of airports, which are explored in the simulations. We decided to define two types of games:

- The 'total' game: in this game, at each iteration, we first draw an airport at random, following the distribution of the number of regulations computed in section 3.2. We then draw a regulation at random (linked to this airport in the data) and solve it.
- The 'partial' game: in this game, we choose an airport first. Then at each step, we draw at random a regulation (linked to this airport in the data).

The partial game is exactly the type of game ('multi-sided', see below) that we used in the previous deliverable D5.1 for LFPG, but we apply it all airports in this deliverable. We will then aggregate the results per airport etc. The total game can be seen as an aggregation of partial games, except for two facts:

- The frequency of appearance of regulation per airport is taken into account in the total game,
- When using the CM, the calibration process is different in the two types of games, see below.

Compared to D5.1, we do not simulate 'single-sided' games, where one player was singled out and studied more in details, but only 'multi-sided' ones, where all players use the same (pure) strategy. This is because this deliverable focuses on having a more realistic computation of the key performance indicators (KPIs), as opposed to the exploration of behaviour performed in D5.1. For the same reason, we will not explore the exact impact of the approximation archetype function on the results as in D5.1[5], even though we are still simulating some mechanism with the true costs and some approximated ones (with the 'Jump3' archetype, see D5.1).

As explained in D5.1, the CM, contrary to other mechanisms, has some free parameters that need to be calibrated. We follow the same simplified procedure as in the previous deliverable, calibrating only one of the parameters (the jump) while assuming an honest behaviour from the airline for the other ones (in particular the margin).

We performed two different calibrations for the two types of games:

- Total game: we perform the calibration on the total game without differentiation across airports. This means that the default jump out of the calibration takes only one value, that is used across all airports in the subsequent simulations.
- Partial game: we perform the calibration on each individual airport, which leads to have as many parameters as the number of airports.

We use this procedure in order to test how tailored the default parameter should be to each airport. Indeed, in the first case, it is possible that airlines will systematically lose or gain from regulations at a given airports (because the global default jump is much higher or smaller than the one we would have obtained in the second case). This may be an issue for some airlines, but allows to have a much more dynamic system, where airlines can choose to lose in some airports to order to gain credits that will be spent in another one.

3.3.4 Indicators

Compared to D5.1, we use almost the same indicators. As primary indicator, we used in D5.1 the relative gain in cost with respect to UDPP, i.e. the cost in the final allocation minus the cost in the FPFS allocation, divided by the latter. Because this indicator is very sensitive to the initial FPFS cost (for instance if it's null…), we added two other indicators. The first one is simply the absolute gain in cost, which tells us how much volume is gained by the airlines. The second one is the gain in cost per flight, which indicates the how much is gained per flight, thus enabling us to detect scaling effects.

Likewise, we added another equity metrics with respect to D5.1. Indeed, ironically, we used in D5.1 two equity metrics that were directly linked to the two indicators mentioned above that we are adding to this deliverable. Indeed, the equity indicators in D5.1 were:

$$EQ1 = 1 - \sum_i \sum_j abs(c_i - c_j) \Big/ \sum_i \sum_j abs(c_i + c_j)$$

$$EQ2 = 1 - \sum_i \sum_j abs(c_i/n_i - c_j/n_j) \Big/ \sum_i \sum_j abs(c_i/n_i + c_j/n_j)$$

[5] As a reminder, in some mechanisms, the airlines can communicate their true costs to the NM, or an approximated version of them, which is more realistic in general to assume. The CM is designed to make the most out of this approximation, is thus based on this process and cannot receive true costs by design.

These definitions are coming directly from the GINI indicator, which is widely used in economy, and reads, for a given quantity x:

$$G(x) = \frac{\sum_i \sum_j abs(x_i - x_j)}{2N \sum_i x_i}$$

To the absolute values in the denominators, the EQ1 and EQ2 can thus be computed as:

$$EQ1 = 1 - G(\delta c)$$

$$EQ2 = 1 - G(\frac{\delta c}{n}),$$

Where δc is the absolute cost saved by an airline with respect to FPFS, and $\delta c/n$ is the cost saved per flight, i.e. the two indicators that we mentioned above that we are going to report. Hence, since we do not have an equity indicator for the third one, the efficiency or relative cost, it is natural to add it and we thus consider the following indicator in this deliverable:

$$EQ3 = 1 - G(\frac{\delta c}{c}),$$

With $\delta c/c$ the relative gain in cost with respect to FPFS.

Co-funded by the European Union

4 Calibration and results

In this section we examine the output of the simulations performed with our model. We start by calibrating the CM, the only mechanism that needs calibration.

4.1 CM calibration

The calibration process of the credit mechanism has already been described in D5.1. Here we show the most important part, the calibration of the default jump parameter. We have two ways of doing this, depending on the flavour of the game that we are using:

- With the partial game, we compute one optimal parameter for each individual airport. We then have a collection of parameters that can be used in two ways: either we use an average for all subsequent simulations, or we load the parameter corresponding to the airport each time we simulate the latter. In the following we chose the second option.
- With the total game, we can only have one optimal parameter for all simulations, by construction. This parameter is, by design, less adapted to the use of the CM for each airport individually, but should be the best one possible on average for the entire set of airports.

We are particularly interested in studying the impact of using only one parameter instead of a set of parameters adapted to each airport, both in terms of overall efficiency and equity between airlines. This is explored in section 4.2.5.

4.1.1 Stable credit trajectory and optimal jump parameter

As a reminder, we mention again that the impact of the default jump parameter is to set to average level of credits in the system. If the parameter is too small, airlines will easily make some credits and thus the CM reverts to being just the GLOBAL algorithm. If it is too big, the airlines cannot spend any credits and will thus always be honest. In this case the CM algorithm reverts back to the GLOBAL algorithm with honest agent.

Figure 21 shows an illustration of this principle. In this instance we simulated 500 iterations of the CM with different default jump parameters. The different trajectories can be bundled in three categories, as explained in D5.1: absorptive (credits tend to 0 in the long run), explosive (credits tend to infinity in the long run), and stable, the one we are interested in. In order to find it, we follow the same principle than in D5.1, where we compute the slope, average credits, and standard deviation of the number of credits in the system[6].

[6] Note that in D5.1 we tracked only the credits of the major player, whereas here we track the sum of the credits across airlines.

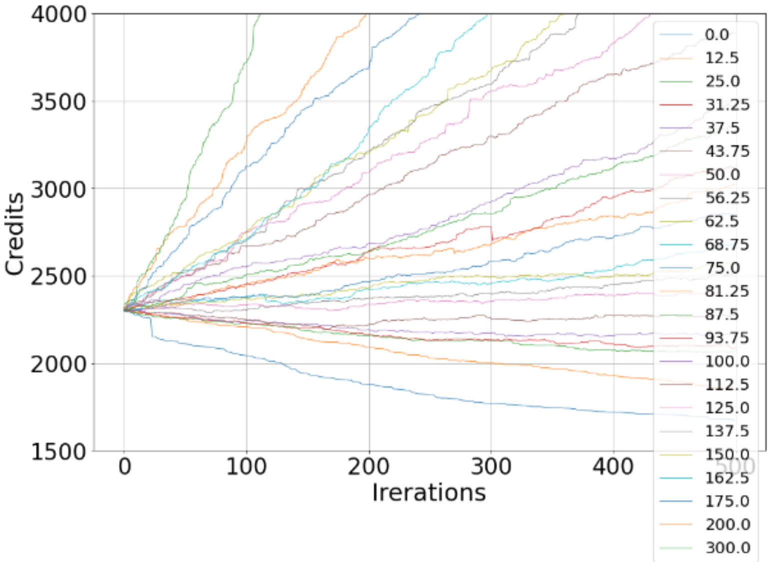


Figure 21: Evolution of credits in a total game with different values of the default jump parameter.

Figure 22 and Figure 23 shows the results of the procedure for a partial game with the EDDF airport and for the total game. In each of these figures, we run 100 iterations (resp. 500) of the partial and total game for different values of the default jump parameter and we compute the metrics mentioned above. We then track the minimum of the standard deviation over the average, which seems to be the most robust way of finding the stable trajectory in credits. Note how the results of a single airport are much noisier than for the total game. This is due to the smaller number of iterations in the partial game, something that will be improved in a future iteration of this work.

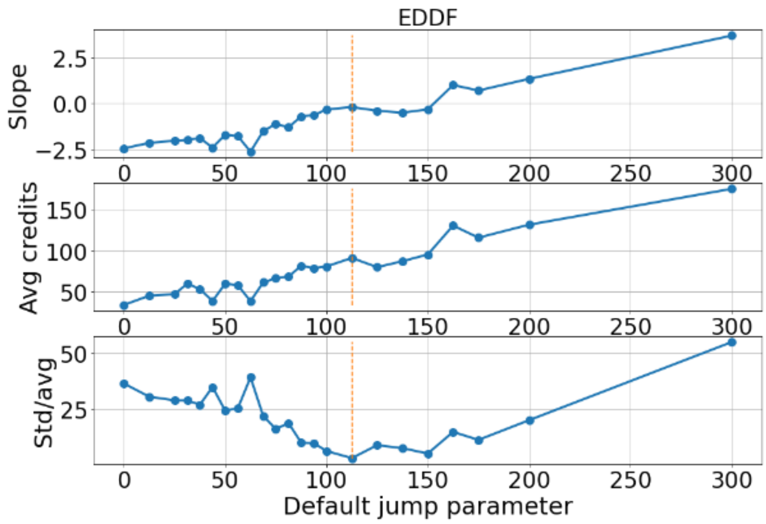


Figure 22: Slope, average credits and standard deviation over average of different credit trajectories for the partial game with airport EDDF. The orange line materialises the optimal default jump parameter for this game.

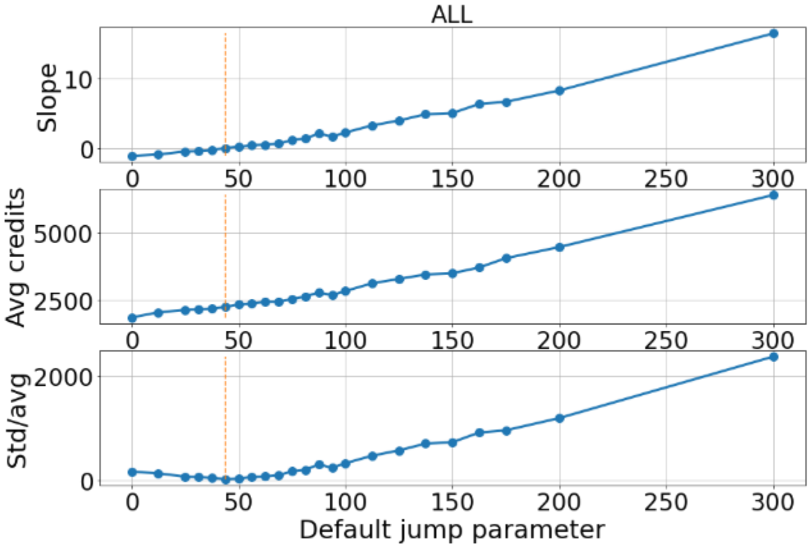


Figure 23: Slope, average credits and standard deviation over average of different credit trajectories for the total game. The orange line materialises the optimal default jump parameter for this game.

4.1.2 Distribution of optimal parameters and relationship with airport size

As explained previously, after the calibration in the partial games, we obtain one value of the optimal parameter for each airport; it corresponds to the value for which the credits to not inflate or deflate in the long run. These values can be really different, as shown in Figure 24.

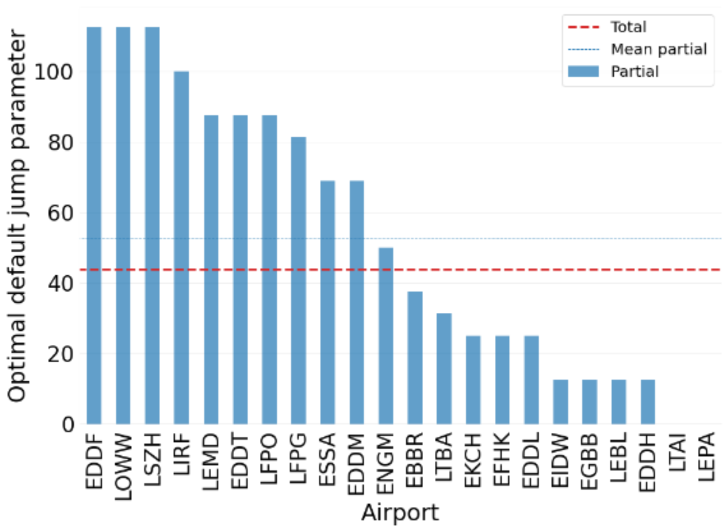


Figure 24: Value of the optimal default jump parameter found for each airport. The dashed blue line corresponds to the average across airports. The dashed red line corresponds to the value calibrated on the total game.

Indeed, this figure show that the default parameter can be as small as 0 in the case of LTAI and LEPA or 120 in the case of EDDF, LOWW, and LSZH. It is also interesting to note that the average of these values (materialised by the blue in the figure) do not coincide with the value obtained after the calibration of the total game. Thus, one cannot simply take the first one for a total game.

One reason for which different airports have different calibrated parameters is the distribution of the size of the regulation at each airport. Indeed, big airports will tend to have big regulations, as shown in section 3.2, which distorts the number of credits needed to have a stable credit amount on average. Figure 25 shows the optimal value as function of the average size of the regulation at the corresponding airport. There is a clear tendency to have higher optimal values for bigger regulations, as expected.

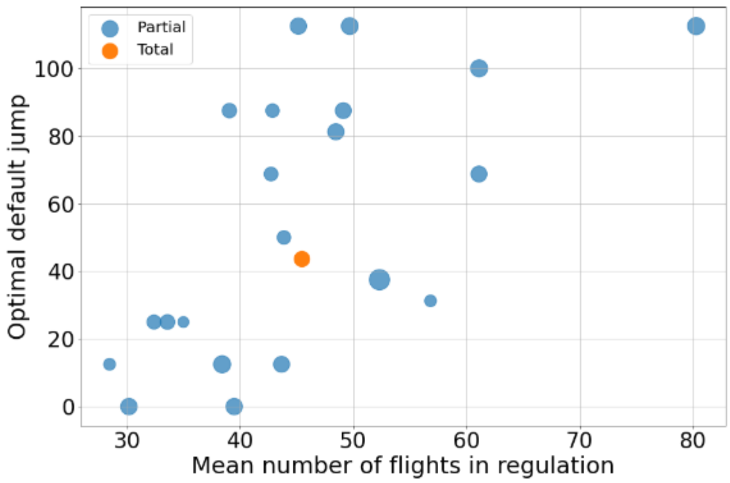


Figure 25: Optimal default jump as function of the average number of flights in the regulation at the corresponding airport. The orange circle corresponds to the total game. The size of the circle is proportional to the size of the airport (number of movements).

These results are not only important to have reliable results with the simulations on the CM, but also have important practical effects for the potential implementation of the mechanism. We will discuss this more in details in the conclusions.

4.2 Results

We now show the results obtained on the partial/total games, the different mechanisms, and the different types of agents implemented. We start by examining the results of the partial games mostly, and we will only highlight at the end the differences found when the total game is used.

An important aspect of this work, even more so than in D5.1, is the difficulty to render all aspects of the simulations, even with a small number of indicators like we have, because of the numerous potential aggregation processes that one can use to study the results. Indeed, the indicators, for instance efficiency, can be aggregated unconditionally, or by airline, or by airport, or by airline and then by airport, or by type of airlines, or by size of airports etc. In order to simplify the analysis, we focus first on an aggregation per airline, then an aggregation per airport, to go to the overall computation of the indicators. At the end we explore the dependence of the indicators on different features of the regulations, agent decisions etc.

Contrary to D5.1, we do not focus only on efficiency and equity, for reasons that will become clear in the following sections. We also add two other metrics: the total cost saved by the mechanisms (always with respect to FPFS) and the cost saved by flight (again, with respect to FPFS). Moreover, we use added another equity metrics, similar to the other ones used in D5.1, as explained in section 3.3.4.

4.2.1 Results per airline

We start by having a look at the impact on individual airlines. In this section and the next one, we focus on the comparison between UDPP, NNBOUND with true costs, and GLOBAL with true costs. Figures for all mechanisms can be found in Appendix D and Appendix E.

4.2.1.1 Example on EDDF

Similarly to what has been done for D5.1 for LFPG, one can study the impact of some mechanisms on the airlines present at this airport. Figure 26 show the absolute save cost, the saved cost per flight, and the efficiency, i.e. the saved cost relative to the FPFS cost, at EDDF when using UDPP. Without any surprise, we find that the absolute savings are made by a major player at this airport, DLH, and that other airlines barely make any savings. However, when we study the cost saved per flights, we start to see some airlines making some gains, in particular SAS. Interestingly, we see that the relative savings are much more distributed, with quite a few airlines being over 10% of efficiency, while DLH remains the airline with the best relative savings. Thus, it seems that major airlines tend to do the best absolute savings (which is obvious), the best savings per flight (which denotes a scaling effect) and also the best relative savings, even though other airlines make some significant ones.

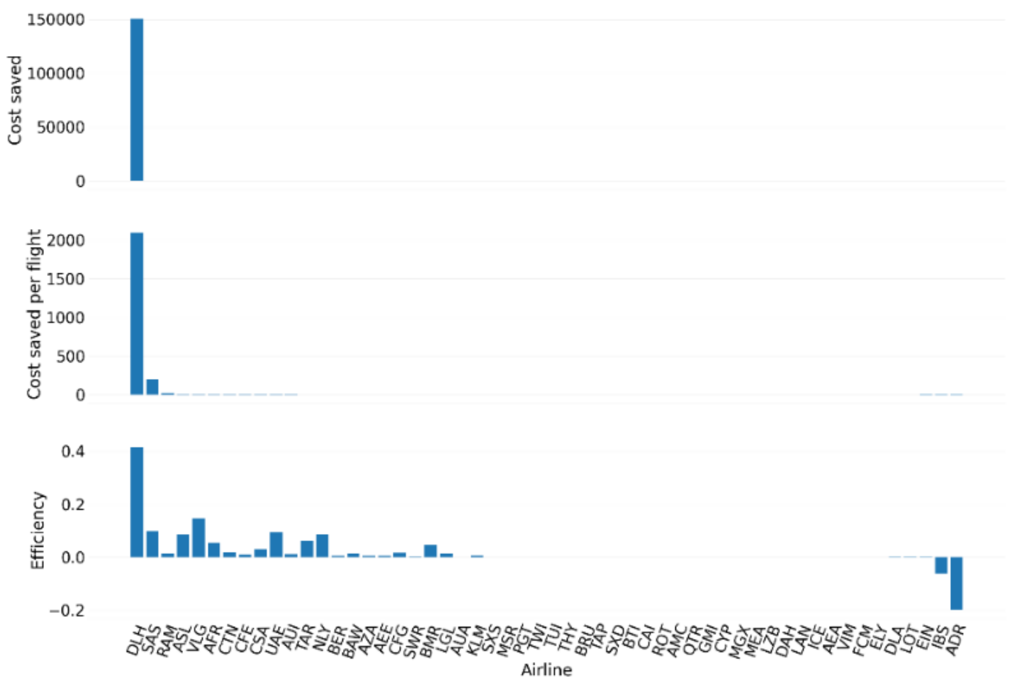


Figure 26: Absolute saved cost, saved cost per flight, and efficiency of UDPP for individual airlines present at EDDF. The airlines are sorted by decreasing absolute savings.

This situation is interesting to compare with the one in Figure 27 and Figure 28, where we present the same results obtained respectively with NNBOUND (with true costs) and GLOBAL (with true costs too). While the absolute savings are more or less the same, the cost saved per flight is much more equitable in the NNBOUND case, with quite a long tail. The efficiency itself is even better distributed, with several airlines over 50% of efficiency, higher than the major player. Also, as expected, it is worth noting that no airline loses from the NNBOUND allocation (by design).

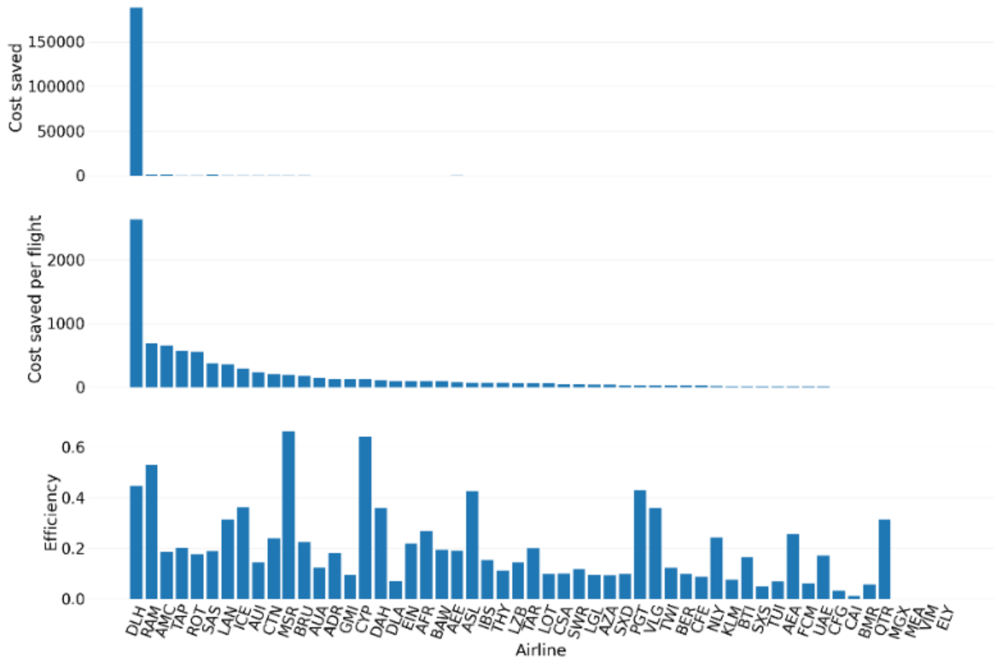


Figure 27: Absolute saved cost, saved cost per flight, and efficiency of NNBOUND (with true costs) for individual airlines present at EDDF. The airlines are sorted by decreasing absolute savings. Note that the order of airlines is thus slightly different from Figure 26.

The picture for GLOBAL shown in Figure 28 is more complex. First, it is now obvious that some airlines (roughly half) lose in terms of saved cost, which is particularly visible in the savings per flight. This is because GLOBAL relaxes the non-negative constraints of NNBOUND on the savings per airline. It is also striking to see that the efficiencies follow quite a different pattern from NNBOUND with most airlines having a null efficiency, a couple of airlines with negative ones and a few ones with positive ones, which can be quite high. This is a good illustration of the limits of this indicator. Indeed, if comparing the savings made by the mechanism to the FPFS cost might make some sense, it is also dangerous when the latter are small. In this case, the efficiency appears very high, with only a moderate gain in real euros. We will come back to this in the following subsections and in the conclusions, but this is the reason why we believe that tracking the gains per flight makes more sense than the efficiency when it comes to individual airlines, in particular when computing equity metrics.

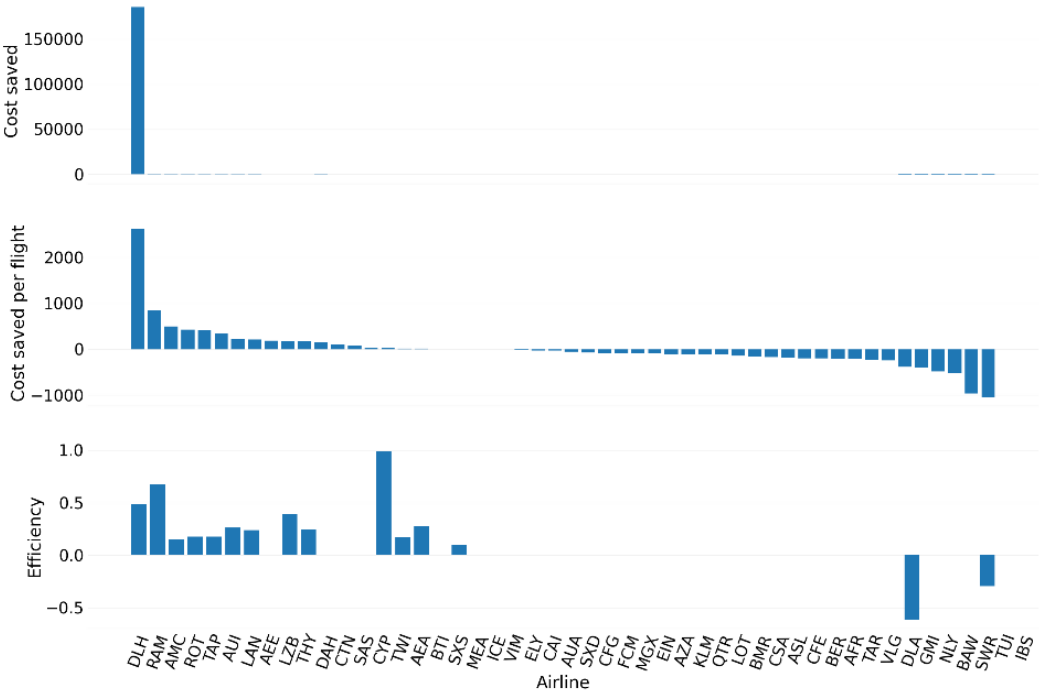


Figure 28: Absolute saved cost, saved cost per flight, and efficiency of GLOBAL (with true costs) for individual airlines present at EDDF. The airlines are sorted by decreasing absolute savings. Note that the order of airlines is thus slightly different from Figure 26 and Figure 27.

4.2.1.2 All airports

Compared with the results in D5.1. we are interested in looking at what happens for several airports at the same time. Indeed, if there is often a major player at a given airport (at least at hubs), in Europe overall the situation is a lot more balanced. Hence, the natural advantage given by some mechanisms to big players may be less important. Moreover, it is likely that non-linearities play a strong role, i.e. the situation where an airline has the same number of flights in each airport is very different from the situation where all its flights are located at the same airport (which is the case in general).

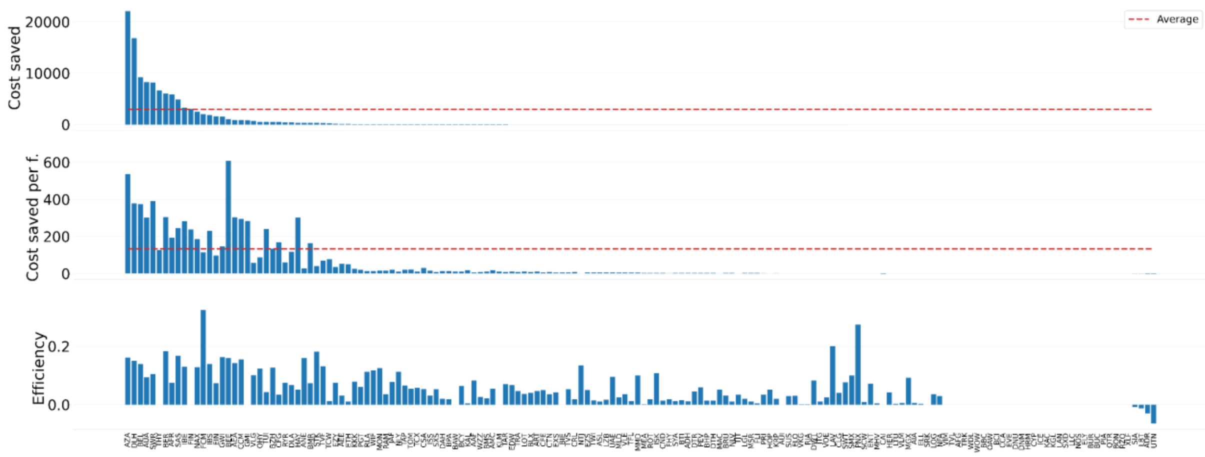


Figure 29: Absolute saved cost, saved cost per flight, and efficiency of UDPP for individual airlines across all airports. The airlines are sorted by decreasing absolute savings. The dashed red line is the average across all airlines. An enlarged version of this figure and the next ones are available in Appendix D.

Figure 29 shows the same kind of results than for individual airport, but this time aggregated over all airports[7]. It is still clear from the absolute gain that the major airlines will, by design, make the most important gains. However, the picture is much more balanced and the cost saved per flight displays quite a lot of airlines saving a significant amount of money per flight.

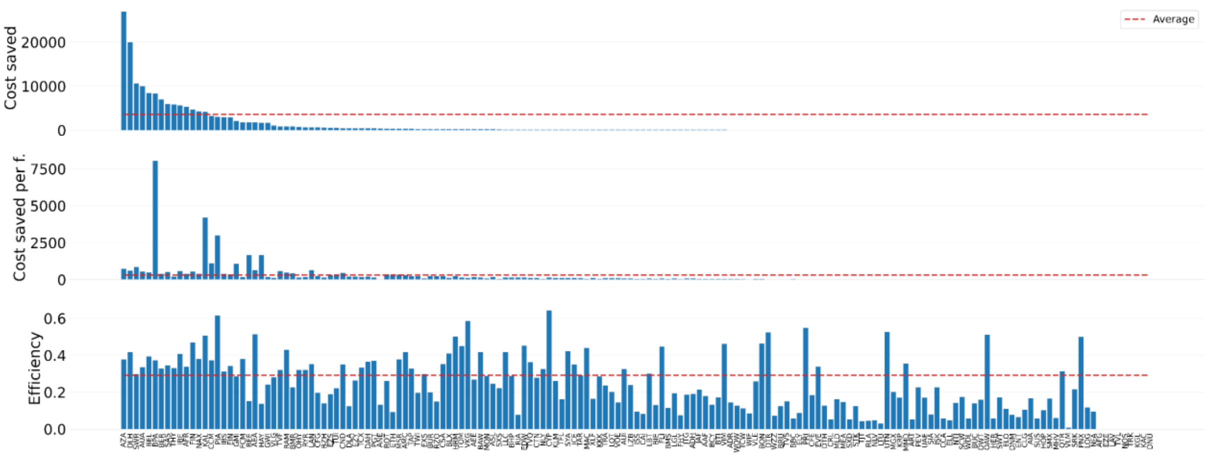


Figure 30: Absolute saved cost, saved cost per flight, and efficiency of NNBOUND (with true costs) for individual airlines across all airports. The airlines are sorted by decreasing absolute savings. The dashed red line is the average across all airlines.

Figure 30 shows the same results for NNBOUND and the comparison is interesting to make. Indeed, the cost saved per flight seems to be very unbalanced in this case, with a few airlines making very high gains (more than 2000 euros per flights, a lot more than the 600 euros maximum in the UDPP case). However, the picture is terms of efficiency is very balanced, with quite high efficiencies for most of the airlines. In terms of equity, it is thus interesting to realise that the perceived fairness of this mechanisms may be very different based on the metrics selected, savings per flight or relative savings…

The picture for GLOBAL in Figure 31 is also quite different. While we see some similar features for the cost saved per flight, we notice once again that some airlines lose significantly from the mechanism (roughly half once again). More importantly, the efficiencies seem to have very different values, which again raises the issue of using this metrics more systematically, in particular for equity metrics.

[7] There is a bigger version of this figure and the following ones, as well as all the mechanisms simulated, in Appendix D.

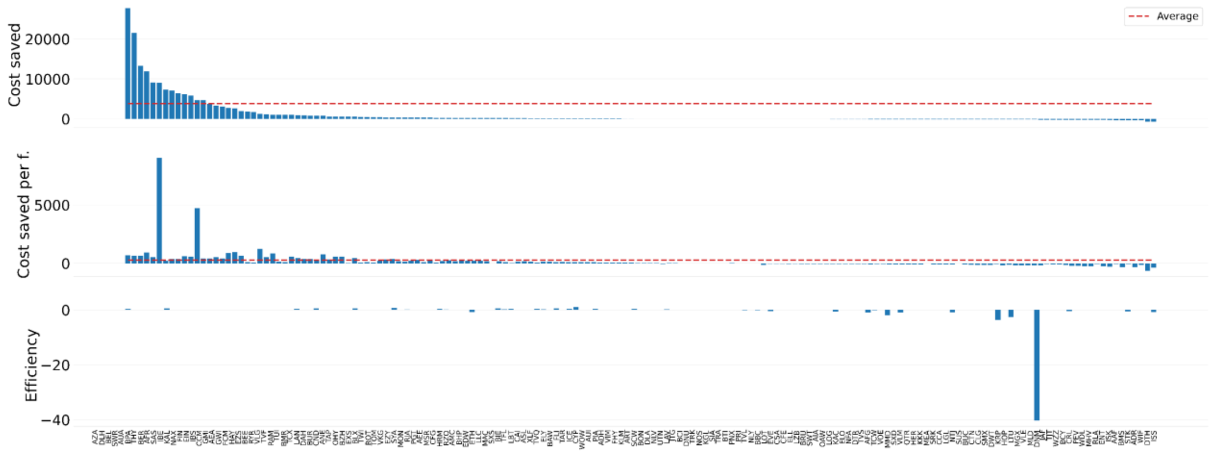


Figure 31: Absolute saved cost, saved cost per flight, and efficiency of GLOBAL (with true costs) for individual airlines across all airports. The airlines are sorted by decreasing absolute savings. The dashed red line is the average across all airlines.

4.2.2 Results per airport

While the picture per airline is important, in particular for fairness and equity, aggregating the results per airport has the benefit to show the randomness of the system and the fact that focusing on a particular airport, like we did in D5.1, gives a very partial view of the situation.

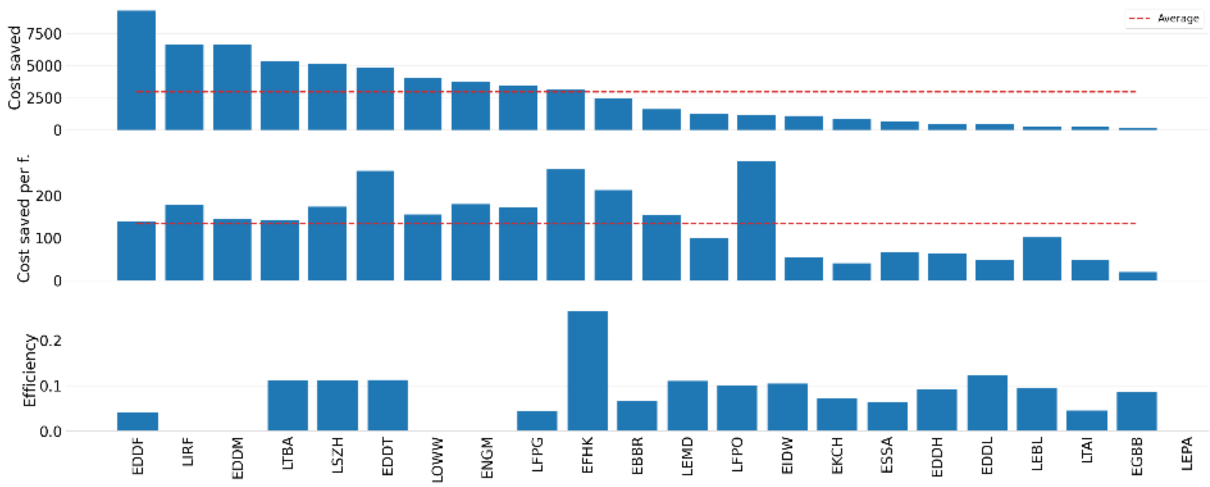


Figure 32: Absolute saved cost, saved cost per flight, and efficiency of UDPP at individual airports. The values are averaged across airlines present at the airport. The airports are sorted by decreasing absolute savings. The dashed red line is the average across all airports.

Figure 32 shows the average absolute savings, savings per flight, and efficiency across airlines present at each airport. Note that these figures do not show the 'objective' efficiencies because of the average across airlines (on the contrary of section 4.2.4), but rather the perceived experience of airlines at each airport (same remark for absolute savings and savings per flight)[8]. Looking at the plot, it is quite clear that the mechanism has very different impacts in different airports. Interestingly, it seems that cost saved per flight are fairly well distributed, and that average efficiencies are usually around 10%[9].

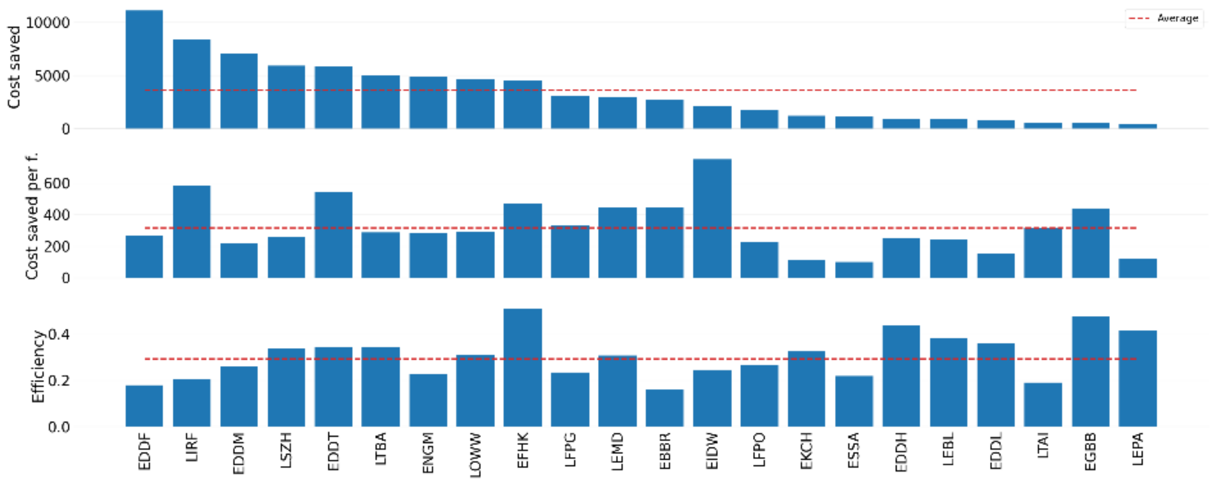


Figure 33: Absolute saved cost, saved cost per flight, and efficiency of NNBOUND (with true costs) at individual airports. The values are averaged across airlines present at the airport. The airports are sorted by decreasing absolute savings. The dashed red line is the average across all airports (the blue bars).

Once again, one can make the comparison with similar plots for NNBOUND and GLOBAL, in Figure 33 and Figure 34. Even though the absolute savings, cost saved per flight, and efficiencies are notably higher in these two figures, there is no important difference in the distribution, with however a tendency to have slightly more distributed savings across airports.

Also, it is interesting to note that with this kind of aggregation, the efficiency appears negative for GLOBAL. This is because a few airlines have very negative efficiency in this case, most probably because their initial FPFS costs were very small. This is exactly the reason why we believe that considering the cost saved per flights as a metrics may be preferable to the efficiency when the level of aggregation is low, like here.

[8] Figures for all the mechanisms can be found in Appendix E.

[9] Again, this is not the total efficiency of the mechanism at this airport, which would be the total saved cost over the total FPFS cost instead of the average of this metrics across airlines

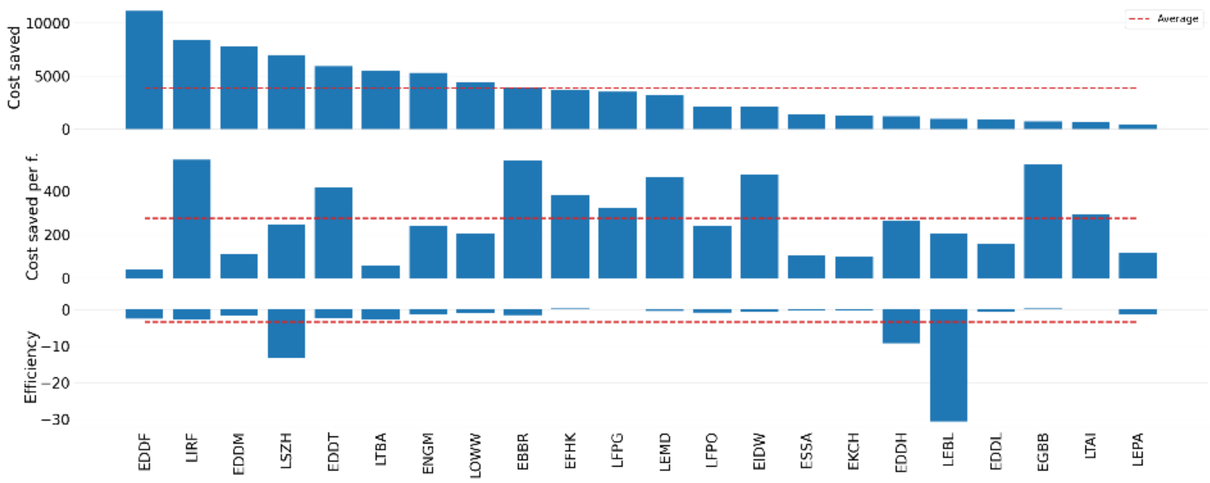


Figure 34: Absolute saved cost, saved cost per flight, and efficiency of GLOBAL (with true costs) at individual airports. The values are averaged across airlines present at the airport. The airports are sorted by decreasing absolute savings. The dashed red line is the average across airports.

We continue the analysis of the aggregated values per airport and airlines in the next section, where we analyse more in depth the relationship between the indicators presented here and some characteristics of the airlines and airports.

4.2.3 Analysis of efficiency vs size of airline, size of regulation, number of airlines.

It is clear from the previous sections that the mechanisms have very different impact on different airlines and at different airports. In this section we try to go deeper into the analysis to see which factors contributes the values of the indicators used above.

4.2.3.1 Focus on a few airport

We first focus on a few airports. First, it is quite clear that a few variables should have some impact on the results of the mechanisms: number of flights in regulation, number of flights per airline, etc. For instance, we show in Figure 46 the savings per flight as a function of the regulation size (in number of flight) at a given airport, EDDF. It is clear that when the regulation is bigger, the potential savings per flight are higher, given than the queue is bigger and delays are higher overall.

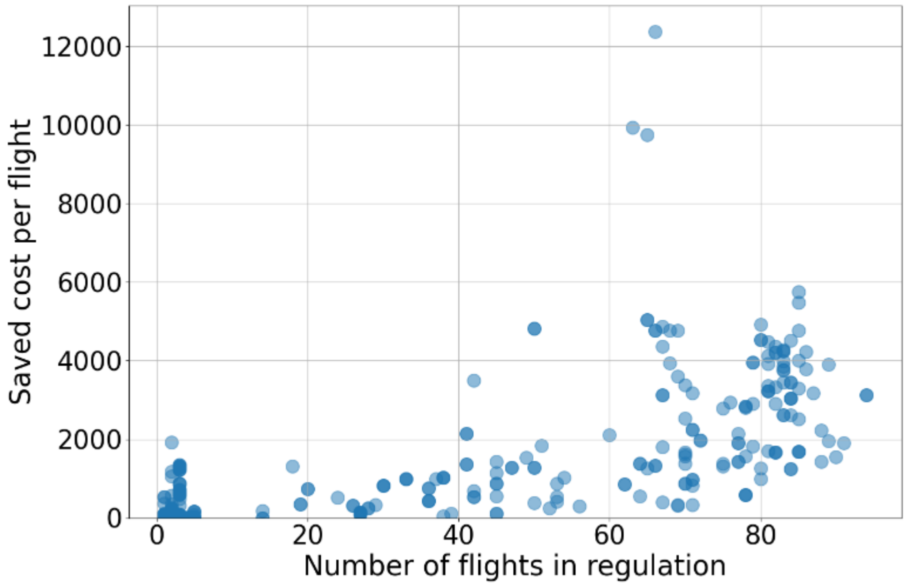


Figure 35: Saved cost per flight for each airline in individual regulations as a function of the number of flights in regulation at EDDF for UDPP.

However, we are also interested in seeing if scaling effects exist when we focus on relative savings. Figure 46 shows the corresponding plot. Interestingly, there seems to be a trend in this case too, as least for bigger regulations, where the efficiency increases with the number of flights in the regulation. This should be expected with at least with UDPP, where by definition an airline can only make gains when it has more flights so that they can swap their slots.

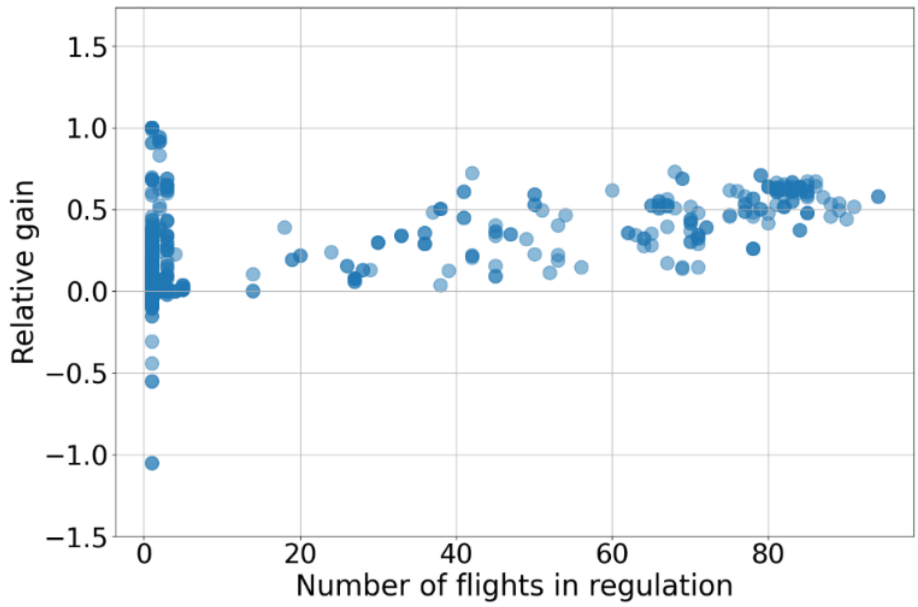


Figure 36: Relative saved cost (efficiency) for each airline in individual regulations as a function of the number of flights in regulation at EDDF for UDPP.

4.2.3.2 Unconditional relationships between indicators and features

The next step is trying to find general relationships between the indicators and some features, regardless of the airports, airlines etc. Here we start doing some simple unconditional relationships, using quantiles in features to have smooth relationships.

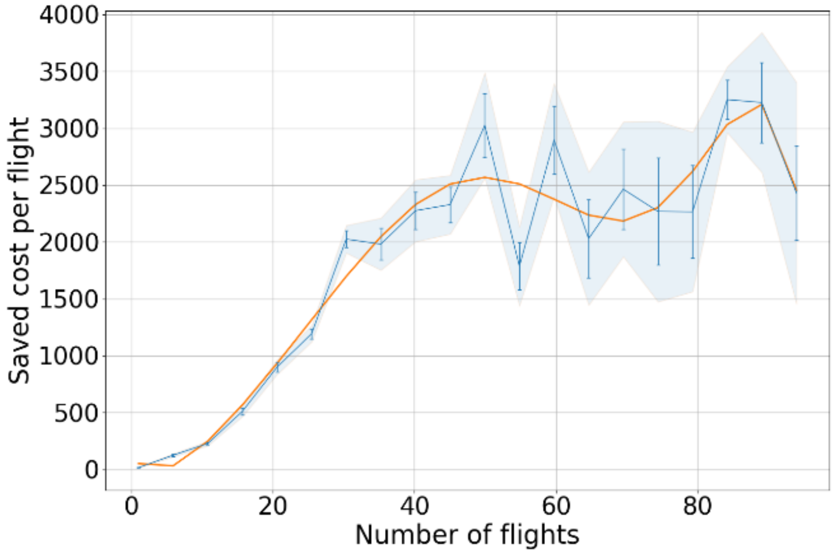


Figure 37: Saved cost per flight as a function of the number of flights of the airline involved in the regulation (UDPP). The indicator is averaged on quantiles. The error bars are standard errors, and we show the 0.05/0.95 confidence intervals with shaded curves. The orange line is just a smooth polynomial regression drawn to guide the eye.

Figure 37 shows the result of this procedure with UDPP, using the number of flights of a given airline in the regulation. It is clear from the plot that higher number of flights in the regulation leads to higher savings per flight, which a non-linear relationship (increase, plateau, second increase, drop at the end). However, we should be careful about overinterpreting this plot, as can be shown with the analysis of Figure 38.

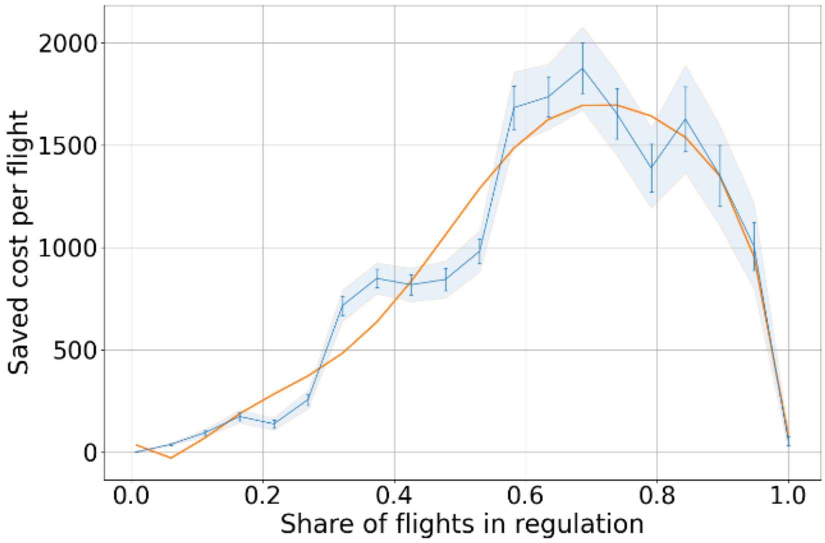


Figure 38: Saved cost per flight as a function of the share of flights of the airlines involved in the regulation (UDPP). The indicator is averaged on quantiles. The error bars are standard errors, and we show the 0.05/0.95 confidence intervals with shaded curves. The orange line is just a smooth polynomial regression drawn to guide the eye.

Indeed, in this case we show the same indicator, but as function the share of flights that the airline has in the regulation. The naïve expectation would be that the higher the share, the higher the savings per flight, but we can see that it is far from try. In fact, it looks like the savings per flight drops to 0 when the share tends to 1. This is actually due to a simple statistical effect. Indeed, regulations where airlines have a very high share are usually very small regulations. Hence, by construction, the airlines cannot make important savings in these cases. This issue leads us to perform a full regression, presented in the next section.

Finally, we show another example of such a plot, using the GLOBAL mechanism with honest agents in Figure 39. While the general trend is the same than in Figure 37, some differences appear, that may have far reaching consequences when it comes to equity for instance. Indeed, in this case it seems that the savings increase more monotonously with the size of the regulation, with only a drop for very big regulations.

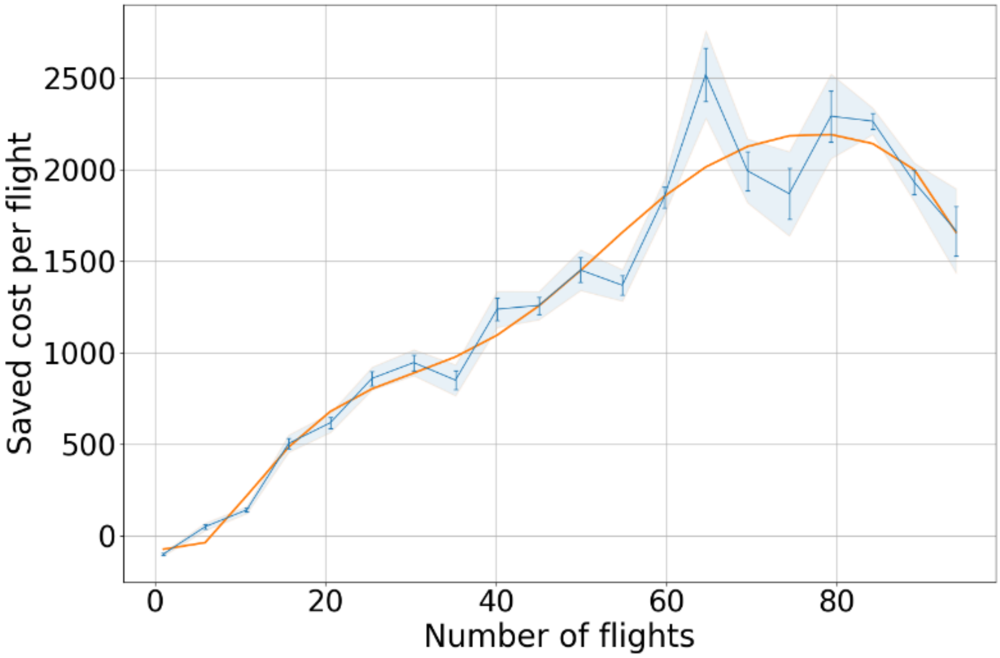


Figure 39: Saved cost per flight as a function of the share of flights of the airlines involved in the regulation (GLOBAL_APPROX). The indicator is averaged on quantiles. The error bars are standard errors, and we show the 0.05/0.95 confidence intervals with shaded curves. The orange line is just a smooth polynomial regression drawn to guide the eye.

4.2.3.3 Regression

The previous considerations lead us naturally to try to understand the conditional relationship between indicators and features, e.g. to understand the impact of the size of the regulation, all other variables being held constant. In order to do this, we use a simple OLS procedure, using linear relationships between saved cost per flight or efficiency and the following features:

- average FPFS cost per flight in regulation (regulation feature),
- standard deviation of FPFS cost per flight in regulation (regulation feature),
- standard deviation of number of flights per airline (regulation feature),
- number of flights in regulation (regulation feature),
- number of airlines in regulation (regulation feature),
- share of flights of airline in regulation (airline feature),
- FPFS cost of airline (airline feature),
- number of flights of airline (airline feature).

We perform this regression for each airport independently and we compute the average and standard errors of the regression coefficients. Figure 40 shows the result of this procedure for the efficiency, and Figure 41 for the saved cost per flight.

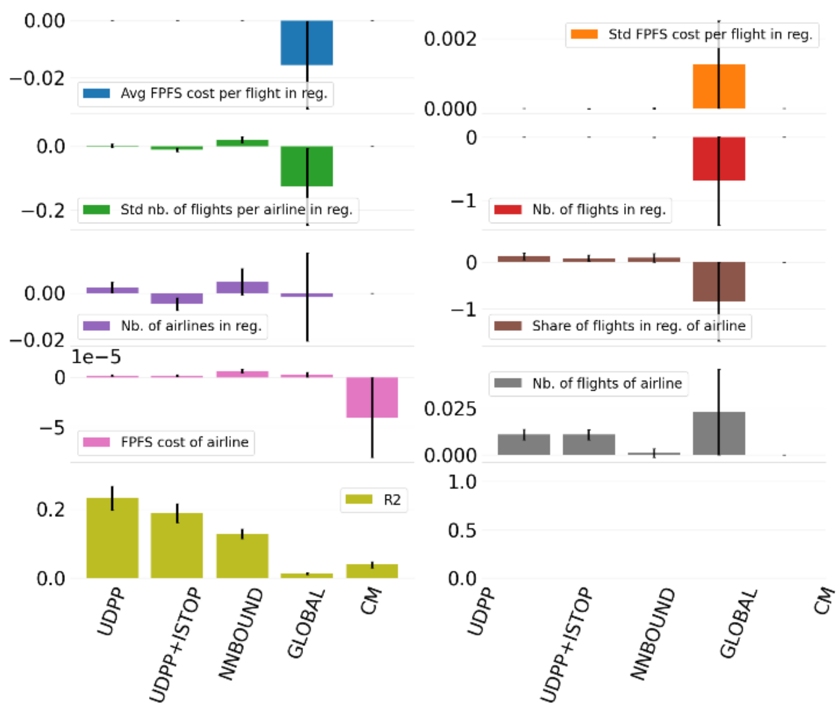


Figure 40: Coefficient values for different features when regressing the efficiency for each individual airport. The R2 plots shows the average and standard errors of the coefficient of determination of the regression.

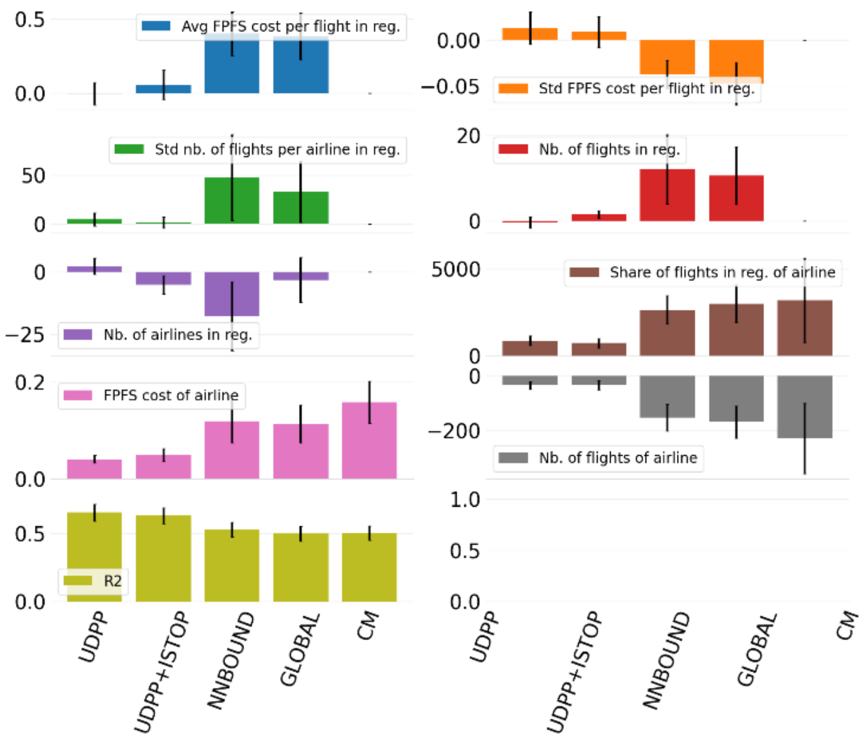


Figure 41: Coefficient values for different features when regressing the saved cost per flight for each individual airport. The R2 plots shows the average and standard errors of the coefficient of determination of the regression.

While the picture is fairly complex, we can extract a few conclusions from these plots:

- The coefficients for the efficiency are mostly null, i.e. the corresponding features do not have a significant effect on the value of the efficiency. Only the number of flights of the airline seems to be important, and only for UDPP and ISTOP. This means that these two mechanisms are much more sensitive to the scaling effects than the other ones when it comes to efficiency. In other words, having a high number of flights in a regulation is important to have a high efficiency, something that is natural for UDPP and ISTOP, since they rely on the number of possible intra-airline swaps.

- The coefficients of determination for the efficiency are notably very low. This definitely linked to the fact that there is a higher variance when using the efficiency.

- On the contrary of the efficiency, most coefficients are significant for the saved cost per flights.

- More specifically, for UDPP and ISTOP, the initial FPFS cost of the airline, the number of flights of the airline, and the share are all significant. It is interesting to note that the second one has a negative coefficient but the latter has a positive one.

- Re. NNBOUND and GLOBAL, it looks like the savings are increasing with the average initial FPFS cost, which seems normal, but decreases with the standard deviation. This result is not obvious, since intuitively we would have expected the savings to be positively impacted when the standard deviation is high in the UDPP/ISTOP case, but fairly neutrally otherwise.

- Likewise, it is surprising to note that the coefficients of the share of flights are higher for GLOBAL and NNBOUND compared to UDPP/ISTOP. This means that airlines make in general proportionally more money when they have a higher share of flights in the regulation in the former mechanisms compared to the latter ones. However, one might have expected intuitively the contrary, since UDPP at least relies on intra-airlines swaps.

- Finally, we note that the saved cost per flight is more easily captured by the OLS than the efficiency, with much higher R2 coefficients.

4.2.4 Aggregated impact of mechanisms

While the aggregation per airlines and airports provides some very important insights, we are interested here in looking at aggregated metrics over all airports, which should be the main indicators on which to base the decision to further develop some mechanisms. Similarly to D5.1, we present results on average savings first and then we look at equity metrics. However, contrary to the last deliverable, we include average savings per flight as aggregated metrics, and not only the efficiency, Indeed, for reasons already explained previously, we believe that the efficiency metrics may not be a stable enough indicator to assess the performance of the mechanisms.

4.2.4.1 Saved cost per flight

We start by showing in Figure 42 the aggregated savings per flight (in euros) for the different mechanisms and agent type. It is first worth noting that UDPP provides already a high level of savings with this metrics with around 900 euros of savings per flight involved in a regulation. UDPP was already noted for its high efficiency in D5.1. Using ISTOP on top of UDPP seems to have no effect when using true costs, and an adverse effect when using approximated cost functions. This is in line, but more pronounced, than the result found in D5.1, in which we found that ISTOP had a very small effect, even when using true costs. NNBOUND seems to add roughly 200 euros of savings per flights, which is significant, but fails to deliver some benefits over UDPP when approximations are used.

GLOBAL, which represents the best case from the total cost point of view, is very close to NNBOUND with the true cost, and is severely impaired by approximation and gaming effects. Interestingly, the bounded behaviour of some agents seems to improve slightly the situation. As already noted in D5.1, it might be that behavioural biases tend to decrease other the impact of other effects, like errors due to the approximation, or gaming effects. It is also worth noting that GLOBAL is barely above UDPP as soon as true costs are not used.

Finally, it is clear that the credit mechanism performs quite badly, in contrast with what we have found in the previous deliverable for LFPG alone. The reason for this is unclear at the moment and will require more effort to be understood.

While the errors bars, which represent errors on the mean value of the indicator, are fairly small because of the number of simulations performed, it is also important to realise the extent of the diversity of the indicators across airlines, something that we already touched upon in the previous sections. In Figure 43 we show the distributions corresponding to the above cases. In these figures (cut in the y axis), it is important to note the very high variance of the GLOBAL and CM mechanisms. While UDPP, ISTOP, and NNBOUND have values which can be already very different from each other, the two other mechanisms have very fat distributions, representing very different situations for different airlines. Note also that for NNBOUND and honest agents the real distribution (the box plot) goes well into the negative domain, showing how the approximation destroys the non-negative constraints[10] from this mechanism.

These figures, in particular Figure 42, represents the most important results of this deliverable, as it encapsulates our most accurate estimation of the impact on the mechanisms and behaviours on the system.

[10] On the declared costs, not the real ones!

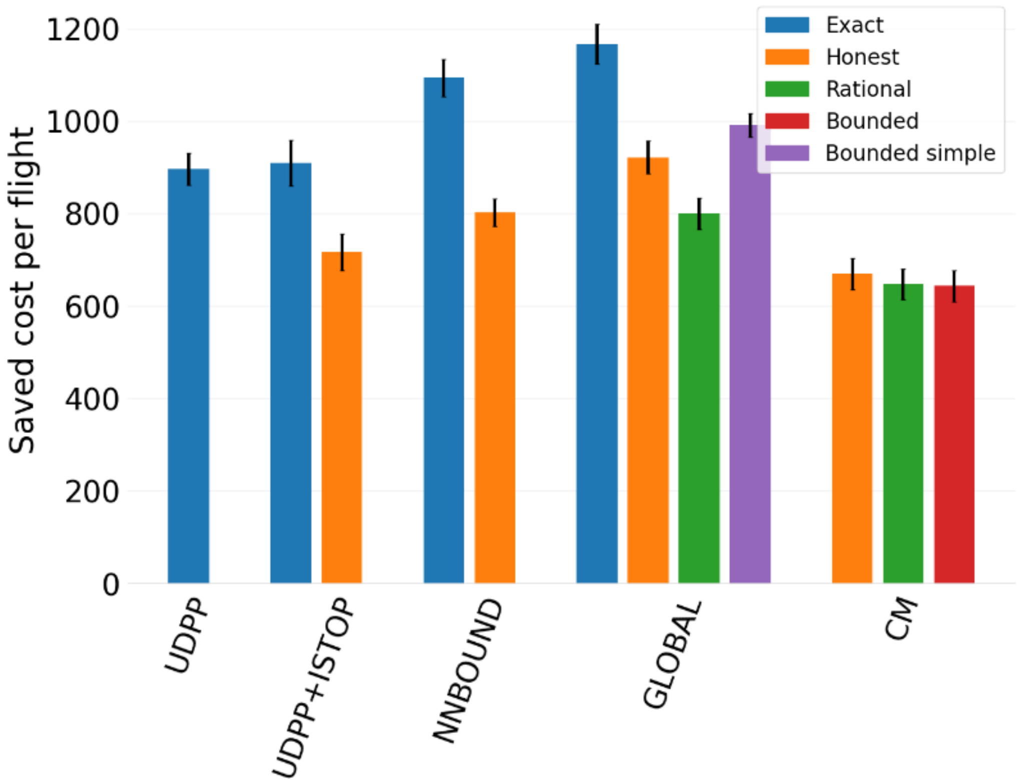


Figure 42: Aggregated cost saved per flight across all airlines and airport for each mechanism and the difference agent types implemented. Exact: the mechanism uses true costs; honest: the mechanism uses the approximation of the cost function provided by an honest agent; rational: the mechanism uses the approximation of the cost function provided by a rational agent; bounded: the mechanism uses the approximation of the cost function provided by a bounded agent; bounded simple: the mechanism uses the approximation of the cost function provided by an bounded-simple agent (see deliverable D5.1 and section 3.3 for the meaning of the agent types). The error bars are standard errors.

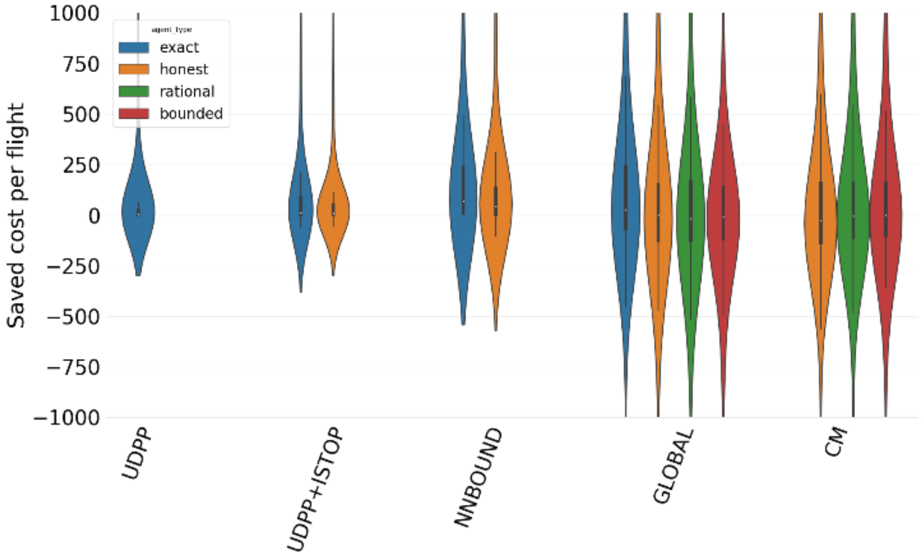


Figure 43: Violin plots (box plots+ Kernel Density Estimates) for the savings per flight in each setting.

4.2.4.2 Efficiency

We then look at similar plots, this time for the efficiency (the relative cost), in the same way than we did for D5.1, in Figure 44 and Figure 45. First, on the mechanisms with the true costs, we have the same conclusions than in the previous section: UDPP is very high, ISTOP does not have a significant effect, NNBOUND and GLOBAL are above but the improvement of the latter over the former is fairly small. The conclusions regarding the approximations seem to hold as well, with the GLOBAL one barely raising to the level of UDPP. Gaming seems to have a similar effect on this indicator, but behavioural biases do not seem to improve the situation here. This fact, like the previous one, is not clear and would require more study.

In any case, the only significant difference seems to lie in the CM, where we find some decent efficiencies on average (comparable to UDPP), among the best in fact when considering approximated costs. This is more in line with what we found in D5.1, where we used indeed the same indicator, the efficiency. We can also note the insensitivity of the CM with respect to gaming and behavioural effects, a fact which was noticed also in D5.1, which is in contrast with the GLOBAL mechanism, and a very desirable characteristic obviously. Finally, we also note that high values of the efficiencies overall (above 40%) compared to the ones found for LFPG in D5.1 (around 15-23%). In particular, we note again that UDPP alone has a very high efficiency on average, around 45%.

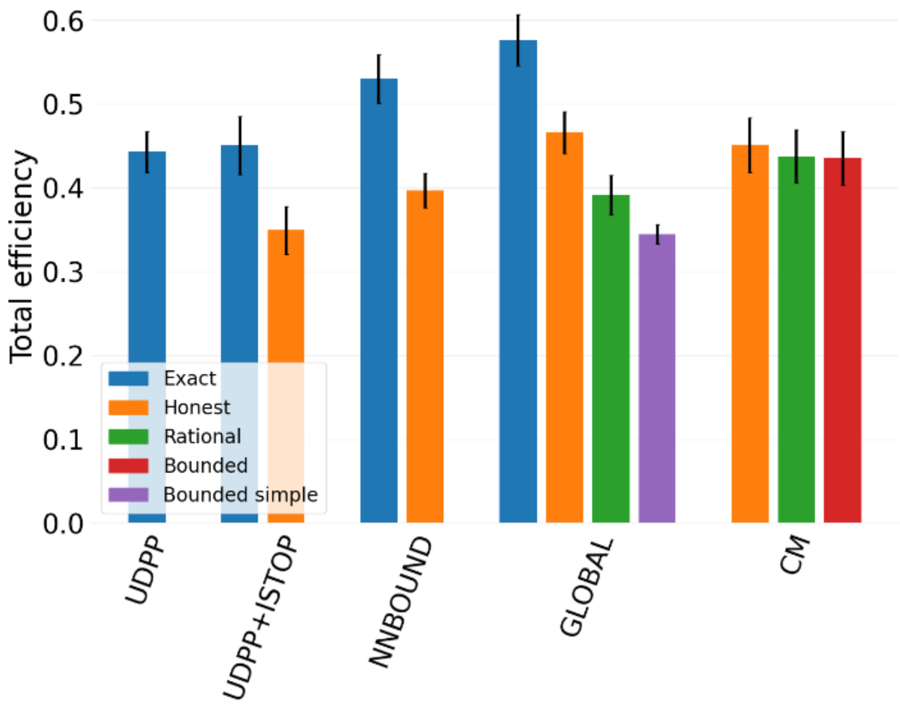


Figure 44: Aggregated efficiency across all airlines and airport for each mechanism and the difference agent types implemented. See Figure 42 for the legend explanation.

In Figure 45 we also show the distribution of values of efficiency for different airlines to give an idea of the variance. Similarly to the cost per flight, the latter is much higher for GLOBAL than UDPP, ISTOP, and NNBOUND, and even more so for CM. Note also how the efficiencies of NNBOUND are notably higher than the ones for UDPP and ISTOP, much more so in the fully aggregated version from Figure 44. This can be seen as a good characteristic of NNBOUND, where the perceived efficiency of the mechanisms for most airlines is close to the real, total one.

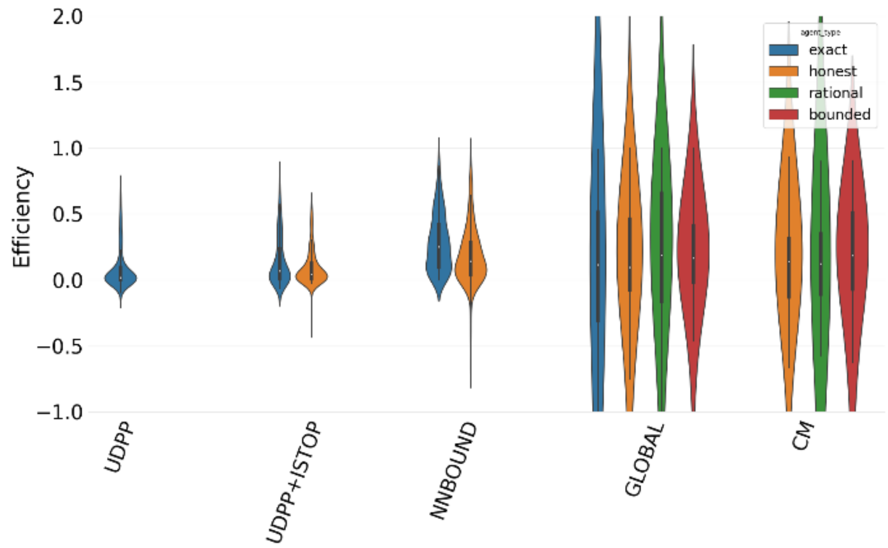


Figure 45: Violin plots (box plots+ Kernel Density Estimates) for the efficiency in each setting.

4.2.4.3 Equity

Finally, we want to study what happens with the indicators that we used previously to track the differential impact of the mechanism on different airline, i.e. the equity (or fairness). In order to keep consistency, we show three indicators, presented in section 3.3.4, related to the three indicators presented above. Indeed, as a reminder, each corresponds to the same formula (1-Gini coefficient) applied to the different indicators: absolute cost, cost per flight, and efficiency. With respect to D5.1, only EQ3, corresponding to the inequality in efficiency, is new.

Figure 46 shows the aggregated equity indicators. The first important fact to notice is that the first indicator is very low (and correspondingly the error bars are bigger), meaning that the inequality in absolute cost is extremely high. This is expected given the distribution per airline shown in previous sections, where usually one player makes almost all the gains in absolute terms. The absolute values for EQ2 and EQ3 are much higher oevrall, indicating that the inequality in cost per flight and efficiency are much smaller, which was again expected given the previous plots.

Just like we noticed in D5.1, NNBOUND seems to be the champion of equity, whatever the indicator we use, with a special emphasis on EQ3 where it reaches almost 60%. The situations are quite different for GLOBAL and CM. While these mechanisms seem to have a similar degree of equality than the other mechanisms with EQ1, they are notably lower in terms of the other indicators. CM is particular has very low scores in equity, even smaller than GLOBAL, which is bad news. Indeed, the main point of the CM mechanism was to allow some global optimisation while keeping some degree of equity thanks to the credits.

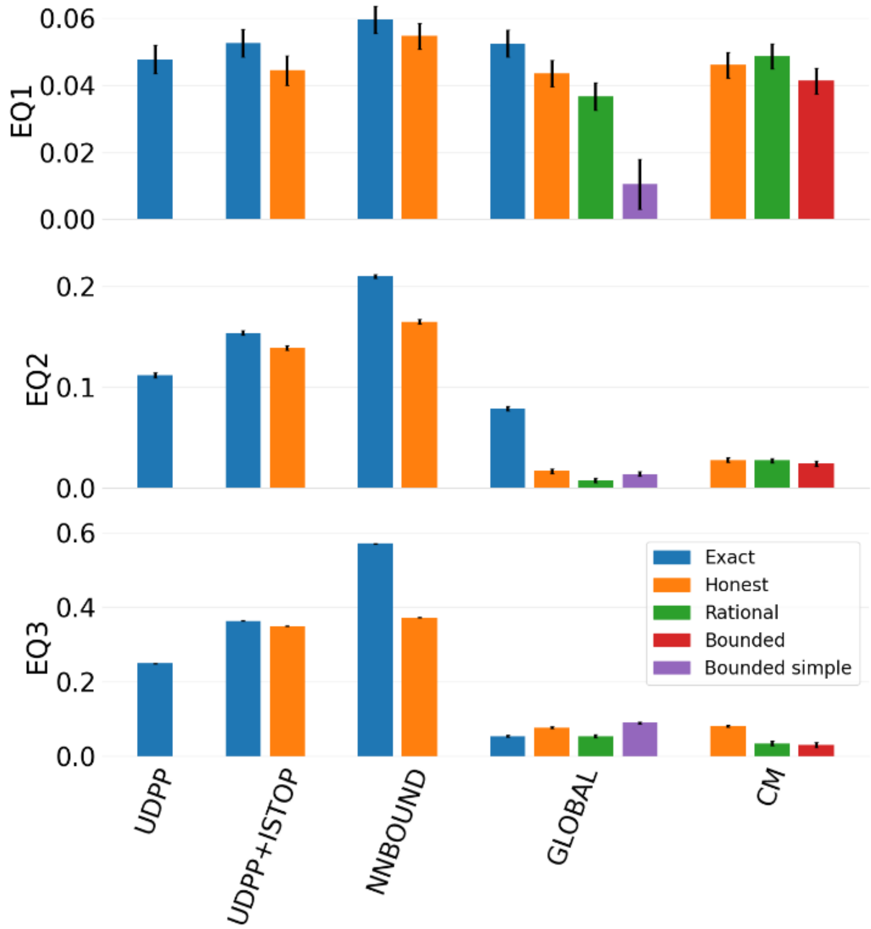


Figure 46: Aggregated equity indicators for different mechanisms and agent types.

Finally, it is striking to see that the approximation process in general has an adverse effect on equity. This effect is highly non-trivial and should be studied more in details. We will come back to this point in conclusions regarding the consequences in terms of mechanism design.

4.2.5 Difference between total and partial games

In this last section, we present two plots that shows that the calibration of the CM is far from trivial. Indeed, in section 3.3 we defined two types of games, 'total' and 'partial', depending on how the calibration was made. In the first one, we performed the calibration on the entire dataset, which means that the default jump parameter is then the same in every simulation for every airport. In the second one, a calibration is made airport per airport, leading to one default parameter for each airport.

The intuitive expectation on the results is that the second type of calibration is more suited and would yield some better results. The plots in Figure 47 show that it is not always the case. Indeed, in Figure 47 we see that the cost saved per flight is much smaller in partial games (using the second option for the calibration) than in total game. A similar trend, although much less pronounced, can be seen for the efficiency. Hence, we find that the mechanism is working better when the calibration is done on all airports at the same time.

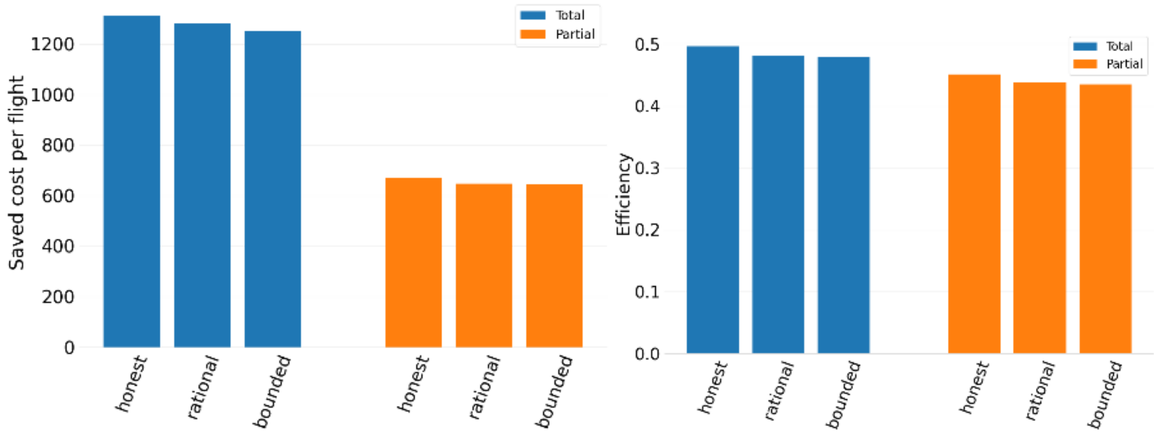


Figure 47: Saved cost per flight (left) or efficiency (right) with the CM in partial and total games.

However, it is important to point out that the efficiency of the mechanism is not the only metrics we care about. In Figure 48 we show indeed the equity metrics in the total and partial games. There are significant differences between these two games, with the partial game in general being fairer than the total game. It is probably because in the total game, the relative inadequacy of the calibration allows some airlines to pile up credits and thus brought to have an efficiency closer to the GLOBAL mechanism. We thus conclude by being reminded the fundamental tension existing between the pure economic efficiency and the fairness among actors, with the CM designed to find a good balance between.

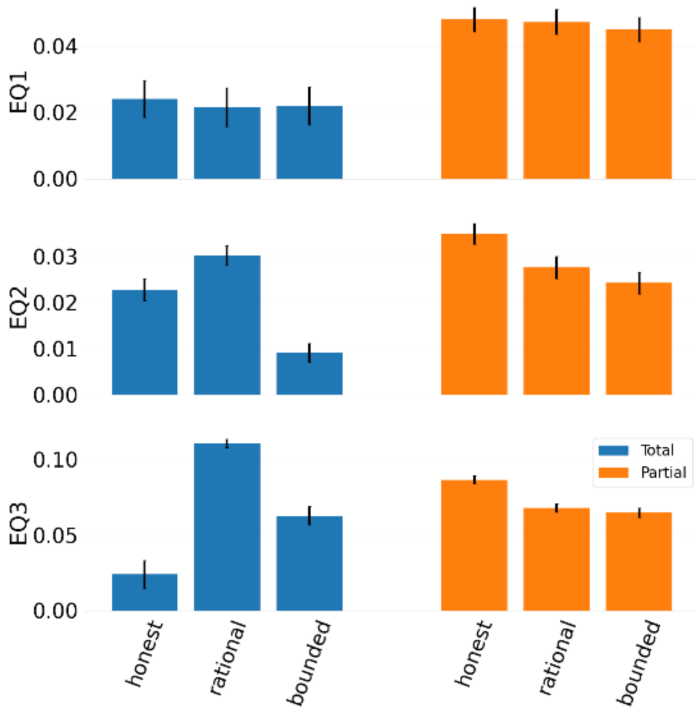


Figure 48: Equity metrics for the partial and total games for CM.

5 Conclusions

In this deliverable we have explored in depth the BEACON mechanisms with the help of small 'games' performed on a dataset compiling regulation information coming from historical data, and human-in-the-loop simulations. The games are played by different types of agents, honest, rational, or bounded. They are played at different airports, which, contrary to D5.1 where we had a similar setting otherwise, allows to have much more robust results, as well as more in-depth insights on the inner workings of the mechanisms in different situations.

The human-in-the-loop simulations allowed us to test our experiments on a setup where real humans could interact with the mechanisms. We designed an interface displaying all the relevant information that an airline could take into account when resolving regulations, and we gathered some important feedback from the participants. We saw that the two mechanisms tested attracted different opinions across the participants, with quite a good satisfaction overall regarding the easiness of use. On top of that, the Mercury simulator was successfully converted into a real-time simulator able to interact with human users, a feature that will likely be reused in the future. However, we note that the number of participants was small and some improvements can be made in terms of interface and learning curve. We are thus interested in developing further this line of research, which allows to gather important information on the behaviour of real humans as opposed to their AI counterpart.

Regarding the small fast-time games, we used similar experiments than in D5.1, only expanding the geographical scope, focusing on getting realistic results for the KPIs, and exploring the relationships existing between the KPIs and the airlines', airports', or regulations' characteristics.

A first important issue when expanding the geographical scope is the calibration of the CM. Indeed, because the calibration implies testing many past regulations to find the right level for the default parameter, using different airports immediately raises the question of whether the parameter should be calibrated for each airport independently, which may be difficult to do for a future implementation, or on all airports at the same time, which might make it less performant in general and in specific airports. We have shown that the calibration itself has a sizable impact on the results, most likely because a poor calibration (in the total game) may lead to a high number of credits for some airlines, which improves the overall efficiency (because the game comes closer to a GLOBAL mechanism) but impairs the equity.

Moving on to the results per se, we decided to track only three indicators (absolute cost, cost per flight, relative cost a.k.a efficiency) but we showed that the picture is already very complex. Indeed, depending on the level of aggregation and exploring the relationship between several variables, we demonstrated several features of the mechanisms, some of which counter-intuitive. We showed how some airlines gain a lot more than others in absolute term from some mechanisms, but also how efficiency was generally better distributed, especially when using the NNBOUND mechanism.

We showed how the analysis per airport brought some questions about the impact of the environment on the mechanism efficiency, i.e. how the exact structure of the regulation is crucial to understand how the mechanisms behave. This led us to compute unconditional non-linear relationships between some features of the regulation or the airline and the indicators themselves. We showed how statistical effects can blurry the picture, rendering naïve views obsolete. Thus, we studied more in details conditional relationships that brought into light the complexity of the analysis. Indeed, it seems that the expected gains for a given airline, in a given regulation bear complex relationships with variables like the total number of flights, the share of flights of the airlines, etc. All these relationships may or may not be important to study more in details the differences between airlines and how well such mechanisms may be adopted by them.

Looking at fully aggregated data, we tried to paint a picture that summarises our findings in terms of mechanisms and agent types. We found that UDPP brings already very high savings, even more so than what we found in D5.1, with around 45% of relative savings, or 900 euros per flights. ISTOP seems to bring no significant benefits over this value, and its notably lower when approximated cost functions are used. NNBOUND fares very well by adding 200 euros per flight of savings with respect to UDPP, close to the theoretical maximum given by GLOBAL, at least with true costs. Both are however lower or just at the level of UDPP when using approximations, contrary to what we found in D5.1. This raises the question of how to go beyond UDPP, given that approximations seem to be unavoidable to some extent in any realistic setting (see also the detailed discussion on this point in conclusions of D5.1 and D6.2). We find that the CM fares only just like UDPP in the best cases, which is not surprising given that it relies on an approximation at its heart. Finally, we find that in general gaming effects and behavioural biases degrade the impact of the mechanism in general, impairing it by several percentage points.

As a final conclusion, we summarise our most important findings below:

- Mechanisms are very sensitive to the cost approximation. Any mechanism using approximations should thus reflect on how to overcome this issue (conclusion in line with D5.1).
- ISTOP brings very little or no benefit on top of UDPP on average, even when used with true costs (conclusion partly in line with D5.1, where we find that ISTOP had some mild benefits).
- CM does not improve the situation with respect to UDPP, even in terms of equity, for which it was designed (conclusion in opposition to D5.1 where we found that CM was significantly more efficient than UDPP).
- NNBOUND seems to bring a significant improvement over UDPP while improving equity over it and over GLOBAL (conclusion in line with D5.1). We thus suggest to keep in mind this mechanism for any future design.
- UDPP overall is very efficient on average (around 45% of cost saved) (conclusion in line but stronger than in D5.1, where the efficiency was around 15%). Only highly optimised mechanism with access to true costs can mildly improve the situation over it.
- Given the previous fact, we raise the question of whether further improvements over UDPP are desirable or even feasible with significant gains. We also discuss this point further in D5.1 and D6.2.
- Behavioural and gaming effects have an important impact on the system and should not be underestimated. We suggest any future line of research to take them into account.
- CM seems to have a lesser sensitivity to both gaming and behavioural effects than its GLOBAL counterpart.

- Finally, we find that the calibration of the mechanism is crucial, and directly impacts how well CM works in terms of pure saved costs and how fair it is.

EUROPEAN PARTNERSHIP
Co-funded by the European Union

6 References

[1] Grant Agreement No 893100 BEACON Annex 1 Description of the Action

[2] BEACON Consortium Agreement

[3] BEACON D1.1 Project Management Plan, July 2020

[4] BEACON D2.1 Data Management Plan, December 2020

[5] BEACON D2.2 Database structure and data elaboration, January 2022

[6] BEACON D3.1 High-level modelling requirements, December 2020

[7] BEACON D3.2 Industry briefing on updates to the European cost of delay, September 2021

[8] BEACON D4.2 Final model results, July 2022.

[9] BEACON D5.1 First tactical model and results, October 2022.

[10] BEACON D8.1 H - Requirement No. 1, May 2021.

[11] BEACON D8.2 POPD - Requirement No. 5, May 2021.

[12] BEACON D8.3 POPD - Requirement No. 7, May 2021.

[13] U.S. Department of Justice and the Federal Trade Commission, Horizontal Merger Guidelines, 2010. Accessible here (retrieved 12/09/2022): https://www.justice.gov/atr/public/guidelines/hmg-2010.pdf

EUROPEAN PARTNERSHIP

Co-funded by the European Union

7 Acronyms

Acronym	Meaning
AFR	Air France
AI	Artificial Intelligence
BE	Behavioural Economics
CDG	Charles de Gaulle airport
CM	Credit Mechanism
FPFS	First Plan First Served
HD	Hyperbolic Discount
HHI	Herfindahl-Hirschmann Index
HITL	Human-in-the-loop
HMI	Human machine interface
HVU	high-volume users
ISTOP	Inter-airline Slot Trading Offer Provider
KDE	Kernel Density Estimate
KPI	key performance indicator
LVU	low-volume user
NM	Network Manager
NNBOUND	Non-negative bounded optimisation
PT	Prospect Theory
SUS	
UDPP	User-Driven Prioritisation Process

Appendix A Volume user categories

This annex presents an alternative categorisation of the airspace users in terms of their size in regulation (number of flights), with 3 categories instead of the 2 presented in section 3.2.5. In this case, we define a Low-Volume User (LVU) as being an airline with a share of flights smaller than 1%, while Medium-Volume Users (MVU) have between 1 and 5% and High-Volume Users have more than 5%. Figure 49 and Figure 50 show the results of this categorisation, in number and share of airlines respectively.

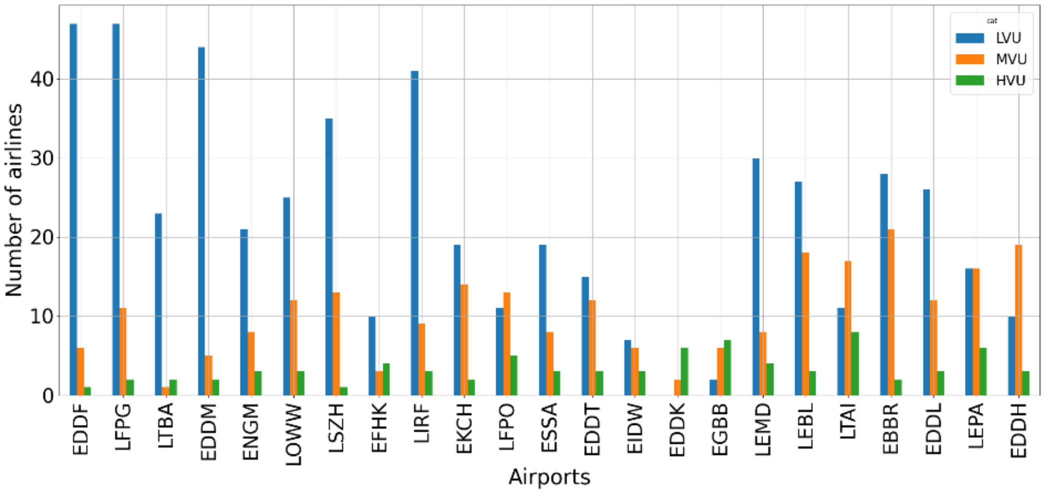


Figure 49: Number of airlines in each category (with 3 categories) in each airport.

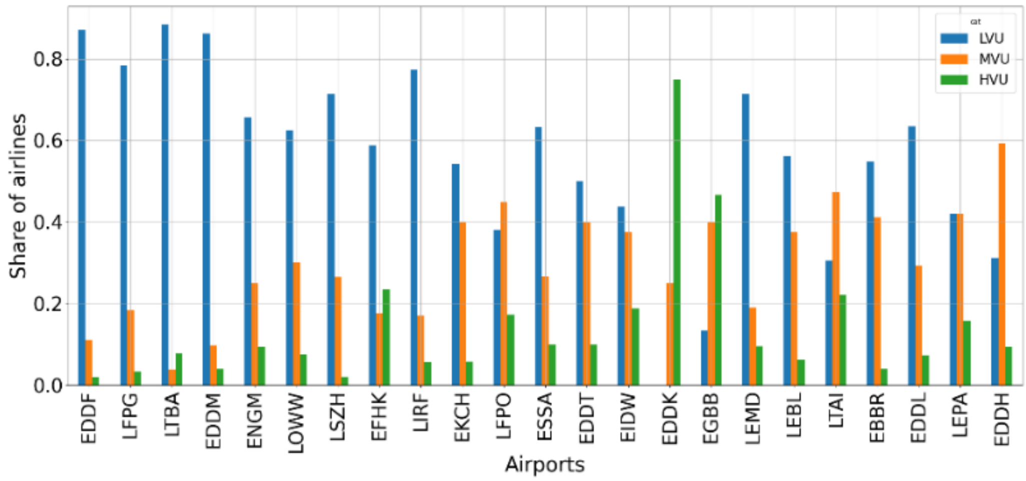
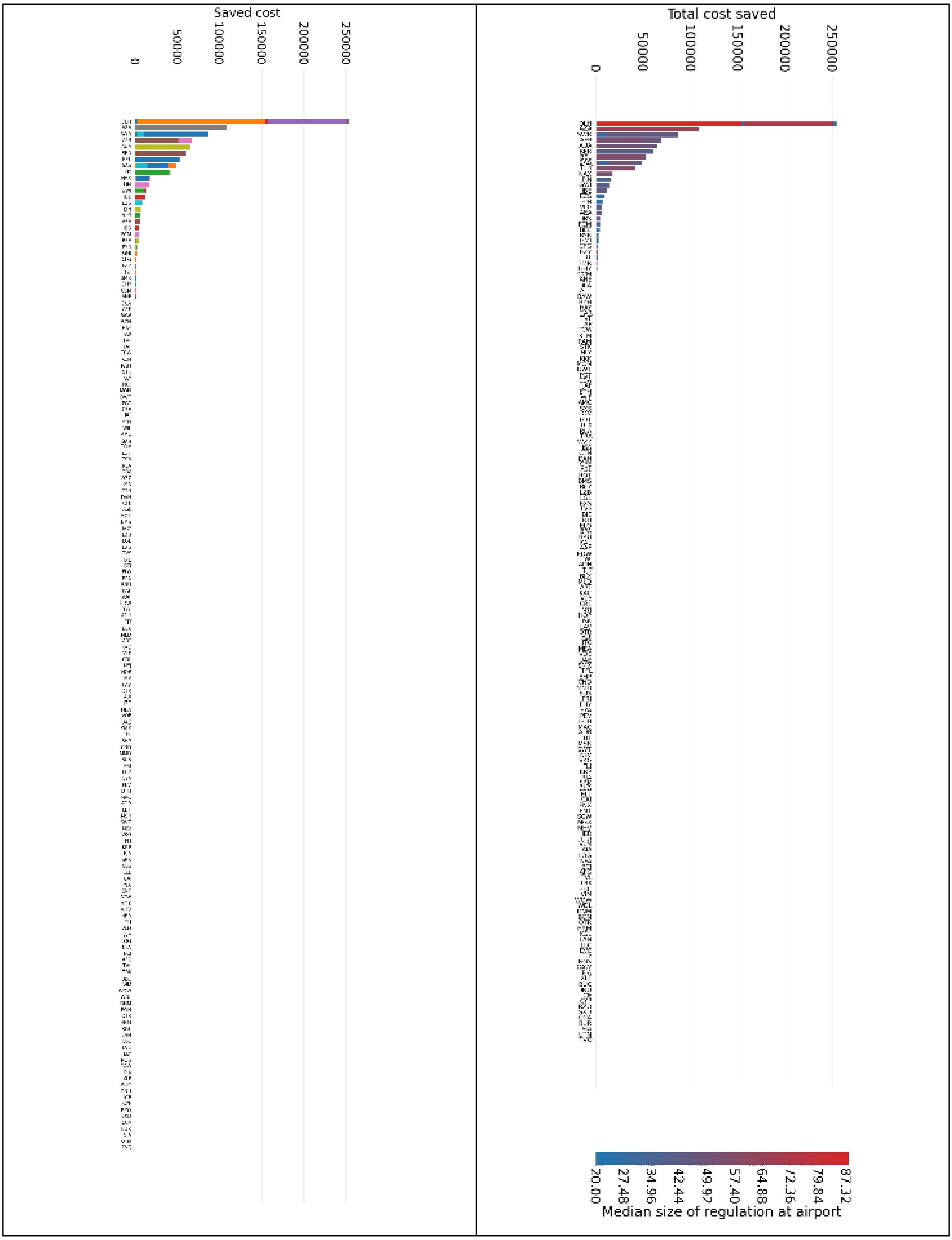


Figure 50: Share of airline in each category (with 3 categories) in each airport.

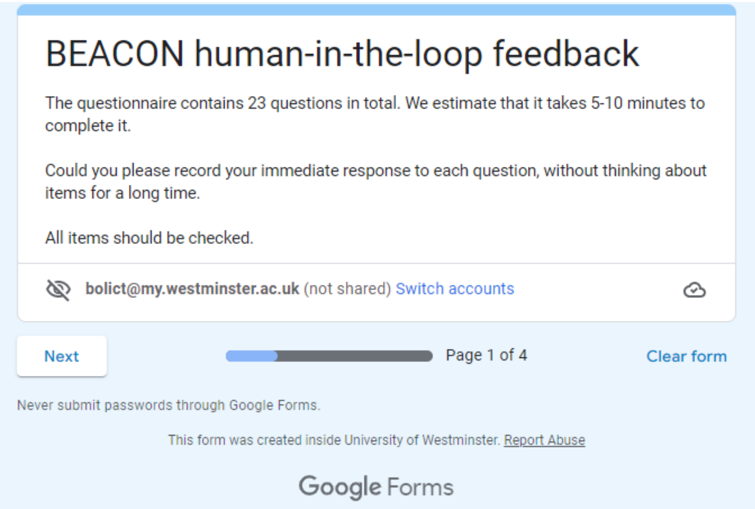
Appendix B Distribution of gain per airports

This annex presents the contribution in total saved cost of each airport for each for UDPP.



Appendix C HITL questionnaire

The after-HITL questionnaire is presented here.



BEACON human-in-the-loop feedback

The questionnaire contains 23 questions in total. We estimate that it takes 5-10 minutes to complete it.

Could you please record your immediate response to each question, without thinking about items for a long time.

All items should be checked.

bolict@my.westminster.ac.uk (not shared) Switch accounts

Next Page 1 of 4 Clear form

Never submit passwords through Google Forms.

This form was created inside University of Westminster. Report Abuse

Google Forms

Credit mechanism

1. I think that I would like to use this mechanism frequently. *

Strongly disagree 1 2 3 4 5 Strongly agree

2. I found the mechanism unnecessarily complex. *

Strongly disagree 1 2 3 4 5 Strongly agree

3. I thought the mechanism was easy to use. *

Strongly disagree 1 2 3 4 5 Strongly agree

4. I think that I would need the support of a technical person to be able to use this mechanism. *

Strongly disagree 1 2 3 4 5 Strongly agree

5. I found the functions in the mechanism were well integrated. *

Strongly disagree 1 2 3 4 5 Strongly agree

6. I thought there was too much inconsistency in this mechanism. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

7. I would imagine that most people would learn to use this mechanism very quickly. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

8. I found the mechanism very cumbersome to use. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

9. I felt very confident using the mechanism. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

10. I needed to learn a lot of things before I could get going with this mechanism. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Back Next Page 2 of 4 Clear form

ISTOP mechanism

1. I think that I would like to use this mechanism frequently. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

2. I found the mechanism unnecessarily complex. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

3. I thought the mechanism was easy to use. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

4. I think that I would need the support of a technical person to be able to use this mechanism. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

5. I found the functions in the mechanism were well integrated. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

EUROPEAN PARTNERSHIP

Co-funded by the European Union

6. I thought there was too much inconsistency in this mechanism. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

7. I would imagine that most people would learn to use this mechanism very quickly. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

8. I found the mechanism very cumbersome to use. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

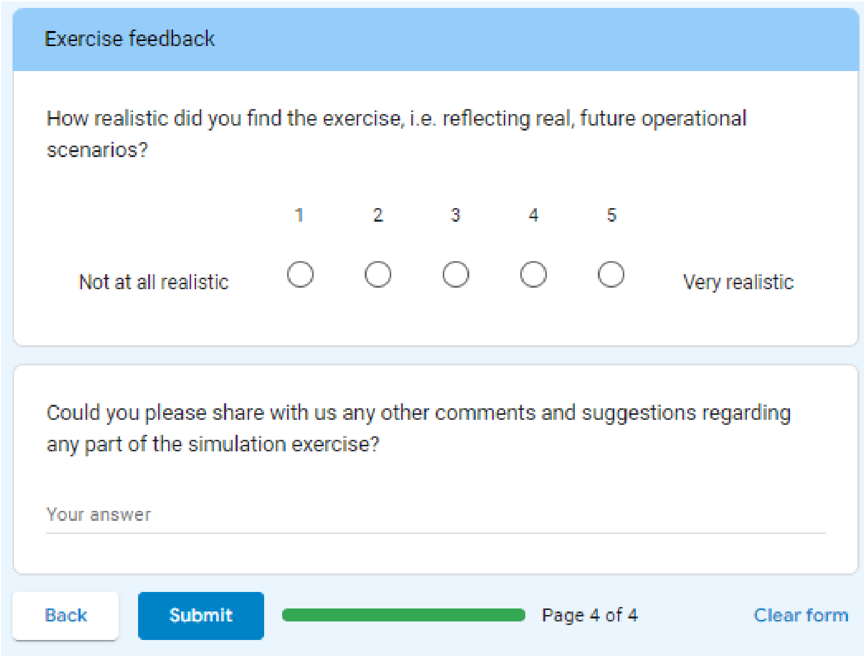
9. I felt very confident using the mechanism. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

10. I needed to learn a lot of things before I could get going with this mechanism. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Back Next Page 3 of 4 Clear form



Exercise feedback

How realistic did you find the exercise, i.e. reflecting real, future operational scenarios?

	1	2	3	4	5	
Not at all realistic	○	○	○	○	○	Very realistic

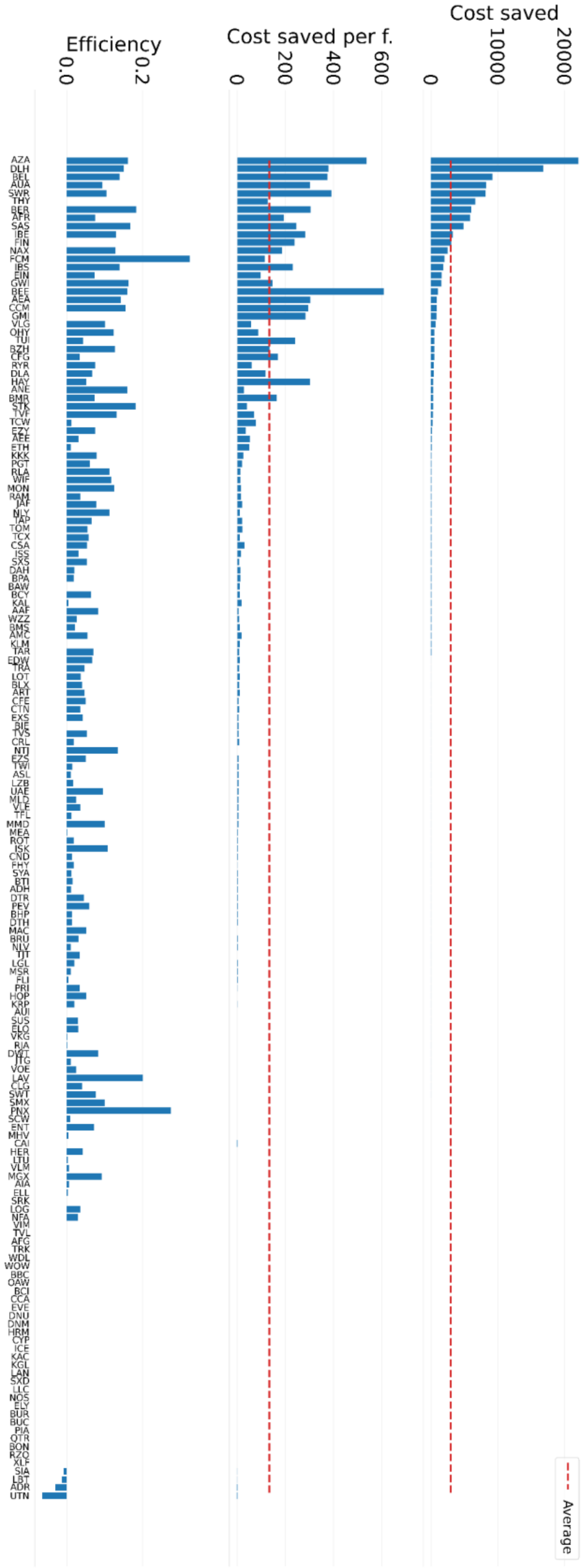
Could you please share with us any other comments and suggestions regarding any part of the simulation exercise?

Your answer

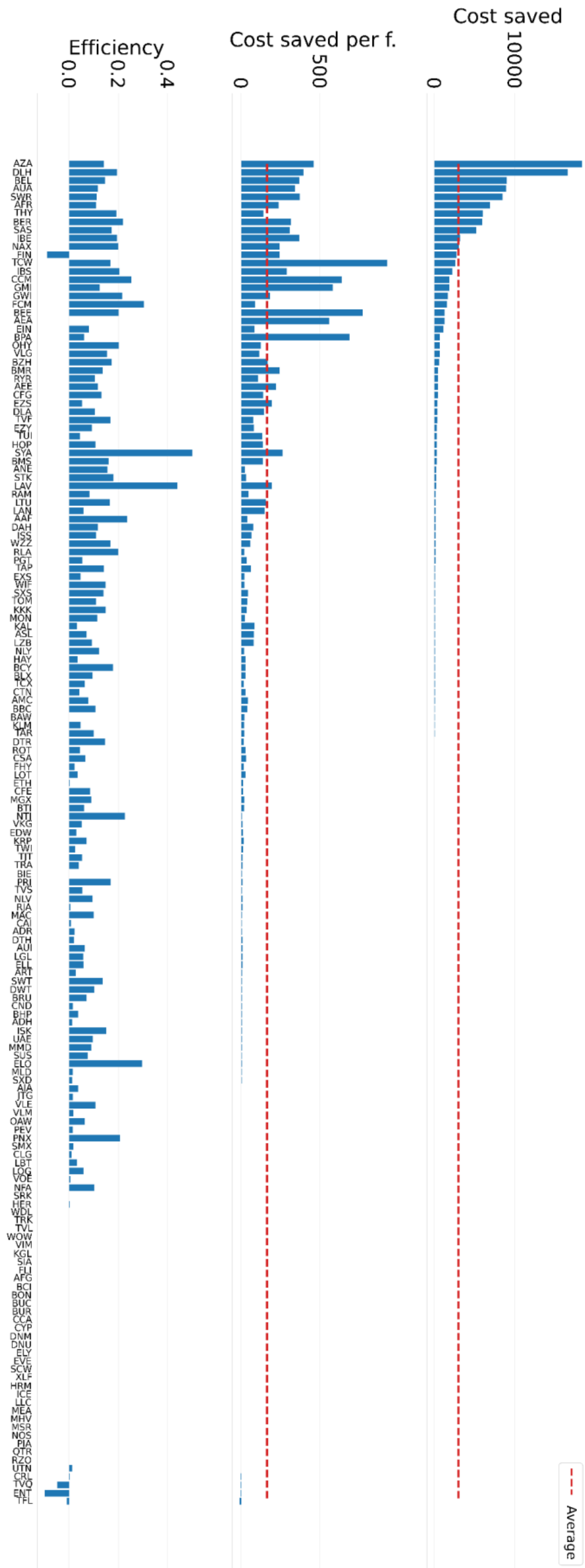
Back Submit Page 4 of 4 Clear form

Appendix D Indicators per airline for all mechanisms

In this appendix we show all the results per airline, some of which may be found in section 4.2.1.

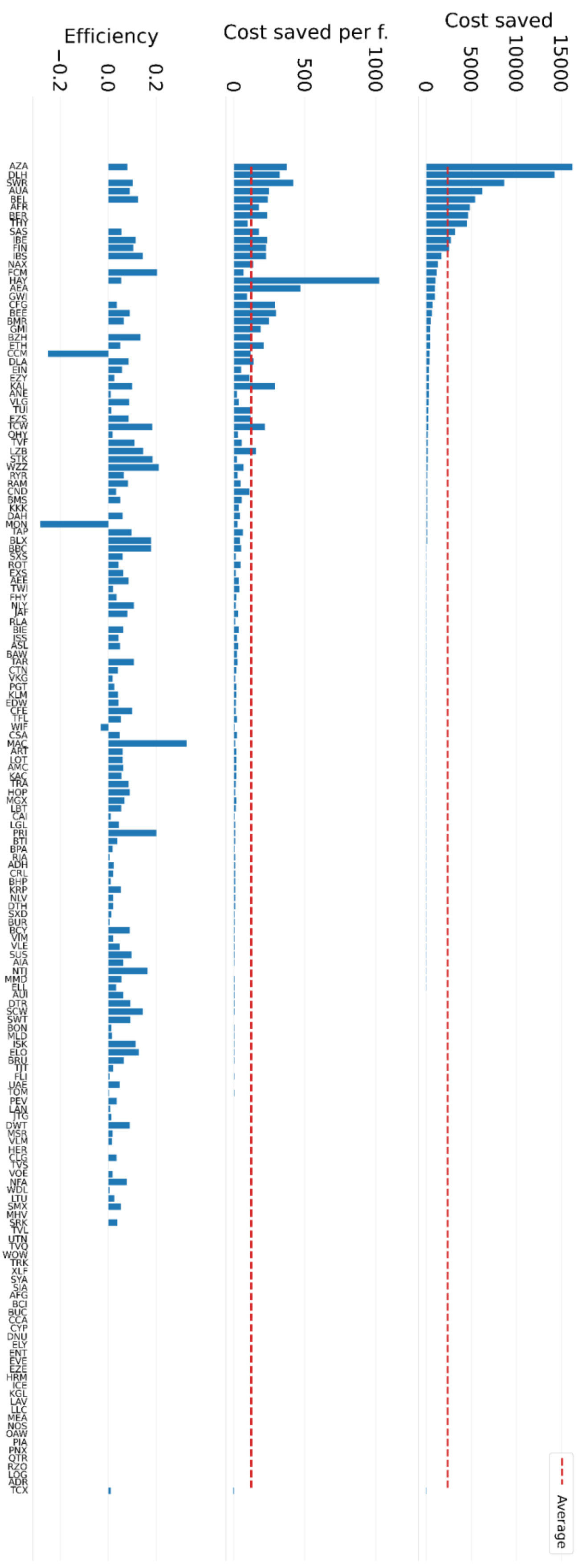


UDPP

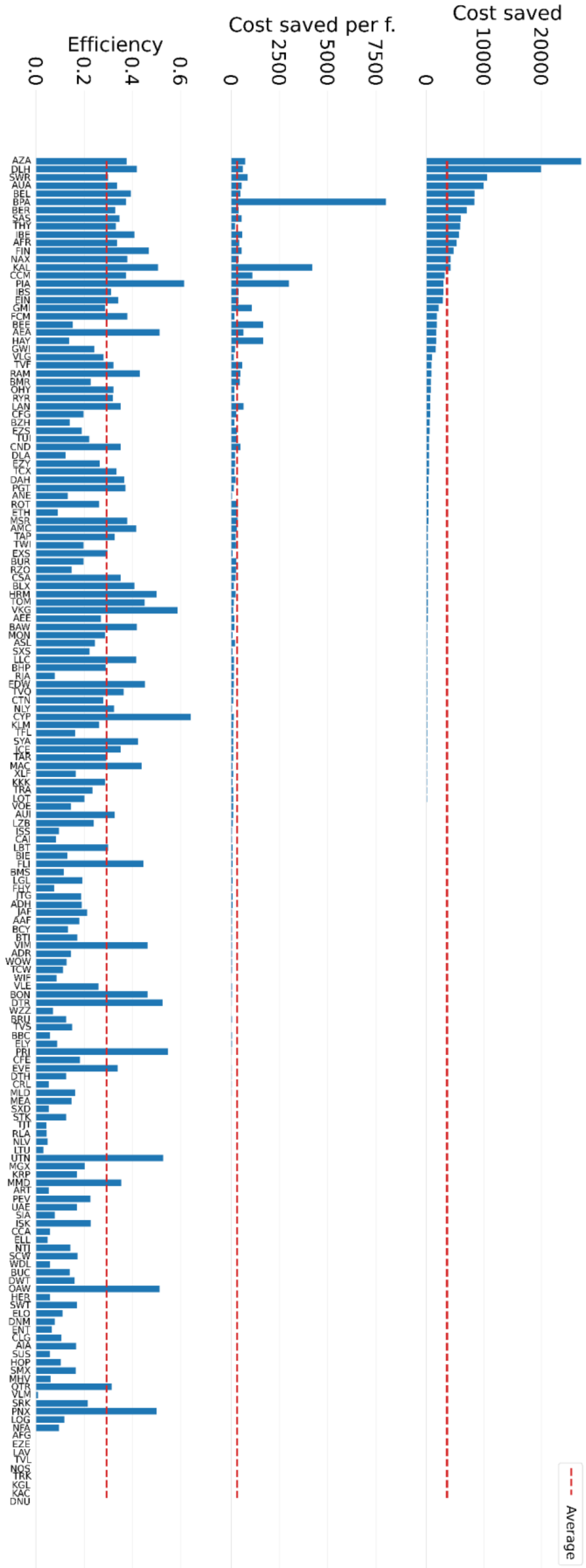


UDPP+ISTOP_TRUE

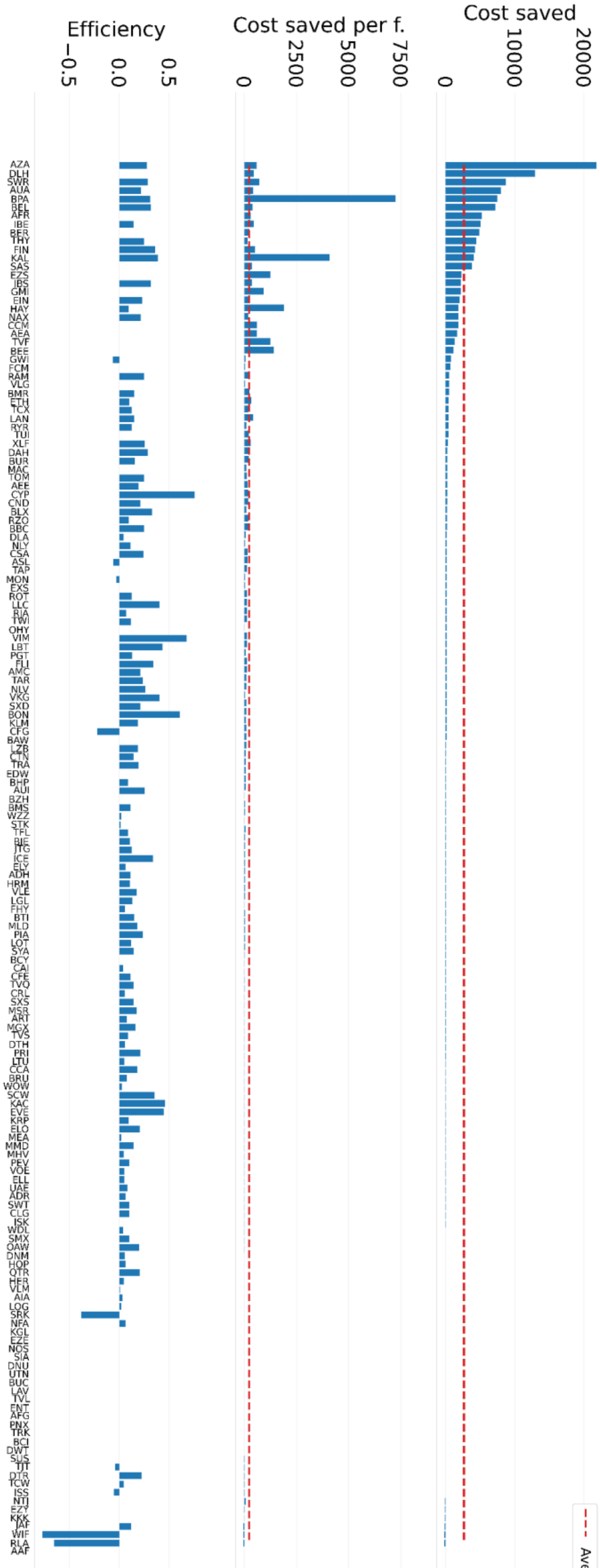
Co-funded by the European Union



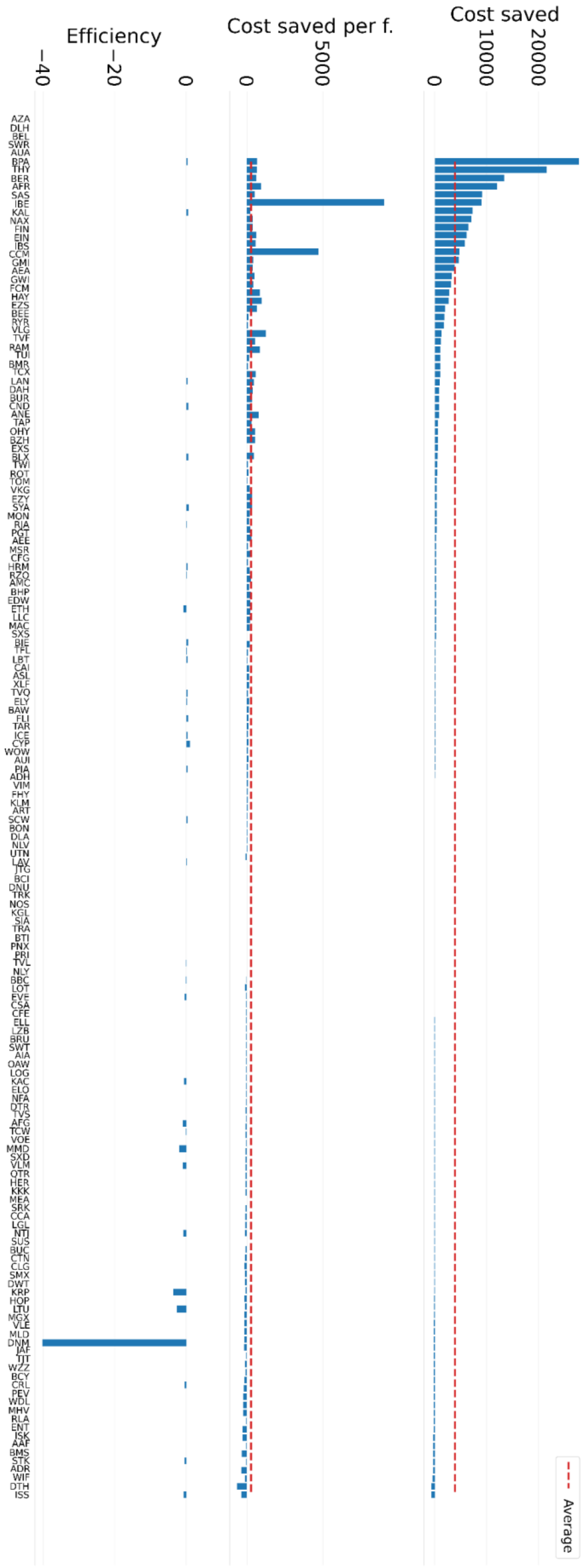
UDPP+ISTOP_APPROX



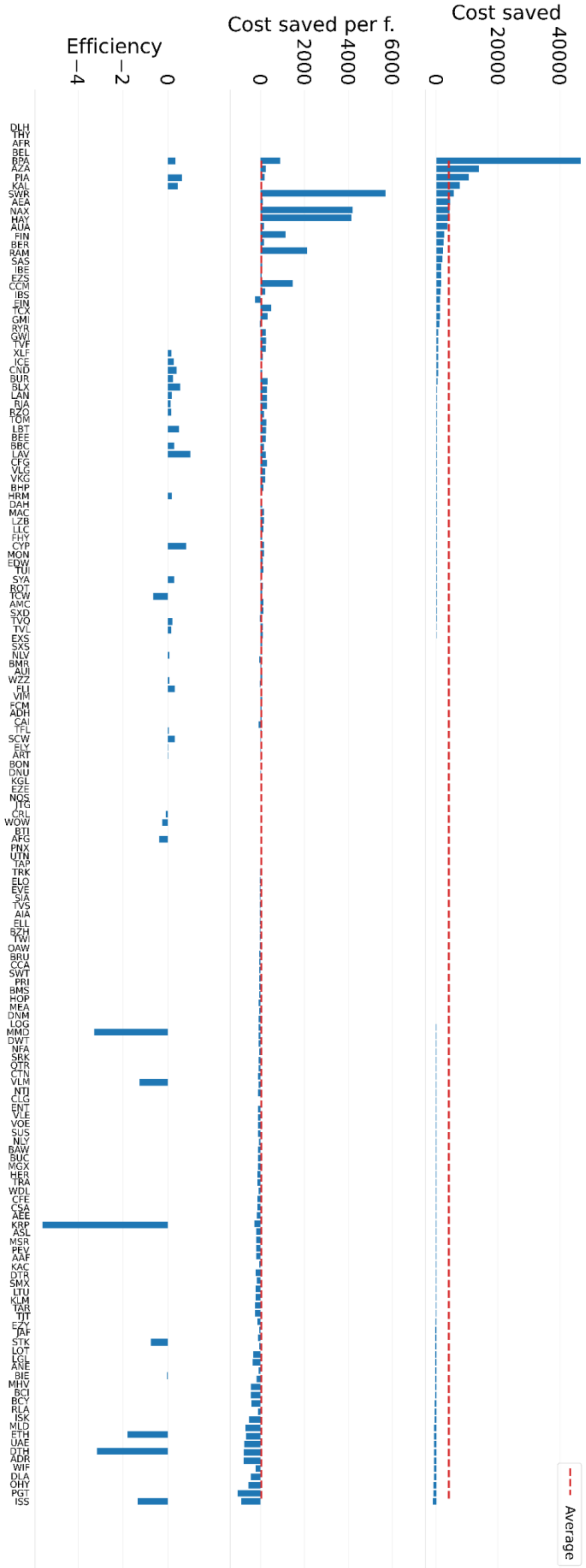
NNBOUND_TRUE



NNBOUND_APPROX


Efficiency
Cost saved per f.
Cost saved
−40
−20
0
0
5000
0
10000
20000
Average

GLOBAL_TRUE

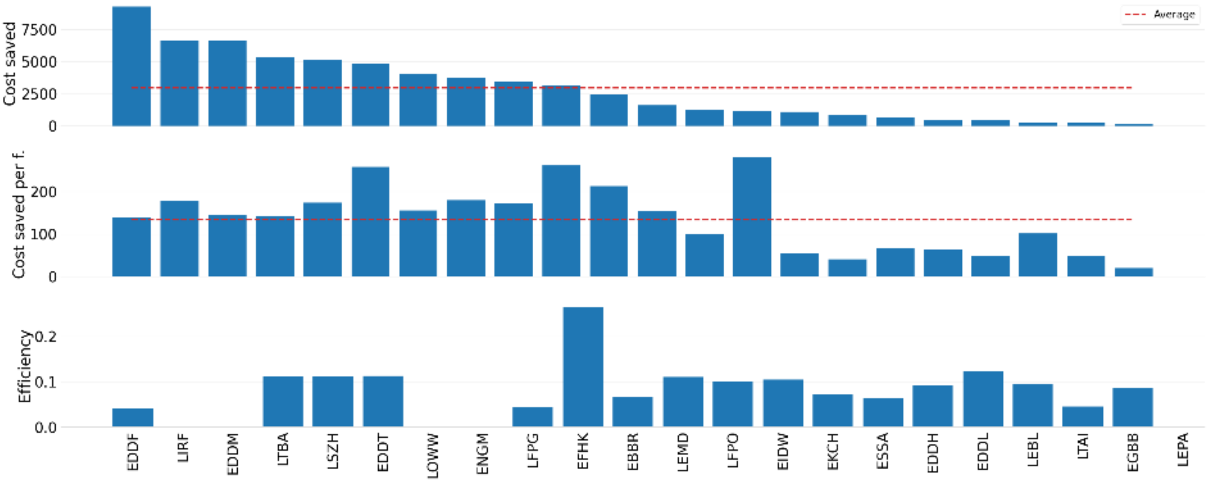


GLOBAL_APPROX

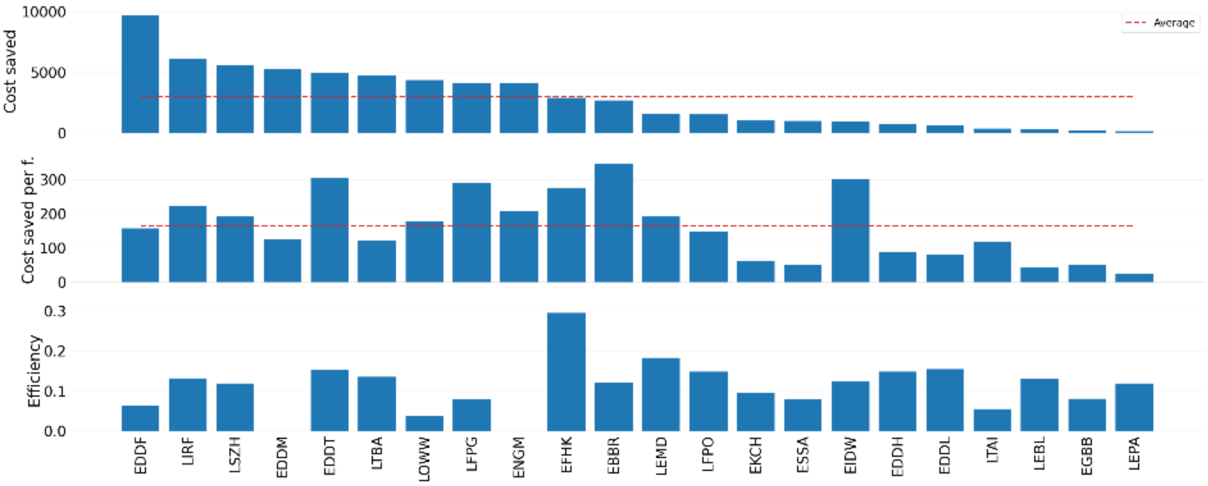
Appendix E Indicators per airports for all mechanisms

In this appendix we show the distributions of indicators per airport, some of which are presented in section 4.2.2. All results have been obtained with the honest agents.

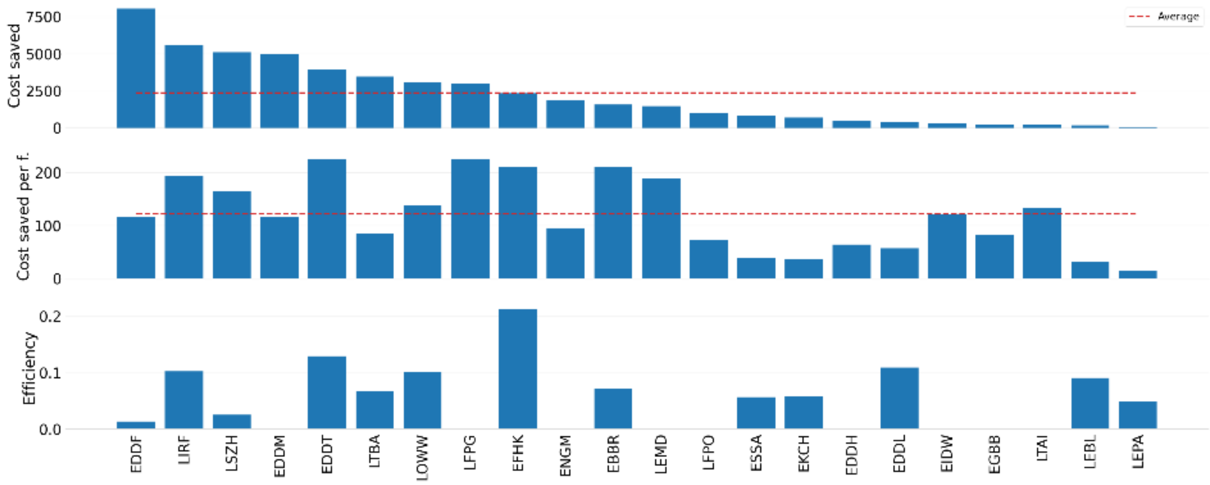
E.1 UDPP



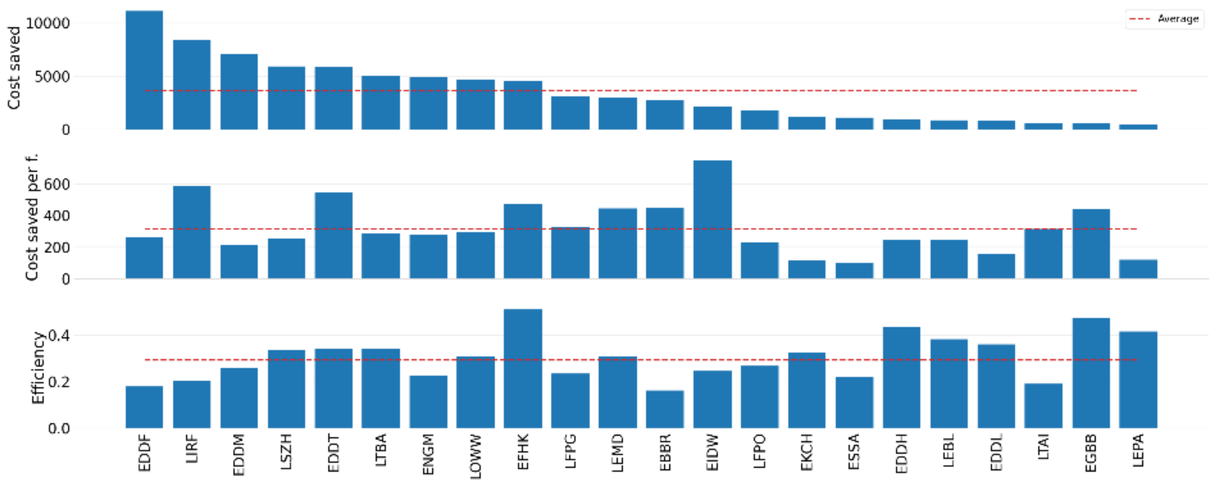
E.2 UDPP+ISTOP_TRUE



E.3 UDPP+ISTOP_APPROX



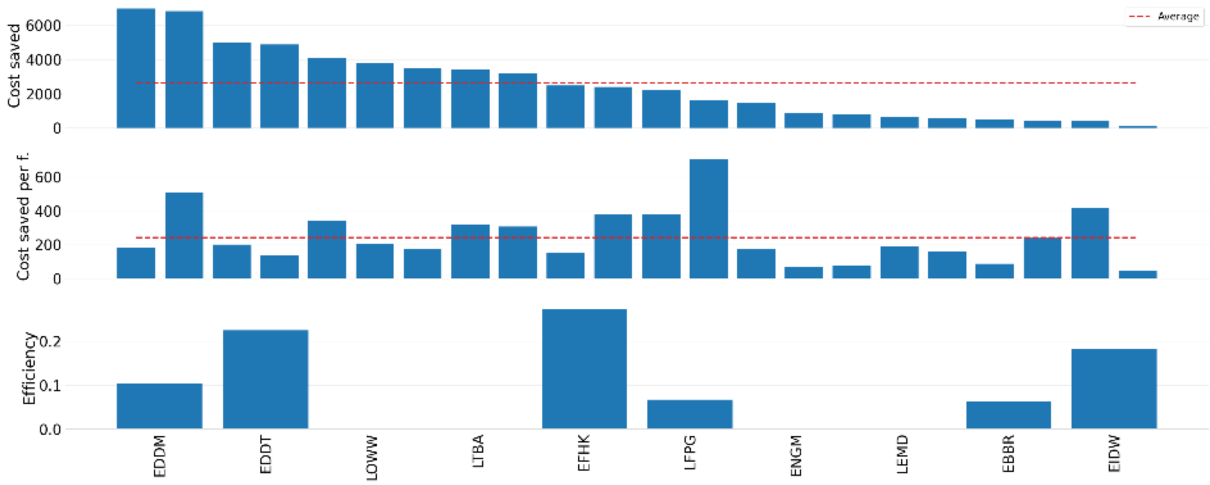
E.4 NNBOUND_TRUE



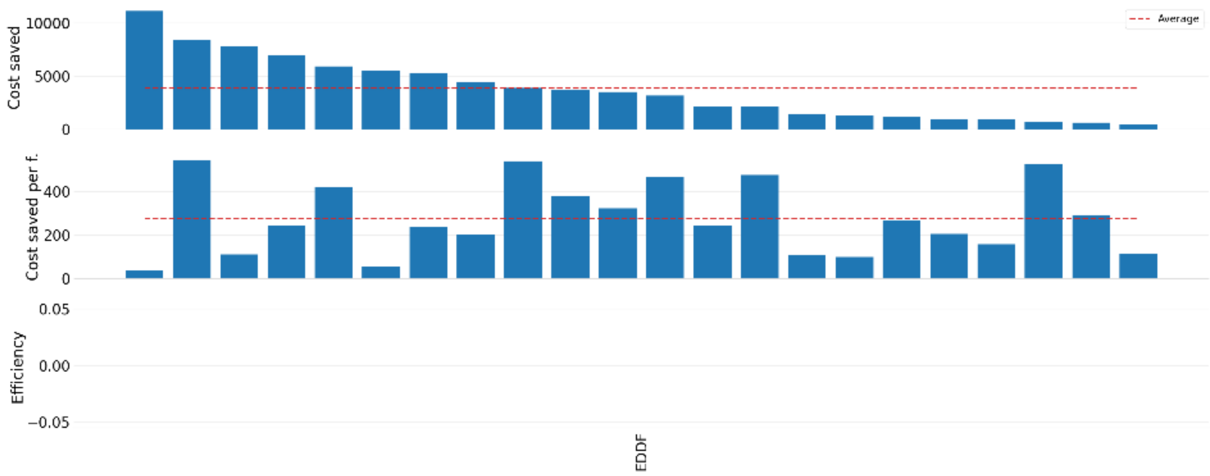
EUROPEAN PARTNERSHIP

Co-funded by the European Union

E.5 NNBOUND_APPROX



E.6 GLOBAL_TRUE



E.7 GLOBAL_APPROX