D5.2 Investigative case studies results

Deliverable 5.1

Domino

Grant:	783206
Call:	H2020-SESAR-2016-2
Topic:	SESAR-ER3-06-2016 ATM Operations, Architecture, Performance and Validation
Consortium coordinator:	University of Westminster
Edition date:	31 July 2019
Edition:	01.00.00

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Fabrizio Lillo / Università di Bologna	Project member	25 July 2019
Piero Mazzarisi / Università di Bologna	Project member	25 July 2019
Silvia Zaoli / Università di Bologna	Project member	25 July 2019
Luis Delgado / University of Westminster	Project member	25 July 2019
Gérald Gurtner / University of Westminster	Project member	25 July 2019
Damir Valput / Innaxis	Project member	25 July 2019

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Andrew Cook / University of Westminster	Project coordinator	26 July 2019
Graham Tanner / University of Westminster	Project member	26 July 2019

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

Name/Beneficiary	Position/Title	Date
Andrew Cook / University of Westminster	Project coordinator	30 July 2019

Rejected By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
N/A		

Document History

Edition	Date	Status	Author	Justification
01.00.00	30 July 2019	Release	Domino Consortium	New document for review by the SJU

Domino

NOVEL TOOLS TO EVALUATE ATM SYSTEMS COUPLING UNDER FUTURE DEPLOYMENT SCENARIOS

This deliverable is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 783206 under European Union's Horizon 2020 research and innovation programme.

Abstract

This deliverable presents the results from the analysis of the model executing the investigative case studies. The document focuses on the validation activities and the results for the three mechanisms modelled in Domino in the unitary case studies.

The three mechanism are: 4D Trajectory Adjustment, which focuses on the use of dynamic cost indexing and wait-for-passengers rules; Flight Prioritisation, which considers the possibility of slot swapping at ATFM regulations; and Flight Arrival Coordination, which models different optimisation approaches E-AMAN could consider. Each mechanism has three levels of implementation: Level 0 (with current capabilities), Level 1 (with more advanced features) and Level 2 (more explorative). The traffic is set on a given day (12 September 2014) considering flights and passengers' itineraries. Two levels of delay are considered: default and stressed. In total 14 scenarios have been modelled and analysed.

This deliverable presents the use of classical and network metrics (centrality and causality) on the outcome of the whole European level agent-based model. The model still requires further development and adjustment, but results show that it is already capable of capturing complex interactions among the ATM elements. Finally, the network metrics are already presenting their potential to capture non-direct interactions between elements in the system.

The results have been shared with experts and airspace users at two workshops. The feedback obtained and the results of the analysis and validation activities will be considered for the final version of Domino.

The opinions expressed herein reflect the authors' views only. Under no circumstances shall the SESAR Joint Undertaking be responsible for any use that may be made of the information contained herein.

Table of Contents

List of figures

List of tables

Executive summary

This deliverable presents the first results of the full Domino model. It focuses on the investigative case studies and on the impact of each of the three mechanisms selected and modelled in Domino.

The model focuses on the traffic of the 12th September 2014 considering flights and passengers' itineraries. This operational environment has been used for the calibration activities which are presented in this document. The calibration of the model has been performed only approximately for this deliverable, due to foreseen changes in the code. The model has been adjusted considering historical values and, in particular, delay statistics from EUROCONTROL CODA [1]. Some shortcomings of this calibration are highlighted and will be considered for the final version of the model (e.g., higher cancellation rate than desired; lack of earlier-than-scheduled departures in the model, while present in the historical data).

The three mechanisms that are considered in Domino are:

- 4D Trajectory Adjustment (4DTA), which focuses on the use of dynamic cost indexing and waiting for passenger rules at hubs;
- Flight Prioritisation (FP), which considers the possibility of slot swapping when ATFM delay is assigned due to regulations at the arrival airport;
- Flight Arrival Coordination (FAC), which considers the implementation of E-AMAN at selected airports with different prioritisation strategies to create the arrival sequence.

Each mechanism is modelled with three incremental levels of complexity:

- Level 0, aimed at capturing the current operational environment;
- Level 1, with improvements that could be achieved in the short term in the context of SESAR;
- Level 2, with more advanced features to provide exploratory results.

The detailed description of the mechanisms are provided in D3.1 (Architecture definition [2]).

Finally, two levels of system stress (i.e., delay) are considered:

- default, where delay in the system is similar to a nominal operating day;
- stressed, where delay has been artificially increased (by selecting high ATFM regulations, and lower airport capacity), in order to assess how the mechanisms perform in these situations. This is more aligned with expected traffic growth and capacity-demand stresses in the system.

With these considerations, a total of 14 scenarios are analysed and presented in this deliverable:

- default: all three mechanisms at Level 0;
- 'unitary': all three mechanisms each at Level 1 and 2, with the other mechanisms kept at Level 0;
- the corresponding 7 extra scenarios, to those above, all in the stressed case.

Some scenarios where more than one mechanism is activated above Level 0 have been computed (e.g., tactical scenario where 4DTA and FAC are implemented at a higher level). However, due to the complexity of the interactions between the mechanisms and the focus here on understanding how the mechanisms perform and the capabilities/limitations of the model and the metrics, these are not presented in this deliverable.

In Deliverable D5.1 - Metrics and analysis approach [3], classical metrics were analysed but more importantly, new network related metrics were presented: centrality, able to indicate how a node in the network is important for the connectivity of passengers, and causality, reflecting indirect causality relationship between nodes. In this deliverable, the usage of these metrics and their computational approach is presented for the different scenarios mentioned above at a European-wide level. In this deliverable, these metrics already show their potential and highlight their possible evolution towards more operational indicators. The model is very detailed and low-level, which allows us to obtain a fine representation of many standard (e.g., delay) and advanced metrics (e.g., cost or passenger trip duration).

The summary of the main findings for all the mechanisms is that in the stressed case, **all advanced levels of all mechanisms improve the cost impact on the airlines with respect to Level 0**. For some mechanisms, this is achieved with trade-offs with other KPIs (e.g., a worsening of the average arrival delay). This is expected, as different flights might experience different costs of delay and hence those will be prioritised when focus is put on cost, rather than delay.

Passengers might experience longer trip times under some mechanisms. This indicates that focusing on airline cost does not necessarily benefit the passenger experience as airline and passenger objectives might not be fully aligned (in particular, when the cost of fuel is considered, the mechanisms might trade delay for fuel savings).

The centrality of airports tends to improve with the introduction of the mechanisms. Interestingly, these effects are sometimes identified even if the airport does not implement the mechanism itself (e.g., airports benefiting from others having an advanced Flight Arrival Coordination in place).

With respect to causality, results are mixed across the mechanisms and scenarios. In some cases, improvements are observed where causality links decrease, but in others the system gets ‘tighter’ (more strongly coupled), leading to a worsening of these performances.

In general, mechanisms provide a better relative outcome with respect to maintaining current (Level 0) implementations, under the stressed scenarios. This could be expected, as if there is not much delay (as in the default case), then the mechanisms do not provide a significant benefit.

All these results indicate that, whilst the model needs some finer calibration and improvements, it is already able to capture the intricate effects arising from the massive number of interactions and the tight connection of the system elements, and to reflect the impact of changes in the ATM environment.

Focusing on the different mechanisms, their main results can be summarised as follows:

- ***4D Trajectory Adjustment***

    - The implementation of advanced levels of 4DTA (Level 1 and Level 2) leads to a general slowing down of flights (or not speeding up that much), relative to the default scenario. This drives an increase in delay (for flights and passengers) and is due to the explicit consideration of fuel costs when deciding whether to recover delay.
    - While passenger delays increase, when the mechanism is implemented at Level 1, the metrics evaluating the preservation of passenger itineraries (i.e., the number of passengers with a modified itinerary and the passenger centrality metrics) show that passengers arrive more often at their destination using their scheduled itinerary than in the baseline. The mechanism is effective at prioritising and maintaining those connections.
    - At Level 2, however, speeding up is reassessed at top of climb, allowing the speed to be reduced with respect to the planned speed, thus saving fuel. The potential reduction in fuel costs is traded against passengers' connectivity. Around 15% of flights which initially (at gate) decide to speed up, at top of climb decide to take advantage of the possibility to slow down and save fuel instead. This might be triggered due to most up to date information on passenger expected missed connections and costs and a better estimation of the expected inbound arrival delay.
    - The propagation of costs of delay among airports, as measured by the density of causal links among airports, diminishes. However, in most cases this decrease of cost propagation does not seem to eliminate the bidirectional propagation patterns measured by the reciprocity, which potentially increases costs by propagating them in a loop (also known as 'back propagation').
- ***Flight Prioritisation***
    - At a system level, the impact of this mechanism on the costs experienced by airlines is negligible. This mechanism only applies to around 2% of the flights (i.e., the ones which arrive at an airport that is regulated due to ATFM). Results show some benefits when restricted to flights that arrive at an airport implementing an ATFM regulation.
    - In the default scenario, the experience of passengers is worsened. However, in the stressed cases, there is an improvement of passenger metrics. The economic interest of airlines is better aligned with passengers' convenience in the high delay cases.
    - In the stressed scenarios, there is an improvement of trip centrality loss in the airports affected by ATFM regulations.
    - At a system level, the mechanism does not have an impact on the propagation of delay (measured through causality).
- ***Flight Arrival Coordination***
    - In the scenarios where the FAC mechanism is implemented, a decrease of average delay is seen at Level 1, but the mechanism is less effective on delay reduction at Level 2. This is expected, as in Level 2 the Arrival Coordination mechanism aims to minimise the airline costs and not necessarily the delay, while Level 1 focuses on the total delay (arrival and reactionary) at the airport. If the analysis is restricted to airports that implement the mechanism, the improvements are, as expected, larger, particularly for passenger associated costs.
    - Passenger delays decrease, especially in the stressed case. The disruption of passengers' itineraries, as measured by the number of passengers with modified itineraries and by passenger centrality loss, sees an overall improvement.

- Network effects are identified for this mechanism. Passenger centrality loss is improved for airports that do not implement the mechanism, even if they are larger. This indicates that introduction of FAC has positive externalities also on airports that do not implement it.
- There is an overall decrease in the level of causality measured by a diminished number of propagation channels at the European level. Once again, some externalities can be quantified on the impact of the mechanism on airports that don't implement it. In most cases, the reduction in the level of causality has a small effect on the bidirectional propagation patterns measured by the reciprocity, which potentially enhances delays, through back propagation.

The information presented in this deliverable has played an important role in the validation, dissemination and feedback activities of Domino. Domino participated in a workshop with airspace users and organised a dedicated full day workshop with ATM experts to present the model, the calibration and some of the results. More information about these activities can be found in D6.3 - Workshop results summary [4].

All the information gathered, results from the validation, shortcomings identified with the model from the validation and the analysis of the results, limitations and possible evolution of the metrics (particularly focusing on the practicalities of the operational usage of the new network metrics) and the selection of relevant scenarios to be modelled for the final version of Domino, will be considered in D3.3 - Adaptive case studies description. In this way, D3.3 will become the blueprint of the activities that will be undertaken until the end of the project, and the information presented in this deliverable is one of the key drivers of these activities.

1 Introduction

This deliverable presents the calibration of the first version of the model and the results of the analysis of key scenarios identified in the investigative case studies. These results have been used to:

- obtain feedback from stakeholders (some results were presented in two different workshop activities - see D6.3 Workshop results summary [4]),
- along with the feedback, identify changes that need to be done to the metrics, the model and the scenarios for the final version of Domino.

This introduction presents the key elements of the deliverable starting by the investigative case studies, the model calibration activities and the differentiation between classical and network metrics, and presents the structure of the document.

1.1 Investigative case studies

In Deliverable D3.1 Architecture definition [2] the different elements that will be modelled in Domino were identified. This included three mechanisms with different levels of implementation:

- 4D Trajectory adjustment (4DTA): focusing on airspace users management of disruption by modifying, adapting their flights (i.e., dynamic cost indexing and waiting-for-passenger at hub);
- Flight prioritisation (FP): this mechanism considers different alternatives to manage the delay assigned by ATFM regulations at destination airports (e.g., no swap allowed, swapping within intra-airline flights and swapping slots inter-airline);
- Flight Arrival Coordination (FAC): FAC mechanism tries to capture different implementation of extended arrival managers (E-AMAN), from focusing on the runway arrival throughput, to considering reactionary delay at the airport or airline expected costs.

As described in D3.2 Investigative case studies description [5], two different delay levels are considered: baseline and stressed, where delay is significantly increased, cases.

This deliverable will focus on the results of unitary case studies to understand the impact of the mechanisms in isolation. See Section 2 for more details on the scenarios modelled and analysed.

1.2 Model calibration

The ABM model described in D4.1 Initial model design [6] has been implemented. There are some effects that are explicitly modelled (e.g., propagation of delay due to reactionary delay), but others are the results of a higher abstraction (e.g., actual taxi time or turnaround times). These parameters are stochastically sampled from distributions that need to be calibrated.

This has been done considering the baseline scenario which represents the current (2014) operations and considering the analysis from different data sources to use them as calibration targets. Section 3 of this deliverable present these calibration activities and outcomes.

1.3 Classical and network metrics

As described in Deliverable D5.1 Metrics and analysis approach [3], classical metrics (delay, cost) have been defined in the past and used considering different stakeholders (namely flights and passengers). Domino goes beyond those metrics identifying and defining network metrics which focus on centrality and causality. The different metrics that have been computed for the different scenarios are summarised before presenting the scenarios results in Section 4 of the deliverable.

1.4 Structure and contents of this deliverable

Section 2 condenses the information on the scenarios modelled and analysed in this deliverable. Section 3 describes the results of the calibration activities carried out in the baseline scenario. The data sources used for these validation activities are also identified.

The results of the investigative case studies described in Section 2 are presented in Section 4. In order to facilitate the readability of the document this section has been structured as follows: first the metrics that have been computed and the main assumptions for their estimation are described (Section 4.1). Then a summary of the key results is presented (Section 4.2). This summary highlights the main findings for the different scenarios and will be sufficient for readers interested in this comparison between scenarios. The detailed analysis of the different scenarios is then presented in Section 4.3, where for each mechanism implemented the results on flight and passenger delays, cost, centrality losses and causality analysis are described for the baseline and stressed cases.

Conclusions of these analysis (beyond the comparison between the scenarios) are summarised in Section 5. The deliverable closes with Section 6 where next steps and look ahead is presented, Section 7 with the references and the acronyms grouped in Section 8.

2 Scenarios modelled and analysed

The results presented in this deliverable have been obtained using several scenarios. In D3.1, the Domino project has produced a roadmap for the simulations, in order to prioritise the scenarios. In this deliverable, we focus mainly on producing results for the isolated mechanisms, in order to test and understand them, with a special emphasis on the Level 2 implementation. As a reminder, for each mechanism, three levels of implementation are defined in Domino following this rationale:

- Level 0: representing current (2014) situation,
- Level 1: advanced capabilities,
- Level 2: exploratory approach.

Two level of congestion are considered in the simulations:

- Default scenario: congestion level based on 2014 data,
- Stressed scenario: highly congested system.

Table 1. Scenarios parameters considered

Congestion level	4DTA	FP	FAC
• Default • Stressed	• Level 0 • Level 1 • Level 2	• Level 0 • Level 1 • Level 2	• Level 0 • Level 1 • Level 2

Combining these features (mechanism and congestion level (as shown in Table 1)), this deliverable analysed the following scenarios:

1. Default scenario (with all mechanisms at Level 0)
2. Default scenario + 4DTA Level 1
3. Default scenario + 4DTA Level 2
4. Default scenario + FP Level 1
5. Default scenario + FP Level 2
6. Default scenario + FAC Level 1
7. Default scenario + FAC Level 2

and the same number for the stressed scenario, for a total of **14 scenarios**.

In the following, we describe more in detail how we built some of the scenarios, and in particular how we implemented the different mechanisms.

2.1 Default scenario

The default scenario is based on a typical day of 2014, the 12th of September (the rationale for the selection of this day has been described in [2]). Individual passenger itineraries are modelled for this day, as well as scheduled flights (including type of aircraft and scheduled departure). These two components are kept constant throughout all the simulations for any scenario. To this baseline of schedules and itineraries, stochastic delays are added:

- for taxi-in, taxi-out, airborne delay, non-ATFM delay, distributions from an independent analysis of data are modelled.
- for ATFM delay, we use an historical sample of regulations from one year of data. Regulations are either modelled by explicitly assigning flights to slots (when the regulations are applied at arrival airports) or based on a probabilistic model (to capture weather and non-weather related airspace regulations). For the regulations at airports we draw at random the regulations which were issued at a given historical day at airports and model them in Domino. The days from which to draw the airport regulations are ranked by the number of regulations issued and a random day between the 20th and the 80th percentile is selected.

2.2 Stressed scenario

We built a 'stressed' scenario in order to see how the implemented mechanism will respond to this operational environment. To achieve this, we increase the level of delay in the system at different points: airborne, taxi-in and taxi-out, and also by selecting higher quantiles for the ATFM regulations explicitly modelled at airports.

This creates an environment where most flights experience delay, which explores the performance of the mechanisms as 'delay and cost management tools', as opposed to 'smoother of operation management tools' in the default scenarios. As an indication, the average delay in these scenarios is roughly three times the delay in the default one (~30 minutes versus ~10 minutes).

2.3 4DTA mechanism

4DTA mechanism deal with the tactical management of delay by airlines on their flights. Two sub-mechanisms are considered: modifying the cost index of a flight to recover part of the expected delay and/or actively delaying a flight to wait for connecting passengers.

2.3.1 Level 0

At Level 0, we use rules of thumb that serve as an approximation of the current practices in the airline industry for the tactical management of flight delay and waiting for passengers at the hub. Two sub-mechanisms are considered:

- determining the cost index of a flight (before take-off), i.e., an increment or not on the cruising speed.
- deciding whether a flight waits for delayed connecting passenger

The specific parameters of these mechanisms have been calibrated according to the feedback received from a number of experts in the industry.

Cost index is calculated before the take-off (i.e., at push back) and it is fixed throughout the flight. To decide on its speed, the flight uses the information about its departure delay. At "pushback ready", the departure delay is assessed by comparing estimated off-block time (EOBT) with scheduled off-block time (SOBT):

$$delay_{departure} = EOBT - SOBT$$

The attempted delay recovery is then performed according to the probability distribution shown on Figure 1. If the estimated departure delay is smaller than 15 minutes, the flight does not try to recover it. If it is larger than 60 minutes, the flight will try to recover as much delay as possible (up to 5 minutes) by selecting a higher speed than the planned one. Upon the consultation with the experts, we decided to never recover the delays of up to 5 minutes as it does not make economic sense. Lastly, the decision on recovering any delay between 15 and 60 minutes is made stochastically according to the linear probabilistic distribution on Figure 1.

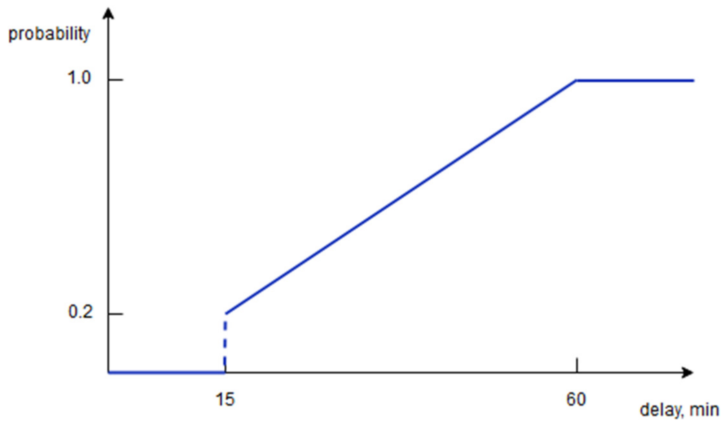


Figure 1. Probabilistic distribution for deciding on the delay recovery depending on the estimated departure delay.

Naturally, maximum delay that can be recovered is limited by maximum additional fuel available. At Level 0, in order to make the application of this rule more aligned with the current practices, the flight never speeds up to the maximum possible speed; rather, the speeding up is capped at 90% of the maximum velocity.

Note that a change on cruise speed will imply also a change on the TOD, generally increasing the cruise and reducing the slower descend.

Wait for passengers is performed 5 minutes before the "pushback ready" event, and it is triggered by a special event called "pax check event". At that moment, we run a check that inspects which passengers are not at the gate ready for boarding, and we estimate how much time they need to make it to the gate. For this estimate, we use the information on their current position: If they are/were arriving on a connecting flight, the in-block time of their previous flight is used (real or estimated, depending on the status of the flight). In addition, the average minimum connecting time is taken into account for the calculation of their estimated at-gate time. The average minimum connecting time has been pre-calculated for each airport and it depends on the type of connection the passenger is making: domestic - domestic, domestic - international, etc.

Finally, the flight decides to wait for any passenger with a flexible ticket whose at-gate time is estimated to be at most 15 minutes later with respect to the flight's expected push back time.

2.3.2 Level 1

The **cost index** at Level 1 is assessed at the **top of climb (TOC)**. Firstly, the arrival delay of a flight is estimated by comparing the current estimate of **EIBT** and **SIBT** (instead of the departing delay as in Level 0). According to that estimate, the flight performs a potential delay recovery by looking at two types of costs:

- **fuel cost**: the cost of the extra fuel that would be needed in order to recover a delay (fully or partially);
- **time cost**: the cost of unrecovered delay, which includes non-passenger costs (maintenance and crew costs), as well as passenger costs (compensation, soft costs and the costs due to the effects of reactionary delays).

The two cost functions are estimated on the range of the estimated arrival delay, and the delay recovery is performed with a time resolution of 1 minute (i.e., the flight can only decide to recover a rounded number of minutes). The recovery decision is made by observing the total cost, i.e., the sum of the fuel and time cost. The flight chooses a recovery time (in minutes) that expects to minimise the total cost (on the domain ranging from 0 minutes to estimated arrival delay).

Unlike at Level 0, there is no limitation on the maximum velocity that the flight can choose in order to recover delay, as the decision is purely driven by the cost and the objective to find the optimal solution given the estimated costs.

At Level 1, **wait for passengers** is performed at the same time stamp as in Level 0: 5 minutes before "pushback ready". However, unlike at Level 0, at Level 1, the flights make more informed decisions on whether to wait for a certain passenger by looking at the costs of each decision.

For each group of passengers (passenger with the same estimated at-gate time), two types of costs are estimated:

- **waiting cost**: the cost of waiting a passenger group p_i for n minutes with respect to EOBT. This wait would essentially delay the EOBT for n minutes, and thus this is the cost of delaying the flight (EOBT) for n minutes.
- **not-waiting cost**: the cost of not waiting a group of passengers and having to take care of them. This cost includes different types of care that the airline needs to provide to the passengers that missed their connecting flight for no fault of their own: duty of care, compensation cost and transfer cost. Transfer cost is calculated by searching for alternative itineraries for stranded passengers, estimating the cost of each itinerary and choosing the least expensive one. Additionally, this also includes soft costs - the cost airline will suffer due to a potential future loss of passengers or reputation.

The final decision is made by looking at all waiting and not-waiting costs (across all the delayed passenger groups), and the wait time that minimises the total additional cost.

2.3.3 Level 2

Level 2 couples the assessment of cost index and wait-for-passenger decision process via a unified cost function. In that way, the optimisation is improved by relying on the sum of all the costs outlined in the previous section (Level 1), including the possibility to recover a part of delay by speeding up and spending more fuel than planned. There are two points during a flight when delay is assessed and 4DTA mechanisms potentially applied: at "pax check event" (5 minutes before "pushback_ready") and the top of climb.

At **5 minutes before a flight is ready for push back**, a joined assessment of departure delay (and its potential recovery) and wait for passenger options is performed. Similarly as in Level 1, we assess current estimated departure delay and consider recovery options by speeding up (changing cost index before departure) through assessing the cost of those options. At the same time, the check for missing passengers is performed, and waiting costs and not-waiting costs are calculated for each passenger group (see Level 1). All those estimated costs are added and observed on the domain of recoverable delay (from 0 to the maximum number of minutes a flight can recover by speeding up and using the extra fuel available). The decision that minimises the total cost is taken, and according to it, cost index might be changed (speeding up) and a number of passenger groups waited for. I.e., a decision to wait and recover delay is performed before push back.

Example. Let's assume that the currently estimated departure delay of the flight is 20 minutes, and there are 2 passenger groups estimated to be late for boarding with delays of 10 and 15 minutes (w.r.t. updated push back time, i.e., waiting 10 minutes will allow the first group of passengers arrive to the flight, waiting 15 minutes will ensure that both group of passengers can board the plane). In this case, all the costs are assessed on the domain of possible delays ranging from 0 (recovering all of the delay, which would require significant increase in velocity and thus spending a large amount of additional fuel, if all 20 minutes can be recovered by speeding up) to 35 minutes (meaning the flight is waiting for both passenger groups and deciding not to recover any delay).

At the **top of climb**, the assessment of expected arrival delay potential speeding up is done as in Level 1. The novelty when compared to the previous level is the ability for a flight to slow down at the top of climb if the (currently) expected arrival time (**EIBT**) is at least 15 minutes before the scheduled arrival time (**SIBT**). In this case, the flight decides to slow down by adding x minutes to the flight, where $x = SIBT - EIBT - 15$.

2.4 FP mechanism

The Flight Prioritisation mechanism deals with the potential swap of ATFM slots at regulations defined at arrival airports by airlines.

2.4.1 Level 0

At Level 0, no swap is performed by the airlines. The delay that is assigned due to ATFM regulations at arrival is performed by the flight who receives it.

2.4.2 Level 1

At Level 1, swaps between flights can be performed if they belong to the same airline. Every time a departure flight plan is submitted (whether it is the first flight plan for this flight or not), the airline estimates if a swap can be done with this flight if ATFM delay has been assigned. The conditions considered are the following:

- both flights must be in the same regulation at their arrival airport;
- the estimated cost of the swap must be negative (i.e., the swap has a positive impact overall).

The cost of the swap is estimated using the following rule: if $COBT_1$ is the controlled off-block time of the first flight, $COBT_2$ the time of the second flight, and cost1 and cost2 the delay cost function of the first and second flight, then we compute:

$$cost_1(COBT_2) + cost_2(COBT_1) - (cost_1(COBT_1) + cost_2(COBT_2))$$

i.e., the total cost of delay if the COBT are swapped minus the cost of delay if they are not. The cost function used for this mechanism is different from the one used for 4DTA at TOC, since the estimation of the cost happens on the ground, before the departure of the flight. More specifically, the cost function is evaluated using as delay the controlled off-block time minus the scheduled one, and includes the following components:

- non-pax cost (maintenance and crew);
- passenger soft cost;
- duty of care;
- passenger compensation.

Moreover, the delays for passengers are computed using the updated information on their next flights and worst case scenarios. In particular, when a passenger is expected to miss their next flight. A full 12 hours delay with overnight compensation is considered.

Some network effects are also taken into consideration using, again, the worst case scenario. We gather information on the next flights using the same aircraft and consider that all of these flights will have the same cost than the current one.

Note that in theory we could estimate exactly which flight will be impacted by the propagation of delay, in terms of aircraft and/or connecting passengers, since all information is known to the airlines (or can be requested from another airline). However, this information takes too long to compute, especially for passengers, which increases by a large factor the time of the simulation. As a consequence, in this deliverable the above heuristics is used for the decision-making process of swapping the slots.

2.4.3 Level 2

At Level 2, flights can be swapped among difference airlines. The mechanism works otherwise exactly the same. The airline starts by checking all flights in the same regulation at the arrival airports, and then requests some information from another airline, if needed, in order to compute the total cost of the swap. The same cost function is applied, and the swap is performed is the cost is negative.

It is clear that in reality, different airlines will never share their true cost, first because it is sensitive information, and more importantly because they have no incentive to not inflate their own reported cost. In the model, we consider that there is a market mechanism (e.g., credit system, auction) in place which allows us to have an efficient market for swaps and thus do a swap only if it is beneficial in average (for instance by giving back some money to the airline delaying its flight). This market mechanism might or might not be feasible in reality, which is another interesting research question. Thus, the model should be considered as a best case in this regard.

2.5 FAC mechanism

The Flight Arrival Coordination mechanism focuses on the sequencing done by the extended arrival manager at airports. This mechanism is only implemented in airports which have or are expected to operate an E-AMAN system (24 airports as according to the SESAR Pilot Common Project [7]): EBBR, EDDB, EDDF, EDDL, EDDM, EGCC, EGKK, EGLL, EGSS, EHAM, EIDW, EKCH, ENGM, ESSA, LEBL, LEMD, LEPA, LFMN, LFPG, LFPO, LIMC, LIRF, LOWW and LSZH.

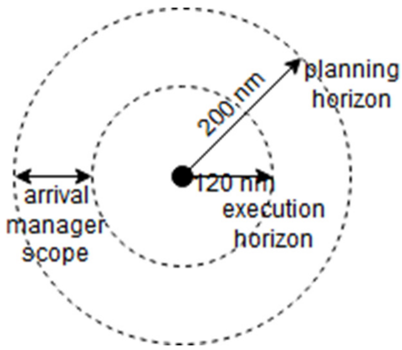


Figure 2. Extended Arrival Manager scope.

For the above mentioned airports two horizons are defined around them: planning horizon (200 nm from the airport) and a tactical or execution horizon (120 nm from the airport) (see Figure 2). These distances are in accordance with the expected extension of the arrival managers from 100-120 nm to 180-200 nm [7]. When flights enter the planning horizon, all the flights which are located in the scope of the arrival manager, i.e., between the 120 nm and 200 nm radii around the airport, are re-optimised, i.e., assigned to the slots which are either planned or available, considering a given optimisation function which depends on the Level of the mechanism. The flight which triggers this optimisation, i.e., the one which enters the arrival manager, receives the amount of delay that it is expected to experience. This ensures that the best sequence is maintained within the arrival manager with respect to the optimisation function, and that the flight can slow down to absorb part of the delay saving some fuel if delay is expected. However, as the amount of delay that can be absorbed is very limited, only the flight which enters the arrival manager considers this speed and TOD variation. When flights arrive at the execution horizon, the sequence is re-optimised and the final arrival slot is assigned to the flight. If delay is needed, this will be done as holding. The arrival capacity at the airport is considered at both horizons.

For the airports which are not listed above, a simple arrival manager located at 100 nm from the airport is considered, and a first-in first-out approach modelled. The assigned delay will hence be done as holding. This ensures that the arrival capacity at the airport is not exceeded.

2.5.1 Level 0

In Level 0, current principles applied on E-AMAN systems are considered: The Flight Arrival Coordination tries to minimise the amount of holding delay that will be carried out at the TMA by minimising the total holding delay. The FAC is focused on the maximisation of the arrival throughput at the runway. No information from the airlines is taken into account when applying this mechanism. When a flight enters the planning horizon, the first slot available in the sequence from the flight estimated landing time is assigned. In a similar manner, once the flight enters the execution horizon, the first available slot is assigned and the holding delay computed.

2.5.2 Level 1

In Level 1, the arrival manager tries to minimise the delay that the airport will generate. This includes the arrival delay but also the potential reactionary delay. In order to achieve this, when the flight enters the planning horizon, the FAC requests from the flight the expected total delay that the flight will experience for each available slot (i.e., arrival delay + expected reactionary delay). The flight provides this information considering the EIBT for each landing slot (i.e., arrival delay) and adding the expected reactionary delay. The expected reactionary delay is computed by the flight by requesting to the AOC the time from which delay will be propagated. To do this, the AOC considers the SOBT of next flight of the same aircraft and the expected turnaround time (i.e., $time\ from\ which\ delay\ will\ be\ propagated = SOBT_{next\ flight} - expected\ turnaround\ time$). The expected reactionary delay is then estimated by the flight for each possible slot as $\max(0, slot_{time}(landing_{time}) + expected\ taxi_{in} time - time\ delay\ propagated)$.

The objective function of FAC Level 1 is then the minimisation of the expected total delay (i.e., arrival delay + expected reactionary delay) for all the flights in the scope of the FAC. Figure 3 presents the messages interchanged between the FAC, Flight and AOC agents when a flight enters the planning horizon.

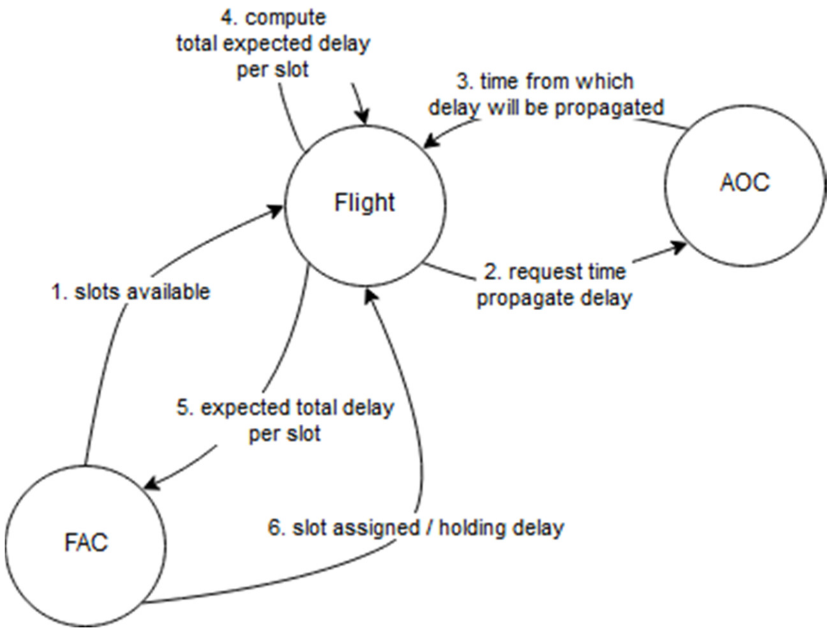


Figure 3. FAC Level 1 messages at planning horizon

When the flight enters the execution horizon, the optimisation is recomputed but considering the same expected delays for the different slots. I.e., the FAC does not have to request any further information from the flight.

2.5.3 Level 2

In Level 2, the same principle applies in the FAC, but in this case, instead of the expected total delay for each slot, the expected cost is considered. In this case the flight will:

- Request the expected cost of delay from the AOC for each slot.
- Compute the potential fuel savings that could be done by absorbing part of the delay reducing the speed.
- Compute the cost due to the holding fuel needed for each slot.

Consider these three points to compute an expected total cost of each slot.

The FAC will then minimise the total expected cost for all the flights in its scope. Figure 4 presents the different messages exchanged.

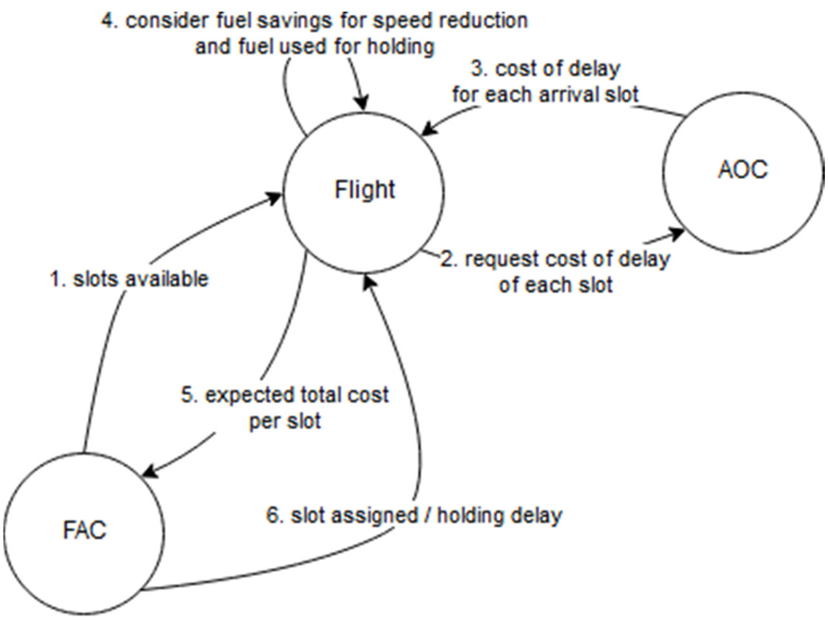


Figure 4. FAC Level 2 messages at planning horizon

As in the Level 1 implementation, when the flight enters the execution horizon, the same optimisation is performed with the already provided information.

3 Model calibration

This section presents the process by which the model has been calibrated and the quality of the calibration. The current calibration has been performed considering the average values coming out of the simulations with the corresponding values available on the empirical dataset. Part of the validity of the model lies in the fact that the distributions obtained as output of the model match also the empirical data. A more complete calibration will be performed for the final version of the model.

3.1 Data used for calibration

There are many processes in the model that are represented explicitly such as turnaround times and hence the propagation of reactionary delay. These processes lead to the output delay from the simulation. Some processes, are based on statistical distributions. For example, the average wind encountered by flights during the en-route phase, or the amount of ATFM delay assigned to a flight due to an en-route regulation. Table 2 presents some of the key processes that are modelled in Domino, and how their distributions have been adjusted.

Table 2. Processes model in Domino with distributions.

Process	Distribution	Based on
Taxi-in	LogNormal distribution considering mean, standard deviation and modifier to consider baseline or stressed scenarios.	IATA Summer Season 2010 from CODA [8]
Taxi-out	LogNormal distribution considering mean, standard deviation and modifier to consider baseline or stressed scenarios.	IATA Summer Season 2010 from CODA [9]
Climb uncertainty	Normal distribution minutes	Analysis DDR difference between planned and executed trajectories (m2, m3) from DCI4HD2D Project [10]
Cruise	Normal distribution nm	Analysis DDR difference between planned and executed trajectories (m2, m3) from DCI4HD2D Project [10]

Wind	Empirical probability distribution function for planned wind during the cruise. Used average wind between regions. No noise added on execution.	For each ANSP to ANSP origin and destination airport consider the difference between requested speed and observed average ground speed for cruise segments from DDR2 analysis (AIRAC1409) [11]
Turnaround time	Exponential distribution considering: • Minimum turnaround time based on airport size, aircraft wake and type of airline (REG, CHT, LCC, FSC) Lambda which depends on scenario (Default or High delay)	Analysis of turnaround times performed in POEM project and used in ComplexityCosts project [12]
Probability ATFM delay	When regulation is explicit at airport, the regulations are based on a given historical day. The days are selected based on their percentile ranked by number of regulations at airport in the day. There is a minimum and maximum percentile to be considered for baseline and stressed scenarios. For regulations in the airspace there are two probabilities one for regulations due to weather and another for regulations due to any other reason.	Based on analysis of DDR2 (AIRAC1313-1413 excluding days with industrial actions) [11]
ATFM delay	Empirical probability distribution function for regulations due to weather and regulations for other reasons.	Based on analysis of DDR2 (AIRAC1313-1413 excluding days with industrial actions) [11]
Non-ATFM delay	Exponential distribution with different lambda as a function of scenario: baseline, stressed	-

Passenger connecting times	LogNormal distribution Considering minimum connecting times per airport and type of connection (between national flights, from national to international and between international flights), sigma and percentile of passengers who connect in less than the minimum connecting time.	Based on analysis of minimum connecting times at ECAC airports originally performed in POEM project [13]
Variation of cruise length due to DCI	Normal distribution nm	Analysis of Performance using Airbus PEP [14]

Then, there are some parameters that can be further adjusted/calibrated, in order to adjust the performance of the model to existing historical datasets. Table 3Table 4 presents the data sources used to identify the target values for key indicators. Note that we are basing the scenarios on the traffic of a given day. Therefore, we have been able to reconstruct the schedules of that day with the execution of the flights from DDR2, and use this information for the validation of the model.

Table 3. Calibration parameters considered for targets.

Parameter	Source
Departure delay	Reconstructed schedules compared with AOBT from DDR2 (m3)
Arrival delay	Reconstructed schedules compared with estimated AIBT from DDR2 (m3)
Delay distribution per reason (Reactionary, en-route, capacity, weather)	CODA 2017 report [1]
Flight plan length (nm)	DDR2 (m1)
Flight plan duration (min)	DDR2 (m1)
Flight execution length (nm)	DDR2 (m3)
Flight execution duration (min)	DDR2 (m3)
Taxi-in	Reconstructed schedules compared with planned taxi times from DDR2 (m1)
Taxi-out	Take off time - AOBT estimated form DDR2 (m3)
Gate-to-gate time (min)	DDR2 with estimated taxi times

Cancellation rate	CODA 2017 report [1]

Finally, Table 4 present the parameters that have been adjusted. Some of these parameters have been chosen based on expert judgment, whereas others have been calibrated using some of the above metrics.

Table 4. Parameters from the distributions that have been adjusted in the model as part of the calibration process.

Process	Parameter	Possible values
Turnaround time	Lambda of exponential distribution	• Default delay scenario value • High delay scenario value
Climb uncertainty	Extra climb minutes	• Value adjusted for calibration in baseline scenario
Airport capacity modifier	Reduction of airport capacity	• Default scenario value: no reduction • High delay scenario value: half capacity reduction
Airport capacity adjustment	Adjustment of capacity at airport considering demand lower due to non-inclusion of non-passenger commercial flights	Value adjusted based on baseline scenario analysis of DDR2 flights
Non-ATFM delay	Parameter in exponential distribution, typical delay	• Default delay scenario value (calibrated on DDR2 data) • High delay scenario value
Probability of ATFM delay explicit at airport	Day selection minimum and maximum percentile considered	• Default delay scenario values (0.3-0.8) • High delay scenario values (0.8-1)
Taxi time modifier	Increment of taxi time	• Default scenario value: no increment • High delay scenario value: half increment
Passenger connecting time	Percentage of passengers who made the connection in less time than the MCT Sigma for the LogNormal distribution	Value adjusted for calibration in baseline scenario

Cancellation rate	Ratio of cancelled flights per day	Value adjusted for calibration in baseline scenario

3.2 Results of calibration

3.2.1 Delay structure

We first investigate the structure of the various delays and travel times produced by the model. Given the small number of free parameters in the model, the calibration is performed by mainly tuning iteratively the average of some of the delays injected into the model to match the average of other metrics that can be computed on the empirical data for comparison (see Table 3). Most of these distributions have been obtained using some analyses empirical data, as explained in the previous section. In practice, we used the cancellation rate and the non-ATFM delay distribution to calibrate the model.

The cancellation rate has initially been set to 2%, which is the consistent with some high cancellation rate months observed in 2014 and 2017 [15] [1]. This value is relatively high and will be reviewed in the next model version. However, after the calibration of this parameter, the explicit cancellation of flights due to curfew was included in the model. This leads to an unexpected high cancellation rate of around 3.9%. The inclusion of explicit cancellations means that the statistical probability of cancelling a flight needs to be reviewed in next versions of the model.

Once the cancellation rate is fixed, the distribution of the delay due to non-ATFM causes has been adjusted.Table 5 presents the summary of the results of the calibration. This table shows the departure and arrival delays ("arrival positive delay" is computed considering negative delays as null), gate-to-gate times, flying distances, taxi times, gate-to-gate time difference between scheduled and actual trajectories, and holding times. The results from the simulations are compared with the metrics computed form the historical data (with schedules reconstructed) for the 12th of September 2014 (as described in the previous section (see Table 3)). The variability of the averages observed in the simulations is also reported in the table. Note that all simulations results are computed on the baseline scenario with 100 iterations.

Table 5. Model calibration summary.

Indicator	Empirical data (based on historical data of the 12SEP14)	Simulation (100 iterations)	Difference	Error in %	Variability in simulation (std)
Departure delay	11.4	10.9	-0.5	-4.3	25.7
Arrival delay	6.6	5.6	-1.0	-15.6	30.1
Arrival positive delay	11.6	10.9	-0.7	-5.6	10.9

Scheduled G2G time	159.5	159.5	0	0	143.8
Actual G2G time	154.7	154.6	-0.1	0	137.8
Scheduled flying time	137.3	135.2	-1.9	-1.5	135.5
Actual flying time	137.1	135.9	-1.2	-0.9	135.6
Scheduled flying distance	960.6	965.6	5.0	0.5	1091.3
Actual flying distance	948.5	958.6	10.1	1.1	1094.4
Actual taxi-out time	12.5	12	-0.5	-3.8	7.2
Actual taxi-in time	5.7	5.5	-0.2	-3.2	5.5
G2G time diff.	-4.8	-5.3	-0.5	-11.4	15.6
Holding time	NA	0.4	-	NA	1.1
Cancellation rate	2%	3.9%	1.9%	95	NA

Generally speaking, the calibration results show that flights in the model tend to have faster trips. First, looking at the scheduled times and distance, we can see that if the scheduled distances are slightly larger in the simulations compared to reality, the scheduled times are slightly smaller. This indicates respectively that the airlines choose trajectories which are too long, but then that they operate the trajectories faster than in the historical data. Some fine tuning to the trajectory choice and trajectory generator is thus likely necessary for the final version of the model. This could also be linked with necessary adjustments on the wind aloft model and on a potential lack of holding at the arrival airport.

Focusing on delays, it seems that flights experience less delays in the simulations than in the historical data. This is due in part to the taxi-out time, which drives the departure delay down (by half a minute), and to the operation of the trajectory itself, probably due to the performance model. The total average arrival delay is short by 1 minute, roughly 16% of its target value. Note that, interestingly, the issue seems to be on the negative delays, since the "arrival positive delay" is in average short of only 5%.

Table 6. Distribution of delays among the main reasons of delay.

Type of delay	Empirical data (based on [1])	Simulation	Difference
Reactionary	5.1	3.1	-2
Turnaround	4.1	4.8	0.7
En-route	0.9	1.9	1
Capacity	0.8	0.4	-0.4
Weather	0.2	0.3	0.1

Table 6 presents the distribution of main delay reasons between the following categories: reactionary delay, turnaround, en-route, capacity and weather. The model assigns the total delay of the flight to the category which produces the greatest contribution to the delay at departure. These values are compared with the historical data reported in [1]. These empirical values are estimated by multiplying the positive arrival delay reported in Table 5 by the rations found in [1]. Note that we cannot compare too precisely the simulation with the empirical data: first, in the model we flag only the most penalising reason, whereas in reality CODA also computes fractional delays for the different reasons; and more importantly, we used proportions of reasons that have been computed an entire year here, whereas the proportions for the 12th of September 2014 can be different.

It is, however, interesting to see that the main reason for the low delays may be due to reactionary delays, where a difference of 2 minutes (over 5) has been observed. This could indicate that the buffer we considered for the flights is too large, that the turnaround times are too short in the model, or just that the arrival delay is lower and hence less delay is propagated.

Table 7. Comparison of delay in the simulation and empirical data: departure delay, en-route delay, and taxi delay.

Type of delay	Empirical data (based on historical data of 12SEP14)	Simulation	Difference
Departure delay	11.4	10.9	-0.5
En-route delay	-0.2	0.8	1.0
Taxi delay	-4.6	-6.1	-1.5
Total (arrival delay)	6.6	5.6	-1.0

Finally, in order to understand which phase is the most important to understand the differences in the arrival delay, we show in Table 7 the comparison between simulations and empirical data for the departure delay, the en-route delay, and the taxi delays (adding taxi-in and taxi-out). From this table, we see that the main contributor to the difference in the arrival delay is the taxi phase, where strong negative delays tend to decrease the total. On the other end, the en-route delay is a bit too high, which might be linked to the longer paths selected by the airlines, as explained previously. Overall, it seems that the distributions of actual taxi times need to be re-adjusted to have a better calibration.

3.2.2 Distributions

In order to validate the model, it is important not only to compare average values, but also the distributions themselves. Here we focus on a few metrics: the departure delay, the arrival delay, the gate-to-gate time, the difference between scheduled and actual flying distances, and the difference between scheduled and actual flying times.

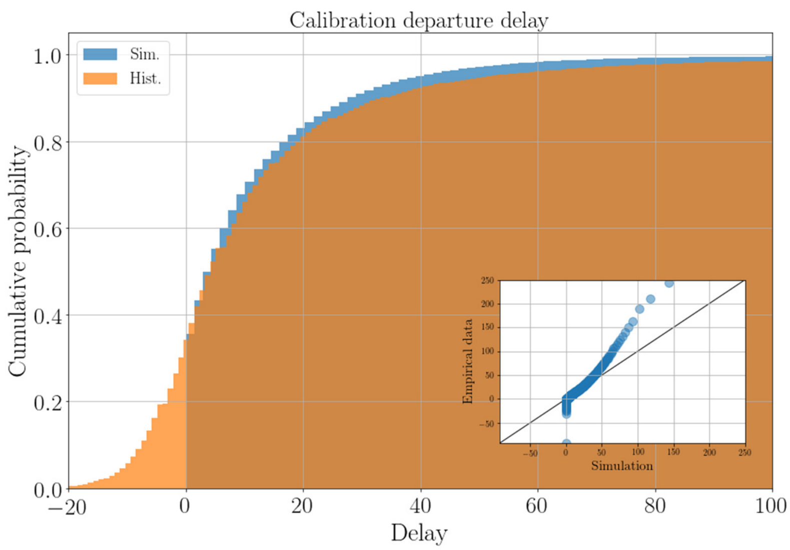


Figure 5. Cumulative distribution of departure delay in the simulations (blue) and the empirical data (orange). In inset we show the QQ-plot obtained with these two distributions.

Figure 5 allows us to compare the distributions of departure delays. Even if the average delay is only off by half a minute (see previous section), the distribution shows that there are other factors that should be considered. In particular, historical data presents negative departure delay (i.e., flights do their AOBT (push back) before their SOBT). The model does not allow for this to happen as flights can only leave their gate at the SOBT the earliest. This means also that higher delays are also experienced in the historical data than in the simulated one. This should be considered in the final version of the model allowing the departure before SOBT in some cases.

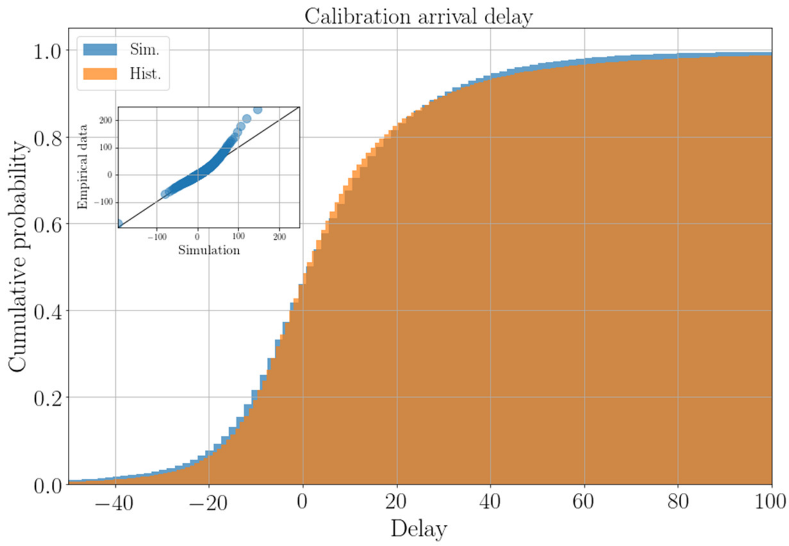


Figure 6. Cumulative distribution of arrival delay in the simulations (blue) and the empirical data (orange). In inset we show the QQ-plot obtained with these two distributions.

Figure 6 allows us to compare the cumulative distributions for arrival delay. Here the agreement is better than for departure delay. It is interesting to observe that in the historical data the right tail is fatter than in the simulated results (this is especially visible in the QQ-plot). It is not clear why this is the case, but it might be related to the taxi-in delay, which is probably underestimated in the model. This indicates that some flights experience higher delay than in the model.

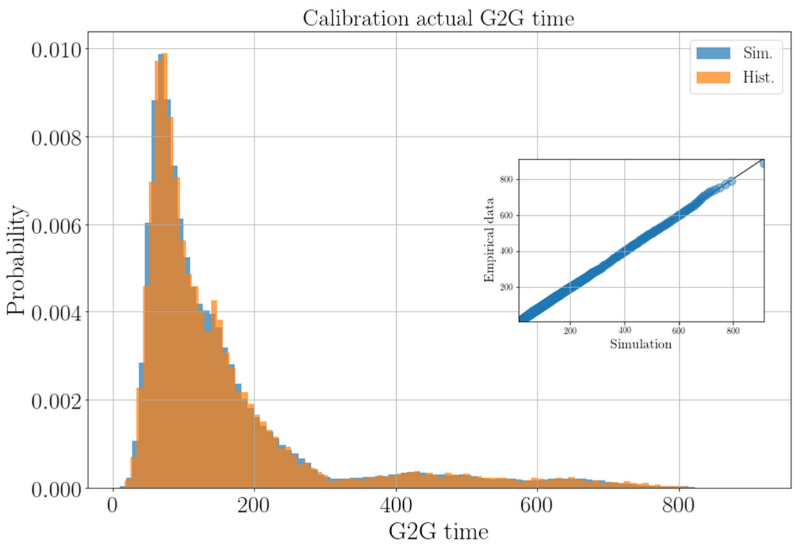


Figure 7. Distribution of gate-to-gate time in the simulations (blue) and the empirical data (orange). In inset we show the QQ-plot obtained with these two distributions.

The actual gate-to-gate times seem to match pretty well in the simulations and the empirical data, as shown in Figure 7. Since the main driver for this time is the flying distance and time, this is expected.

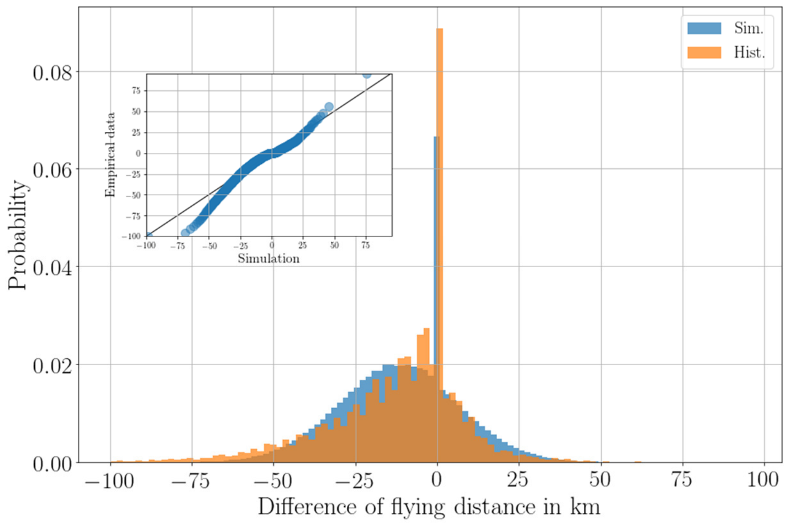


Figure 8. Distribution of difference in flying distances (scheduled minus actual) in the simulations (blue) and the empirical data (orange). In inset we show the QQ-plot obtained with these two distributions.

More importantly than the absolute values of the flying distances and times, is the comparison between the scheduled flight plan distance and the one experienced during the execution of the flight. This is depicted in Figure 8. A first comment on this figure is that the empirical differences in distance seems more 'clustered' than the simulation ones. This could be due to the fact that flights are changing their flight plans in a more systematic way than in the model, for instance shortening their flight plan using always the same new trajectory for a given origin-destination pair. The selection of trajectory may be a lot more random in the simulations, which leads to a smoother distribution.

The second comment is that the empirical distribution has much fatter tails than the simulations', especially on the left side. Since the trajectories in the simulations are limited in number, due to the fact that we used clusters of trajectories, it could happen that these 'typical' trajectories do not allow very short or very long trajectories. In reality, a few flights are able to shorten or lengthen significantly their flight plan.

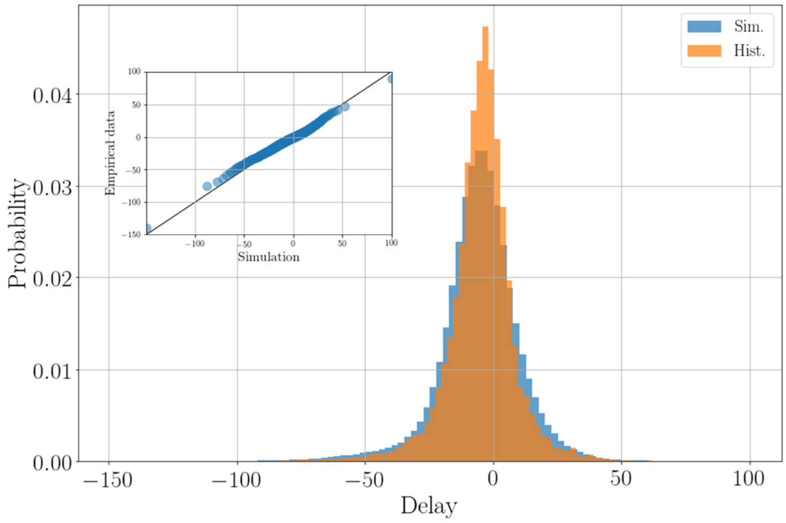


Figure 9. Distribution of travelling time difference (between scheduled and actual trajectories) in the simulations (blue) and the empirical data (orange). In inset we show the QQ-plot obtained with these two distributions.

Finally, we show in Figure 9 the difference between the scheduled flight plan time (from take-off to landing) and the actual one. This difference is the consequence of the difference in distance depicted in Figure 8 and of the additional noise delay happening during the flight. Interestingly, if the difference in distance has fatter tails in the empirical data, the difference in time is actually slightly thinner compared to the simulations. This could indicate some small active adjustments from the pilot to stick to schedule which are not properly captured by the model. This could be adjusted by modifying slightly the distributions used to create some delay during the flight. Note that despite the relatively high negative and positive delays, the simulated distribution is only off by half a minute in average compared to the empirical one.

4 Investigative case studies results

4.1 Metrics

This section is dedicated the explanation of the metrics that are used to analyse the model's outputs and how they are computed. First, we consider some baseline metrics related to delays, passengers and costs. Secondly, we consider the advanced network metrics of centrality and causality that were presented in Deliverable 5.1 [3]. In the case of centrality, an additional centrality metric, which was not presented in the earlier deliverable and is more focused on the passengers' itineraries, has been added.

4.1.1 Delay metrics

For delay, we report the following baseline metrics, averaged over 100 iterations of the agent-based model per scenario:

- Mean departure delay of flights
- Number of flights with departure delay > X minutes
- Total departure delay of flights with departure delay > X minutes
- Mean departure delay of flights with departure delay > X minutes

- Mean arrival delay of flights (early arrivals are counted as negative delays)
- Mean arrival delay of delayed flights (early arrivals are counted as 0)
- Number of flights with arrival delay > X minutes
- Total arrival delay of flights with arrival delay > X minutes
- Mean arrival delay of flights with arrival delay > X minutes

where X=15, 60, 180

- Mean gate-to-gate delay of flights (obtained as the difference between the scheduled and the realised gate-to-gate time)
- Mean per-passenger gate-to-gate delay (for each flight, the delay is divided by the number of passengers on the flight. The result is averaged over all flights)

- Number of cancelled flights
- Mean reactionary delay (computed as the mean delay of flights whose main reason for delay is reactionary)
- Number of flights with reactionary delay

For each of these metrics we show the average, the first and third quartile of their distribution over the iterations (see Section 4.3). The statistical significance of the differences in the averages for different scenarios are tested via a two-sample T-test with 5% confidence level, testing the null hypothesis that the difference between the two samples of size 100 has mean zero.

Note that we consider the tail statistics (on delays larger than a threshold) for departure and arrival delay because often changes are observed in the tail of the distribution, while the bulk is unchanged. For the same reason, the delay distributions in the baseline and advanced scenarios are compared by means of QQ-plots, highlighting possible deviations in the tail. In a QQ-plot, the samples from two distributions are compared by plotting the n-th quantile of the first against that of the second. If the samples come from the same distribution, the points will fall on a 1:1 line. Deviations from the 1:1 line highlight where the two distributions differ from each other. For example, if the QQ plot is shifted above the 1:1 line, it means that the distribution whose quantiles are represented on the y axis is shifted to the right with respect to the other one. If the right end of the QQ plot deviates from the 1:1 line by going below it, it means that the y-axis distribution has a thinner right tail. Conversely, if it goes above the 1:1 line it means that the y-axis distribution has a fatter tail. To compute QQ-plots, all iterations are considered at the same time (e.g., a sample is made of the departure delays of all flights in all 100 iterations of one scenario).

Flights with more than 300 minutes of delay, resulting from a fat-tailed distribution of the randomly-generated ATFM delays (to be corrected in the future implementations of the model), are not considered in the analysis. They amount, on average, to less than 1‰.

4.1.2 Passengers metrics

We consider the following metrics related to passengers, averaged over 50 iterations of the model (due to technical limitations):

- Average passenger delay. The passenger delay is computed as the arrival delay at the passenger's final destination and is averaged over all passengers.
- Average positive passenger delay. The average is computed counting early arrivals as zeros. This metric is interesting because of its related metrics for the flight delays, used for instance by CODA. Note that if for a flight the early arrivals are not beneficial, and thus can be set to 0, they are probably important for passengers, whose utility always decreases with the length of the trip (this is captured in the previous average passenger delay metric).
- Number of passengers with delay > X minutes.
- Total delay of passengers with delay > X minutes.
- Mean delay of passengers with delay > X minutes.

where X=15, 60, 180

For each of these metrics we show the average over 50 iterations, and the first and third quartile of its distribution over the iterations (see Section 4.3).

4.1.3 Cost metrics

For the costs, we report the following baseline metrics, averaged over 100 iterations of the model:

- Average excess cost of fuel. It is the extra cost with respect to the planned cost of fuel, which can also be negative if fuel was saved.
- Average cost of compensation.
- Fraction of flights paying compensation.
- Average cost of transfer.
- Fraction of flights paying transfer.
- Average duty of care cost.
- Fraction of flights paying duty of care.
- Average soft costs.
- Fraction of flights paying soft costs.
- Average non-pax costs (crew+maintenance).
- Fraction of flights paying non-pax costs.
- Average total excess cost.

All costs are in euros. Average costs are computed on all flights, including those that did not experience that kind of cost (counted as zero). Each metric is computed for each iteration and then averaged over 100 iterations, and the first and third quartile of its distribution over the iterations are reported under the result in tables. As for the delay analysis, flights with more than 300 minutes of delay are excluded. QQ-plots are also shown, to highlight changes in the tails of the delay distribution.

In the model outputs the number of cancellations is too high with respect to what was observed in the data for the simulated day, as seen in the validation section. The cancelled flights make up a large percentage (~60%) of the non-fuel costs, and given that the mechanisms do not act explicitly on cancellations, they could cover the effects of the mechanisms. Therefore, we also present the cost metrics on a sample where the cancelled flights have been excluded.

4.1.4 Centrality metrics

We consider two types of centrality metrics. The first is trip centrality, which was introduced in Deliverable 5.1 [3]. The outgoing trip centrality of an airport counts all the "potential" itineraries having that airport as the origin, while the incoming trip centrality counts those having that airport as a destination. Potential itineraries are all the sequences of any number of flights that can be potentially taken one after the other, given their schedule. An itinerary of n legs is weighted α^n, where $\alpha < 1$, so that itineraries made of more legs are counted less. For all the results shown in this deliverable $\alpha = 0.02$, except for the results of the FAC mechanism, for which a comparison of results with $\alpha = 0.02$, and $\alpha = 0.2$, is shown. Note that, due to how the metric is computed (see Deliverable 5.1 for details [3]), no upper or lower limit for the connecting time is considered, so that two flights can be taken in sequence as long as the second one departs later than the arrival of the first. Trip centrality can either count only the itineraries made of legs of the same airline or alliance, corresponding to setting $\varepsilon = 0$ (see Deliverable 5.1 [3]), or count also the itineraries using two or more airlines or alliances. For all the results presented in this deliverable we used $\varepsilon = 0$, therefore the walks counted are those within an alliance or within an airline (for airlines that do not belong to any alliance).

The second centrality metric that we consider is passenger centrality, a new metric which was not introduced in the previous Deliverable 5.1 [3]. In the computation of passenger centrality each itinerary contributes to the outgoing or incoming centrality of an airport an amount which corresponds to the number of passengers on that itinerary. Therefore, the outgoing passenger centrality of an airport corresponds to the number of passengers that depart from that airport (either as their first departure or taking a flight connection there) and are directed to another destination, either with a direct flight or with connections. The incoming centrality of an airport, instead, corresponds to the number of passengers that land in that airport, either as their final destination or to take a connection.

For both types of centralities, we are not interested in their absolute value but in the *loss of centrality* between the scheduled and the actual network, quantifying the damage to the network connectivity due to delays and cancellations. For trip centrality, as explained in Deliverable 5.1 [3], the centrality in the actual network is computed by using the actual network structure, which accounts for the delays and cancellations, and by excluding from the counting the new itineraries that become possible due to delays. An airport's centrality in the actual network is therefore always smaller than its centrality in the scheduled network, therefore losses are always positive. The loss of outgoing trip centrality of an airport measures the loss of potential outgoing itineraries that are not feasible anymore (due to delays or cancellations), therefore quantifying the decrease in the potential to go from that airport to the rest of the airport network. For example, the cancellation of a flight from airport i to airport j will cause a loss of $\alpha = 0.02$ in the outgoing trip centrality of airport i. If in airport j it was originally possible to connect to another flight, the same cancellation will also cause an additional loss of α^2, and so on for longer itineraries that are lost due to that cancellation. Similarly, if the flight from i to j is delayed, and therefore the potential connection with a subsequent flight departing from j is not feasible anymore, this will cause a loss of α^2, plus other losses from the longer itineraries depending on that missed connection. The loss of incoming trip centrality, similarly, measures the loss of potential incoming itineraries that are not feasible anymore (due to delays or cancellations), therefore quantifying the diminishing of the potential to arrive to that airport from the rest of the airport network.

For passenger centrality, in the actual network we only count passengers that reach their destination using their scheduled itinerary. The actual outgoing passenger centrality of an airport corresponds to the number of passengers that were counted in the scheduled outgoing passenger centrality and that manage to follow their scheduled itinerary. If, for example, N incoming passengers miss their connection in airport i, and are rebooked to another outgoing flight, airport i will have a loss of outgoing centrality amounting to N. The same loss would apply if N passengers depart late from i and miss their next connection at another airport. Therefore, the loss of outgoing passenger centrality of an airport accounts both for the passenger that experience a disruption in that airport and for those that experience problems downstream (missed connections or cancellations). This is different from the loss of outgoing trip centrality, which does not account for missed potential connections in the airport itself. The actual incoming passenger centrality of an airport, instead, corresponds to the number of passengers that were counted in the scheduled incoming passenger centrality and that manage to follow their scheduled itinerary up to that airport. If they miss their connection there, there is no loss of incoming centrality. Therefore, the loss of incoming passenger centrality can be interpreted as the damage to airport i (in terms of passengers that cannot reach it using their scheduled itinerary) caused by issues upstream.

The information on which passengers miss their connections is included in the output of the model. However, in the current version of the model this information was stored in a way which made it extremely slow to retrieve. Therefore, in the present deliverable we assume that a passenger misses a connection if the incoming flight lands less than 20 minutes before the departure of the connecting flight. This is a lower bound for connecting times, which can be larger depending on the airport and due to stochasticity present in the model, therefore we obtain an optimistic estimate of the passenger centrality loss. This issue will be solved in the next version of the model, therefore in Deliverable 5.3 we will be able to use the real missed connections in the computation on passenger centrality loss.

In order to interpret the results, the differences between Trip and passenger centrality should be kept in mind. First, the connections that are considered in passenger centrality are only those that passengers actually used, and therefore that imply a cost if disrupted. For this reason, they are explicitly considered by the mechanisms when costs are minimised. Instead, trip centrality considers also potential connections that are not used by passengers (at least in the considered day). These connections are not considered explicitly by the mechanisms. However, the mechanisms can still have an indirect effect also on the connections that are not used by passengers: for example, a general reduction of delays would improve the preservation of these connections as well. Secondly, while the loss of outgoing passenger centrality of an airport reflects also the missed connections in the airport itself, the loss of outgoing trip centrality does not. Therefore, the former would be interesting both for a hub and for a regional airport, while the latter would be mostly interesting for a regional airport, the passenger departing from which are often going to take connections elsewhere.

For both types of centrality, incoming and outgoing, we look at the average centrality loss on the entire network. Losses are averaged over all airports for each iteration, and then the average over 100 iterations and the quartiles are reported. Note that, when comparing a scenario with the corresponding baseline, a smaller average centrality loss represents an improvement.

Note that for trip centrality, when the centrality loss is averaged over the entire network, the loss of incoming centrality equals exactly the loss of outgoing centrality. In fact, each loss of outgoing centrality corresponds to an equal loss of incoming centrality of another airport. Therefore, in this case, we will refer to it as "Average trip centrality loss". Instead, when referring to a subset of the airports, incoming and outgoing centrality losses can be different.

4.1.5 Causality metrics

We consider two types of causality metrics, Granger Causality (GC) in mean and Granger Causality in tail. These metrics are applied to a set of subsystems whose evolution is described by some time series of states, e.g., the state of delay of the airport defined as the average departure delay of flights within some time window. The first GC metric detect the presence of a propagation channel for some process, e.g., delay propagation, between two subsystems, e.g., airports, while the second GC metric is similar, but considering only propagation of 'extreme' events for the process, where 'extreme' means events lying in the right tail of the distribution of states, e.g., occurrence of large departure delays which identifies the state of congestion of an airport. For the detailed description of the statistical tests of Granger causality, both in mean and in tail, see the Deliverable 5. [3].

In the causality analysis presented here, we study the network of airports, i.e., the nodes of the network, over which some propagation process takes place. In particular, we consider that each node-airport is described by

1. its state of delay, i.e., the average departure delay of flights within a time window of one hour;
2. its state of the cost of delay, i.e., the average cost of delay of flights departing from the airport within a time window of one hour;
3. its state of congestion, i.e., an extreme event for the state of delay, where an event is defined as ‘extreme’ when its value is above the 90%-percentile of the distribution of the states of delay in the corresponding baseline scenario. Thus, a binary time series is obtained, whose state is 1 if the event is extreme, 0 otherwise;
4. its state of extreme cost of delay, i.e., an extreme event for the state of the cost of delay.

Then, we apply the pairwise Granger causality test, both in mean and in tail, to any couple of node-airports in the ECAC airspace (by excluding very small airports in terms of traffic size) in order to detect the presence of a (directional) causal link, thus constructing the corresponding network of causal relations. Note that the whole 100 iterations of the Agent Based Model (ABM) developed within the Domino project are used to build a single instance of the causality networks. In fact, each iteration represents a time series of 24 states for each node (i.e., 24 hours per day). Thus, 100 iterations are used to construct a time series of 240 states for each node, in order to obtain statistically significant results.

Given a causality network, we measure some network metrics (selected in Deliverable D5.1 [3]) which signal the presence of important linkage structures for the propagation of both delays and costs. The first and the simplest metric is the link density of the network, i.e., the number of observed links divided by the total number of possible couples of nodes: it is a measure of the overall level of causality of the system. Then, we consider the clustering coefficient and the number of feedback triplets which measure how much interconnected is the system and the presence of amplifying loops for delay propagation in a loop (back propagation), respectively. Finally, reciprocity measures how many airports are mutually linked in the causality network, thus representing subsystems for delay amplification in a loop because of the presence of return-trips. The measured value of such network metrics depends on the degree of connectivity of the network. Since link density may vary from one scenario to another one, for the sake of fair comparison between scenarios, we consider the over-expression of the described metrics with respect to some random baseline, here represented by the randomisation of links or, equivalently, the Erdos-Renyi graph with the same link density. In particular, the adopted measure for the over-expression is the ratio between the observed value and the expected value for the Erdos-Renyi model (with the same link density of the considered causality network).

In terms of Granger causality, a given Domino mechanism at some level of implementation represents an improvement for the system when we observe:

1. a decreasing of the link density, thus indicating a lower level of causality because of less propagation channels;
2. a diminishing number (in absolute value) of the network metrics describing subsystems which amplify propagation of either delays or costs, especially for the extreme events (which

are associated explicitly to the occurrence of large positive values, i.e., large delays and high costs);

3. a negative variation of the over-expression of the network metrics with respect to the corresponding baseline, which signals the disruption of such amplifying subsystems.

In conclusion, GC in mean applied to the state of delay of the airports detects the channel of delay propagation, but considering both small and large delays. When restricting to the extreme positive events, GC in tail identifies the channels of propagation, but for the state of congestions of the airports. Then, a similar analysis can be applied for the cost of delay. Since the cost of delay is a non-trivial function of the delay itself, the causality network for the cost of delay may display a different structure of the corresponding causality network built with the states of delay. For instance, the same delay propagated to the periphery or to some hub airport has significantly different impact in terms of costs.

4.1.6 Restricted samples

The FP and the FAC mechanisms do not affect all airports and flights, but only a subset. For this reason, in addition to the metrics computed on the entire set of airports or flights, we also perform analyses on restricted samples, in order to highlight the direct effects of the mechanisms and also to evaluate its externalities, i.e., the indirect effects on flights or airports which the mechanism does not affect directly.

In the FAC case, the mechanism is only active in 24 airports [7]: EBBR, EDDB, EDDF, EDDL, EDDM, EGCC, EGKK, EGLL, EGSS, EHAM, EIDW, EKCH, ENGM, ESSA, LEBL, LEMD, LEPA, LFMN, LFPG, LFPO, LIMC, LIRF, LOWW and LSZH. For the computation of costs and delays, we consider the restricted sample of the flights that arrive to one of these airports. For the centrality, we compare the average centrality losses in the set of 24 airports in which FAC is implemented to those in the rest of the airports. For the causality metrics, we consider the causality links within the FAC airports, within the non-FAC airports, and the one from one restricted set to the other.

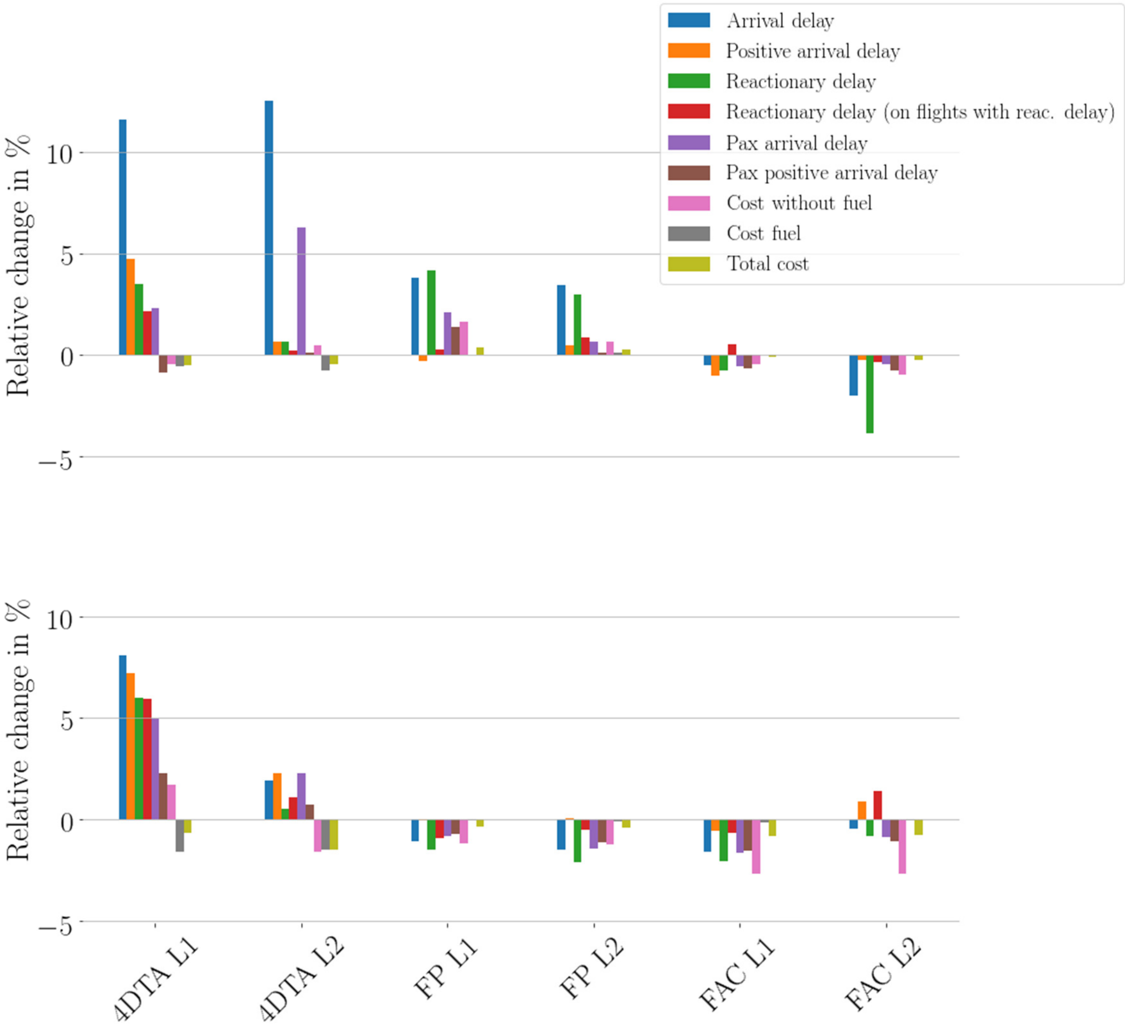
In the FP case, flight swaps can only be performed on flights that are subject to a regulation. Therefore, we consider a set of 50 iterations in which the same set of regulations have been applied, and we identify the set of airports that are subject to at least one regulation, and the set of flights that arrived at one of these airports and whose scheduled in-block time falls during a regulation. These, in fact, are the flights which can potentially apply the FP mechanism. The delay and cost metrics are then computed on the restricted sample of flights said above. For the centrality metrics, we compare the average centrality losses in airports that were subject to at least one regulation with those of airports without regulations.

4.2 Summary of key results

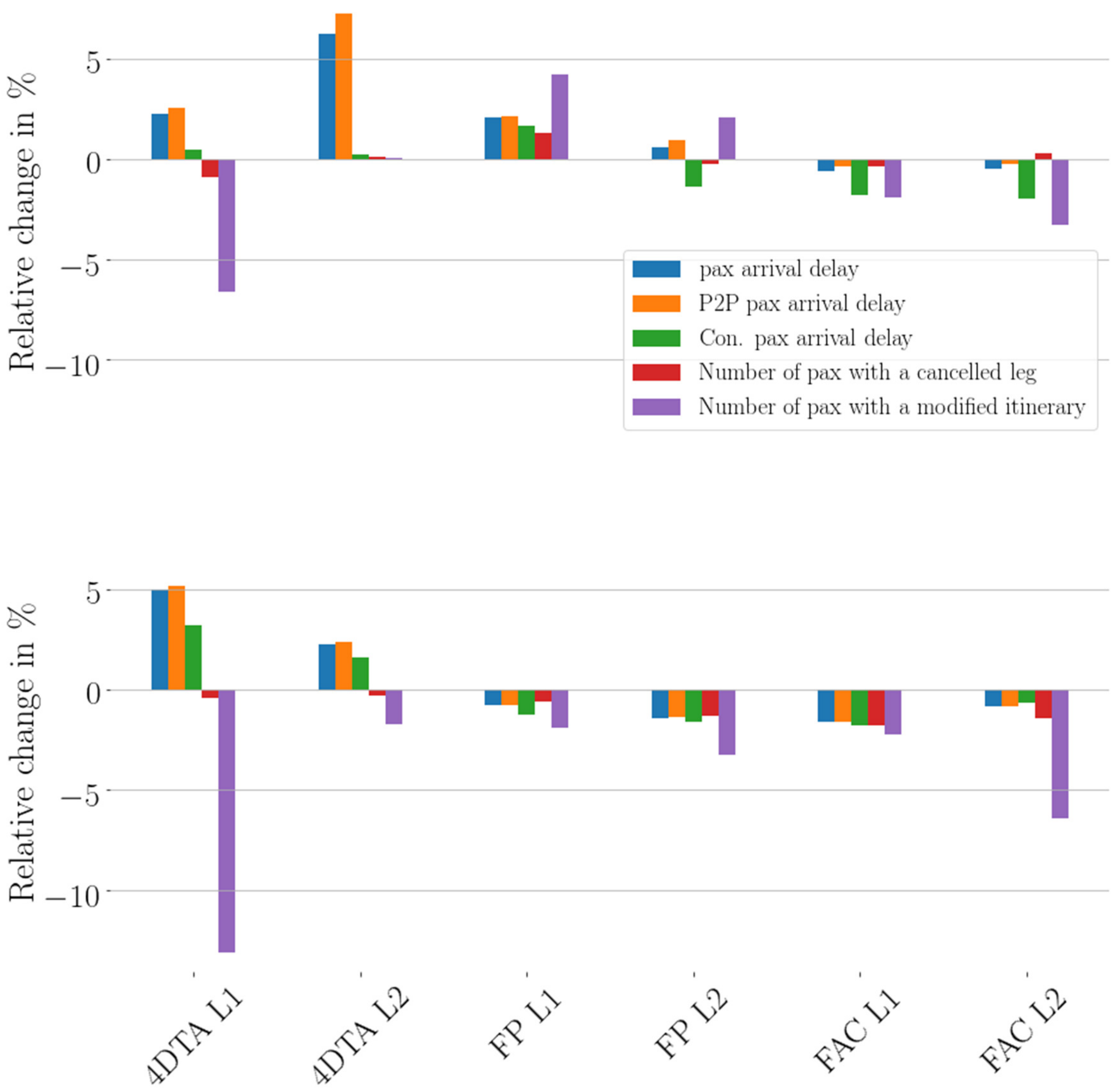
An overview of the key results from the analysis of the outputs of the model are presented in this section. The full results are detailed in Section 4.3.

In Figure 10 presents the percentage change of some of the metrics mentioned in Section 4.1 for the scenarios where the mechanisms are implemented (at Level 1 or 2) with respect to their corresponding baseline scenario (the default baseline for the default scenarios, and the stressed baseline for the stressed ones). As explained in Section 4.1, the metrics are averaged over 100 iterations of the model, or 50 in some specific cases. The results are grouped in four categories:

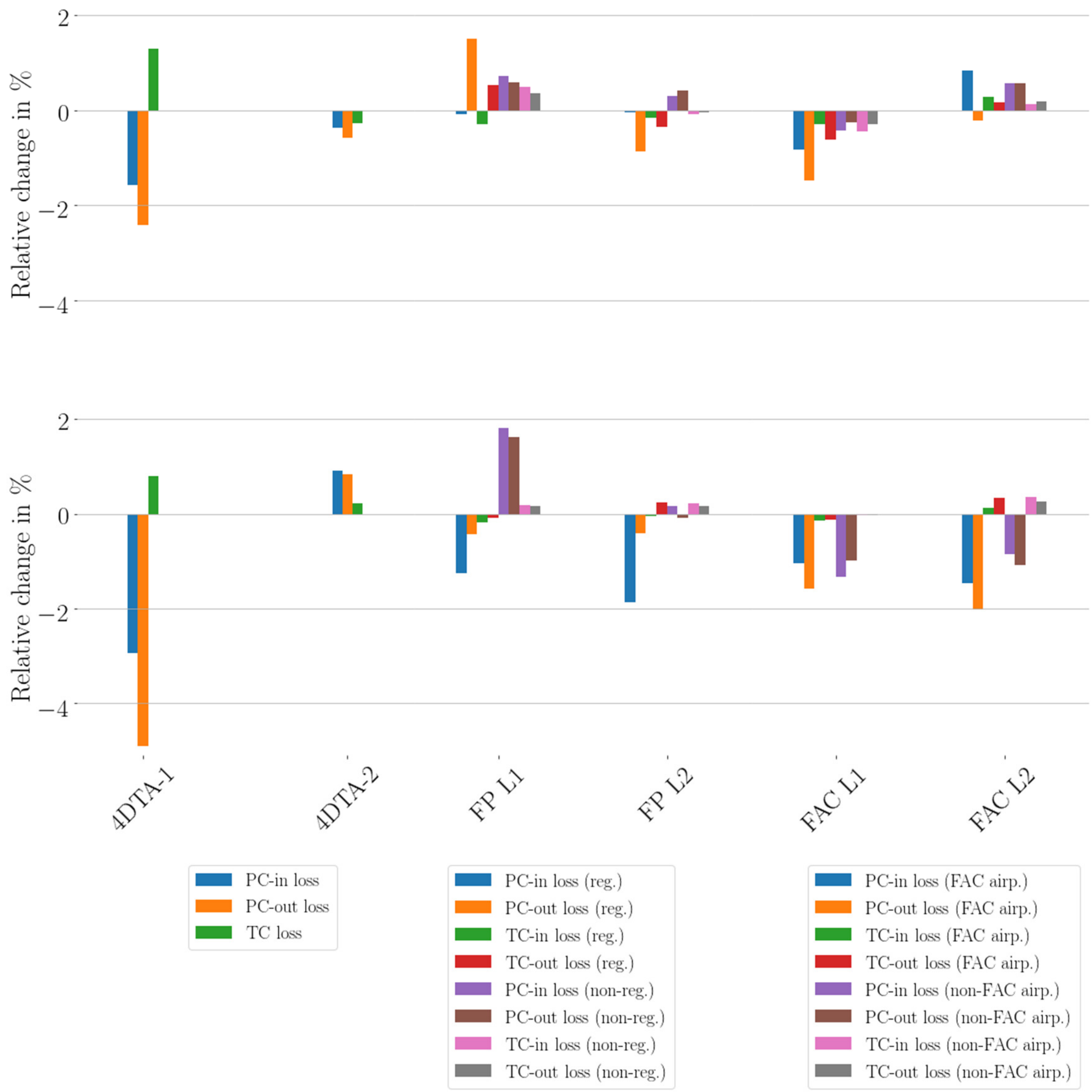
- airline centric metrics: flights delays and airline cost of delay (Figure 10.a);
- passenger centric metrics: delays, missed connections, itineraries disrupted (Figure 10.b);
- centrality metrics: passenger and trip centrality (Figure 10.c);
- causality metrics: density and reciprocity in mean and in tail (Figure 10.d).



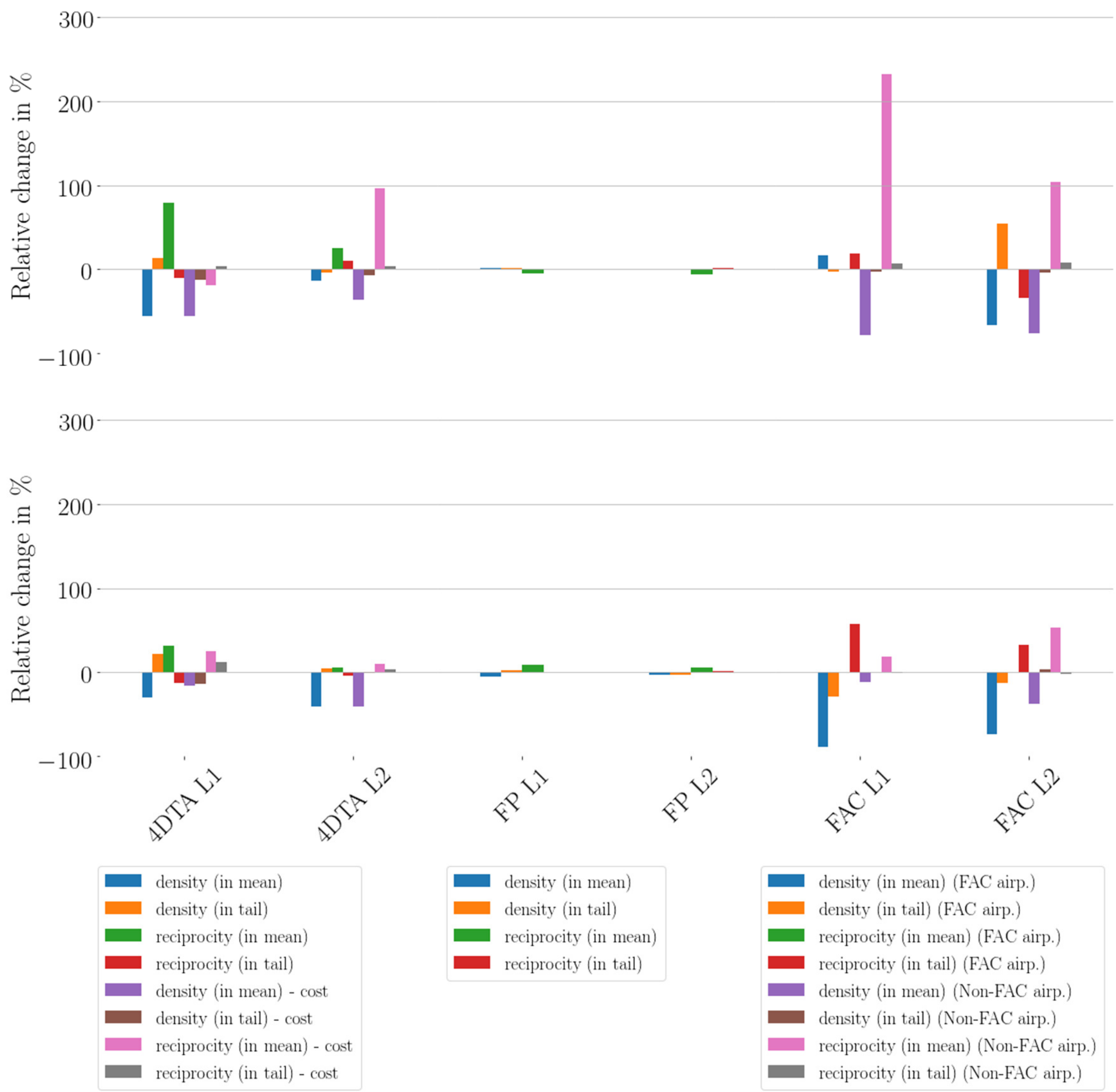
a) airline centric metrics (flight delays and costs)



b) passenger centric metrics (delays, missed connections, itineraries disrupted)



c) centrality metrics



d) causality metrics

Figure 10. Summary of results changes with respect to baseline (top) and stressed scenarios (bottom).

4D Trajectory Adjustments

Airline centric metrics

In the scenarios where the 4DTA mechanism is implemented, the minimisation of fuel costs by not speeding up to recover delay or even slowing down implies an increase of all types of delay (flight and passenger), driven by an increase in the gate-to-gate delays. The increase in departure delays might be also partly driven by the increased use of wait-for-passengers. Correspondingly, the total cost diminishes, the decrease being almost completely driven by the decrease in the cost of fuel.

Passenger centric metrics and centrality

While the passenger delays increase, when the mechanism is implemented at Level 1 the metrics evaluating the preservation of passenger itineraries (i.e., the number of passengers with a modified itinerary and the "passenger centrality" metrics) show that passengers arrive more often to their destination using their scheduled itinerary than in the baseline. Given the general increase of delays and the fact that the increased loss of trip centrality tells us that potential connections (i.e., all connections that are possible, not only the ones actually used by passengers) are increasingly disrupted with respect to the baseline, the improvement from the passengers point of view must be due to the increased use of wait-for-passengers and to the better evaluation of the possibility to speed up to preserve passenger connections.

This improvement from the passengers' point of view is however absent or very reduced when the mechanism is implemented at Level 2. This is probably due to the additional possibility to slow down that is evaluated at TOC in Level 2. In some cases, it is more economical for the airline to save fuel than preserve passengers' itineraries. Around 15% of flights initially deciding to speed up later decide to take advantage of the possibility to slow down that is evaluated at TOC. This speed change could be triggered for a reassessment of the impact of the expected delay on passengers and considering an updated expected inbound delay.

Causality metrics

The propagation of costs of delay among airports, as measured by the density of causal links among airports, diminishes. However, in most cases this decrease of cost propagation does not seem to eliminate the bidirectional propagation patterns measured by the reciprocity, which potentially increases costs through back propagation.

4.2.1 Flight Prioritisation

Airline centric metrics

In the scenarios where the FP mechanism is implemented, a negligible effect on the delays and costs is seen at the whole system level. This is understandable, as the mechanism only can be applied to 1-2% of the flights (i.e., the ones which arrive to an airport which is regulated due to ATFM). Additionally, the flight swapping reduces the delay of a flight at the expense of increasing that of another flight, so it is expected that few differences are seen in the average delays.

Passenger centric metrics and centrality

Regarding passengers, passenger delays and itinerary disruption seem to experience a worsening in the baseline case, but an improvement in the stressed case. This suggests that when there are large delays in the system, the economic interest of the airlines is better aligned with the passengers' convenience.

Figure 10.c compares the average centrality loss in the set of airports subject to regulations to that in the airports not subject to any, as explained in Section 4.1. We find that in the stressed scenarios the formers have an improvement in terms of loss of passenger centrality that the latter do not have.

Causality metrics

The study of the propagation of the state of delay among airports shows no impact by the FP mechanism at the systemic level. Overall, however, very small changes are observed with respect to the baseline, proving that FP is a mechanism whose impact are mostly local, having no significant relevance at the whole system level.

4.2.2 Flight Arrival Coordination

Airline centric metrics

In the scenarios where the FAC mechanism is implemented, a decrease of average delays is seen at Level 1, but the mechanism is less effective on delay reduction at Level 2. This is expected, as in Level 2 the Arrival Coordination aims to minimise the airline costs and not necessarily the delay, while Level 1 focuses on the total delay (arrival and reactionary) at the airport.

As explained in Section 4.1, we also considered the sample restricted to the flights arriving in airports with FAC implemented. We find that the improvements are, in percentage, larger in this restricted sample. A small decrease in passenger-related costs is seen, more important for Level 2 and larger, in percentage, in the restricted sample. No significant changes are seen in the fuel costs, which is expected given that the FAC mechanism acts only on the final part of the flight, which cannot affect strongly the excess fuel use. Nevertheless, some fuel is saved by absorbing part of the delay by reducing the speed. The amount of delay that can be absorbed by this technique, in the model, and hence the impact on fuel, is very small.

Passenger centric metrics and centrality

Passenger delays decrease, especially in the stressed case. In the default case, the decrease is much more important for connecting passengers, while in the stressed case it is similar for direct and connecting passengers. The disruption of passengers' itineraries, as measured by the number of passengers with modified itinerary and by passenger centrality loss, has an overall improvement. When comparing the passenger centrality loss in airports with FAC and airports without, the improvements are larger in airports with FAC, but also diffuse to the rest of the system. This indicates that introduction of FAC has positive externalities also on airports that do not implement it.

Improvements in the preservation of potential itineraries, as measured by the loss of trip centrality, are similar to the improvements in passenger centrality loss in the default scenarios, but much smaller in the stressed scenario. A possible explanation of this is that in the default case, when delays are small, the decrease of average delays obtained by FAC is effective in preserving not only passenger itineraries, but also potential ones, i.e., flights are maintained closer to their schedules. In the stressed case, instead, the decrease of delays is less effective in preserving itineraries, given that delays are larger, but passenger itineraries are actively targeted by the optimisation mechanism, as they incur in an explicit cost which is tried to be minimised, and therefore they are still preserved.

Therefore, similarly to the 4DTA case, the difference in the Passenger centrality and trip centrality results (one showing improvements and other worsening) is due to the active preservation of passengers' itineraries by the mechanism.

Causality metrics

The causality analysis of the scenarios where FAC mechanism is implemented, shows an overall decreasing of the level of causality, measured by a diminished number of propagation channels.

It is worth noting that the aggregate behaviour is driven by the reduction in the number of causal links in both the subgraphs of airport implementing the FAC mechanism and airports without it, as well as in the bipartite causality structure of airports with FAC propagating delays to airports without the mechanism. This externality quantifies the impact of implementing the FAC mechanism in a subset of airports characterised by high traffic, thus being more important for delay propagation. However, in most cases the reduction in the level of causality has small effect on the bidirectional propagation patterns measured by the reciprocity, which potentially increases delays through back propagation.

4.3 Detailed analysis of results

In the following sections the detailed analysis of the model outputs are presented. The results are organised as follows: One sub-section is dedicated to each mechanism (4D Trajectory Adjustments, Flight Prioritisation, Flight Arrival Coordination); for each of these, we first present the results of baseline metrics regarding flight and passenger delays and costs, and then the results of the centrality and causality metrics. All metrics used are explained in Section 4.1. Results are shown for the default and for the stressed scenarios, and all mechanisms are compared to the corresponding baseline scenario. For the stressed scenarios, the stressed baseline scenario is used for the comparisons.

4.3.1 4D Trajectory Adjustment

4.3.1.1 Flights delays

The following tables report the statistics relative to departure delay, arrival delay, gate-to-gate delay, cancellations and reactionary delay. Figure 11 summarises the percentage changes of some of the delay statistics with respect to the baseline, to ease the comparison.

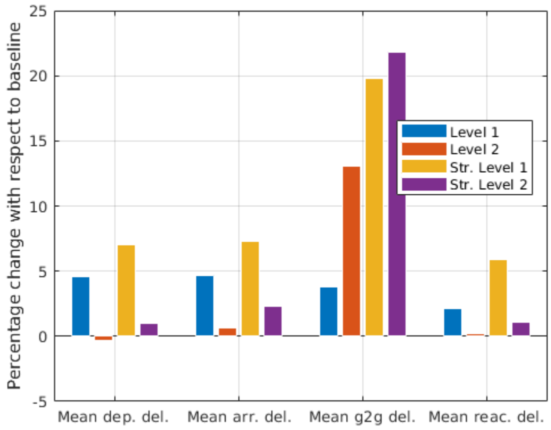


Figure 11. Percentage change with respect to the corresponding baseline of the mean departure delay, mean arrival delay (counting early arrivals as zero), mean gate-to-gate delay and mean reactionary delay for 4DTA mechanism.

Overall, the 4DTA mechanism increases all types of delays, both in average and in number. The percentage increase with respect to the baseline is especially remarkable for the gate-to-gate delay.

In the default case, 4DTA at Level 1 slightly increases both the average departure and the arrival delays. Although it reduces the number of flights in the tail (less flights with delay >180), it increases the number of delays larger than 15 minutes. According to a two-sample T-test with 5% confidence level performed on the two samples of 100 mean delays of each iteration, these differences are statistically significant. The mean reactionary delay and the number of flights with reactionary delay both increase, the latter in a statistically significant way.

4DTA at Level 2 slightly decreases the average departure delay and slightly increases the average arrival delays (counting early arrivals as zero). The average arrival delay including the early arrivals increases, as expected because in this scenario flights that estimate to arrive early slow down. The number of flights with delays larger than 15 minutes has only a slight increase, but on average the delays >15 minutes are smaller and the improvements in the tail are more important than in the Level 1 scenario. According to a two-sample T-test with 5% confidence level performed on the two samples of 100 mean delays of each iteration, the differences are not statistically significant. The mean reactionary delay and the number of flights with reactionary delay both increase, the latter in a statistically significant way.

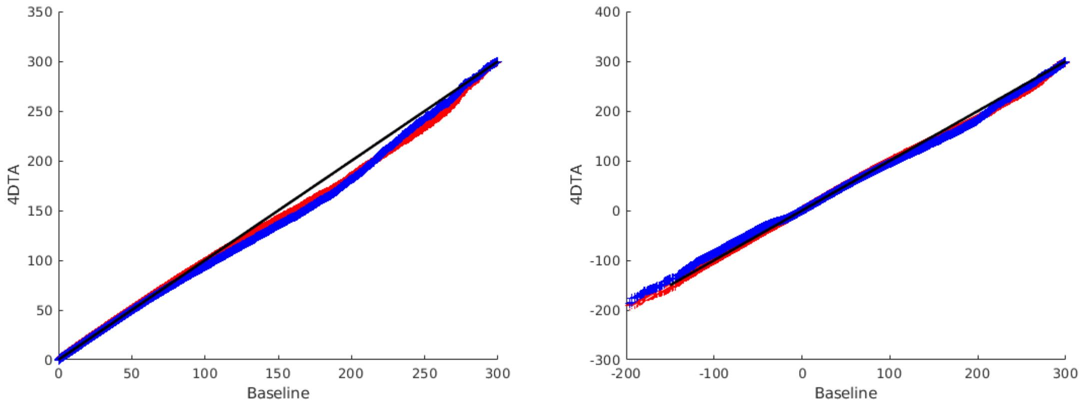


Figure 12. QQ-plots comparing departing delay distribution in the default baseline scenario and in the default scenarios with 4DTA. Left: QQ-plot of departure delay; left: QQ-plot of arrival delay. Red: Level 1, Blue: Level 2. The black line is the 1:1 line.

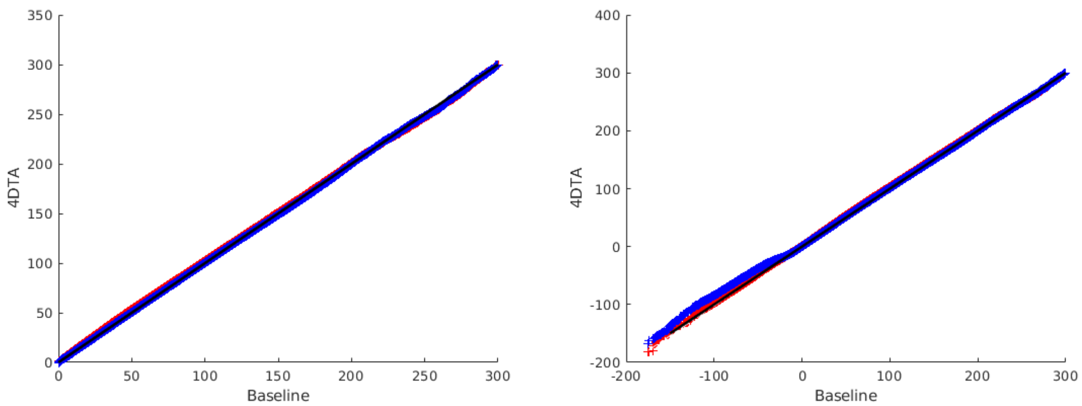


Figure 13. QQ-plots comparing arrival delay distribution in the stressed baseline scenario and in the stressed scenarios with 4DTA. Left: QQ-plot of departure delay; left: QQ-plot of arrival delay. Red: Level 1, Blue: Level 2. The black line is the 1:1 line.

From the QQ plots in Figure 12, we see that 4DTA reduces the delays above 100 minutes, both in departure and arrival. Level 1 seems to be more effective on larger delays (>200 minutes) with respect to Level 2, which seems to be more effective on delays around 150 minutes. Level 2 reduces the negative delays in arrival (early arrivals), as expected.

In the stressed case, the positive effect on the tails that was seen in the default case is barely noticeable: only the tail of delays larger than 180 minutes has an improvement in the Level 2 scenario. Also in the QQ-plot (Figure 13), the thinning of the distribution for large delays is not observed any more. All types of delay increase on average, as well as the number of flights with departure and arrival delays larger than 15 minutes and 60 minutes. According to a two-sample T-test with 5% confidence level performed on the two samples of 100 mean delays of each iteration, the differences in the average departure delays and in the average arrival delays (counting early arrival as zeros) are statistically significant, except for the average departure delay for Level 2. The number of flights with reactionary delays increase significantly.

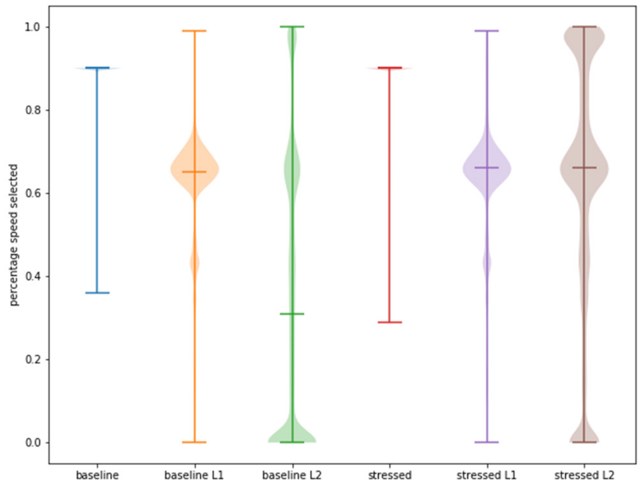


Figure 14. Violin plot of the percentage speed selected by flights (either before departure (Level 0, 1, 2) and at TOC (Level 2)) in the different scenarios. The horizontal lines represent the median and the quartiles of the distribution. A percentage speed of 1 means that the flight decides to speed up as much as possible (a maximum of 0.9 is allowed in the baseline scenarios (Level 0)), while percentages close to 0 mean the flight decides to slow down to the minimum possible speed (not allowed in the baseline).

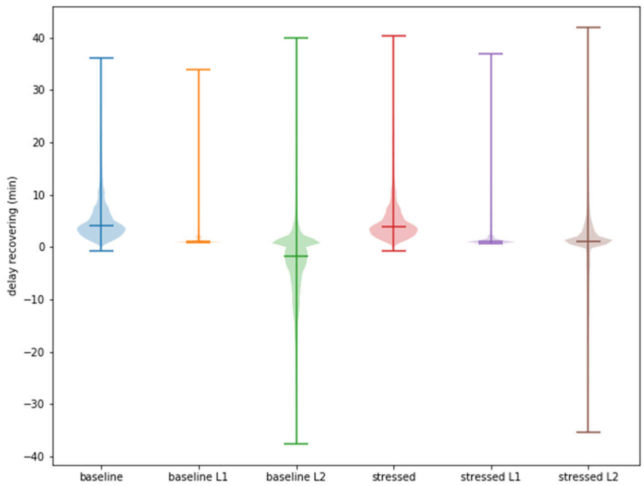


Figure 15. Violin plot of the delay expected to be recovered by changing speed in the different scenarios. The horizontal lines represent the median and the quartiles of the distribution.

The reason for this general increase of all delays lies in the different aptitude towards speeding up and fuel saving of the different scenarios. Figure 14 shows the percentage speed selected by flights when applying DCI (either at gate, or also at TOC in Level 2 scenarios), and Figure 15 shows the minutes of delay that are expected to be recovered with this choice.

In the baseline scenarios, flights most often decide to speed up as much as possible to recover delay, as the cost of fuel is not considered in the decision, and as much as 5 minutes of delay are expected to be recovered on average. In Level 1 scenarios, the aptitude is more careful with respect to fuel consumption, as total cost is considered, therefore the percentage speed selected are typically smaller, and on average only 1 minute of delay is expected to be recovered. In Level 2 scenarios, additionally, flights are also allowed to slow down, and in fact we see that many flights choose percentage speeds close to zero, implying the creation of more delay (this happens especially when the system is not stressed, while in the stressed case hardly any new delay is created). On top of that, we also see flights choosing extremely high percentage speed. Note, however, that many of these extreme speeding up decisions are taken at gate and reconsidered at TOC, therefore their effect is not seen in the results. In Level 2, the mechanism evaluates delaying the flight at the gate to wait-for-passengers and speed up conjointly. Therefore, when at the gate, the solution considered optimal is to wait for some passengers and then speed up to a very high speed. However, once the situation is re-evaluated at TOC with the passengers already on-board, the management is more conservative for fuel and speed more similar to Level 1 are selected. Level 2 shows larger estimated delay recovery than Level 1, driven by this joint assessment at the gate.

In conclusion, the general increment of delays stems from the fact that in Levels 1 and 2 of the 4DTA mechanisms flights choose to speed up less with respect to the baseline, or even to slow down in Level 2. This increases the gate-to-gate delay, and consequently the arrival, reactionary and departure delays. The increase in departure delays could even be partly due to an increased use of the wait-for-passenger. However, this the information on how many flights decide to wait for passenger was not saved in the current model version, therefore this hypothesis cannot be confirmed. This information will be saved and analysed in the next version of the model. Finally, note that the behaviour observed might indicate also that the Level 0 strategy is too aggressive when delay is recovered and the rule of thumb is using fuel consumptions that are too high.

Table 8, Table 9, Table 10 and Table 11 present the results for flight departure delay, flight arrival delay, flight gate-to-gate and cancellations and reactionary delay statistics respectively for the 4DTA mechanism for the three levels of implementation in the default and stressed scenarios.

Table 8. Departure delay statistics for the 4DTA scenarios where it is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean departure delay (min)	**10.5**	**11.0**	**10.5**	**25.9**	**27.7**	**26.1**
	10.1	10.6	10.1	25.2	27.0	25.5
	10.7	11.2	10.7	26.3	28.2	26.6

#flights with dep.del>15 min	**6143.9**	**6536.5**	**6174.5**	**13560.2**	**14345.4**	**13643.4**
	6037.5	6419	6079.5	13411	14252.5	13544.5
	6239	6653.5	6272.5	13700.5	14457	13774.5
Total dep delay >15 min	**200178.1**	**212282.1**	**198568.3**	**614651.3**	**664525.9**	**620008.2**
	189484.1	202108.7	189795.9	596452.8	645840.4	603559.0
	204517.8	219114.4	206148.7	626200.1	679045.3	633010.0
Mean delay of flights with dep. del. >15 min	**32.9**	**32.5**	**32.2**	**45.33**	**46.3**	**45.4**
#flights with dep. del >60 min	**450.6**	**476.0**	**432.8**	**3013.1**	**3387.3**	**3065.8**
	366.5	395.5	368	2834.5	3239.5	2952.5
	490	530.5	486	3124	3521	3194
Total dep. delay >60 min	**40172.9**	**41194.8**	**37198.4**	**274106.6**	**305022.7**	**278014.2**
	29543.6	31658.0	29662.0	255304	285237.4	261200.8
	43460.1	44313.6	41228.1	285799.5	320625.4	291322.8
Mean delay of flights with dep. del. >60 min	**89.2**	**86.6**	**85.9**	**91.0**	**90.1**	**90.7**
#flights with dep.del>180 min	**18.8**	**13.6**	**12.9**	**83.4**	**84.2**	**80.3**
	4	3	3.5	51.5	54.5	49.5
	12.5	12	11.5	94	105.5	97.5

Total dep. delay >180 min	**4161.9**	**3007.1**	**2906.0**	**18017.0**	**18118.7**	**17391.3**
	777.3	600.4	753.6	11078.2	11656.2	10381.0
	2734.5	2711.2	2440.8	20037.1	22434.7	21550.9
Mean delay of flights with dep. del. >180 min	**221.3**	**220.5**	**224.9**	**216.1**	**215.3**	**216.6**

Table 9. Arrival delay statistics for 4DTA scenarios where it is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean arrival delay (min)	**5.18**	**5.86**	**5.84**	**29.09**	**31.55**	**30.05**
	4.81	5.46	5.5	28.43	30.86	29.4
	5.35	6.13	6.08	29.56	32.08	30.54
Mean arrival delay of delayed flights (min)	**10.49**	**10.99**	**10.56**	**30.87**	**33.12**	**31.58**
	10.1	10.62	10.24	30.22	32.43	30.94
	10.64	11.2	10.79	31.33	33.64	32.07
#flights with arr.del>15 min	**6614.13**	**6960.16**	**6696.11**	**15930.43**	**16634.67**	**16099.35**
	6504	6839	6615.5	15800	16538.5	16021.5
	6674	7045.5	6784	16035	16723	16191.5
Total arr. delay >15 min	**220541.86**	**233849.61**	**222649.26**	**758486.18**	**819811.72**	**777518.1**
	210083.98	224242.83	214234.72	740924.41	802431.3	761954.34
	225008.08	239974.11	229238.37	770880.74	835011.18	793228.29

Mean delay of flights with arr. del. >15 min	**33.34**	**33.6**	**33.25**	**47.61**	**49.28**	**48.29**
#flights with arr. del >60 min	**523.45**	**566.69**	**526.58**	**3944.73**	**4501.68**	**4137.13**
	440.5	490.5	466.5	3797.5	4359	4010.5
	565.5	607.5	574	4067.5	4631.5	4277.5
Total arr. delay >60 min	**45928.83**	**48349.54**	**44604.46**	**360514.68**	**409531.33**	**377380.37**
	35987.14	39552.64	37640.9	340020.67	390113.28	359110.5
	50130.78	51100.06	48361.74	372987.32	425905.81	393059.91
Mean delay of flights with arr. del. >60 min	**87.74**	**85.32**	**84.71**	**91.39**	**90.97**	**91.22**
#flights with arr. del. >180 min	**19.51**	**14.77**	**13.95**	**108.15**	**111.83**	**105.88**
	4.5	4	4	72	77.5	74
	14.5	13.5	12	124.5	130	124
Total arr. delay >180 min	**4323.19**	**3245.96**	**3138.04**	**23307.19**	**24037.01**	**22840.83**
	989.74	989.02	940.06	15350.88	16417.31	15576.3
	3228.24	3033.68	2722.33	26928.5	27652.4	26676.01
Mean delay of flights with arr. del. >180 min	**221.59**	**219.77**	**224.95**	**215.51**	**214.94**	**215.72**

Table 10. Gate-to-gate delay statistics for 4DTA scenarios where it is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean gate-to-gate delay (min)	**-5.3**	**-5.1**	**-4.6**	**3.3**	**3.9**	**4.0**
	-5.4	-5.2	-4.7	3.2	3.85	3.9
	-5.3	-5.1	-4.6	3.3	3.93	4.0

Table 11. Cancellation and reactionary delay statistics for 4DTA scenarios where 4DTA is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
# cancelled flights	**1087.95**	**1075.06**	**1083.42**	**1160.20**	**1157.05**	**1165.21**
	1026.50	1027.50	1033.00	1096.50	1098.50	1103.50
	1136.00	1121.00	1121.00	1185.00	1207.50	1199.50
Mean reactionary delay (min)	**17.72**	**18.10**	**17.75**	**38.44**	**40.72**	**38.86**
	15.11	15.48	14.89	35.07	37.22	35.50
	16.90	16.91	17.08	39.07	41.27	38.84
# flights with reactionary delay	**2888.71**	**2986.47**	**2919.47**	**7197.35**	**7464.97**	**7308.00**
	2823.50	2922.00	2871.00	7114.00	7384.00	7233.50
	2947.00	3044.00	2966.00	7268.50	7528.00	7363.00

4.3.1.2 Passengers delays

Table 12 reports metrics related to the passenger arrival delay for different scenarios linked to the 4DTA mechanism

Table 12. Passenger indicators statistics for 4DTA mechanism.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean delay	17.2	17.6	18.3	40	42	40.9
25% perc. delay	-9.4	-9.1	-9.1	4.7	5.8	5
75% perc. delay	15.2	15.9	15.6	44.8	47.9	46.2
Positive mean delay	23.6	23.9	23.9	42.4	44.1	42.9
Positive 25% perc. delay	0	0	0	4.7	5.8	5
Positive 75% perc. delay	15.2	15.9	15.6	44.8	47.9	46.2
Number of pax with delay>15	859784.1	892452.8	877714.2	2005583.6	2084142.6	2030892.7
Total delays with delay>15	74470449.7	75315812.7	75319356.9	138623108	144701628.5	140463044.9
Mean delay with delay>15	86.6	84.4	85.8	69.1	69.4	69.2
Number of pax with delay>60	188560.8	192402.6	192295.4	561065.1	614159.1	583253.5
Total delays with delay>60	55486051.8	55286443	55739860.1	91227358.5	95640736.3	92477631.4

Mean delay with delay>60	294.3	287.3	289.9	162.6	155.7	158.6
Number of pax with delay>180	142665.4	141242	142924.6	158099.6	157273.1	155240.2
Total delays with delay>180	51765630	51136593.1	51790382.8	56374057.5	56048900.5	55472940.6
Mean delay with delay>180	362.8	362	362.4	356.6	356.4	357.3

Considering the average delays passengers experience worse or similar performance when the mechanism is in place. The number of passengers delayed by the different threshold tend to increase or stay constant when the mechanism is implemented, with the exception of the very high delays in the stressed scenario, in which case 4DTA seems to reduce the number of affected passengers. Note that even in this case, the average delay in this high range is still constant, denoting the fact that some passengers are not delayed any more, but that the situation for the ones who are does not improve.

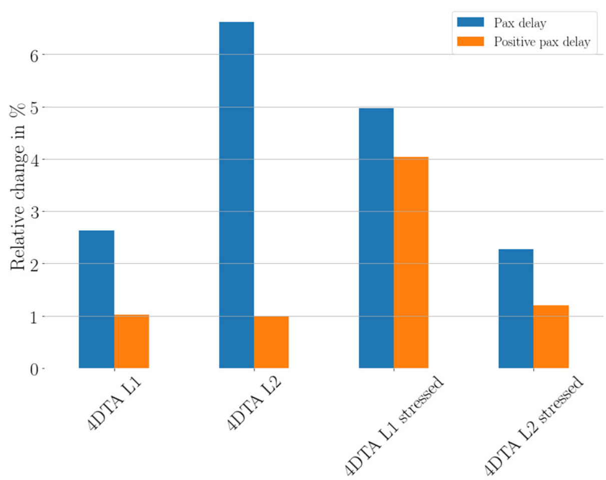


Figure 16. Change in passenger delay for 4DTA: the mean delay (blue) and the mean positive delay (orange) with respect to baseline scenarios (default on the left, stressed on the right).

The relative changes can be appreciated in Figure 16 for the global average. In all scenarios, the average delay increases, reaching up to 7%. It is interesting to note that the positive delay is not increasing as much as the total delay (including negative delay, i.e., early arrivals). This would mean that the mechanism tends mainly to increase the negative delays, i.e., passengers are arriving earlier than scheduled. This is in line with what is observed for the flight arrival delays and is due to the fact that, especially at Level 2, the mechanism allows planes to slow down in order to save fuel, which tends to reduce in absolute value the negative delays. If flights are expected to arrive earlier than their schedule, the airline decides to save fuel by slowing down the flight and still arrive on-time.

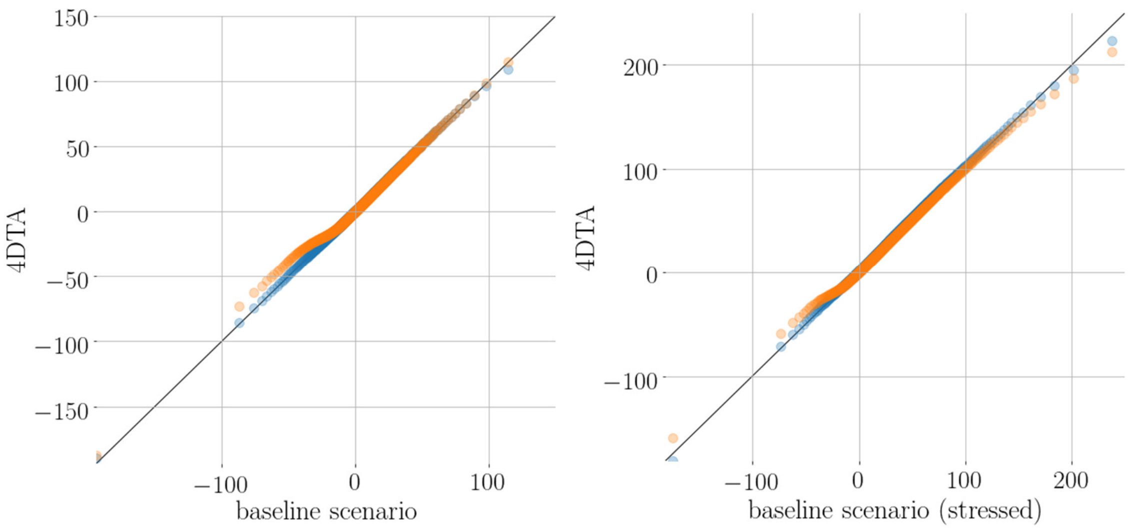


Figure 17. QQ-plot of arrival passenger delay in 4DTA in the baseline scenario (left) and stressed scenario (right). The blue circles correspond to the 4DTA at Level 1, the orange ones to Level 2. The plots have been cut at 150 and 200 minutes respectively.

To inspect further this fact, Figure 17 shows two QQ-plots, which help to understand the differences in scenarios. In the non-stressed case, the implementation of the 4DTA at Level 1 seems to have very little effect on the distribution of delay. On the contrary, Level 2 has a visible effect on the left tail of the distribution, squeezing it towards the right. This is consistent with the previous figure.

In the stressed scenarios, the mechanism seems to have a bigger impact. On the left side of the distribution, the effect is similar, with a squeeze of the distribution at Level 2. On the right side, it is now more obvious that both levels have an effect. The tail is indeed squeezed by the implementation of the mechanism, even more so at Level 2 than Level 1. This means that the situation of highly delayed passengers is improved by the mechanism.

In summary, the effect of 4DTA for passengers is the following:

- for non-stressed scenarios, 4DTA increases the delays mainly in the negative part, particularly at Level 2. This drives the average delay up by 2-7% but its positive part only by 1%. I.e., if as cost of fuel is considered, if flights are going to arrive earlier than scheduled, they are slowed down.
- for stressed scenarios, the same effect on the negative delays can be seen, particularly at Level 2. Moreover, 4DTA improves the situation for very high delays, with fewer passengers experiencing them.

4.3.1.3 Costs

In this section we discuss the impact of the 4DTA mechanism on the different kinds of costs airlines incur due to extra fuel consumption and delay.

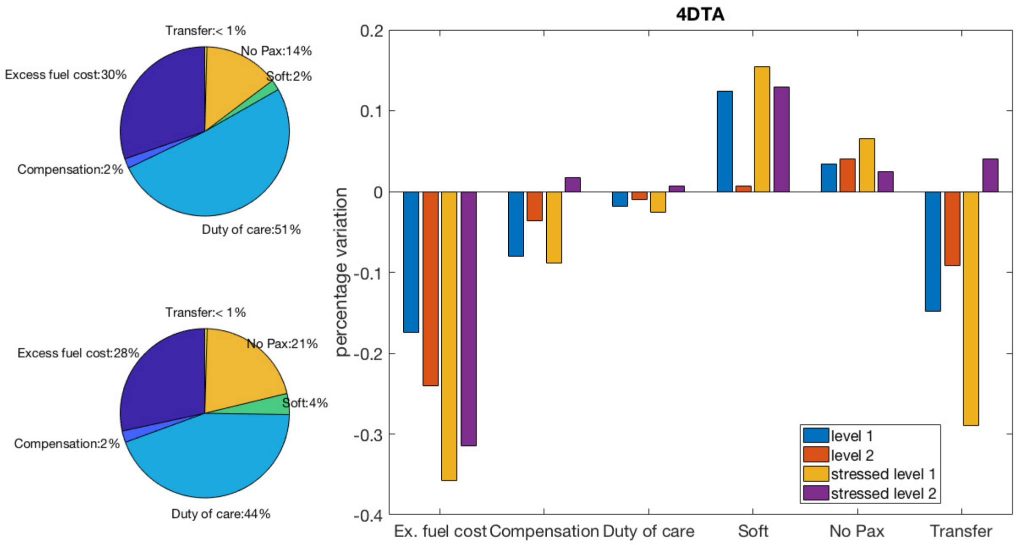


Figure 18. Cost of delay for 4DTA. Left: Percentage of the total cost represented by each type of cost considered in the analysis (upper pie: default case, lower pie: stressed case). Right: Percentage variations of the different kinds of costs with respect to the corresponding baseline scenario.

As shown in Figure 18, the 4DTA mechanism tend to decrease the overall cost per flight, in particular by decreasing the cost of fuel with respect to the baseline scenario, for both default and stressed cases. In fact, the baseline behaviour for delay recovery is based on speeding up the flight in order to decrease the arrival delay up to 5 min. On the contrary, the dynamic cost indexing implemented in the 4DTA mechanism, both at Level 1 and Level 2, compute the cost by including considerations on flight connections as well as the optimal speed in order to preserve them. The cost of fuel is estimated and considered. Hence, connections can be preserved without consuming more fuel than was necessary. This is what we observe in Figure 18, where a percentage decrease of the excess cost of fuel is coupled with a decrease (on average) also of the passengers' costs, i.e., 'transfer', duty of care, and compensation. In this context, we observe instead an increase of the soft and non-passengers' costs: since these types of costs are increasing functions of delays, their increased value with respect to the baseline is expected. Nevertheless, the profit in saving the excess fuel governs the overall cost per flight.

The mechanism in Level 1 and Level 2 speed up flights which provide a reduction on delay costs (e.g., preserving connection of passengers) but reduces the occasions where delay is recovered (at the expense of extra fuel consumption) without providing additional savings (e.g., connections which are already missed or ensured even if flight is delayed).

4DTA: detailed analysis of costs

Table 13 and Table 14 show the detailed values obtained by considering the flights simulated by the ABM. The samples of flights over which averages are computed are built as explained in the captions. In particular, Table 14 show the values associated to the different types of costs without considering in the analysis the flights which have been cancelled. This is to better capture the impact of the 4DTA mechanism on normal operations. However, in both cases, the results are qualitatively the same.

Table 13. Detailed costs by considering all flights 4DTA (with the exception of flights with delays larger than 300 min).

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Excess cost of fuel	**136.0**	**112.3**	**103.4**	**182.9**	**117.6**	**125.4**
	135.0	111.5	102.2	181.2	116.1	124.2
	137.1	112.9	104.5	184.7	118.6	126.8
Cost of compensation	**8.3**	**7.7**	**8.0**	**14.1**	**12.9**	**14.4**
	7.3	6.7	7.3	12.8	11.7	13.3
	9.1	8.3	8.5	14.7	13.8	15.0
Fraction of compensations	**1.1%**	**0.9%**	**1.1%**	**2.7%**	**2.3%**	**2.7%**
Costs of 'transfer'	**1.8**	**1.6**	**1.7**	**2.9**	**2.1**	**3.0**
	1.0	1.0	0.9	2.1	1.3	1.9
	2.4	1.8	2.1	3.5	2.5	3.5
Fraction of transfer costs	**0.1%**	**0.1%**	**0.1%**	**0.3%**	**0.2%**	**0.3%**
Duty of care	**229.8**	**225.6**	**227.5**	**284.9**	**277.5**	**286.7**
	219.8	216.3	218.7	270.0	265.8	273.1
	237.2	233.5	232.7	290.0	284.2	293.5
Fraction of 'duty of care'	**13.5%**	**13.2%**	**13.3%**	**20.0%**	**20.0%**	**20.2%**
Soft costs	**9.1**	**10.2**	**9.2**	**25.9**	**29.9**	**29.3**
	3.6	3.9	3.8	10.3	11.4	10.7
	14.0	14.7	14.3	40.9	44.5	42.2
Fraction of soft costs	**46.4%**	**47.0%**	**46.4%**	**57.5%**	**57.7%**	**57.3%**

Non passenger costs	**64.4**	**66.5**	**67.0**	**134.0**	**142.7**	**137.4**
	62.3	64.5	65.5	129.9	138.9	134.2
	64.9	67.4	67.9	136.9	145.9	139.9
Fraction of non passenger costs	**91.1%**	**91.3%**	**91.5%**	**93.2%**	**93.4%**	**93.4%**
Total excess cost	**449.5**	**423.9**	**416.7**	**644.8**	**582.8**	**596.1**
	436.6	412.7	407.7	619.8	563.2	578.8
	457.9	432.6	426.1	658.3	597.4	609.2

Table 14. Detailed costs 4DTA considering flights with delays<300 min, excluding cancelled flights.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Excess cost of fuel	**141.6**	**116.8**	**107.6**	**190.8**	**122.7**	**130.9**
	140.3	115.9	106.4	188.9	121.5	129.6
	142.6	117.5	108..8	192.8	123.4	132.3
Cost of compensation	**1.5**	**1.0**	**1.4**	**6.9**	**4.8**	**7.1**
	1.2	0.8	1.2	6.2	4.1	6.3
	1.5	1.0	1.5	7.4	5.2	7.7
Fraction of compensations	**0.6%**	**0.4%**	**0.6%**	**2.1%**	**1.6%**	**2.1%**
Costs of 'transfer'	**0.3**	**0.1**	**0.2**	**1.3**	**0.3**	**1.3**
	0.2	0.1	0.2	1.1	0.2	1.1
	0.4	0.2	0.3	1.5	0.4	1.5

Fraction of transfer costs	**5e-2%**	**3e-2%**	**5e-2%**	**0.2%**	**8e-2%**	**0.2%**
Duty of care	**39.7**	**36.5**	**37.7**	**75.9**	**69.1**	**77.2**
	35	32.4	35.1	68.0	62.6	70.7
	41	38.2	39.5	80.1	74.8	80.5
Fraction of 'duty of care'	**11.8%**	**11.5%**	**11.6%**	**18.2%**	**18.2%**	**18.5%**
Soft costs	**7.3**	**8.3**	**7.4**	**24.6**	**28.7**	**27.9**
	2.9	3.2	3.1	9.8	10.9	10.3
	11.1	12.0	11.5	38.8	42.6	40.3
Fraction of soft costs	**46.1%**	**46.8%**	**46.2%**	**57.5%**	**57.7%**	**57.3%**
Non passenger costs	**67.0**	**69.3**	**69.7**	**139.8**	**148.9**	**143.4**
	64.8	67.2	68.1	135.4	144.5	140.2
	67.4	69.9	70.6	142.7	152.2	146.2
Fraction of non passenger costs	**94.8%**	**95.0%**	**95.2%**	**97.2%**	**97.4%**	**97.4%**
Total excess cost	**257.6**	**232.1**	**224.1**	**439.5**	**374.6**	**387.9**
	248.7	225.4	218.4	420.4	357.3	374.0
	262.9	235.8	228.5	454.7	389.1	399.9
	142.6	117.5	108..8	192.8	123.4	132.3

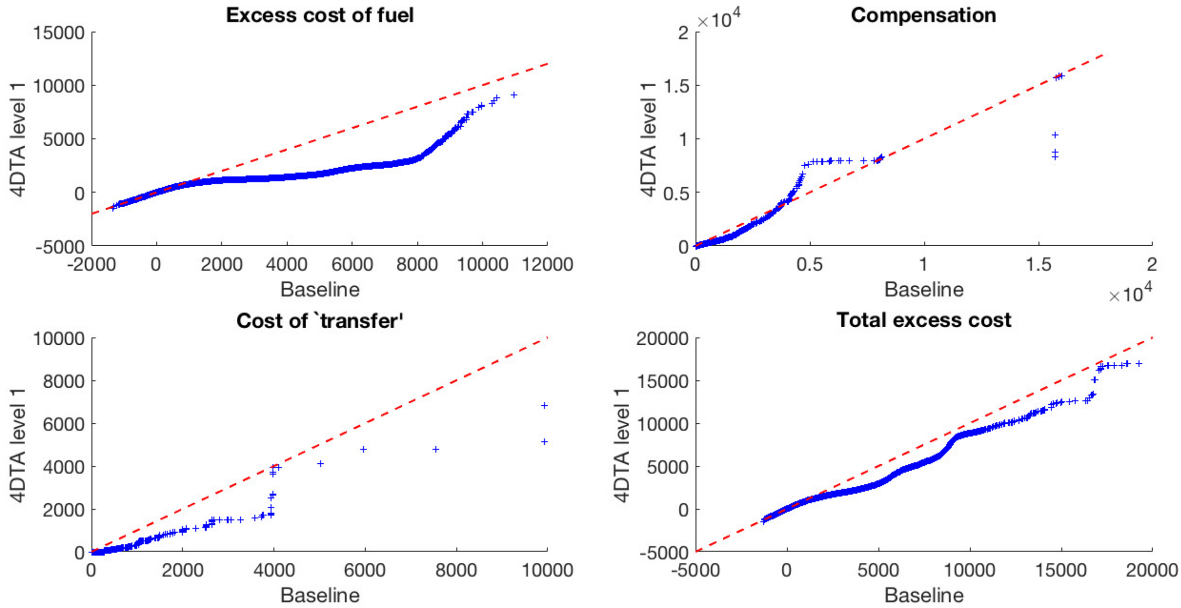


Figure 19. QQ-plots for 4DTA Level 1 baseline comparing the distribution of costs (the type of cost is shown in each title of plots).

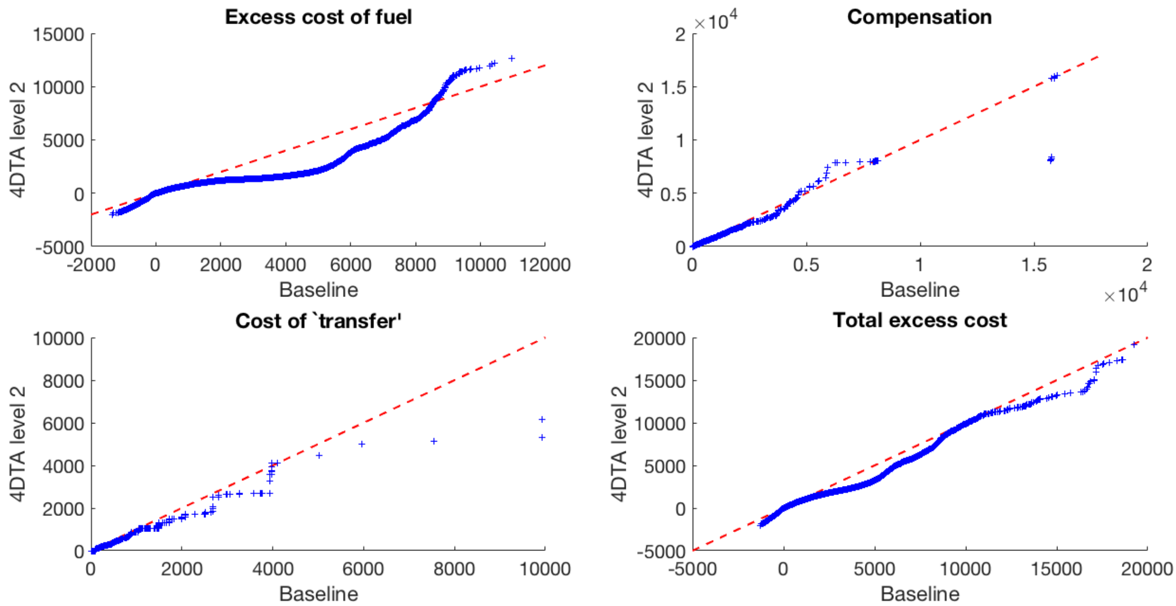


Figure 20. QQ-plots for 4DTA Level 2 baseline comparing the distribution of costs (the type of cost is shown in each title of plots).

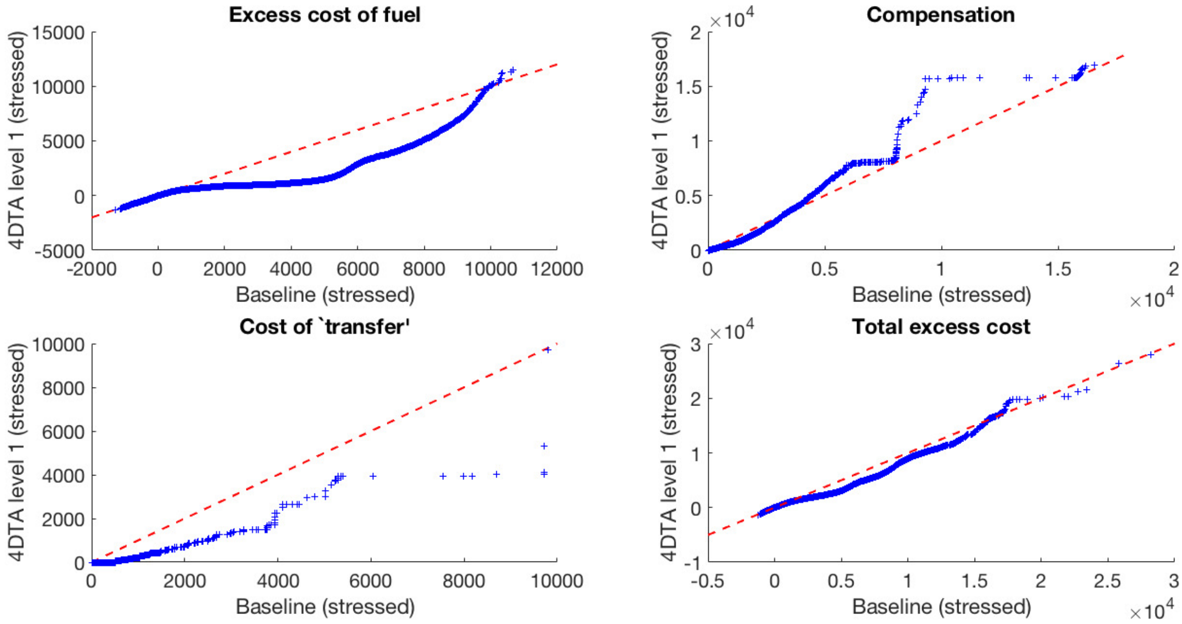


Figure 21. QQ-plots for 4DTA Level 1 stressed comparing the distribution of costs (the type of cost is shown in each title of plots).

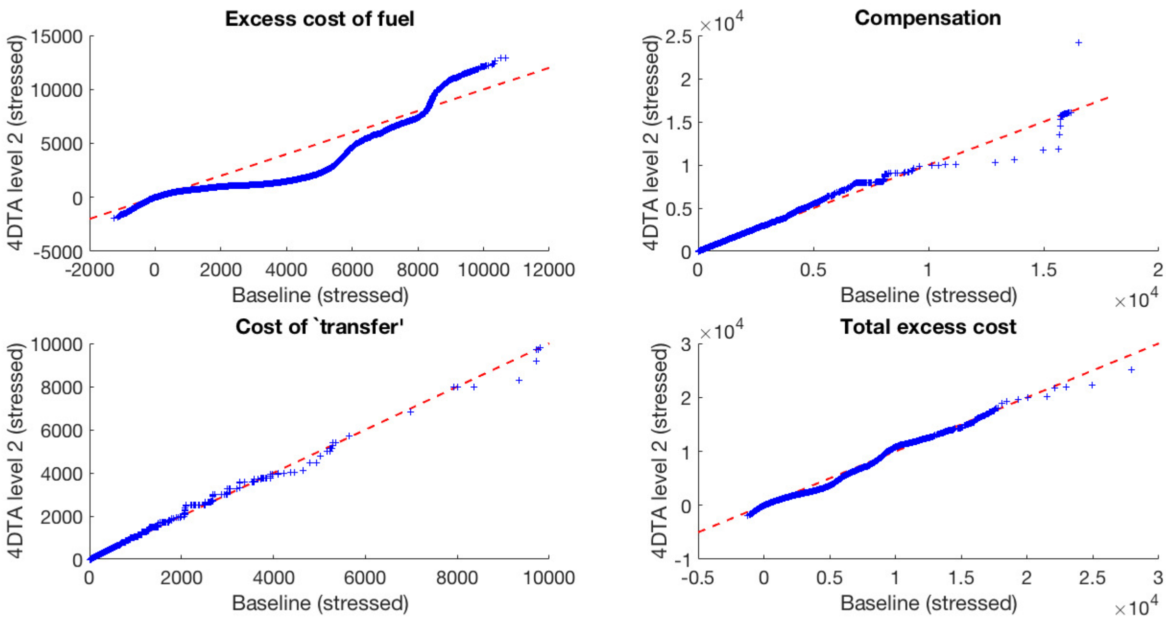


Figure 22. QQ-plots for 4DTA Level 2 stressed comparing the distribution of costs (the type of cost is shown in each title of plots).

Figure 19, Figure 20, Figure 21 and Figure 22 show the QQ-plots to characterise the distribution of each type of costs, and how they change when we move from the baseline scenario, both default and stressed, to the scenario with implemented 4DTA.

With the exception of the stressed scenario with 4DTA implemented at Level 2 (results are difficult to interpret and more research is necessary). We observe a clear pattern for the cost of transfer whose distribution is characterised by a thinner right tail, i.e., smaller probability of extreme positive events or, equivalently, high costs. A similar behaviour is observed for the distribution of the excess cost of fuel. The pattern describing the cost of compensation is less evident, whose distribution displays however a smaller mean with respect to the baseline. In any case (except stressed 4DTA at Level 2), the impact of the Domino mechanism on the total cost describes an overall improvement because the distribution of the sum of all the components shows a thinner right tail with respect to the baseline, thus pointing out a decreasing of the occurrence of extreme cost events.

Comments

According to the results about mean costs (diversified by type) shown in the previous Tables, we point out that the 4DTA mechanism tend to reduce significantly the total cost per flight by reducing mainly the excess cost of fuel at the expense of soft and non-passengers costs of delay, which, on the contrary, display an increase when we move from the baseline scenario (both in the default and stressed cases) to the scenario with 4DTA mechanism implemented, especially at Level 1.

This is also confirmed by looking at the distributions of costs (QQ-plots), especially in the default case: the benefit of introducing the 4DTA mechanism (at both Level 1 and Level 2) is particularly related to a thinner distribution for the excess cost of fuel.

4.3.1.4 Centrality losses

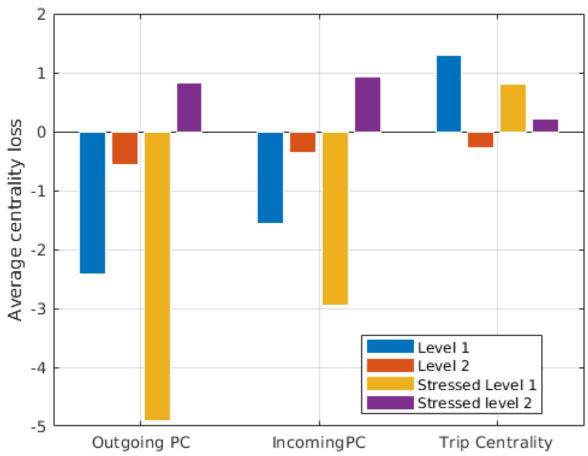


Figure 23. Percentage change in the average centrality loss in 4DTA scenarios implemented at Level 1 and 2, with respect to the baseline. Three types of centralities are considered: passenger centrality (outgoing and incoming) and trip centrality (in this case, the average incoming and outgoing centralities coincide).

Figure 23 presents the average centrality loss changes for passenger and trip centrality for the mechanism in Level 1 and Level 2 for both the default and stressed scenarios with respect to their respective baseline with 4DTA at Level 0.

Passenger centrality

Table 15. Average passenger centrality loss, 4DTA, incoming and outgoing. in the scenarios in which 4DTA is implemented and in the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Average incoming passenger centrality loss	**186.03**	**183.12**	**185.37**	**201.48**	**195.55**	**203.34**
	177.93	176.11	177.55	192.20	187.85	197.12
	194.01	191.24	192.31	206.50	205.32	210.32
Average outgoing passenger centrality loss	**206.25**	**201.27**	**205.09**	**232.88**	**221.45**	**234.83**
	196.81	192.87	196.35	223.26	211.89	227.48
	215.26	209.62	212.82	238.94	231.80	242.47

When the 4DTA mechanism is implemented at Level 1, the average passenger centrality loss (incoming and outgoing) is slightly decreasing, both in the default and in the stressed scenarios (Table 15). The percentage decrease with respect to the corresponding baseline (see Figure 23) is larger for the outgoing centrality and in the stressed case. The decrease is statistically significant for the outgoing centrality loss (default and stressed) and for the incoming centrality loss in the stressed scenario. The box plots in Figure 24 show that in the stressed case the Level 1 mechanism limits the outliers with large average centrality loss.

When the mechanism is implemented at Level 2, instead, there is only a very slight, non-significant decrease of the average passenger centrality loss in the default case, and a slight, non-significant increase in the stressed case.

The reason for the smaller effect of the mechanism at Level 2 in preserving passengers' connectivity might be due to the second application of DCI at TOC. In fact, at TOC flights have the additional possibility to slow down, which they did not have at gate. Therefore, flights that had decided to wait for passengers and then speed up, as that was the most convenient choice at gate, might now find more convenient to slow down and save fuel, even if this means disrupting some passenger connections.

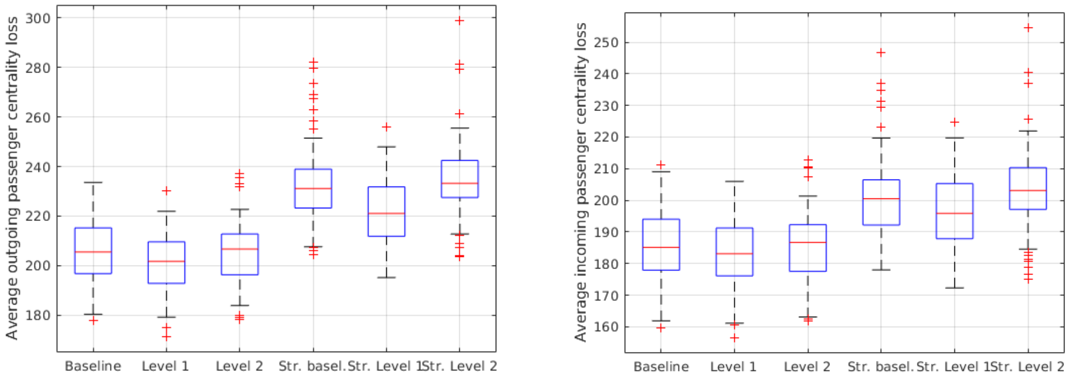


Figure 24. Average passenger centrality loss 4DTA Box plots comparing the average outgoing (left) and incoming (right) passenger centrality loss in the scenarios where 4DTA is implemented at Level 1 and 2 and in the corresponding baselines. For each scenario, the average centrality losses in each of the 100 iterations of the model are considered. implemented at Level 1 and 2, with respect to the baseline. Three types of centralities are considered: passenger centrality (outgoing and incoming) and trip centrality (in this case, the average incoming and outgoing centralities coincide).

Figure 24 compares the average outgoing and incoming passenger centrality loss in the scenarios where 4DTA is implemented at Level 1 and 2 and in the corresponding baselines.

Trip centrality

When 4DTA is implemented at Level 1, there is a small but statistically significant increase in the average trip centrality loss, both in the default and in the stressed cases. No statistically significant change is seen at Level 2 (see Table 16 and Figure 25).

Table 16. Average trip centrality loss in 4DTA scenarios and in the corresponding baselines passenger centrality loss, 4DTA, incoming and outgoing. in the scenarios in which 4DTA is implemented and in the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Average trip centrality loss	**0.53**	**0.54**	**0.53**	**0.73**	**0.74**	**0.73**
	0.52	0.53	0.52	0.73	0.73	0.73
	0.54	0.54	0.54	0.74	0.74	0.74

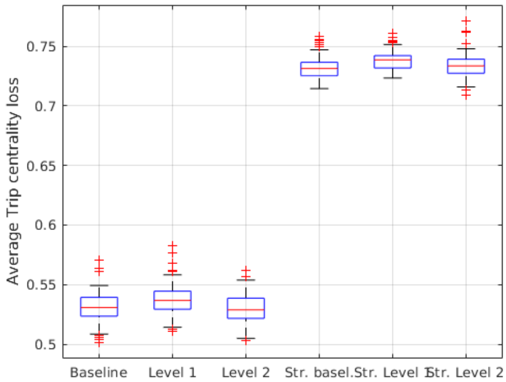


Figure 25. Trip centrality loss comparison 4DTA Box plots comparing the average trip centrality loss in the scenarios where 4DTA is implemented at Level 1 and 2 and in the corresponding baselines. For each scenario, the average centrality losses in each of the 100 iterations of the model are considered.

The fact that, when 4DTA is implemented at Level 1, the loss of Passenger centrality decreases while the loss of trip centrality increases with respect to the baseline tells us that the mechanism, while introducing more delays, limits the disruption of passengers' itineraries (e.g., by using wait-for-passengers). In fact, the increase of trip centrality loss is a consequence of the increased number of delayed flights, causing disruptions in the potential itineraries that are considered by trip centrality. However, the itineraries that are actually used by passengers are considered in the mechanism optimisation, as their disruption implies costs. Therefore, such itineraries are increasingly preserved with respect to any of the potential itineraries that trip centrality considers. As a result, passenger centrality losses decrease with the Level 1 mechanism.

In summary, the effects of the 4DTA mechanisms are:

- The passengers' itineraries are increasingly preserved at Level 1, but not at Level 2, as shown by the passenger centrality losses. This is in agreement with the number of passengers having a modified itinerary as shown in Section 4.2.1: for 4DTA at Level 1 there is a reduction with respect to the baseline but there is not a significant change in Level2.
- The itineraries that are not explicitly considered in the optimisation process are slightly less preserved in all scenarios, with respect to the baseline, as shown by the trip centrality losses.

4.3.1.5 Causality analysis

In this section, we show the causality analysis applied to the scenarios where 4DTA mechanism is implemented and we compare the results with the corresponding baselines. In this section, we study the Granger causality (both in mean and in tail) networks built with simulated data. In particular, we consider both propagation of delays and propagation of costs of delay. Since the 4DTA mechanism aim to reduce the costs each airline incur, by means of a dynamic cost computation during the tactical phase, the study of cost propagation may reveal patterns not observed for delay propagation because of the non-trivial dependence of the cost of delay from the delay itself.

Causality in mean

Table 17.Metrics for the Granger causality in mean network 4DTA (delay), for the baseline scenario (both default and stressed) obtained for the state of delay.

Delay	Baseline default scenario	Baseline stressed scenario
Link density	0.0043	0.0034
Mean degree	1.09	0.87
Clustering coefficient	0.52	0.53
Over-expression of feedback triplets	908.3	353.7
Reciprocity	0.28	0.23

Table 18. Metrics for the Granger causality in mean network 4DTA (cost of delay), for the baseline scenario (both default and stressed) obtained for the state of cost of delay

Delay	Baseline default scenario	Baseline stressed scenario
Link density	0.0063	0.0053
Mean degree	1.6	1.4
Clustering coefficient	0.43	0.46
Over-expression of feedback triplets	454.2	237.4
Reciprocity	0.23	0.20

In this sub-section we focus on the study of Granger causality in mean.

Table 17 and Table 18 show the value of the network metrics, for both default and stressed baseline scenarios. In Table 17, we show such metrics in the case of the state of delay, while Table 18 is for the state of cost of delay. Notice that, in both cases, link density decreases from the default case to the stressed one. That is, the higher delays of the stressed scenario tend to be less correlated, thus the endogenous process of propagation of both delays and costs becomes less important.

When we aim to compare the Granger causality network in mean for the two baseline scenarios for the state of delay and the state of cost of delay, we can use the Jaccard index to assess the similarity between the two networks. The Jaccard index is the ratio between the size of the intersection and the size of the union for the two sets of causal links defining the two causality networks. A high value of this index suggests that the two networks display the same channels of propagations, whereas small values indicate that two networks are significantly different. In this case, the Jaccard index is equal to 0.4978 for the default case and 0.3995 for the stressed case, thus revealing a partial superposition for the propagation channels of both delays and costs.

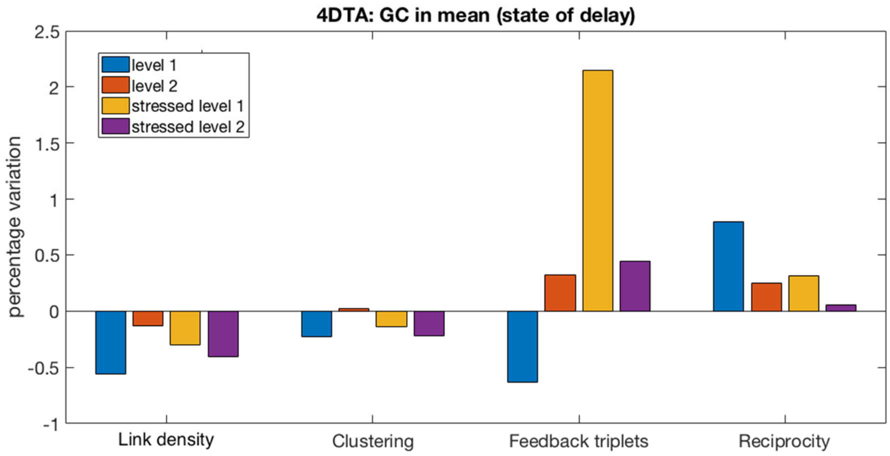


Figure 26. Percentage variation of Granger causality in mean for 4DTA for the selected metrics of causality in mean network, i.e., link density, clustering coefficient, number of feedback triplets, and reciprocity coefficient, when we compare the scenario with 4DTA mechanism with the corresponding baseline (for both default and stressed case). (In order to compare metrics between two different scenarios, we consider the measured value of each metric but normalised by the expected value for the random case. This is done for a fair comparison, because link density changes from one scenario to another).

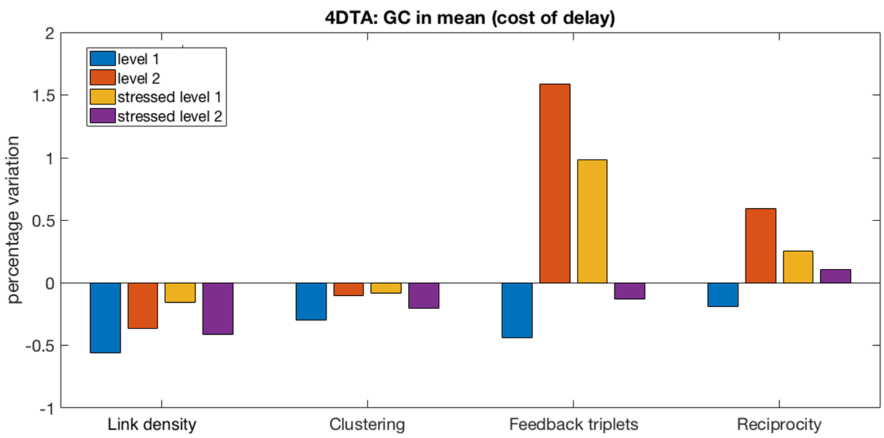


Figure 27. Percentage variation of Granger causality in mean cost of delay for 4DTA in the selected metrics, i.e., link density, clustering coefficient, number of feedback triplets, and reciprocity coefficient, of the Granger causality in mean network but built with the state of cost of delay when we compare the scenario with 4DTA mechanism with the corresponding baseline (for both default and stressed case). (In order to compare metrics between two different scenarios, we consider the measured value of each metric but normalised by the expected value for the random case. This is done for a fair comparison, because link density changes from one scenario to another).

Figure 26 and Figure 27 show the percentage variations of the network metrics for the scenarios where 4DTA is implemented with respect to their baseline values shown in Table 17 and Table 18. Hence, we can obtain some insights on the impact of the 4DTA mechanism at the global network level. Considering the case of Granger causality in mean, the impact of 4DTA is quite equivalent for the propagation of both delays and costs. We can notice that 4DTA, at any level of implementation and in both default and stressed cases, is associated with a decreasing level of causality, i.e., smaller link density. For the state of delay, this is coupled with an overall decrease of the clustering coefficient and smaller number of feedback triplets, but at the default Level 1. In any other case, we notice an increase of the remaining metrics. In particular, the over-expressions of feedback triplets and reciprocal links suggest that the 4DTA mechanism tends to reduce on average the number of propagation channels for the delay, but the subsystems of delay amplification are less affected (this is proved numerically below for reciprocity in the Granger causality in tail). A similar behaviour is observed also when studying the propagation of the cost of delay.

Causality in tail

In this sub-section we focus on the case of Granger causality in tail, i.e., we study the channels of propagation of extreme delay events or extreme cost events (see Table 19 and Table 20).

The Jaccard index for the two Granger causality in tail networks built with simulated data of the baseline scenarios is equal to 0.1627 for the default case and 0.3018 for the stressed case, thus suggesting the presence of different channels of propagation, in particular for the default case.

Table 19. Metrics for the Granger causality in tail 4DTA network for the baseline scenario (both default and stressed) obtained for the state of congestion.

Delay	Baseline default scenario	Baseline stressed scenario
Link density	0.1594	0.3460
Mean degree	40.5	87.9
Clustering coefficient	0.43	0.56
Over-expression of feedback triplets	3.7	1.9
Reciprocity	0.21	0.28

Table 20. Metrics for the Granger causality in tail 4DTA extreme events network for the baseline scenario (both default and stressed) obtained for the state of extreme event for the propagation of the cost of delay.

Delay	Baseline default scenario	Baseline stressed scenario
Link density	0.229	0.3442
Mean degree	56.6	87.4

Clustering coefficient	0.43	0.55
Over-expression of feedback triplets	2.4	1.9
Reciprocity	0.21	0.27

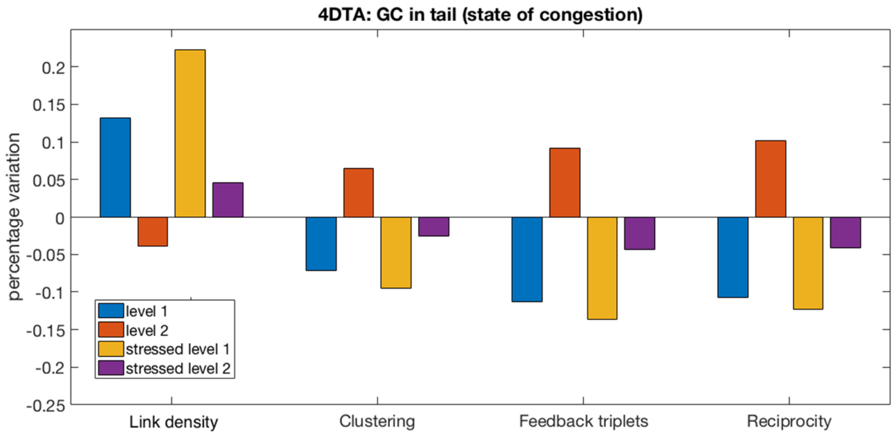


Figure 28. Percentage variation of the selected metrics of the Granger causality in tail 4DTA, i.e., mean degree, clustering coefficient, number of feedback triplets, and reciprocity coefficient, when we compare the scenario with 4DTA mechanism with the corresponding baseline (for both default and stressed case). (In order to compare metrics between two different scenarios, we consider the measured value of each metric but normalised by the expected value for the random case. This is done for a fair comparison, because link density may vary from one scenario to another).

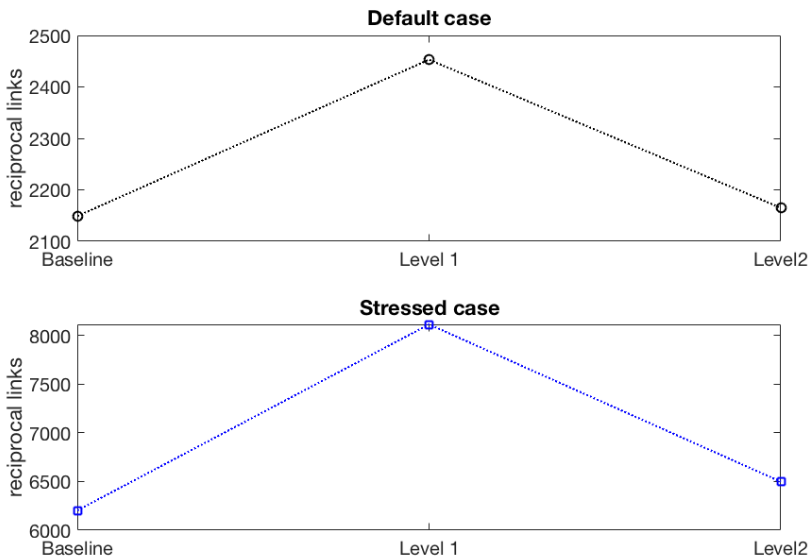


Figure 29. Number of reciprocal links (4DTA) at the different levels of implementation for both the default and stressed case. (GC in tail network built with the state of congestion).

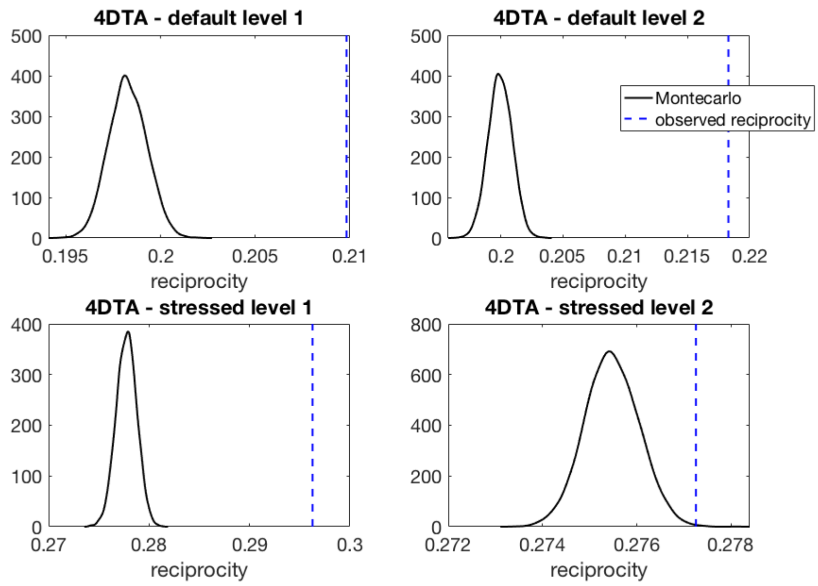


Figure 30. Distribution of reciprocity for 4DTA delay according to the Monte Carlo simulations compared with the observed reciprocity in the case of Granger causality in tail networks built with the state of congestion.

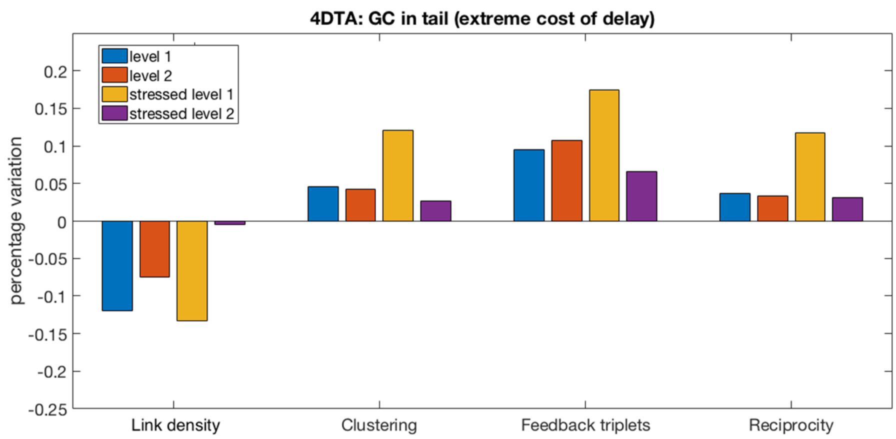


Figure 31. Percentage variation of the selected metrics of Granger causality in tail 4DTA, i.e., mean degree, clustering coefficient, number of feedback triplets, and reciprocity coefficient, of the Granger causality in tail network but built with the state of cost of delay when we compare the scenario with 4DTA mechanism with the corresponding baseline (for both default and stressed case). (In order to compare metrics between two different scenarios, we consider the measured value of each metric but normalised by the expected value for the random case. This is done for a fair comparison, because link density may vary from one scenario to another).

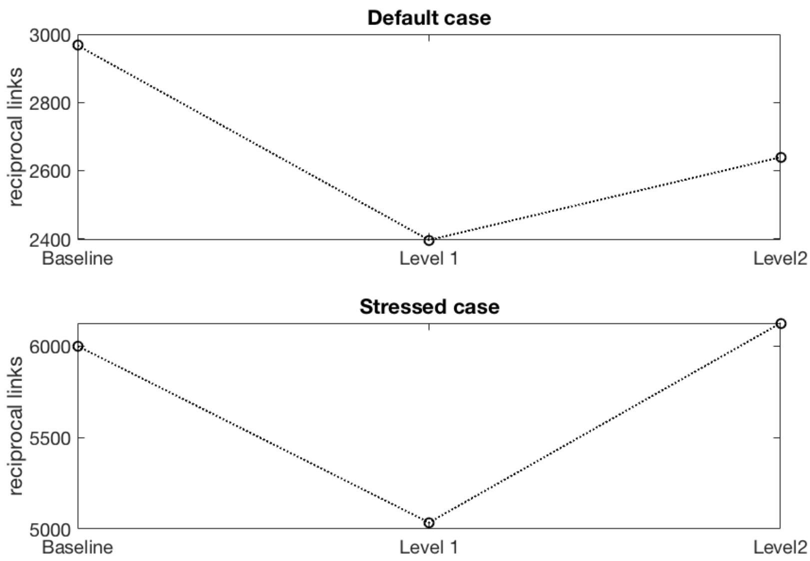


Figure 32. Number of reciprocal links at the different levels of implementation 4DTA, for both the default and stressed case. (GC in tail network built with the state of 'extreme' cost of delay)

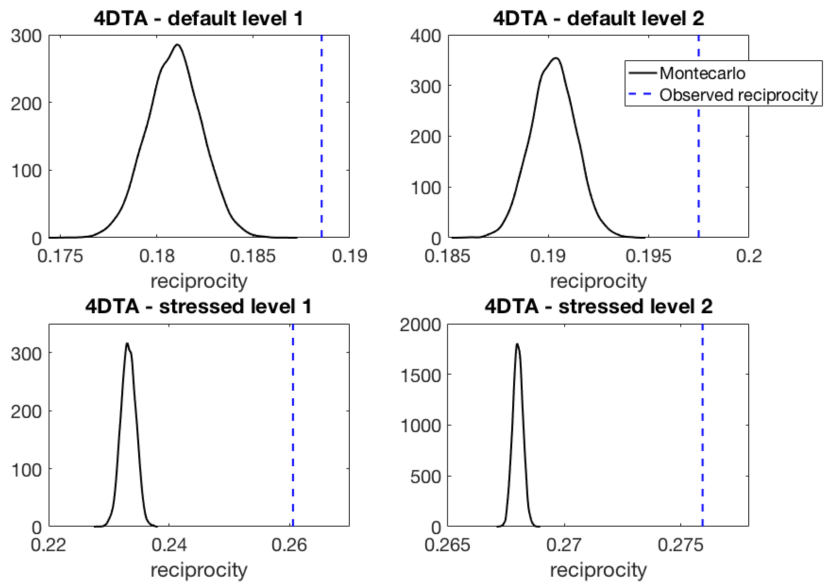


Figure 33. Distribution of reciprocity for 4DTA cost of delay according to the Monte Carlo simulations compared with the observed reciprocity for the 4DTA scenarios in the case of Granger causality in tail networks built with the state of 'extreme' costs of delay.

Figure 28 and Figure 31 show the percentage variations of the network metrics for the scenarios where 4DTA is implemented with respect to their baseline values shown in Table 19 and Table 20. For the propagation of extreme delays among the airports (also referred as the states of congestion of airports), we notice an increase of the level of causality (with the exception a small decrease in the default case at Level 2) measured with the increase of the link density. This result seems to suggest a negative impact of 4DTA. However, notice that not all delays have the same impact on costs. Interestingly, we observe an overall decrease of the level of causality in the networks built with the states of extreme costs of delay. Hence, some 'extreme' delays propagate more through the network, but the corresponding channels are less important when looking at the cost of delay, whose propagation is in turn reduced. For this reason, the 4DTA mechanism represents an overall improvement for the system in terms of costs, at the expense of an increased number of propagation channels for some less important delays. This is consistent with the dynamic cost index implemented at the AOC level, which weights the flight delays in terms of costs, e.g., reducing the propagation of costs between connecting flights (because of connecting passengers) and, possibly, increasing delays to reduce the fuel expense when there is not an significant impact for connections.

In all cases considered in Figure 28 and Figure 31 for both delays and costs, we observe a negative correlation between link density and the over-expression of any other network metric with respect to the random case, i.e., clustering, feedback triplets, and reciprocity. For simplicity, let us focus on reciprocity: in fact, reciprocal links represent the subsystems most simple to analyse. In Figure 29 and Figure 32, we can notice that the number of reciprocal links in the Granger causality in tail networks, for both delays and costs, is in effect correlated with the variation of the link density. Hence, the 4DTA mechanism reduces also the number of subsystems amplifying costs (see Figure 32). However, its over-expression increases. To better understand this behaviour, let us consider this simple Monte Carlo simulation: (i) given the network for the baseline scenario, and (ii) observed a given decrease of link density for one scenario with implemented 4DTA, (iii) we can randomly erase some links from the baseline network in order to target the same number of links of the network for the 4DTA scenario, and, finally, (iv) measure the network metric under investigation (when link density increases from the baseline to the 4DTA scenario, we simply add links, instead of removing them). The result of this simple Monte Carlo experiment for reciprocity is shown in Figure 30 and Figure 33 for the Granger causality in tail networks for both delays and costs. Notice that the expected measure of reciprocity according to Monte Carlo simulations is always smaller than the observed one. This suggests that: (i) the overall number of reciprocal links decreases or increases by following the variation in the total number of links, but (ii) the over-expression of reciprocity is inversely correlated with link density because (iii) some reciprocal subsystems are unaffected from the implemented mechanism.

Comments

The causality analysis for the propagation in mean of both delays and costs suggests that the 4DTA mechanism has the effect of decreasing the mean level of causality by decreasing the number of propagation channels. When focusing on the propagation of 'extreme events', we notice an increased level of causality for delays, but corresponding to an improvement for costs characterised by a decreasing number of propagation channels. Thus, 4DTA represents an overall improvement for the system in terms of costs, but some subgraphs representing amplifying subsystems for cost (and delay) propagation are unaffected from the mechanism.

4.3.2 Flight Prioritisation

4.3.2.1 Flights delays

Table 21, Table 22, Table 23 and Table 24 report the delay statistics for the scenarios where FP is implemented and for the corresponding baselines. When analysing all flights, FP brings no statistically significant changes in the average delays, and no clear effect is seen on the tails either. Such results suggest that almost no consequence is seen at the whole system level, which is understandable given that FP is used by a very small number of flights (1-2%). To check for a possible local effect of FP, we restrict our analysis to the set of flights whose destination airport was subject to a regulation at their scheduled in-block time. For this analysis, we use a new set of model simulations in which the set of regulations applied is the same across iterations and across scenarios (although it differs between the default and the stressed case). The results, shown in Table 25, Table 26, Table 27 and Table 28 are averaged over 50 iterations. Concerning delays, however, even after focusing on the flights subject to a regulation, no effect is seen.

Table 21. Departure delay statistics for the FP scenarios where it is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean departure delay (min)	**10.51**	**10.49**	**10.55**	**25.86**	**25.87**	**25.92**
	10.11	10.12	10.17	25.17	25.14	25.3
	10.69	10.71	10.83	26.27	26.53	26.51
#flights with dep.del>15 min	**6143.88**	**6154.19**	**6173.72**	**13560.2**	**13646.55**	**13655.75**
	6037.5	6054	6102.5	13411	13498	13478
	6239	6270	6253.5	13700.5	13799.5	13810
Total dep delay >15 min	**200178.08**	**199601.84**	**201382.87**	**614651.32**	**614711.9**	**616141.35**
	189484.14	189984.8	191497.07	596452.79	595999.28	598992.81
	204517.78	206610.88	208518.29	626200.08	632825.28	631284.3
Mean delay of flights with dep. del. >15 min	**32.58**	**32.43**	**32.62**	**45.33**	**45.05**	**45.12**

#flights with dep. del >60 min	**450.61**	**446.79**	**456.43**	**3013.1**	**2998.63**	**2999.76**
	366.5	377.5	386	2834.5	2867.5	2862.5
	490	502.5	524	3124	3158.5	3149.5
Total dep. delay >60 min	**40172.87**	**39382.64**	**40683.73**	**274106.64**	**271114.74**	**272446.6**
	29543.56	31006.23	30817	255304	252138.93	254973.72
	43460.13	43587.29	47498.69	285799.53	287599.26	287404.75
Mean delay of flights with dep. del. >60 min	**89.15**	**88.15**	**89.13**	**90.97**	**90.41**	**90.82**
#flights with dep.del>180 min	**18.81**	**17.2**	**18.18**	**83.39**	**76.34**	**83.35**
	4	3	4	51.5	52	54.5
	12.5	11.5	18.5	94	92	98
Total dep. delay >180 min	**4161.89**	**3838.37**	**4031.85**	**18016.96**	**16375.62**	**17891.12**
	777.3	650.36	792.59	11078.22	10966.32	11430.03
	2734.54	2709.86	3994.83	20037.11	19648.26	20792.65
Mean delay of flights with dep. del. >180 min	**221.26**	**223.16**	**221.77**	**216.06**	**214.51**	**214.65**

Table 22. Arrival delay statistics for FP scenarios where it is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean arrival delay (min)	**5.18**	**5.15**	**5.23**	**29.09**	**29.1**	**29.14**
	4.81	4.79	4.85	28.43	28.38	28.52
	5.35	5.37	5.52	29.56	29.78	29.76
Mean arrival delay of delayed flights (min)	**10.49**	**10.46**	**10.54**	**30.87**	**30.86**	**30.9**
	10.1	10.11	10.16	30.22	30.16	30.28
	10.64	10.7	10.81	31.33	31.55	31.49
#flights with arr.del>15 min	**6614.13**	**6611.44**	**6642.49**	**15930.43**	**15993.58**	**16007.64**
	6504	6520	6570.5	15800	15887	15880
	6674	6705.5	6735.5	16035	16118.5	16145.5
Total arr. delay >15 min	**220541.86**	**219680.99**	**221843.79**	**758486.18**	**758686.47**	**759957.83**
	210083.98	210524.58	212417.21	740924.41	739959.42	743231.18
	225008.08	226853.11	229015.14	770880.74	777320.38	775629.12
Mean delay of flights with arr. del. >15 min	**33.34**	**33.23**	**33.4**	**47.61**	**47.44**	**47.47**
#flights with arr. del >60 min	**523.45**	**517.44**	**532.41**	**3944.73**	**3936.19**	**3935.61**
	440.5	445	458.5	3797.5	3792	3798.5
	565.5	568.5	588.5	4067.5	4097.5	4064.5

Total arr. delay >60 min	**45928.83**	**45055.79**	**46713.96**	**360514.68**	**358015.32**	**359009.66**
	35987.14	36030.45	36857.53	340020.67	340589.68	339663.08
	50130.78	49249.83	52775.28	372987.32	373084.63	373454.92
Mean delay of flights with arr. del. >60 min	**87.74**	**87.07**	**87.74**	**91.39**	**90.95**	**91.22**
#flights with arr. del. >180 min	**19.51**	**18.42**	**19.42**	**108.15**	**100.32**	**106.61**
	4.5	4	5	72	71.5	74
	14.5	13	19.5	124.5	122.5	120
Total arr. delay >180 min	**4323.19**	**4104.25**	**4286.89**	**23307.19**	**21480.47**	**22891.87**
	989.74	935.85	1055.41	15350.88	15250.7	15816.34
	3228.24	2960.51	4231.14	26928.5	26278.78	25290.83
Mean delay of flights with arr. del. >180 min	**221.59**	**222.81**	**220.75**	**215.51**	**214.12**	**214.73**

Table 23. Gate-to-gate delay statistics for FP scenarios where it is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean gate-to-gate delay (min)	**-5.33**	**-5.34**	**-5.32**	**3.25**	**3.25**	**3.24**
	-5.36	-5.37	-5.36	3.2	3.2	3.19
	-5.28	-5.31	-5.29	3.3	3.29	3.27

Table 24. Cancellations and reactionary delay statistics for FP scenarios and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
# Cancelled flights	**1087.95**	**1092.14**	**1087.83**	**1160.20**	**1146.14**	**1144.46**
	1026.50	1039.50	1037.00	1096.50	1087.00	1100.00
	1136.00	1127.00	1130.50	1185.00	1194.00	1182.00
Mean reactionary delay (min)	**17.72**	**17.76**	**17.88**	**38.44**	**38.08**	**38.24**
	15.11	15.05	15.08	35.07	34.50	35.29
	16.90	17.37	17.58	39.07	38.09	37.57
# Flights with reactionary delay	**2888.71**	**2893.68**	**2896.96**	**7197.35**	**7217.55**	**7223.23**
	2823.50	2835.50	2850.00	7114.00	7121.50	7140.50
	2947.00	2946.50	2946.50	7268.50	7320.00	7303.50

Table 25. Departure delay statistics for the FP scenarios restricted sample to flights subjects to regulations.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean departure delay (min)	**26.19**	**26.10**	**26.31**	**42.26**	**42.30**	**42.05**
	25.59	25.49	25.54	41.67	41.73	41.63
	26.73	26.50	26.92	42.94	43.09	42.46
Mean per-pax dep. delay (min)	**2.39E-01**	**2.39E-01**	**2.40E-01**	**4.09E-01**	**4.09E-01**	**4.07E-01**
	2.34E-01	2.32E-01	2.33E-01	4.03E-01	4.02E-01	4.01E-01

	2.44E-01	2.45E-01	2.48E-01	4.16E-01	4.14E-01	4.12E-01
#flights with dep.del>15 min	**249.16**	**247.22**	**249.22**	**941.86**	**942.82**	**943.02**
	243.00	241.00	243.00	928.00	935.00	933.00
	255.00	254.00	256.00	955.00	953.00	956.00
Total dep delay >15 min	**13734.43**	**13691.22**	**13800.10**	**56372.64**	**56502.30**	**56158.41**
	13435.05	13311.17	13282.95	55550.95	55408.98	55375.03
	14022.87	13988.40	14245.53	57570.93	57588.33	57053.97
Mean delay of flights with dep. del. >15 min	**55.12**	**55.38**	**55.37**	**59.85**	**59.93**	**59.55**
#flights with dep. del >60 min	**85.04**	**84.02**	**85.10**	**363.64**	**362.20**	**357.88**
	82.00	81.00	83.00	352.00	350.00	351.00
	89.00	87.00	89.00	375.00	374.00	366.00
Total dep. delay >60 min	**8,609.26**	**8,591.46**	**8,703.75**	**36,393.77**	**36,430.72**	**35,920.20**
	8198.12	8103.05	8250.98	34937.05	35064.10	35064.17
	8956.12	8948.52	9136.73	37509.00	37783.82	36884.15
Mean delay of flights with dep. del. >60 min	**101.24**	**102.25**	**102.28**	**100.08**	**100.58**	**100.37**
#flights with dep.del>180 min	**4.34**	**4.74**	**4.78**	**19.24**	**20.40**	**20.34**
	3.00	3.00	3.00	16.00	16.00	17.00
	5.00	6.00	7.00	22.00	23.00	23.00

Total dep. delay >180 min	**916.29**	**998.96**	**1000.95**	**4140.82**	**4406.13**	**4389.06**
	654.92	619.03	590.83	3512.12	3457.98	3759.63
	1144.32	1299.15	1446.68	4707.93	5066.38	5096.18
Mean delay of flights with dep. del. >180 min	**211.13**	**210.75**	**209.40**	**215.22**	**215.99**	**215.78**

Table 26. Arrival delay statistics for FP scenarios restricted sample to flights subjects to regulations.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean arrival delay (min)	**22.24**	**22.25**	**22.48**	**45.82**	**45.89**	**45.68**
	21.69	21.57	21.55	45.34	45.29	45.22
	22.76	23.01	23.42	46.43	46.62	46.08
Mean arrival delay of delayed flights (min)	**24.65**	**24.64**	**24.85**	**46.74**	**46.78**	**46.59**
	24.15	24.08	23.96	46.21	46.12	46.2
	25.11	25.21	25.8	47.4	47.45	47.05
#flights with arr.del>15 min	**248.06**	**246.8**	**249.6**	**1047.46**	**1052.4**	**1052.14**
	243	240	243	1038	1042	1039
	253	254	259	1061	1064	1065
Total arr. delay >15 min	**13483.15**	**13460.99**	**13603.39**	**64042.66**	**64204.65**	**63907.36**
	13165.98	13057.48	12988.93	63052.1	63077.72	63033.47
	13647.5	13794.62	14191.92	65022.37	65274.38	64723.28

Mean delay of flights with arr. del. >15 min	**54.35**	**54.54**	**54.5**	**61.14**	**61.01**	**60.74**
#flights with arr. del >60 min	**81.58**	**80.8**	**81.64**	**412.28**	**410.8**	**409.28**
	78	77	77	400	399	401
	85	83	85	421	421	418
Total arr. delay >60 min	**8298.4**	**8274.97**	**8384.1**	**42018.24**	**42005.81**	**41620.16**
	7848.45	7851.65	7894.52	40748.08	40350.03	40807.85
	8635.57	8598.92	8866.77	43216.88	43565.68	42474.32
Mean delay of flights with arr. del. >60 min	**101.72**	**102.41**	**102.7**	**101.92**	**102.25**	**101.69**
#flights with arr. del. >180 min	**4**	**4.48**	**4.48**	**26.4**	**27.06**	**27.36**
	3	3	3	23	23	25
	5	6	6	29	31	30
Total arr. delay >180 min	**847.2**	**936.96**	**932.56**	**5662.99**	**5850.5**	**5895.19**
	623.6	594.15	591.68	4872.6	4888.2	5310.8
	1056.77	1162.57	1234.95	6232.88	6711.03	6587.98
Mean delay of flights with arr. del. >180 min	**211.8**	**209.14**	**208.16**	**214.51**	**216.2**	**215.47**

Table 27. Gate-to-gate delay statistics for FP scenarios restricted sample to flights subjects to regulations

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean gate-to-gate delay (min)	**-3.95**	**-3.85**	**-3.83**	**3.61**	**3.64**	**3.68**
	-4.22	-4.04	-4.1	3.49	3.46	3.44
	-3.67	-3.63	-3.53	3.79	3.87	3.88

Table 28. Cancellation and reactionary delay statistics for FP scenarios restricted sample to flights subjects to regulations.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
# cancelled flights	**20.94**	**20.62**	**21.10**	**70.76**	**68.64**	**68.54**
	17.00	17.00	18.00	65.00	63.00	63.00
	24.00	24.00	25.00	77.00	74.00	74.00
Mean reactionary delay (min)	**31.82**	**32.42**	**31.34**	**58.79**	**58.69**	**58.31**
	29.28	29.34	29.09	56.90	57.65	56.83
	34.40	35.23	33.63	60.29	60.64	59.26
# flights with reactionary delay	**53.18**	**50.06**	**52.42**	**338.60**	**338.42**	**338.20**
	49.00	46.00	49.00	331.00	327.00	330.00
	57.00	55.00	56.00	346.00	348.00	346.00

4.3.2.2 Passengers delays

In this section we show the same metrics presented in Section 4.3.2.1, with the FP mechanism. These results have also been obtained using 50 iterations.

On the non-stressed scenarios, the implementation of FP seems to have a small increasing effect on the average delay (positive or not). The number of passengers delayed in the various threshold is also reasonably constant. In the stressed scenarios, if the averages seem to decrease slightly, it seems also that the passengers involved in high delays are decreasing in number, especially for the very high delays (>180 minutes) (see Table 29)

Table 29. Passenger indicators statistics for FP mechanism.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean delay	17.2	17.6	17.3	40	39.7	39.4
25% perc. delay	-9.4	-9.4	-9.3	4.7	4.6	4.5
75% perc. delay	15.2	15.2	15.3	44.8	44.5	44.3
Positive mean delay	23.6	24.1	23.8	42.4	42	41.8
Positive 25% perc. delay	0	0	0	4.7	4.6	4.5
Positive 75% perc. delay	15.2	15.2	15.3	44.8	44.5	44.3
Number of pax with delay>15	859784.1	862912.6	865899.5	2005583.6	1998571.9	1994573.5
Total delays with delay>15	74470449.7	75906819.2	74995551.2	138623108	137452785.9	136653497.2
Mean delay with delay>15	86.6	88	86.6	69.1	68.8	68.5
Number of pax with delay>60	188560.8	192384.2	191231.6	561065.1	555273.8	548116.1
Total delays with delay>60	55486051.8	56958103	55917776.3	91227358.5	90093683.9	89200936

Mean delay with delay>60	294.3	296.1	292.4	162.6	162.3	162.7
Number of pax with delay>180	142665.4	146413.4	143547	158099.6	155659	154764
Total delays with delay>180	51765630	53213325.9	51992217.4	56374057.5	55590581.8	55315694.8
Mean delay with delay>180	362.8	363.4	362.2	356.6	357.1	357.4

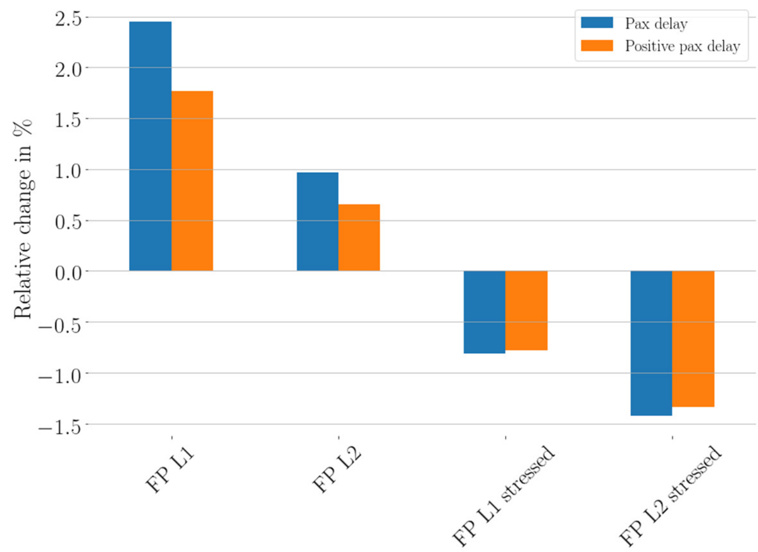


Figure 34. Change in passenger delay for FP: the mean delay (blue) and the mean positive delay (orange) with respect to baseline scenarios (default on the left, stressed on the right).

In Figure 34, we show the small changes in the average delays. Interestingly, when compared to the 4DTA case, the average on the positive delays is quite close to the non-positive one. This indicates that FP, on the contrary of 4DTA, does not impact the negative delays so much. This is due to the fact that FP is designed primarily to avoid big delays (in fact, high cost) before departure. Moreover, even if the variations are small, it is surprising to see that FP seems to have a benefit for passengers only in the stressed case. This may indicate that the airline interest (profit) is only aligned with the passengers' (utility) when the delays are very high.

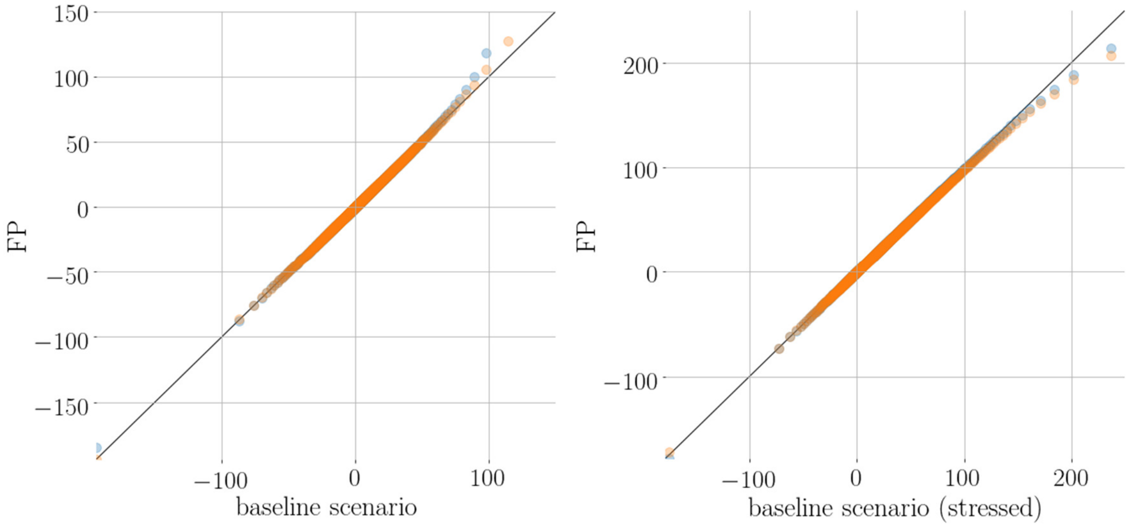


Figure 35. QQ-plot of arrival passenger delay in FP in the baseline scenario (left) and stressed scenario (right). The blue circles correspond to the FP at Level 1, the orange ones to Level 2. The plots have been cut at 150 and 200 minutes respectively.

The QQ-plots in Figure 35 confirm the trends. In the non-stressed scenario, the right tail of the distribution seems to fatten with the implementation of FP (and more with Level 1 than Level 2). In the stressed scenarios the effect is the opposite, and the right tail gets squeezed, demonstrating an improvement of the situation for the very high delays. The effect on the average is thus mainly due to the effect on the right tail.

In summary, the effect of the implementation of FP is the following:

- for non-stressed scenarios, the implementation of FP increases slightly the average delay, and more at Level 1 than Level 2. This is due mainly to the fattening of the right side of the distribution of delay.
- for stressed scenario, the implementation of FP decreases even more slightly the average delay, and more at Level 2 than Level 1. This is due mainly to the squeeze of the right side of the distribution.

4.3.2.3 Costs

In this section we present the impact of the FP mechanism on the different kinds of costs airlines incur because of both excess fuel and delay.

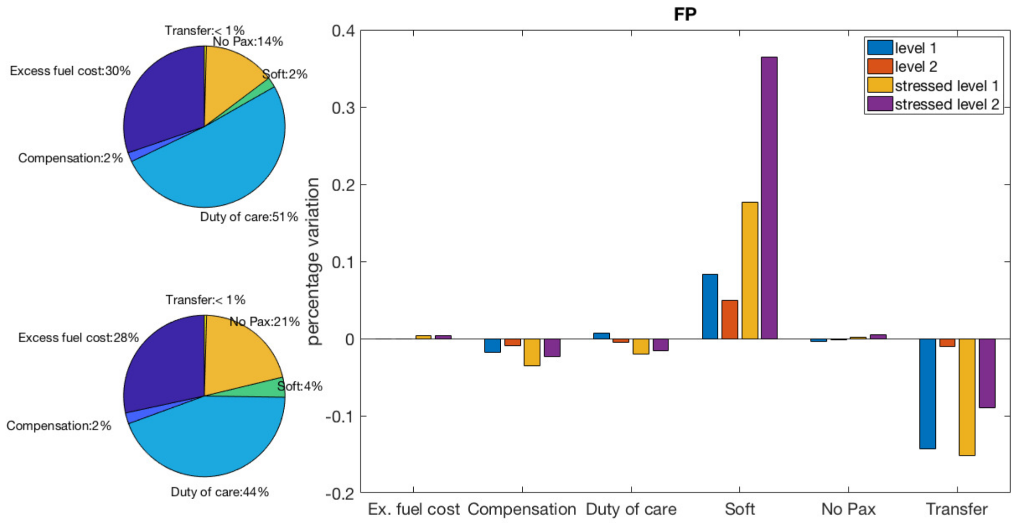


Figure 36. Cost of delay for FP. Left: Percentage of the total cost represented by each type of cost considered in the analysis (upper pie: default, lower pie: stressed); right: Percentage variations of the different kinds of costs with respect to the corresponding baseline scenario. The plots are obtained with 100 iterations of the ABM with regulations sampled randomly from historical observations.

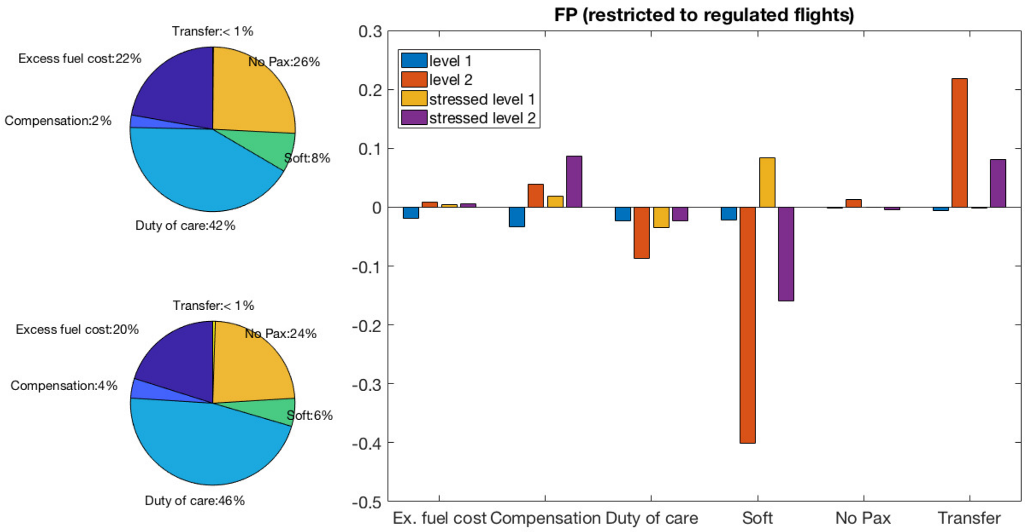


Figure 37. Cost of delay for FP restricted regulated flights. Left: Percentage of the total cost represented by each type of cost considered in the analysis (upper pie: default, lower pie: stressed); right: Percentage variations of the different kinds of costs with respect to the corresponding baseline scenario. The plots are obtained with 50 iterations of the ABM with fixed regulation.

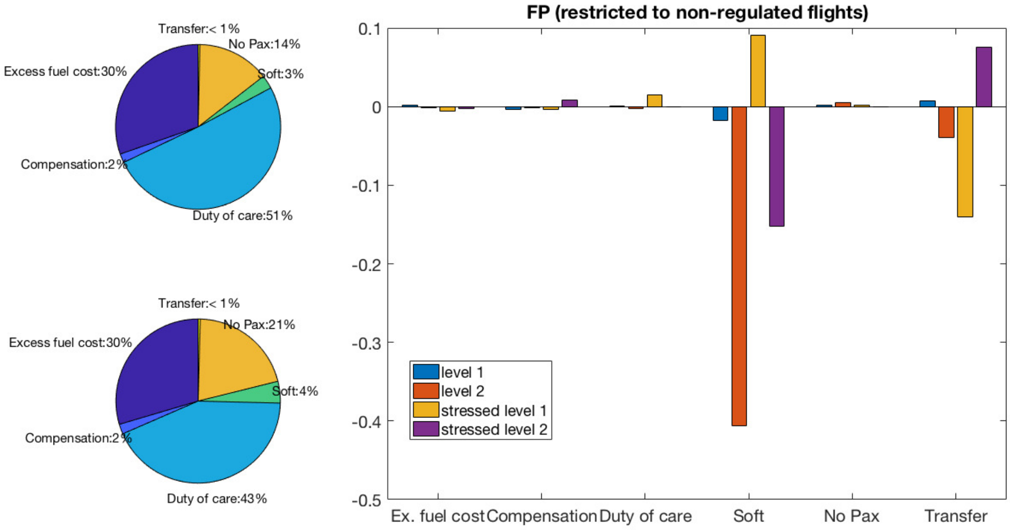


Figure 38. Cost of delay for FP restricted non-regulated flights. Left: Percentage of the total cost represented by each type of cost considered in the analysis (upper pie: default, lower pie: stressed); right: Percentage variations of the different kinds of costs with respect to the corresponding baseline scenario. The plots are obtained with 50 iterations of the ABM with fixed regulation.

Figure 36, Figure 37 and Figure 38 summarise the results, in particular Figure 36 is obtained by considering 100 iterations of the ABM model where regulations are randomly sampled from historical observations, whereas Figures 2 and 3 are obtained with 50 iterations of the ABM model, but with a fixed regulation. The choice of analysing 50 iterations with fixed regulation is to focus on the systemic impact of the FP mechanism on a fixed set of flights in order to exclude spurious effects arising from comparing different samples. The results show that the FP mechanism has no systemic impact overall. In fact, percentage changes in the value of costs from baselines to scenarios with FP are very small, with the exceptions of soft and transfer costs. However, note that these two kinds of costs represent the smallest percentages of the total cost. Moreover, they represent also the smallest costs in absolute value. Thus, the larger percentage changes with respect other costs can be interpreted at all effects as random fluctuations.

FP: detailed analysis of costs

Table 13 and Table 14 show the detailed values obtained by considering the flights simulated by the ABM. The samples of flights over which averages are computed are built as explained in the captions. In particular, Table 14 show the values associated to the different types of costs without considering in the analysis the flights which have been cancelled. This is to better capture the impact of the 4DTA mechanism on normal operations. However, in both cases, the results are qualitatively the same.

Table 30. Detailed costs by considering all flights FP (with the exception of flights with delays larger than 300 min).

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Excess cost of fuel	**136.0**	**135.9**	**136.0**	**182.9**	**183.7**	**183.6**
	135.0	134.6	134.6	181.2	182.1	181.1
	137.1	137.3	137.3	184.7	185.4	185.3
Cost of compensation	**8.3**	**8.1**	**8.2**	**14.1**	**13.6**	**13.8**
	7.3	7.3	7.3	12.8	12.6	12.5
	9.1	8.7	8.7	14.7	14.5	14.7
Fraction of compensations	**1.1%**	**1.1%**	**1.1%**	**2.7%**	**2.7%**	**2.7%**
Costs of 'transfer'	**1.8**	**1.6**	**1.8**	**2.9**	**2.5**	**2.6**
	1.0	1.0	0.9	2.1	1.8	1.9
	2.4	1.9	2.3	3.5	2.9	3.2
Fraction of transfer costs	**0.1%**	**0.1%**	**0.1%**	**0.3%**	**0.3%**	**0.3%**
Duty of care	**229.8**	**231.4**	**228.5**	**284.9**	**279.2**	**280.4**
	219.8	221.1	219.2	270.0	268.0	266.9
	237.2	237.5	234.4	290.0	290.4	288.7
Fraction of 'duty of care'	**13.5%**	**13.6%**	**13.5%**	**20.0%**	**19.7%**	**19.8%**
Soft costs	**9.1**	**9.9**	**9.5**	**25.9**	**30.5**	**35.3**
	3.6	3.6	3.6	10.3	10.5	39.8
	14.0	14.1	14.2	40.9	41.1	41.2
Fraction of soft costs	**46.4%**	**46.4%**	**46.4%**	**57.5%**	**57.5%**	**57.5%**
Non passenger costs	**64.4**	**64.1**	**64.3**	**134.0**	**134.2**	**134.6**

	62.3	62.2	62.7	129.9	130.6	130.9
	64.9	64.6	65.1	136.9	137.8	138.1
Fraction of non passenger costs	**91.1%**	**91.1%**	**91.1%**	**93.2%**	**93.3%**	**93.3%**
Total excess cost	**449.5**	**451.0**	**448.5**	**644.8**	**643.7**	**650.5**
	436.6	435.9	438.1	619.8	627.3	635.9
	457.9	459.7	454.6	658.3	654.4	665.9

Table 31. Detailed costs FP considering all regulated flights (with the exception of flights with delays larger than 300 min).

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Excess cost of fuel	**135.8**	**133.2**	**136.9**	**192.2**	**193.1**	**193.2**
	128.8	127.7	127.6	186.3	187.1	188.4
	140.9	137.8	146.6	196.7	200.7	199.2
Cost of compensation	**14.9**	**14.4**	**15.5**	**37.1**	**37.8**	**40.3**
	12.0	11.4	11.1	33.4	33.5	35.2
	17.9	17.8	19.4	40.5	42.9	44.5
Fraction of compensations	**5.3%**	**5.1%**	**5.1%**	**6.8%**	**6.8%**	**7.0%**
Costs of 'transfer'	**0.9**	**0.9**	**1.1**	**4.9**	**4.8**	**5.2**
	0.2	0.1	0	3.0	3.4	3.7
	1.5	1.1	1.3	5.8	6.8	6.7
Fraction of transfer costs	**0.2%**	**0.2%**	**0.2%**	**0.7%**	**0.7%**	**0.7%**
Duty of care	**256.0**	**250.1**	**233.6**	**443.5**	**428.3**	**432.9**

	222.4	221.3	210.1	407.7	398.9	408.5
	282.8	287.8	255.3	463.8	453.3	459.2
Fraction of 'duty of care'	**21.3%**	**21.2%**	**20.9%**	**33.9%**	**33.9%**	**34.3%**
Soft costs	**47.1**	**46.0**	**28.2**	**52.7**	**57.2**	**44.3**
	51.1	49.1	13.2	19.2	19.4	35.2
	55.1	54.9	52.3	77.3	77.9	44.5
Fraction of soft costs	**61.1%**	**61.6%**	**61.5%**	**66.8%**	**66.9%**	**67.0%**
Non passenger costs	**156.7**	**156.6**	**158.8**	**224.8**	**224.6**	**223.7**
	154.3	153.2	155.1	219.7	220.4	220.4
	160.8	160.7	162.2	229.4	229.0	226.6
Fraction of non passenger costs	**95.1%**	**95.1%**	**95.0%**	**94.9%**	**95.0%**	**95.1%**
Total excess cost	**611.4**	**601.2**	**573.1**	**955.3**	**945.9**	**939.8**
	567.6	563.7	542.3	923.5	908.4	893.9
	645.1	642.3	595.5	986.7	984.4	987.9

Table 32. Detailed costs FP considering flights with delays<300 min excluding cancelled.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Excess cost of fuel	**141.6**	**141.4**	**141.5**	**190.8**	**191.6**	**191.5**
	140.3	140.2	140.2	188.9	189.5	188.8
	142.6	143.0	143.0	192.8	193.5	193.5

Cost of compensation	**1.5**	**1.5**	**1.5**	**6.9**	**6.7**	**6.7**
	1.2	1.2	1.2	6.2	6.0	6.0
	1.5	1.6	1.6	7.4	7.4	7.2
Fraction of compensations	**0.6%**	**0.6%**	**0.6%**	**2.1%**	**2.1%**	**2.1%**
Costs of 'transfer'	**0.3**	0.3	0.3	**1.3**	1.1	1.1
	0.2	0.2	0.2	1.1	1.0	0.9
	0.4	0.4	0.4	1.5	1.3	1.3
Fraction of transfer costs	**5E-2%**	**5E-2%**	**5E-2%**	**0.2%**	**0.2%**	**0.2%**
Duty of care	**39.7**	39.3	39.1	**75.9**	73.6	74.7
	35	34.9	35.1	68.0	66.8	67.5
	41	41.1	40.6	80.1	79.6	80.2
Fraction of 'duty of care'	**11.8%**	**11.8%**	**11.7%**	**18.2%**	**18.0%**	**18.0%**
Soft costs	**7.3**	7.9	7.7	**24.6**	29.0	33.6
	2.9	2.9	3.0	9.8	10.0	37.8
	11.1	11.2	11.3	38.8	39.0	39.2
Fraction of soft costs	**46.1%**	**46.1%**	**46.1%**	**57.5%**	**57.5%**	**57.5%**
Non passenger costs	**67.0**	**66.7**	**66.9**	**139.8**	**139.9**	**140.4**
	64.8	64.6	65.2	135.4	136.2	136.3
	67.4	67.2	67.8	142.7	143.6	144.1
Fraction of non passenger costs	**94.8%**	**94.8%**	**94.6%**	**97.2%**	**97.2%**	**97.3%**

Total excess cost	**257.6**	**257.3**	**257.2**	**439.5**	**442.1**	**448.1**
	248.7	250.1	249.7	420.4	429.0	437.1
	262.9	260.6	261.3	454.7	453.9	459.4

Table 33. Detailed costs FP considering all regulated flights excluding cancelled (with the exception of flights with delays larger than 300 min).

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Excess cost of fuel	**140.5**	**137.7**	**141.7**	**201.7**	**202.3**	**202.4**
	133.1	132.4	131.2	195.3	195.3	197.6
	146.4	143.3	151.2	207.4	209.9	209.1
Cost of compensation	**11.5**	**10.7**	**13.2**	**32.2**	**33.1**	**35.2**
	9.5	8.6	9.0	27.7	29.0	29.7
	13.4	12.3	16.5	35.8	37.8	40.0
Fraction of compensations	**5.1%**	**4.8%**	**4.9%**	**6.3%**	**6.4%**	**6.6%**
Costs of 'transfer'	**0.6**	**0.7**	**0.9**	**3.6**	**3.7**	**4.3**
	0.0	0.0	0.0	2.2	2.2	2.8
	0.9	1.1	1.3	4.6	5.3	5.1
Fraction of transfer costs	**0.2%**	**0.2%**	**0.2%**	**0.6%**	**0.6%**	**0.6%**
Duty of care	**89.0**	**87.7**	**160.9**	**203.7**	**204.6**	**206.5**
	78.6	74.4	82.4	193.2	191.2	197.4
	98.4	98.2	233.4	216.5	214.2	214.8
Fraction of 'duty of care'	**19.9%**	**19.8%**	**20.1%**	**32.5%**	**32.5%**	**33.0%**

Soft costs	**45.7**	**44.7**	**27.7**	**52.1**	**56.7**	**43.8**
	48.9	47.4	12.9	19.1	19.3	18.6
	53.2	53.1	52.7	76.2	77.1	75.5
Fraction of soft costs	**61.0%**	**61.6%**	**61.6%**	**67.1%**	**67.3%**	**67.3%**
Non passenger costs	**162.2**	**161.8**	**161.2**	**235.8**	**235.3**	**234.4**
	159.3	158.6	155.2	231.1	230.5	231.2
	166.1	165.6	166.1	239.5	240.2	236.9
Fraction of non passenger costs	**98.3%**	**98.4%**	**96.7%**	**99.6%**	**99.6%**	**99.6%**
Total excess cost	**449.6**	**443.4**	**434.1**	**729.2**	**735.9**	**726.8**
	431.8	421.5	407.2	695.8	704.7	699.9
	469.7	457.9	465.8	762.3	772.3	759.6

In Table 30, Table 31, Table 32 and Table 33 we show the detailed values of all kind of costs for all scenarios and both default and stressed case. Notice that there are no significant variations, especially in the distributions characterised by the interquartile range.

Comments

From the considered detailed analysis on the different kind of costs, we can conclude that the FP mechanism has no systemic impact on the costs flights face at the day of operations, since FP does not affect both excess fuel consumption and delays when looking at the aggregates.

4.3.2.4 Centrality losses

Passenger centrality

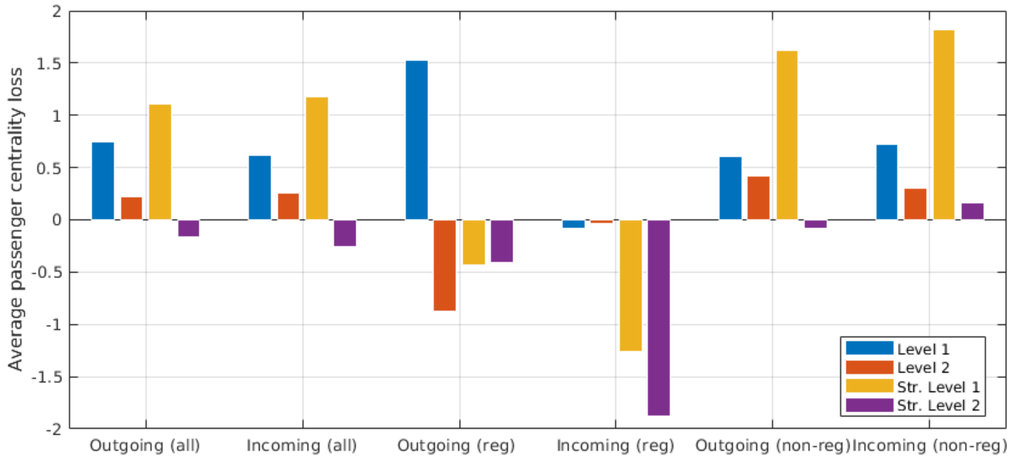


Figure 39. Percentage change in the average centrality loss in FP scenarios for the analysis on all airports, on the airports subject to a regulation and on the airports without.

Table 34 reports the passenger centrality loss statistics obtained averaging losses over all the airports. All losses are slightly increasing, except for Level 2 stressed, where they are decreasing very slightly, as can be also seen in Figure 39, showing the percentage change in the centrality loss with respect to the baseline. All changes are not statistically significant.

Table 34. Average passenger centrality loss for FP, incoming and outgoing. in the scenarios where FP is implemented at Level 1 and 2 and in the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Average incoming passenger centrality loss	**184.94**	**186.09**	**185.42**	**198.72**	**201.07**	**198.20**
	176.76	178.09	175.49	191.24	194.26	193.76
	192.20	194.19	194.93	206.83	209.25	206.19

Average outgoing passenger centrality loss	**204.33**	**205.85**	**204.78**	**228.38**	**230.91**	**228.01**
	195.92	196.86	194.34	220.24	224.11	222.84
	211.56	214.08	215.16	236.74	239.24	235.71

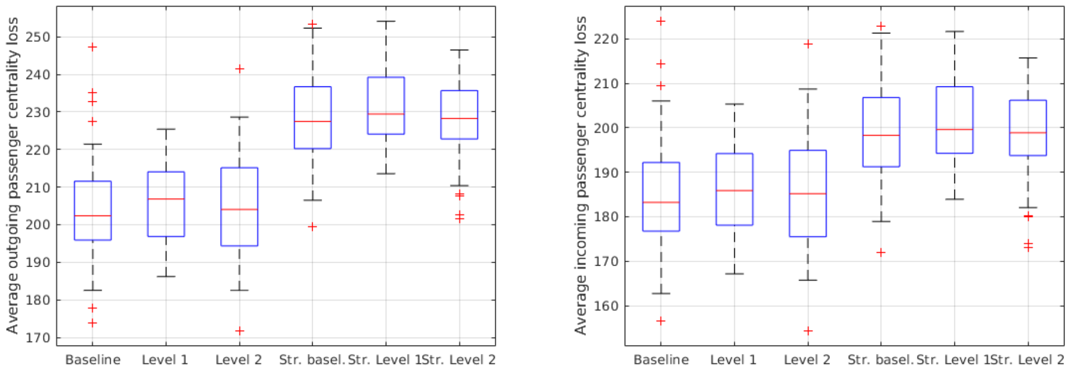


Figure 40. Average passenger centrality loss FP Box plots comparing the average outgoing (left) and incoming (right) passenger centrality loss in the scenarios where FP is implemented at Level 1 and 2 and in the corresponding baselines. For each scenario, the average centrality losses in each of the 100 iterations of the model are considered.

Figure 40 compares the average outgoing and incoming passenger centrality loss in the scenarios where FP is implemented at Level 1 and 2 and in the corresponding baselines.

We then restrict our analysis to the set of airports that were subject to a regulation, and compare it to the set of airport where no regulation was applied, in order to see if the first group has a more important improvement. The centrality losses are reported in Table 35 and the corresponding box-plots are shown in Figure 41, while the percentage change with respect to the baseline is shown in Figure 39. Although the differences are not statistically significant, we remark that in the stressed case airports that were subject to a regulation have on average smaller centrality losses when FP is implemented at Level 1 and 2 with respect to the baseline, while for airports that were not subject to regulations the losses are slightly larger or very similar. In the default case, no improvement is observed except for the outgoing centrality at Level 2. The fact that incoming centrality losses decrease for airports subject to a regulation suggests that flight swapping improves the connections at the departure airport of the swapped flights. If we make a comparison at a single airport level, in the stressed scenarios 60-70% of the 22 regulated airports have, on average, smaller passenger centrality losses than in the corresponding baseline, however the difference in the averages are not statistically significant for any of them. In the default scenarios, instead, such percentage is smaller than 35% for outgoing centrality at Level 1 and close to 50% in the other cases.

In conclusion, the FP mechanism causes a slight increase in the average centrality losses in the entire system, and only in the stressed case we see an improvement for the regulated airports, for which there is a percentage improvement which is not seen in the non-regulated one. However, this difference is not statistically significant.

Table 35. Average passenger centrality loss for FP for airports with regulations, incoming and outgoing, implemented at Level 1 and 2 and in the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Average incoming passenger centrality loss (regulated airp)	**1400.27**	**1399.12**	**1399.78**	**1565.44**	**1545.75**	**1536.12**
	1258.14	1300.64	1281.71	1433.50	1458.27	1435.36
	1487.07	1488.57	1547.64	1633.05	1637.27	1635.95
Average incoming passenger centrality loss (non-regulated airp)	**164.29**	**165.48**	**164.79**	**161.87**	**164.82**	**162.13**
	156.42	159.62	156.18	155.97	157.99	156.04
	171.33	172.20	172.65	168.10	170.42	168.97
Average outgoing passenger centrality loss (regulated airp)	**1877.33**	**1905.96**	**1860.92**	**2161.30**	**2151.92**	**2152.54**
	1767.79	1771.50	1688.93	2037.27	2055.95	2062.82
	2005.79	2040.21	2058.21	2257.36	2243.36	2249.41
Average outgoing passenger centrality loss (non-regulated airp)	**175.91**	**176.97**	**176.64**	**176.26**	**179.12**	**176.12**
	168.95	170.87	167.86	170.01	172.83	170.88

	183.15	184.00	185.27	183.94	184.54	182.05

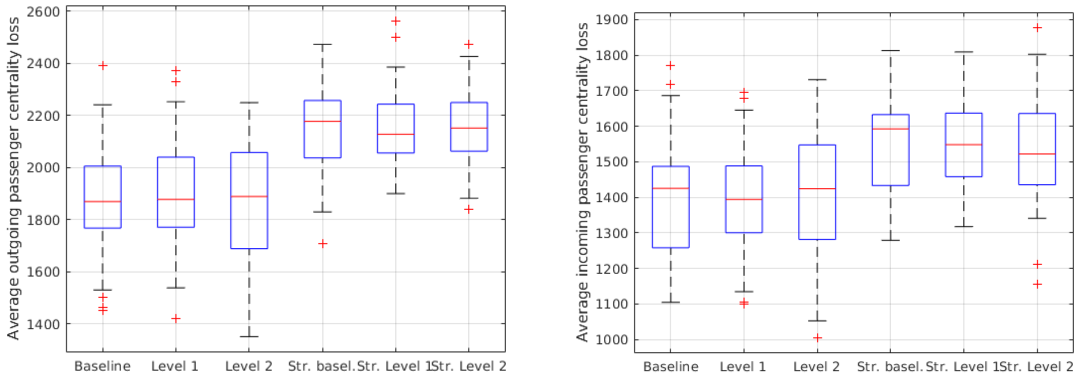


Figure 41. Average passenger centrality loss FP Box plots comparing the average outgoing (left) and incoming (right) passenger centrality loss for airports subject to a regulation in the scenarios where FP is implemented at Level 1 and 2 and in the corresponding baselines. For each scenario, the average centrality losses in each of the 100 iterations of the model are considered.

Figure 41 compares the average outgoing and incoming passenger centrality loss in the scenarios where FP is implemented at Level 1 and 2 and in the corresponding baselines on the restricted dataset (only considering flights which to airports with an ATFM regulation).

Trip centrality

Table 36 reports the trip centrality loss statistics obtained averaging losses over all the airports. All losses are slightly increasing, except for Level 2 default, where they are decreasing very slightly, as can be also seen in Figure 39, showing the percentage change in the centrality loss with respect to the baseline. All changes are not statistically significant.

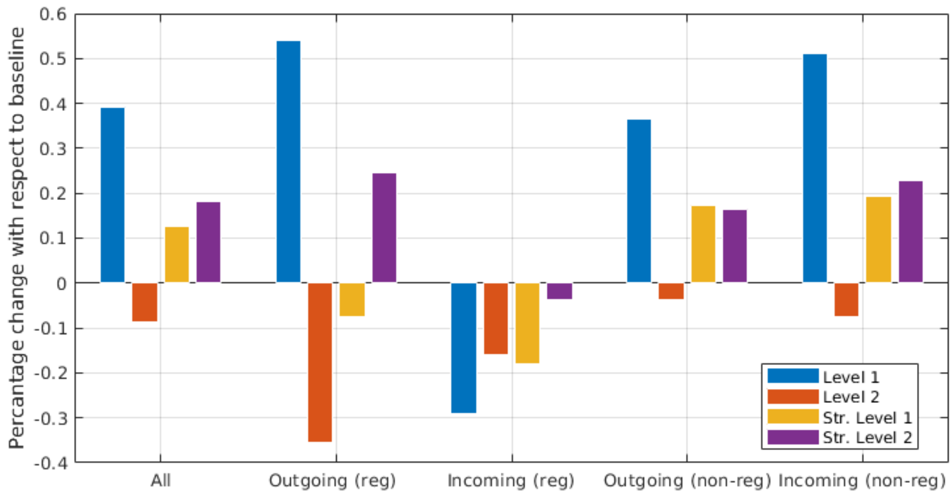


Figure 42. Percentage change in the average centrality loss in FP scenarios for the analysis on all airports, on the airports subject to a regulation and on the airports without.

Table 36. Average trip centrality loss FP at Level 1 and 2 and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Average trip centrality loss	**0.536**	**0.538**	**0.536**	**0.733**	**0.734**	**0.734**
	0.530	0.532	0.529	0.729	0.730	0.729
	0.543	0.545	0.542	0.738	0.736	0.740

We then restrict our analysis to the set of airports that were subject to a regulation, and compare it to the set of airports where no regulation was applied. The centrality losses are reported in Table 37, while the percentage change with respect to the baseline is shown in Figure 42. The airports that were subject to a regulation have on average smaller incoming centrality losses when FP is implemented at Level 1 and 2 with respect to the baseline, while for airports that were not subject to regulations the losses are slightly larger or very similar. For the outgoing centrality, the same is true only in Level 2 default and 1 stressed, while for Level 1 default and 2 stressed we see a worsening with respect to the baseline, which is similar in regulated and non-regulated airports. All differences are still non statistically significant, as in the passenger centrality case.

Table 37. Average trip centrality loss FP restricted to airports with regulations, incoming and outgoing, in the scenarios where FP is implemented at Level 1 and 2 and in the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Average incoming trip centrality loss (regulated airp)	4.779	4.766	4.772	5.050	5.041	5.048
	4.662	4.688	4.661	5.002	5.006	5.005
	4.881	4.856	4.888	5.100	5.076	5.099
Average incoming trip centrality loss (non-regulated airp)	0.464	0.467	0.464	0.617	0.618	0.618

	0.459	0.462	0.458	0.614	0.614	0.613
	0.469	0.473	0.469	0.620	0.620	0.623
Average outgoing trip centrality loss (regulated airp)	5.133	5.161	5.115	5.435	5.431	5.449
	5.040	5.070	5.015	5.381	5.382	5.414
	5.248	5.245	5.223	5.471	5.471	5.491
Average outgoing trip centrality loss (non-regulated airp)	0.458	0.460	0.458	0.606	0.607	0.607
	0.453	0.456	0.453	0.604	0.604	0.602
	0.464	0.465	0.464	0.609	0.610	0.611

In summary, the FP mechanism overall preserves slightly less itineraries, both passenger itineraries and potential ones. In some cases, it brings improvements locally, to the airports that are subject to a regulation.

4.3.2.5 Causality analysis

There are no measurable effects on causality due to the local application of the mechanisms (only applicable to flights which arrive to an airport with an ATFM regulation), and the small number of flights affected. For this reason, no detailed results are presented.

4.3.3 Flight Arrival Coordination

4.3.3.1 Flights delays

Table 38, Table 39, Table 40 and Table 41 present the delay metrics for the scenarios where the FAC mechanisms are implemented.

At Level 1, both in the default and stressed case, FAC brings a small overall improvement in arrival and departure delays, in the number of flights with reactionary delays and in the number of cancellations. At Level 2 it introduces a small overall worsening, except for the arrival delay in the default case and the number of cancellations in the stressed case. All differences are not statistically significant, except for the increase of average delays and number of reactionary delays in Level 2 stressed and for the decrease of cancellations in Level 1 stressed.

In the default case, the mechanism, at both levels, has a positive effect on the tails of the distributions: at Level 1, there are less flights with more than 15, 60 and 180 minutes of delay (both in arrival and departure), and at Level 2 there are less flights with more than 180 minutes of delay, and also with 60 minutes in the default case. The percentage change with respect to the baseline of some of these metrics are plotted in Figure 43. The effect on the tails can be seen clearly in the QQ-plots in Figure 44 as a thinning of the distribution of departure and arrival delays above 100 minutes, more pronounced for the Level 1 scenario (blue line). In the stressed case, only the Level 1 mechanism retains its positive effect on the tail, which is however less pronounced, as the QQ-plots in Figure 45 show.

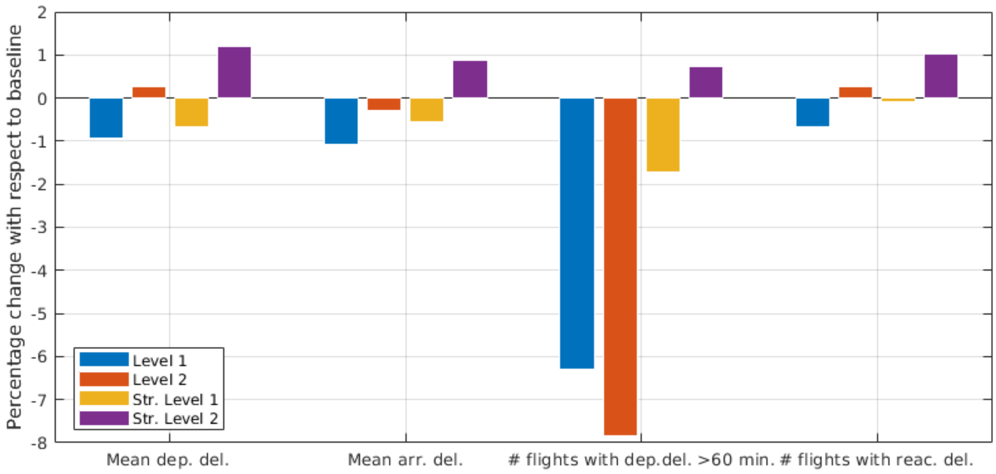


Figure 43. Percentage change flight delay in FAC.

Table 38. Departure delay statistics for the FAC scenarios where it is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean departure delay (min)	**10.51**	**10.41**	**10.54**	**25.86**	**25.69**	**26.17**
	10.11	10.04	10.19	25.17	24.97	25.57
	10.69	10.65	10.81	26.27	26.19	26.7
#flights with dep.del>15 min	**6143.88**	**6129.23**	**6228.15**	**13560.2**	**13549.25**	**13764.01**
	6037.5	6033	6122	13411	13436.5	13634
	6239	6235.5	6339	13700.5	13671	13901

Total dep delay >15 min	**200178.08**	**197700.15**	**199448.39**	**614651.32**	**610699.38**	**622165.35**
	189484.14	187974.18	190696.76	596452.79	592220.32	602788.49
	204517.78	204210.27	206502.9	626200.08	624079.24	635930.78
Mean delay of flights with dep. del. >15 min	**32.58**	**32.26**	**32.02**	**45.33**	**45.07**	**45.2**
#flights with dep. del >60 min	**450.61**	**430.23**	**430.28**	**3013.1**	**2973.82**	**3040.19**
	366.5	359	359	2834.5	2827	2868
	490	479	482	3124	3111	3172
Total dep. delay >60 min	**40172.87**	**37647.05**	**37028.78**	**274106.64**	**269395.67**	**276094.15**
	29543.56	29153.35	29121.52	255304	248352.8	257148.04
	43460.13	43460.46	42091.34	285799.53	284049.43	292990.55
Mean delay of flights with dep. del. >60 min	**89.15**	**87.5**	**86.06**	**90.97**	**90.59**	**90.81**
#flights with dep.del>180 min	**18.81**	**14.84**	**12.05**	**83.39**	**79.14**	**84.18**
	4	2	3	51.5	48	50
	12.5	14	12	94	91	103
Total dep. delay >180 min	**4161.89**	**3302.88**	**2668.46**	**18016.96**	**16929.61**	**18127.95**
	777.3	570.74	613.26	11078.22	10203.23	10740.28
	2734.54	3157.42	2820.81	20037.11	19175.69	22726.38

Mean delay of flights with dep. del. >180 min	221.26	222.57	221.45	216.06	213.92	215.35

Table 39. Arrival delay statistics for FAC scenarios where it is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean arrival delay (min)	**5.18**	**5.1**	**5.21**	**29.09**	**28.94**	**29.41**
	4.81	4.74	4.85	28.43	28.22	28.76
	5.35	5.33	5.55	29.56	29.47	29.87
Mean arrival delay of delayed flights (min)	**10.49**	**10.38**	**10.46**	**30.87**	**30.7**	**31.15**
	10.1	10.04	10.14	30.22	30	30.51
	10.64	10.61	10.75	31.33	31.23	31.62
#flights with arr.del>15 min	**6614.13**	**6580.41**	**6653.3**	**15930.43**	**15924.32**	**16077.09**
	6504	6481.5	6546	15800	15815	15974
	6674	6671.5	6748	16035	16050	16169
Total arr. delay >15 min	**220541.86**	**217639.08**	**219359.62**	**758486.18**	**754579.7**	**766416.78**
	210083.98	208561.33	210704.1	740924.41	737172.32	747264.25
	225008.08	224011.89	226657.29	770880.74	769983.12	780447.22
Mean delay of flights with arr. del. >15 min	**33.34**	**33.07**	**32.97**	**47.61**	**47.39**	**47.67**

#flights with arr. del >60 min	**523.45**	**506.45**	**506.6**	**3944.73**	**3904.3**	**3998.91**
	440.5	440.5	442	3797.5	3759.5	3849
	565.5	560.5	552.5	4067.5	4029.5	4130
Total arr. delay >60 min	**45928.83**	**43656.81**	**43154.69**	**360514.68**	**355631.38**	**364788.53**
	35987.14	35341.03	35604.32	340020.67	333589.21	345240.38
	50130.78	48592.29	48096.23	372987.32	369942.87	381229.23
Mean delay of flights with arr. del. >60 min	**87.74**	**86.2**	**85.18**	**91.39**	**91.09**	**91.22**
#flights with arr. del. >180 min	**19.51**	**15.56**	**13.43**	**108.15**	**102.4**	**107.1**
	4.5	4	4	72	70	68.5
	14.5	14.5	13.5	124.5	115	131.5
Total arr. delay >180 min	**4323.19**	**3469.05**	**2963.92**	**23307.19**	**21965.68**	**23043.86**
	989.74	789.64	830.04	15350.88	14814.76	14521.19
	3228.24	3194.05	3068.4	26928.5	24101.47	28815.49
Mean delay of flights with arr. del. >180 min	**221.59**	**222.95**	**220.69**	**215.51**	**214.51**	**215.16**

Table 40. Gate-to-gate delay statistics for FAC scenarios where it is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2

Mean gate-to-gate delay (min)	**-5.92**	**-5.89**	**-5.89**	**3.25**	**3.26**	**3.26**
	-5.97	-5.95	-5.95	3.2	3.21	3.23
	-5.86	-5.82	-5.83	3.3	3.31	3.31

Table 41. Cancellation and reactionary delay statistics for FAC scenarios where FAC is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
# cancelled flights	**1087.95**	**1075.08**	**1091.28**	**1160.20**	**1131.65**	**1145.29**
	1026.50	1043.50	1039.00	1096.50	1093.00	1088.00
	1136.00	1115.50	1142.00	1185.00	1159.50	1188.50
Mean reactionary delay (min)	**17.72**	**17.81**	**17.66**	**38.44**	**38.19**	**38.98**
	15.11	15.17	14.88	35.07	34.69	35.71
	16.90	17.20	17.14	39.07	38.06	40.12
# flights with reactionary delay	**2888.71**	**2869.36**	**2896.00**	**7197.35**	**7191.48**	**7272.01**
	2823.50	2812.50	2836.50	7114.00	7136.00	7200.00
	2947.00	2924.00	2961.00	7268.50	7264.50	7348.50

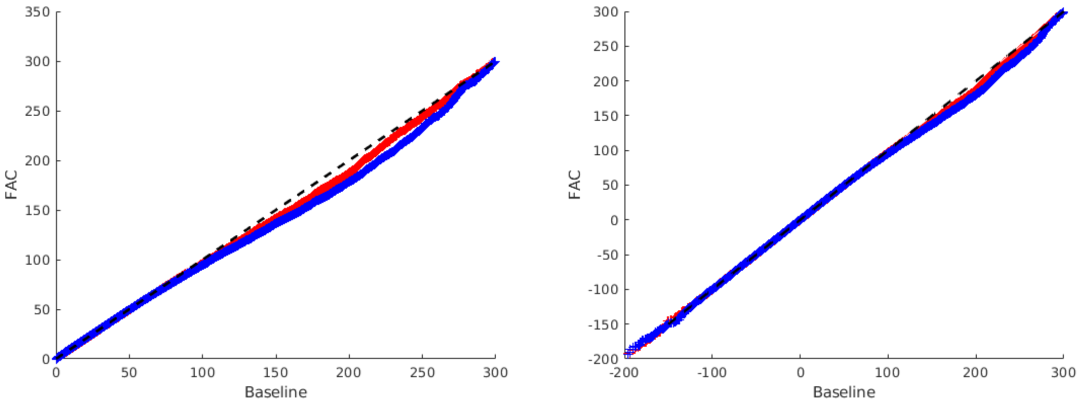


Figure 44. QQ-plots flight delay distributions default scenario FAC Left: QQ-plot of departure delay; right: QQ-plot of arrival delay. Red: Level 1, Blue: Level 2. The dashed line is the 1:1 line.

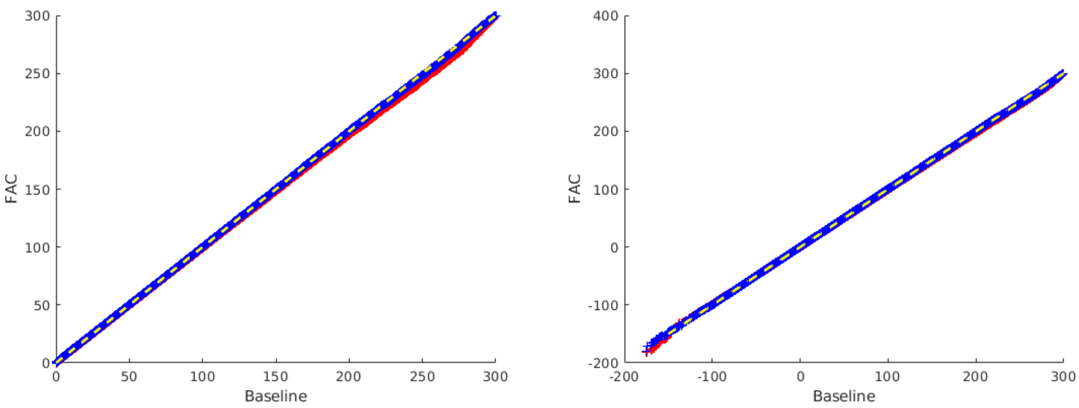


Figure 45. QQ-plots flight delay distributions stressed scenario FAC Left: QQ-plot of departure delay; right: QQ-plot of arrival delay. Red: Level 1, Blue: Level 2. The dashed line is the 1:1 line.

Table 42, Table 43, Table 44 and Table 45 report the same delay statistics for the restricted sample of flights arriving at airports in which FAC has been implemented. Here, we see that both levels introduce improvements to average delays and to the tails in the default case. Level 1 decreases the average delay more than Level 2, although Level 2 has more effect on the tails, as can be noted in Figure 47, where the blue curve is slightly lower than the red one. The thinning of the departure and arrival delay distribution is enhanced with respect to Figure 44, where all flights were considered, showing that the mechanism has a stronger effect on the flights arriving to airports implementing FAC. In the stressed case, only Level 1 manages to diminish the average delays and the number of flights in the tails, while Level 2 increases both, and it also increases significantly the flights with reactionary delay. The thinning of the right tail of the delay distribution for Level 1 is anyway less pronounced than in the default case, as shown in Figure 48. As can be remarked by Figure 46, the percentage improvements with respect to the baseline are larger in this restricted sample than overall.

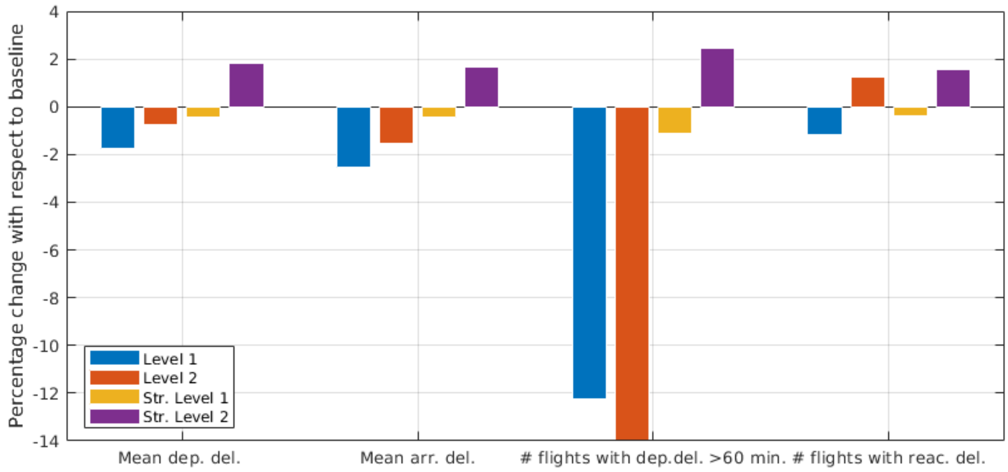


Figure 46. Percentage change flight delay in FAC restricted to airports with FAC.

Table 42. Departure delay statistics for the FAC scenarios restricted to airports with FAC where it is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean departure delay (min)	**10.26**	**10.08**	**10.18**	**25.94**	**25.83**	**26.41**
	9.64	9.5	9.65	24.88	24.8	25.26
	10.58	10.29	10.44	26.79	26.81	27.37
#flights with dep.del>15 min	**2163.29**	**2156.85**	**2187.13**	**4998.1**	**5000.62**	**5088.14**
	2096	2086.5	2130	4925	4930.5	5012.5
	2228.5	2198	2234	5070	5077	5158
Total dep delay >15 min	**71169.03**	**69405.09**	**69920.61**	**226184.78**	**225352.82**	**230692.84**
	64869.26	63754.58	64774.32	215998.07	215191.08	219923.65
	74261.89	71293.85	72031.34	234546.78	234864.6	239620.14

Mean delay of flights with dep. del. >15 min	**32.9**	**32.18**	**31.97**	**45.25**	**45.06**	**45.34**
#flights with dep. del >60 min	**165.84**	**151.57**	**150.83**	**1112.96**	**1104.65**	**1138.75**
	117.5	109	110.5	1029.5	1017	1046.5
	190.5	166.5	166.5	1180.5	1196	1228.5
Total dep. delay >60 min	**15378.37**	**13495.34**	**13231.83**	**100636.78**	**99528.28**	**103113.9**
	9605.83	8627.46	8665.17	89423.9	88289.38	92278.7
	16596.47	14395.37	14004.99	109382.98	109989.87	112181.45
Mean delay of flights with dep. del. >60 min	**92.73**	**89.04**	**87.73**	**90.42**	**90.1**	**90.55**
#flights with dep.del>180 min	**8.95**	**5.9**	**5**	**27.55**	**26.02**	**30.16**
	1	1	0	14	13	13.5
	5.5	4	4	34	33.5	38
Total dep. delay >180 min	**1975.53**	**1307.78**	**1101.69**	**5912.59**	**5512.98**	**6445.49**
	186.91	180.48	0	3073.01	2755.61	2772.2
	1244.85	1028.13	826.81	7334.32	7198.63	8037.48
Mean delay of flights with dep. del. >180 min	**220.73**	**221.66**	**220.34**	**214.61**	**211.87**	**213.71**

Table 43. Arrival delay statistics for FAC scenarios restricted to airports with FAC where it is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean arrival delay (min)	**4.35**	**4.2**	**4.3**	**28.96**	**28.86**	**29.51**
	3.73	3.6	3.74	27.87	27.83	28.46
	4.64	4.39	4.61	29.78	29.86	30.45
Mean arrival delay of delayed flights (min)	**9.28**	**9.04**	**9.14**	**30.35**	**30.21**	**30.85**
	8.67	8.5	8.63	29.25	29.18	29.8
	9.56	9.16	9.35	31.17	31.17	31.78
#flights with arr.del>15 min	**2081.76**	**2055.31**	**2084.73**	**5851.11**	**5846.46**	**5927.99**
	2012	1988.5	2028.5	5794.5	5782	5870
	2141.5	2105.5	2126.5	5924	5918.5	5988
Total arr. delay >15 min	**68949.83**	**66704.18**	**67387.83**	**272320.53**	**271099.58**	**277611.26**
	63136.72	61315.28	62488.61	261784.78	260258.93	267112.42
	71849.07	68316.59	69597.17	279626.55	280983.73	287250.06
Mean delay of flights with arr. del. >15 min	**33.12**	**32.45**	**32.32**	**46.54**	**46.37**	**46.83**
#flights with arr. del >60 min	**164.29**	**150.78**	**150.76**	**1384.97**	**1370.53**	**1421.73**
	116.5	109.5	113	1297	1273	1339
	184	166	164	1456	1457	1510
Total arr. delay >60 min	**15326.04**	**13553.6**	**13370.95**	**125687.31**	**124109.86**	**129288.22**

	9558.08	8961.31	9094.13	114545.69	112169.29	117940.58
	15754.57	14567.05	14213.63	134927.41	134797.84	138487.15
Mean delay of flights with arr. del. >60 min	**93.29**	**89.89**	**88.69**	**90.75**	**90.56**	**90.94**
#flights with arr. del. >180 min	**9.71**	**6.66**	**5.86**	**35.19**	**32.87**	**36.94**
	1	1	2	20	19.5	19.5
	7	5	5	47	43	44
Total arr. delay >180 min	**2143.86**	**1486.55**	**1295.37**	**7548.2**	**7008.74**	**7927.64**
	249.19	234.99	399.1	4184.35	4070.4	4204.78
	1558.53	1237.56	1179	9745.58	8999.88	9311.48
Mean delay of flights with arr. del. >180 min	**220.79**	**223.21**	**221.05**	**214.5**	**213.23**	**214.61**

Table 44. Gate-to-gate delay statistics for FAC scenarios restricted to airports with FAC where it is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Mean gate-to-gate delay (min)	**3.03**	**3.03**	**3.11**	**3.03**	**3.03**	**3.11**
	2.94	2.96	3.03	2.94	2.96	3.03

	3.12	3.11	3.19	3.12	3.11	3.19

Table 45. Cancellation and reactionary delay statistics scenarios restricted to airports with FAC where FAC is implemented and the corresponding baselines.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
# cancelled flights	**380.46**	**375.08**	**383.01**	**387.90**	**378.43**	**379.03**
	357.50	357.50	363.50	356.00	354.00	355.50
	406.00	390.50	400.50	396.50	391.50	392.00
Mean reactionary delay (min)	**26.59**	**25.77**	**25.58**	**48.14**	**47.57**	**48.21**
	23.88	23.69	23.65	46.36	45.93	46.59
	27.33	26.43	26.53	49.21	49.07	49.47
# flights with reactionary delay	**976.22**	**965.00**	**988.25**	**2661.66**	**2651.94**	**2703.17**
	944.00	937.50	956.50	2626.50	2615.00	2669.00
	1004.00	986.50	1015.50	2695.00	2687.50	2737.00

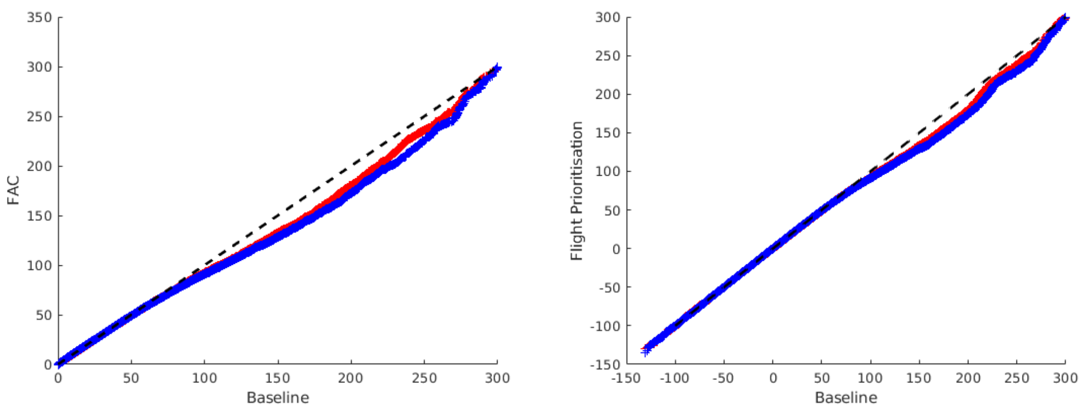


Figure 47. QQ-plots flight delay distributions default scenario FAC restricted airports with FAC Left: QQ-plot of departure delay; right: QQ-plot of arrival delay. Red: Level 1, Blue: Level 2. The dashed line is the 1:1 line.

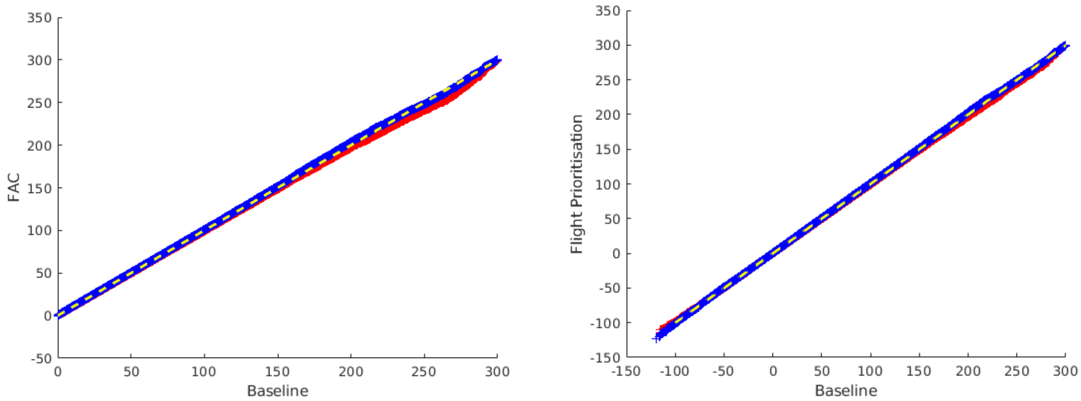


Figure 48. QQ-plots flight delay distributions stressed scenario FAC restricted airports with FAC Left: QQ-plot of departure delay; right: QQ-plot of arrival delay. Red: Level 1, Blue: Level 2. The dashed line is the 1:1 line.

In summary, the FAC mechanism brings the following changes:

- It lowers average delays at Level 1, but increases them at Level 2.
- It decreases the number of delays in the tail (>60 minutes) at Level 1, and also at Level 2 in the default case.
- The improvements are larger in the default case than in the stressed one, where the mechanism does not seem to be very effective.
- At Level 1 the mechanism has more positive effects on delays than at Level 2 which could be linked to the fact that at Level 2 is focused on cost and not on delay.

4.3.3.2 Passengers delays

In this section we show the same metrics presented in Section 4.3.2.2 but for the FAC mechanism. These results have also been obtained using 50 iterations.

Both for stressed and non-stressed scenarios, the impact of FAC is very small. The averages are almost the same, both for the standard delay and its positive counterpart. In the non-stressed scenario, FAC does not seem to have any effect on high delays either, given that the number of passengers delayed more than 15, 60, and 180 are roughly the same across scenarios. In the stressed scenarios on the other end, passengers with very high delays seem to decrease with the implementation of the mechanism (see Table 46).

Table 46. Passenger indicators statistics for FAC mechanism.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2

Mean delay	17.2	17.1	17.2	40	39.3	39.6
25% perc. delay	-9.4	-9.3	-9.3	4.7	4.6	4.8
75% perc. delay	15.2	15.1	15.2	44.8	44.2	44.8
Positive mean delay	23.6	23.6	23.6	42.4	41.7	42
Positive 25% perc. delay	0	0	0	4.7	4.6	4.8
Positive 75% perc. delay	15.2	15.1	15.2	44.8	44.2	44.8
Number of pax with delay>15	859784.1	856507.5	862516.2	2005583.6	1993969.1	2010923.2
Total delays with delay>15	74470449.7	74193796.1	74208659.2	138623108	136258012.7	137320156.7
Mean delay with delay>15	86.6	86.6	86	69.1	68.3	68.3
Number of pax with delay>60	188560.8	188450.7	187441.8	561065.1	548094.8	557709.9
Total delays with delay>60	55486051.8	55301715.3	55103800.7	91227358.5	88832690.4	89556722.7
Mean delay with delay>60	294.3	293.5	294	162.6	162.1	160.6
Number of pax with delay>180	142665.4	142266.7	142262	158099.6	153008	153587.8
Total delays with delay>180	51765630	51556141.3	51479453.7	56374057.5	54771632	54759893.8

Mean delay with delay>180	362.8	362.4	361.9	356.6	358	356.5

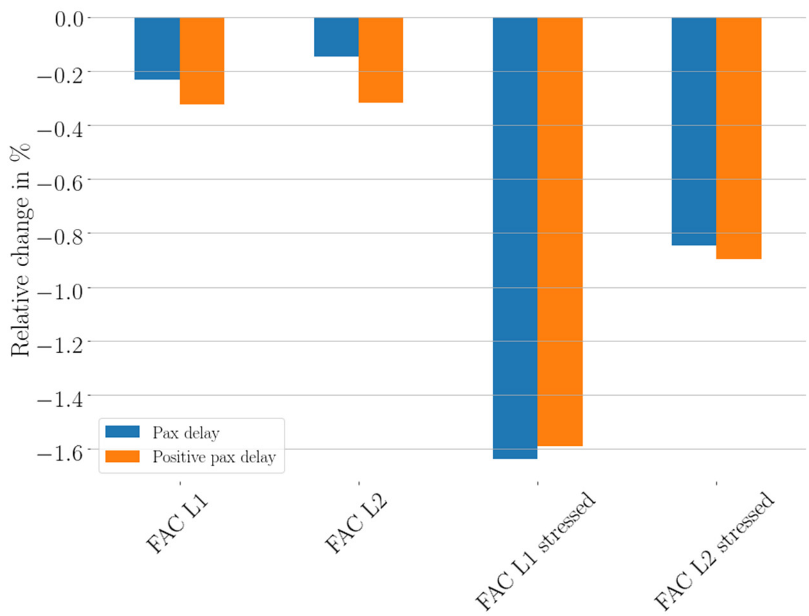


Figure 49. Change in passenger delay for FAC: the mean delay (blue) and the mean positive delay (orange) with respect to baseline scenarios (default on the left, stressed on the right).

Figure 49 shows that indeed the variations in the average delays are very small, even if they are clearly larger in the stressed scenarios. It is also interesting to note that, contrary to FP and 4DTA, the variations on the average of the positive delays seem to be comparable or even greater than the variations on the normal delay. This denotes the fact that FAC has an effect mainly on the positive delays.

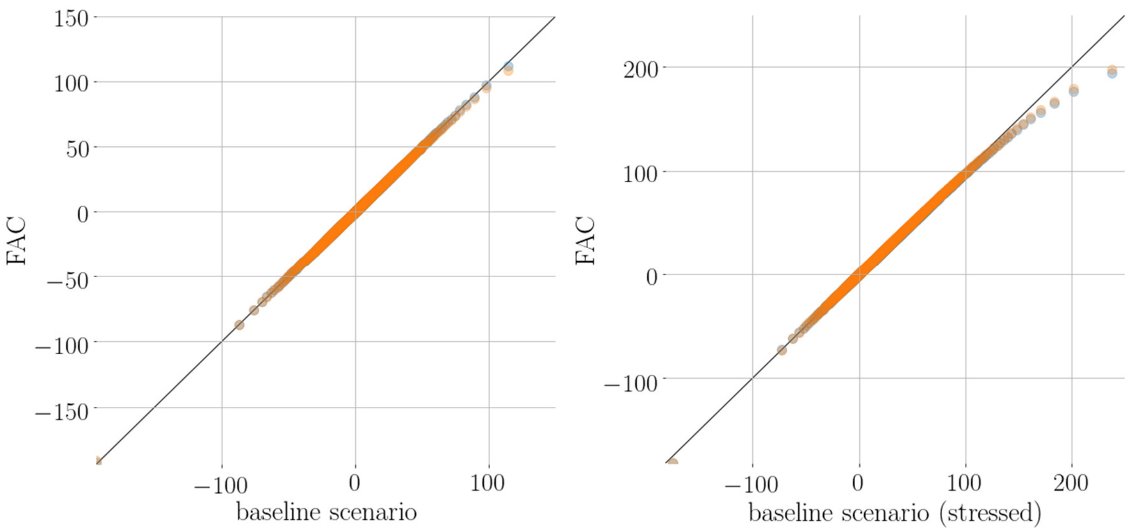


Figure 50. QQ-plot of arrival passenger delay in FAC in the baseline scenario (left) and stressed scenario (right). The blue circles correspond to the FP at Level 1, the orange ones to Level 2. The plots have been cut at 150 and 200 minutes respectively.

Finally, the QQ-plots in Figure 50 shows the above trends. On the no-stressed scenarios, the effect of FAC is negligible, even if a small squeezing trend can be seen on the right side of the distribution. On the other hand, the effect of FAC in the stressed scenarios can be seen very clearly, with a much thinner right tail for the distribution. Note also that Level 1 and Level 2 seem to have similar effects.

In summary, the effect of the implementation of FAC is the following:

- in non-stressed scenarios, the effect is negligible, both on negative or positive delay, and both on the tails and on the bulk.
- in the stressed scenarios, there an improvement which is mainly seen for very high delays, with a corresponding limited effect for the average. Level 1 seems to have a slightly better effect than Level 2.

4.3.3.3 Costs

In this section we discuss the impact of the FAC mechanism on the different kinds of costs airlines incur because of both excess fuel and delay.

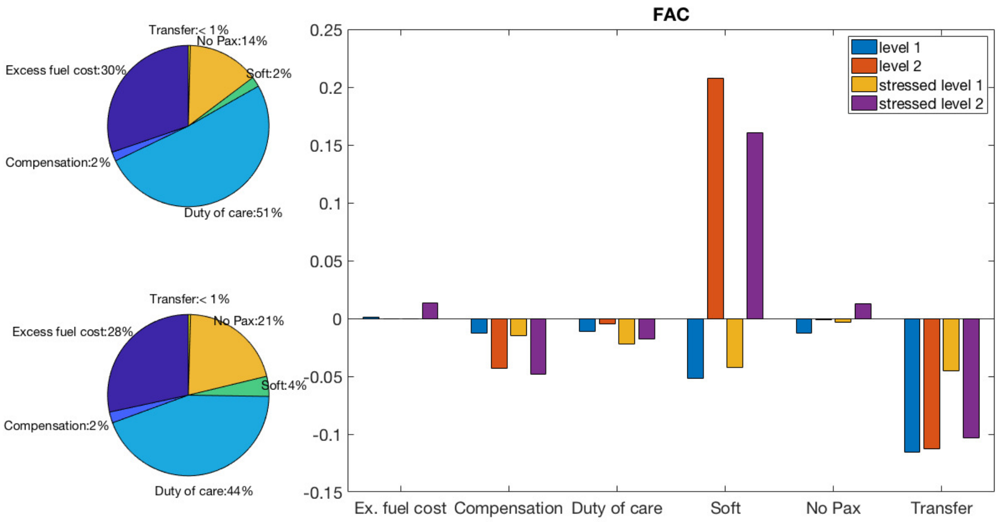


Figure 51. QQ-plot of arrival passenger delay in FAC in the baseline scenario (left) and stressed scenario (right). The blue circles correspond to the FP at Level 1, the orange ones to Level 2. The plots have been cut at 150 and 200 minutes respectively.

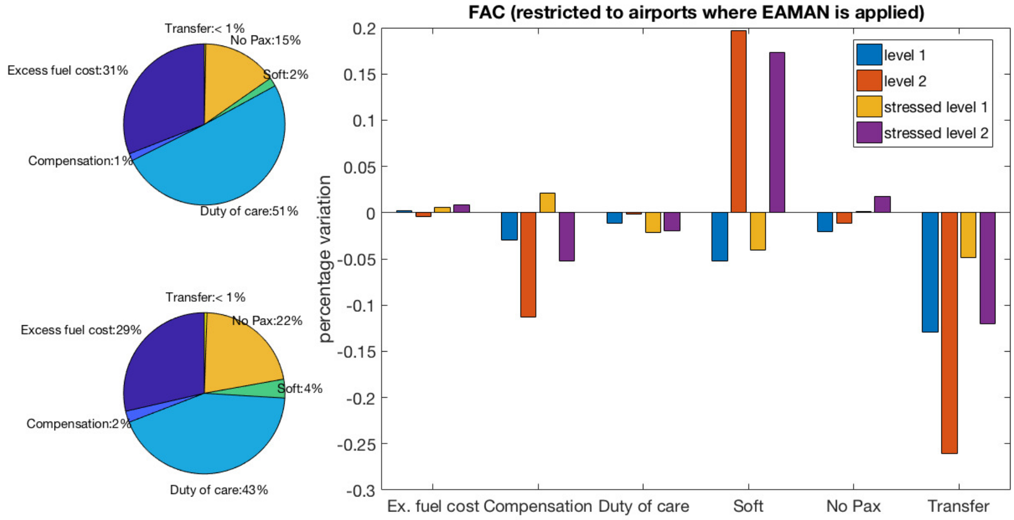


Figure 52. QQ-plot of arrival passenger delay in FAC in the baseline scenario (left) and stressed scenario (right). The blue circles correspond to the FP at Level 1, the orange ones to Level 2. The plots have been cut at 150 and 200 minutes respectively.

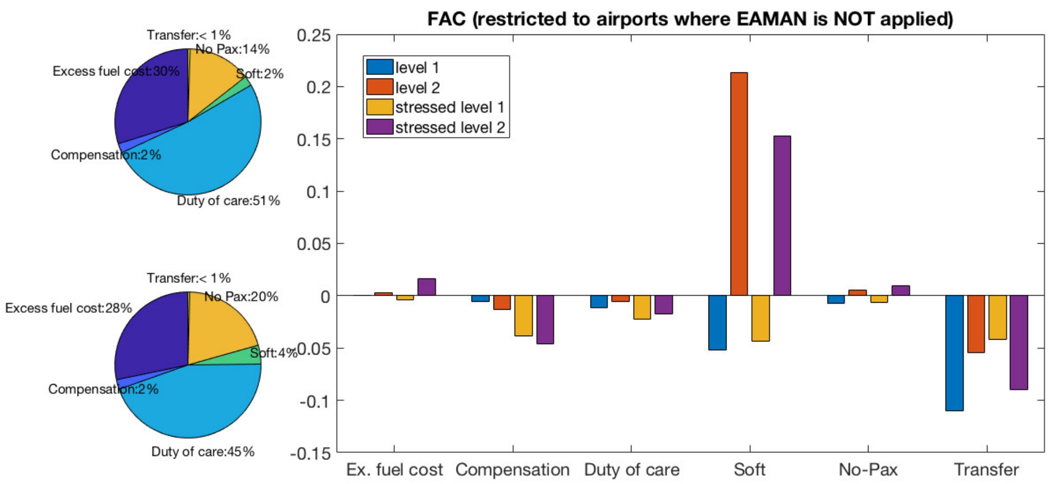


Figure 53. QQ-plot of arrival passenger delay in FAC in the baseline scenario (left) and stressed scenario (right). The blue circles correspond to the FP at Level 1, the orange ones to Level 2. The plots have been cut at 150 and 200 minutes respectively.

The FAC mechanism has no impact on the excess fuel cost but tend to reduce significantly (on average) the costs of compensation and transfer, see Figure 51 and Figure 52. The only outlier is represented by the positive percentage variation of the cost of compensation for the stressed scenario with FAC at Level 2, observed in the restricted sample of flights landing at any airport where E-AMAN is implemented. However, this is a small effect. The decreasing of the two types of costs can be explained in terms of the implemented protocols of E-AMAN. At Level 1, the objective function of the E-AMAN is reducing the reactionary delays of the next flights associated with landing aircraft, thus preserving, e.g., the flight connections. A consequence of preserving connections is reducing the cost of transfer by cutting off the cost of rebooking of passengers. This preserve of connections is implicit. At Level 2, this improvement is also more evident as a consequence of the objective functions including the airline utility, i.e., a priority is assigned to the most costly flights and cost is driven by passenger connections. As a consequence, a compensation and transfer costs reduce further with respect to the corresponding baselines. Also a smaller decrease of the duty of care of passengers can be noticed.

The improvements in terms of costs is observed for both the aggregate system, the smaller sample of flights landing at some E-AMAN airports, and the complementary sample with flights landing at airports where E-AMAN is not applied (Figure 53). Hence, an externality is observed, i.e., the advantages of introducing the FAC mechanism in some airports (usually the ones characterised by higher traffic) propagate to the whole system, thus contributing to the decrease of these types of costs overall.

FAC: detailed analysis of costs

Table 47. Detailed costs for flights with delays<300 min FAC

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Excess cost of fuel	**136.0**	**136.2**	**135.9**	**182.9**	**182.9**	**185.4**
	135.0	135.1	134.7	181.2	181.3	183.6
	137.1	137.4	137.1	184.7	184.9	187.2
Cost of compensation	**8.3**	**8.2**	**7.9**	**14.1**	**13.9**	**13.5**
	7.3	7.3	7.4	12.8	12.7	12.3
	9.1	9.0	8.3	14.7	14.8	14.4
Fraction of compensations	**1.1%**	**1.1%**	**1.0%**	**2.7%**	**2.7%**	**2.6%**
Costs of 'transfer'	**1.8**	**1.6**	**1.6**	**2.9**	**2.7**	**2.6**

	1.0	0.9	0.9	2.1	1.9	1.8
	2.4	2.0	2.1	3.5	3.0	3.1
Fraction of transfer costs	**0.1%**	**0.1%**	**0.1%**	**0.3%**	**0.3%**	**0.3%**
Duty of care	229.8	227.2	228.8	284.9	278.7	279.7
	219.8	219.6	217.5	270.0	267.7	266.2
	237.2	235.2	237.7	290.0	284.6	288.1
Fraction of 'duty of care'	**13.5%**	**13.5%**	**13.4%**	**20.0%**	**19.8%**	**19.8%**
Soft costs	**9.1**	**8.6**	**10.9**	**25.9**	**24.8**	**30.0**
	3.6	3.5	3.7	10.3	12.7	10.6
	14.0	13.8	14.3	40.9	14.8	41.3
Fraction of soft costs	**46.4%**	**46.5%**	**46.7%**	**57.5%**	**57.6%**	**57.5%**
Non passenger costs	**64.4**	**63.6**	**64.3**	**134.0**	**133.5**	**135.7**
	62.3	61.8	62.6	129.9	129.5	131.5
	64.9	64.4	65.6	136.9	136.9	138.5
Fraction of non passenger costs	**91.1%**	**91.1%**	**91.2%**	**93.2%**	**93.3%**	**93.4%**
Total excess cost	**449.5**	**445.5**	**449.7**	**644.8**	**636.7**	**646.9**
	436.6	435.0	437.2	619.8	619.9	632.0
	457.9	452.9	459.8	658.3	651.3	661.5

Table 48. Detailed costs with flights with delays<300 min restricted arriving at airports with FAC.

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2

Excess cost of fuel	**143.8**	**144.2**	**143.2**	**194.9**	**195.9**	**196.5**
	141.6	142.3	141.0	191.8	192.7	193.5
	145.8	146.1	144.9	197.6	199.2	199.5
Cost of compensation	**6.7**	**6.5**	**6.0**	**15.5**	**15.8**	**14.7**
	5.6	5.5.	5.1	13.9	13.6	12.9
	7.4	7.2	12.6	16.9	17.0	16.0
Fraction of compensations	**1.3%**	**1.2%**	**1.2%**	**3.7%**	**3.8%**	**3.5%**
Costs of 'transfer'	**1.4**	**1.2**	**1.0**	**3.6**	**3.4**	**3.1**
	0.9	0.9	0.7	2.9	2.7	2.5
	1.6	1.6	1.3	4.1	3.9	3.6
Fraction of transfer costs	**0.2%**	**0.2%**	**0.2%**	**0.6%**	**0.5%**	**0.5%**
Duty of care	**235.0**	**232.4**	**234.6**	**294.5**	**288.9**	**288.8**
	221.7	218.4	219.9	270.5	268.7	270.3
	243.9	244.4	249.3	304.4	300.7	300.1
Fraction of 'duty of care'	**15.7%**	**15.7%**	**15.3%**	**23.4%**	**23.1%**	**23.1%**
Soft costs	**8.2**	**7.8**	**9.8**	**25.7**	**24.6**	**30.1**
	3.2	3.1	3.4	10.3	10.2	10.6
	12.5	12.2	12.6	405	39.7	41.4
Fraction of soft costs	**44.6%**	**44.8%**	**44.8%**	**58.1%**	**58.4%**	**58.1%**
Non passenger costs	**69.6**	**68.1**	**68.8**	**147.6**	**147.7**	**150.2**
	65.7	64.9	65.9	141.6	142.1	144.4

	71.1	69.5	70.3	153.2	153.3	157.3
Fraction of non passenger costs	**91.0%**	**91.1%**	**91.1%**	**93.6%**	**93.7%**	**93.8%**
Total excess cost	**464.8**	**460.3**	**463.4**	**681.8**	**675.9**	**683.5**
	448.4	441.6	444.1	652.2	648.4	657.8
	471.7	472.8	480.8	700.3	697.9	702.2

Table 49. Detailed costs excluding cancelled flights FAC (with the exception of flights with delays larger than 300 min).

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Excess cost of fuel	**141.6**	**141.7**	**141.6**	**190.8**	**190.7**	**193.3**
	140.3	140.4	140.2	188.9	188.9	191.6
	142.6	143.1	142.8	192.8	192.6	195.4
Cost of compensation	**1.5**	**1.4**	**1.3**	**6.9**	**6.9**	**6.5**
	1.2	1.2	1.1	6.2	6.1	5.8
	1.5	1.5	1.4	7.4	7.4	7.1
Fraction of compensations	**0.6%**	**0.6%**	**0.5%**	**2.1%**	**2.1%**	**2.0%**
Costs of 'transfer'	**0.3**	**0.3**	**0.2**	**1.3**	**1.2**	**1.1**
	0.2	0.2	0.2	1.1	1.0	0.8

	0.4	0.4	0.3	1.5	1.2	1.3
Fraction of transfer costs	**5e-2%**	**5e-2%**	**5e-2%**	**0.2%**	**0.2%**	**0.2%**
Duty of care	**39.7**	**38.4**	**38.1**	**75.9**	**75.1**	**74.7**
	35	35.4	35.0	68.0	68.8	67.7
	41	40.7	40.1	80.1	80.2	80.6
Fraction of 'duty of care'	**11.8%**	**11.8%**	**11.6%**	**18.2%**	**18.1%**	**18.1%**
Soft costs	**7.3**	6.9	8.8	**24.6**	23.6	28.6
	2.9	2.9	3.1	9.8	9.8	10.1
	11.1	11.1	11.5	38.8	38.4	39.3
Fraction of soft costs	**46.1%**	**46.2%**	**46.4%**	**57.5%**	**57.6%**	**57.5%**
Non passenger costs	**67.0**	**66.2**	**66.9**	**139.8**	**139.2**	**141.5**
	64.8	64.4	65.2	135.4	134.9	137.7
	67.4	67.0	68.2	142.7	142.6	144.2
Fraction of non passenger costs	**94.8%**	**94.8%**	**94.9%**	**97.2%**	**97.3%**	**97.4%**
Total excess cost	**257.6**	**254.9**	**257.1**	**439.5**	**436.8**	**445.8**
	248.7	248.3	250.2	420.4	421.7	433.2
	262.9	259.1	262.9	454.7	448.9	458.1

Table 50. Detailed costs excluding cancelled flights restricted arrival airports with FAC (with the exception of flights with delays larger than 300 min).

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2

Excess cost of fuel	**149.4**	**149.7**	**148.8**	**202.5**	**203.5**	**204.1**
	147.3	147.6	146.9	199.2	200.4	201.4
	151.2	151.7	150.9	205.5	206.4	206.7
Cost of compensation	**2.9**	**2.6**	**2.4**	**11.7**	**12.1**	**11.2**
	2.2	2.1	2.0	10.5	10.5	9.5
	3.0	2.8	2.5	12.6	13.3	12.5
Fraction of compensations	**0.9%**	**0.8%**	**0.8%**	**3.4%**	**3.4%**	**3.1%**
Costs of 'transfer'	**0.8**	**0.6**	**0.6**	**3.1**	**2.9**	**2.7**
	0.4	0.4	0.3	2.5	2.3	1.9
	0.9	0.8	0.8	3.5	3.4	3.1
Fraction of transfer costs	**0.1%**	**0.1%**	**0.1%**	**0.5%**	**0.5%**	**0.5%**
Duty of care	**48.6**	**46.4**	**45.2**	**99.3**	**98.5**	**98.3**
	41.9	40.7	39.6	86.3	86.4	85.0
	50.2	49.3	48.4	107.2	109.6	109.7
Fraction of 'duty of care'	**14.1%**	**14.1%**	**13.7%**	**22.0%**	**21.7%**	**21.8%**
Soft costs	**6.3**	**5.9**	**7.5**	**24.4**	**23.4**	**28.7**
	2.5	2.4	2.6	9.8	9.7	10.2
	9.3	9.1	9.6	38.5	37.8	39.3
Fraction of soft costs	**44.2%**	**44.4%**	**44.5%**	**58.1%**	**58.5%**	**58.2%**
Non passenger costs	**72.3**	**70.8**	**71.4**	**153.4**	**153.4**	**156.0**
	68.3	67.5	68.6	146.9	147.0	149.6
	73.9	72.2	73.1	159.4	159.1	163.4

Fraction of non passenger costs	**94.5%**	**94.5%**	**94.6%**	**97.3%**	**97.3%**	**97.4%**
Total excess cost	**280.4**	**276.2**	**275.9**	**494.4**	**493.8**	**500.9**
	266.2	265.7	266.4	472.8	475.6	480.2
	283.9	279.6	282.7	518.4	514.6	515.8

In Table 47 and Table 48, we showed the detailed values of costs used to obtain Figure 51 and Figure 52. Here, we consider the averages over the whole sample of flights, but excluding the ones with a delay larger than 300 min. Furthermore, in Table 49 and Table 50, we restrict further the sample by excluding the cancelled flights in order to capture how the FAC mechanism works for normal operations. We observe qualitatively the same behaviour. Then, in this case, cancellations do not affect crucially the impact of the mechanism. In effect, the FAC mechanism takes place at the arrival phase, whereas cancellations refer to departures. From the analysis of average costs, we are not able to capture some interdependences between the two phases.

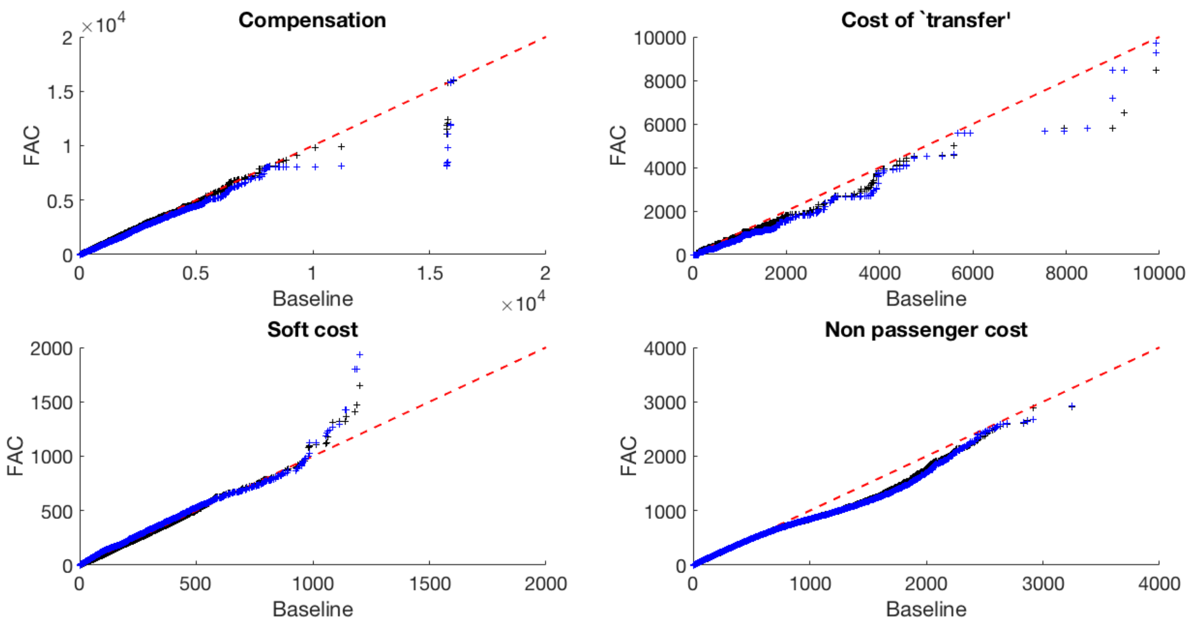


Figure 54. QQ-plot of costs in FAC baseline by comparing the default baseline scenario with the ones with the FAC mechanism but restricting to all flights landing in airports where E-AMAN is applied. Black is for FAC at Level 1, while blue is for FAC at Level.

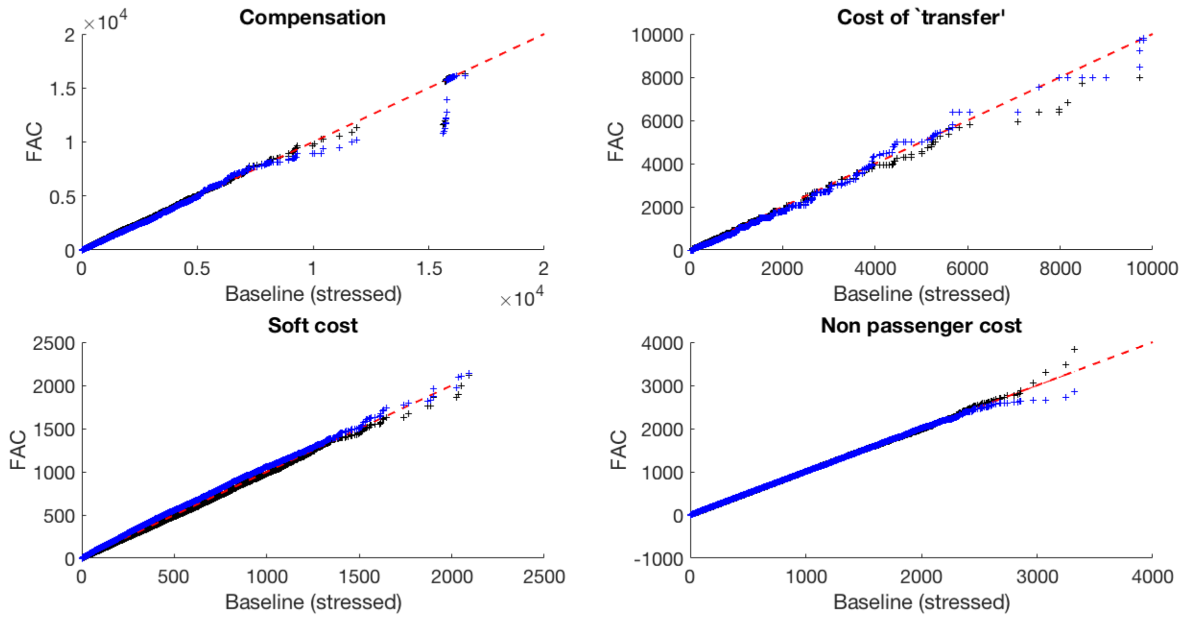


Figure 55. QQ-plot of costs in FAC stressed by comparing the stressed baseline scenario with the ones with the FAC mechanism but restricting to all flights landing in airports where E-AMAN is applied. Black is for FAC at Level 1, while blue is for FAC at Level.

In Figure 54 and Figure 55, we study how the distributions of costs (compensation, transfer, soft, and non-passenger costs) in the subsamples of flights landing at some E-AMAN airports, are affected by the introduction of the FAC mechanism. In the default case (Figure 54), we notice clear patterns, whereas the stressed case shows more homogenous distributions (Figure 55), with the exception of the extreme right tails. In both cases, we observe fewer extreme events for compensation, transfer, and non-passenger costs, consistently with the average values of the distributions. Finally, the reason of the increase of the mean soft cost is less clear: in both default and stressed scenarios, the distributions look quite equivalent, with the exceptions of fatter right tail in the default case, for both Level 1 and Level 2. However, in both default and stressed cases, we observe a decrease of the mean soft cost at Level 1, but an increase at Level 2. We argue that this is due to small fluctuations in the bulk of distributions. The distributions of the other types of costs are quite equivalent when we move from the baseline to any scenario with implemented FAC mechanism at any level.

Comments

According to the detailed analysis of costs, we point out that the FAC mechanism shows a clear advantage for airlines because of the reduction of the passengers' costs, in particular compensation, duty of care, and costs of transfer. At Level 1, this is the consequence of reducing the reactionary delays of flights landing at some E-AMAN airports. A further improvement is observed at Level 2, where flight priorities are considered by E-AMAN. Finally, an externality is observed: the advantage in terms of costs of introducing the FAC mechanism in a small set of airports propagate to the airports where E-AMAN is not implemented, thus improving the system overall.

4.3.3.4 Centrality losses

Passenger centrality

Table 51 reports the average loss of passenger centrality, incoming and outgoing, in the scenarios with FAC implemented and in the corresponding baseline. Here, the centrality losses are averaged over all the airports, and we see small improvements in the centrality loss (i.e., smaller loss), except for Level 2 default. The percentage change with respect to the baseline amounts to 0.5-1.5% (see Figure 56), while the absolute change tells as that, on average over all airports, 1-3 more passengers manage to arrive to their final destination. All changes are not statistically significant, however.

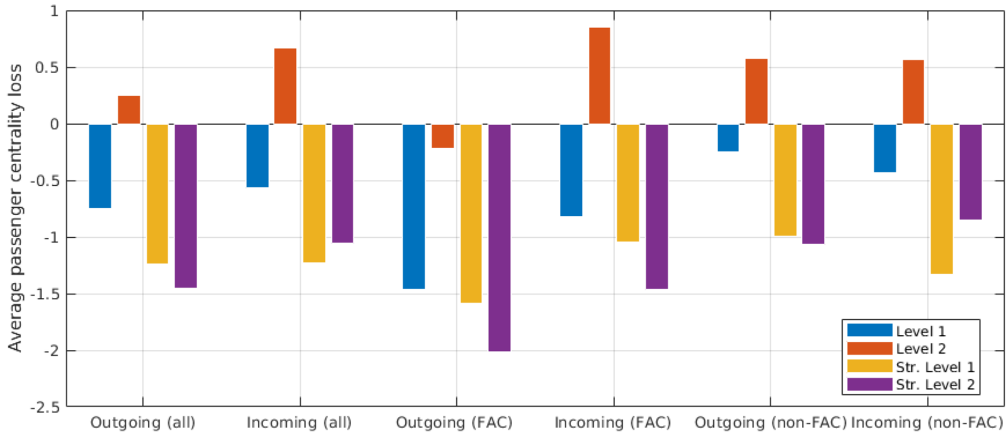


Figure 56. Percentage change average passenger centrality loss FAC with respect to the baseline for the analysis on all airports, on the airports with FAC implemented and on the airports without FAC.

Table 51. Average passenger centrality loss all airports FAC

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Average incoming passenger centrality loss	**186.03**	**184.97**	**187.28**	**201.48**	**199.00**	**199.35**
	177.93	179.51	178.90	192.20	192.39	190.53
	194.01	191.53	195.82	206.50	203.94	209.28
Average outgoing passenger centrality loss	**206.25**	**204.71**	**206.78**	**232.88**	**230.00**	**229.48**
	196.81	197.74	198.36	223.26	222.63	219.21
	215.26	212.68	215.96	238.94	236.14	238.79

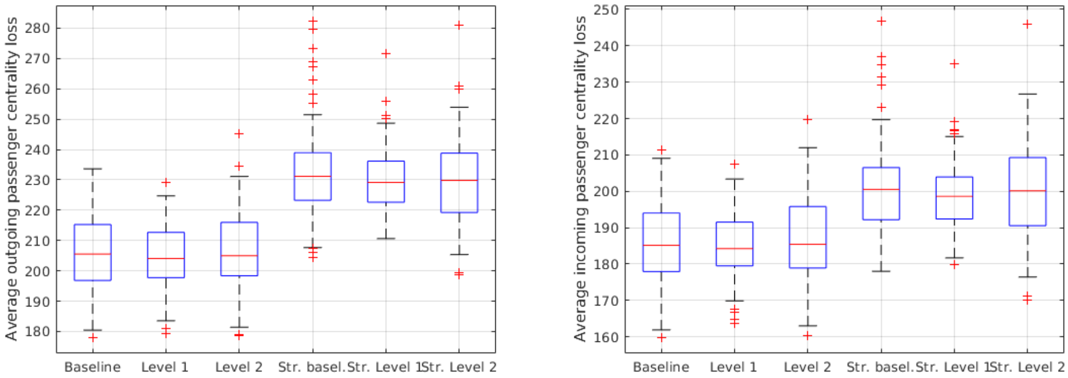


Figure 57. Comparison change average passenger centrality loss FAC outgoing (left) and incoming (right) passenger centrality loss in the scenarios where FAC is implemented at Level 1 and 2 and in the corresponding baselines. For each scenario, the average centrality losses in each of the 100 iterations of the model are considered.

Figure 57 compares the average outgoing and incoming passenger centrality loss in the scenarios where FAC is implemented at Level 1 and 2 and in the corresponding baselines.

Table 52. Average passenger centrality loss all airports FAC airports with FAC

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Average incoming passenger centrality loss (FAC airp)	**2294.29**	**2275.47**	**2313.79**	**2375.99**	**2351.03**	**2341.14**
	2147.29	2157.48	2161.35	2243.21	2232.48	2201.23
	2423.71	2378.42	2432.04	2486.06	2449.94	2453.65
Average incoming passenger centrality loss (non-FAC airp)	**123.87**	**123.34**	**124.58**	**137.37**	**135.55**	**136.20**
	118.02	119.34	119.38	130.96	131.19	129.53
	129.85	128.32	129.04	142.48	140.05	143.15

Average outgoing passenger centrality loss (FAC airp)	**2937.78**	**2894.64**	**2931.30**	**3349.44**	**3296.36**	**3282.05**
	2768.85	2779.27	2776.94	3194.85	3158.48	3128.67
	3105.35	3011.33	3076.79	3478.75	3419.15	3405.71
Average outgoing passenger centrality loss (non-FAC airp)	**125.72**	**125.40**	**126.45**	**140.99**	**139.59**	**139.48**
	120.36	120.96	120.36	134.75	134.87	133.93
	131.06	130.56	132.57	146.17	143.89	145.22

We then restrict our analysis to the set of airports where FAC is implemented, and compare it to the set of airport where it is not implemented, in order to see if there is a "local" improvement. if we look at absolute losses, the improvements are way larger in airports where FAC is implemented rather than where FAC is not implemented (see

Table 52), except for the incoming centrality in Level 2 default, for which there is a worsening. Depending on the scenario, we have, on average in each FAC airport, that 15 to 60 passengers more arrive to their destination following their original itinerary (while for the non-FAC airports results are similar to the all-airports analysis). All changes are, however, not statistically significant.

Since the airports with FAC implemented have a larger flux of passengers, and therefore larger losses, a fairer comparison would be the one of percentage change with respect to the baseline. Figure 56 shows the percentage change of the average centrality loss with respect to the baseline for the analysis on all airports, on the airports with FAC implemented and on the airports without FAC. We see that, especially in the stressed case, also the percentage improvement is larger for the airports with FAC implemented. The difference is more evident for the outgoing centrality, which is expected, given that the FAC mechanism aims to reduce the arrival and reactionary delays, consequently preserving connections at the airport where it is implemented and downstream.

In all cases, Level 2 seems to cause a larger centrality loss than the baseline in the default case, while it brings an improvement in the stressed one. Given that Level 2 focuses on the costs, this could mean that the economic interest of the airline coincided with the passengers' interests only when the system is stressed, as in this case the passenger costs might be comparable to the fuel costs.

In conclusion what we can say is that, concerning the preservation of passengers' itineraries, FAC brings a small overall improvement in the entire network (except Level 2 default), and that, in percentage, this improvement is larger for airports that implement FAC, especially for the outgoing centrality. (However, almost all changes are not statistically significant.)

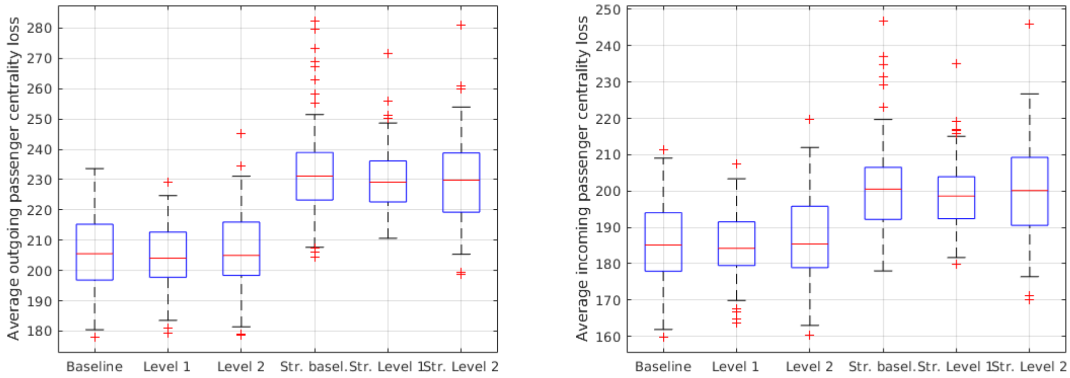


Figure 58. Comparison change average passenger centrality loss FAC restricted flights arrival FAC outgoing (left) and incoming (right) passenger centrality loss in the scenarios where FAC is implemented at Level 1 and 2 and in the corresponding baselines. For each scenario, the average centrality losses in each of the 100 iterations of the model are considered.

Figure 58 compares the average outgoing and incoming passenger centrality loss in the scenarios where FAC is implemented at Level 1 and 2 and in the corresponding baselines on the restricted dataset (only considering airports which implement the mechanism).

Trip centrality

Here we present results concerning changes in trip centrality loss for two different values of α: 0.2 and 0.02. The value of the parameter α determines the relative importance given to walks made of several legs with respect to shorter ones. For example, when $\alpha = 0.2$ a walk made of a single flight weights as much as 5 walks made of two flights, while when $\alpha = 0.02$ it weights as much as 50 walks made of two flights. Therefore, the relative impact of cancellations and connection disruptions on the centrality loss will also differ in the two cases, with the disruption of potential connections having a larger impact when α is larger.

When considering all flights, we see small percentage improvements (i.e., smaller centrality losses) with respect to the baseline in Level 1, and also in Level 2 default for the larger value of α (see Figure 59 and Figure 60). The percentage improvements are larger when α is larger, except for Level 2 default. Note that Level 2 default has a smaller trip centrality loss than the baseline when $\alpha = 0.2$, but a larger one when $\alpha = 0.02$. This might be due to the fact that Level 2 default has an increase of cancellations with respect to the baseline, and this has more impact in the centrality loss when α is small, where it counters the positive effect of the smaller delays in preserving potential connections. However, all changes are not statistically significant. The average centrality losses are reported in Table 53 and Table 54. Note that, when computing the average on all airports, only the loss of outgoing centrality is reported because the incoming is exactly equivalent.

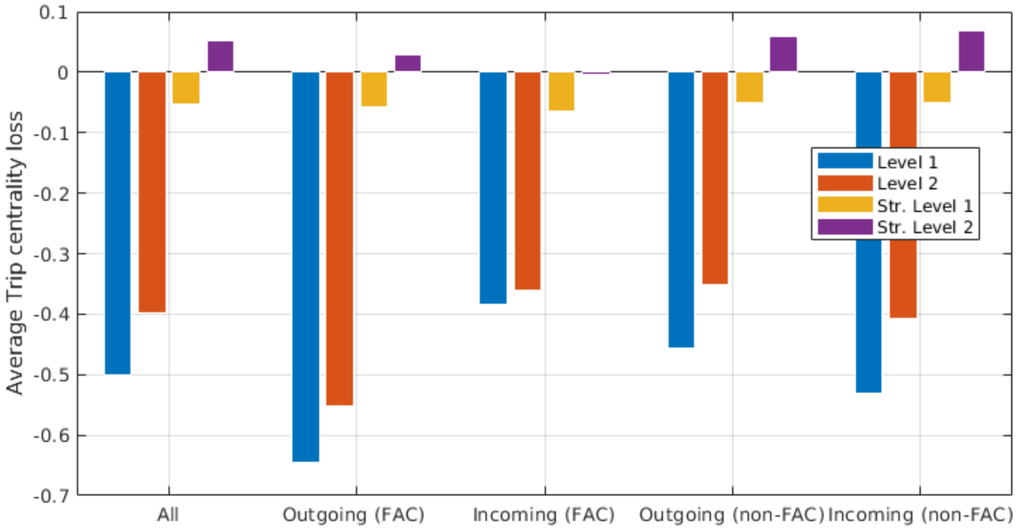


Figure 59. Percentage change of the average trip centrality loss all airports with α=0.2 with respect to the baseline for the analysis on all airports, on the airports with FAC implemented and on the airports without

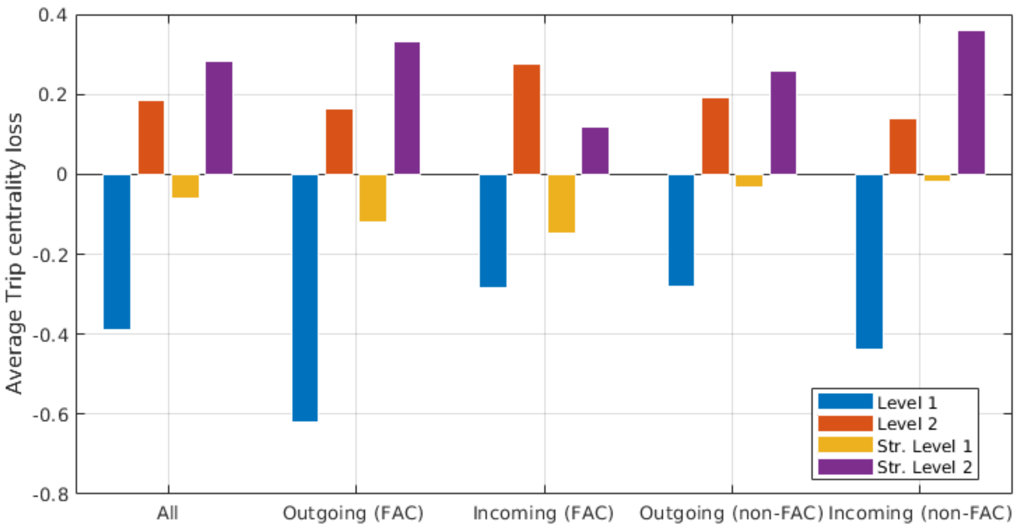


Figure 60. Percentage change of the average trip centrality loss all airports with α=0.02 with respect to the baseline for the analysis on all airports, on the airports with FAC implemented and on the airports without

Table 53. Average trip centrality loss all airports FAC with α=0.2

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Average trip centrality loss	**10681**	**10628**	**10639**	**12801**	**12795**	**12808**
	10509	10464	10462	12732	12724	12745
	10817	10773	10794	12875	12862	12875

Table 54. Average trip centrality loss all airports FAC with α=0.02

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Average passenger centrality loss	**0.531**	**0.529**	**0.532**	**0.732**	**0.732**	**0.734**
	0.523	0.523	0.524	0.725	0.725	0.728
	0.539	0.536	0.539	0.737	0.735	0.739

When comparing the airports where FAC is implemented to those where it is not implemented, at Level 1 FAC is seen to imply a larger improvement in outgoing centrality loss to the airports where FAC is implemented, both in absolute and in percentage, although the changes are very small and not statistically significant (see Tables Table 55 and Table 56, and Figure 59 and Figure 60). At Level 2, for the default case the same effect is seen for =0.2, while in the stressed case and for =0.02 the average centrality losses increase, both in the FAC and non-FAC airports

Finally, we remark that while in the default scenarios the improvements in trip centrality are quite similar to those in passenger centrality, in the stressed scenario the improvements in trip centrality are much smaller. This can be explained by the fact that in the default case, when delays are small, the diminishment of average delays implied by FAC is effective in preserving not only passenger itineraries, but also potential ones. In the stressed case, instead, the diminishment of delays is less effective in preserving itineraries, given that delays are larger, but passenger itineraries are actively targeted by the optimisation mechanism and therefore they are still preserved.

In summary, the FAC mechanism brings a small improvement to potential connections especially at Level 1, and for the outgoing centrality this improvement is more pronounced for airports where FAC is implemented, similarly to what was found for passenger centrality. Differently for passenger centrality, for trip centrality improvements are mostly in the default scenarios, and only small effects are seen in the stressed scenarios.

Table 55. Average trip centrality loss all airports FAC with α=0.2 restricted airports with FAC

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2

Average incoming trip centrality loss (regulated airp)	71056	70581	70656	76465	76301	76467
	69834	69070	69197	75954	75775	75963
	72043	71858	71946	77057	76874	76934
Average incoming trip centrality loss (non-regulated airp)	9655	9609	9619	11085	11082	11092
	9512	9458	9462	11029	11030	11035
	9767	9743	9750	11146	11136	11152
Average outgoing trip centrality loss (regulated airp)	86991	86470	86409	89003	88910	89095
	85460	84990	84565	88398	88332	88552
	88231	88097	88105	89543	89448	89645
Average outgoing trip centrality loss (non-regulated airp)	9385	9339	9351	10747	10742	10751
	9245	9189	9196	10693	10684	10697
	9500	9471	9480	10803	10790	10804

Table 56. Average trip centrality loss all airports FAC with α=0.02 restricted airports with FAC

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2

Average incoming passenger centrality loss (FAC airp)	**5.895**	**5.878**	**5.911**	**8.255**	**8.243**	**8.265**
	5.796	5.814	5.802	8.182	8.170	8.186
	6.011	5.963	6.011	8.315	8.305	8.331
Average incoming passenger centrality loss (non-FAC airp)	**0.373**	**0.371**	**0.373**	**0.510**	**0.510**	**0.512**
	0.367	0.367	0.368	0.506	0.505	0.508
	0.379	0.376	0.378	0.514	0.514	0.516
Average outgoing passenger centrality loss (FAC airp)	**5.921**	**5.884**	**5.931**	**8.384**	**8.374**	**8.412**
	5.827	5.794	5.786	8.295	8.314	8.322
	6.039	5.957	6.049	8.462	8.440	8.505
Average outgoing passenger centrality loss (non-FAC airp)	**0.372**	**0.371**	**0.373**	**0.507**	**0.506**	**0.508**
	0.367	0.366	0.368	0.502	0.503	0.504
	0.378	0.376	0.377	0.510	0.510	0.512

Local analysis in a regional airport

Using centrality metrics, we can answer questions regarding the indirect effect of FAC in airports where it is not implemented. For example, let us consider a regional airport where FAC is not implemented and from which passengers depart to connect in hubs where FAC is implemented. Does such an airport receive an indirect gain because of the implementation of FAC in other airports?

Let us consider the airport of Edinburgh, where FAC is not implemented and which has several flights to and from airports with FAC implemented, among which hubs like Heathrow, Frankfurt, Munich, Madrid, Zurich. Looking at the loss of incoming and outgoing passenger centrality of Edinburgh, we can see if passengers arriving there (either directly of with a connection) and passengers departing from there arrive more often to their destination following their scheduled itinerary when FAC is implemented, with respect to the baseline. The loss of incoming and outgoing trip centrality, instead, tells us whether potential itineraries arriving incoming to Edinburgh or outgoing from it are increasingly preserved when FAC is implemented.

Looking at Table 57 and at Figure 61 we see that in the default scenarios the centrality losses (both passenger and trip centrality, the latter computed with α=0.02) increase with respect to the baseline. Instead, in the stressed scenario the percentage improvements in passenger centrality loss are quite large, especially at Level 2 (while no change is seen for trip centrality).

We can conclude that passengers departing from or arriving to the airport of Edinburgh have a larger probability to arrive to destination with their scheduled itinerary if the system is stressed, but a smaller one if the delays in the system are small, as a result of the implementation of FAC in hubs to which the airports is connected. The "potential connections" counted by trip centrality, which are not addressed explicitly by the mechanism, do not improve even in the stressed case.

Table 57. Average trip centrality loss Edinburgh

Metric	Baseline	Level 1	Level 2	Stressed baseline	Stressed Level 1	Stressed Level 2
Average outgoing passenger centrality loss (EGPH)	**826.47**	**878.17**	**852.41**	**933.39**	**916.79**	**869.03**
	541.00	610.50	563.00	691.50	621.50	573.50
	1037.50	1153.00	1128.50	1197.50	1125.00	1125.50
Average incoming passenger centrality loss (EGPH)	**800.10**	**874.29**	**857.40**	**947.26**	**894.27**	**825.00**
	517.50	539.00	570.00	667.50	584.00	490.00
	937.00	1183.50	1089.00	1205.00	1163.00	1090.00
Average outgoing trip centrality loss (EGPH)	**1.89**	**1.98**	**1.91**	**3.04**	**3.05**	**3.03**
	1.76	1.80	1.72	2.89	2.94	2.90

	1.99	2.13	2.05	3.17	3.17	3.16
Average incoming trip centrality loss (EGPH)	**1.89**	**1.98**	**1.91**	**3.04**	**3.05**	**3.03**
	2.81	2.86	2.85	4.16	4.18	4.13
	3.16	3.25	3.28	4.41	4.43	4.44

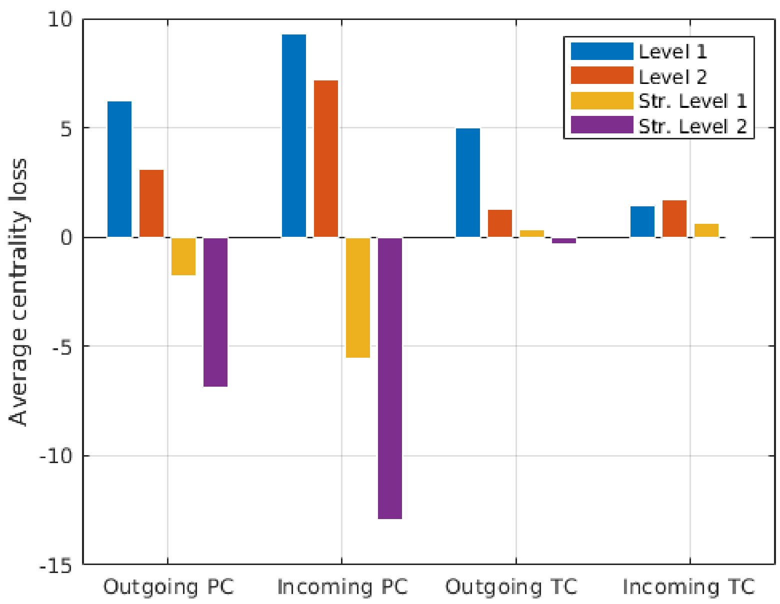


Figure 61. Change of centrality loss with respect to the baseline for the airport of Edinburgh.

4.3.3.5 Causality analysis

In this section, we show the causality analysis applied to the scenarios where FAC mechanism is implemented and we compare the results with the corresponding baselines. In this section, we study the Granger causality (both in mean and in tail) networks built with both the state of delay and the state of congestion of the airports.

Causality in mean

Table 58. Metrics for the Granger causality in mean network FAC (delay) (both default and stressed)

Delay.	Baseline default scenario	Baseline stressed scenario
Link density	0.0043	0.0034

Mean degree	1.09	0.87
Clustering coefficient	0.52	0.53
Over-expression of feedback triplets	908.3	353.7
Reciprocity	0.28	0.23

In this paragraph we show the results obtained for the Granger causality in mean networks.

The selected network metrics measured on the Granger causality in mean network built for the baseline scenario, in both the default and stressed case, are shown in Table 58. We can notice that moving from the default case to the stressed one is associated with a decreasing level of causality (i.e., smaller link density) for the delay propagation mechanism. That is, the higher delays of the stressed scenario tend to be less correlated, thus the endogenous process of delay propagation becomes less important.

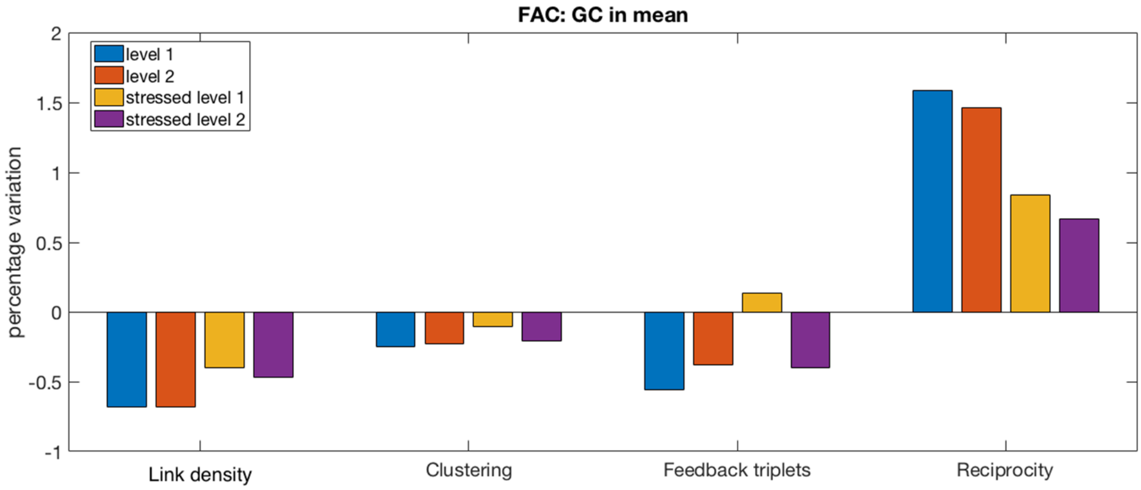


Figure 62. Percentage change Granger causality in mean FAC, i.e., mean degree, clustering coefficient, number of feedback triplets, and reciprocity coefficient, when we compare the scenario with FAC mechanism with the corresponding baseline (for both default and stressed case). (In order to compare metrics between two different scenarios, we consider the measured value of each metric but normalised by the expected value for the random case. This is done for a fair comparison, because link density may vary from one scenario to another).

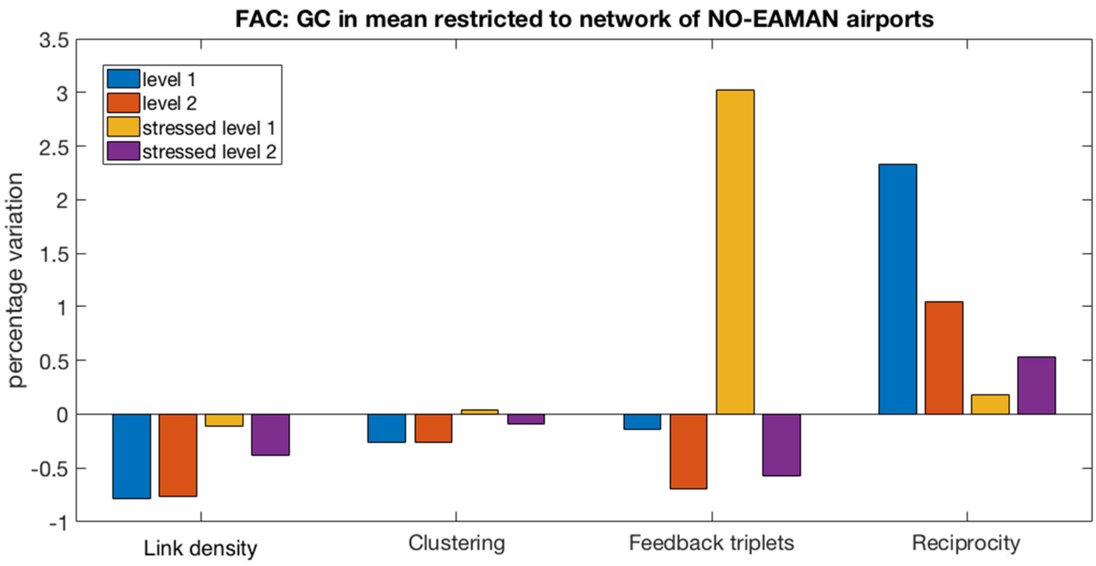


Figure 63. Percentage change Granger causality in mean FAC airports not implanting FAC, i.e., mean degree, clustering coefficient, number of feedback triplets, and reciprocity coefficient, when we compare the scenario with FAC mechanism with the corresponding baseline (for both default and stressed case) but restricting to the subgraph formed by node-airports where E-AMAN is not implemented. (In order to compare metrics between two different scenarios, we consider the measured value of each metric but normalised by the expected value for the random case. This is done for a fair comparison, because link density may vary from one scenario to another

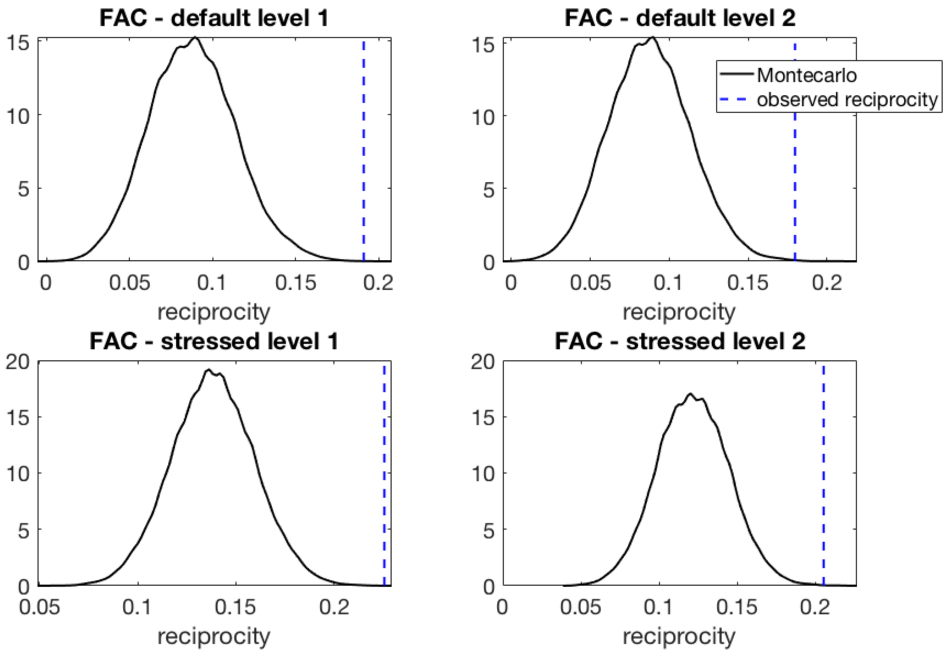


Figure 64. Distribution of reciprocity Granger causality in mean FAC according to Monte Carlo simulations compared with the observed reciprocity for the FAC scenarios in the case of Granger causality in mean networks built with the state of delay.

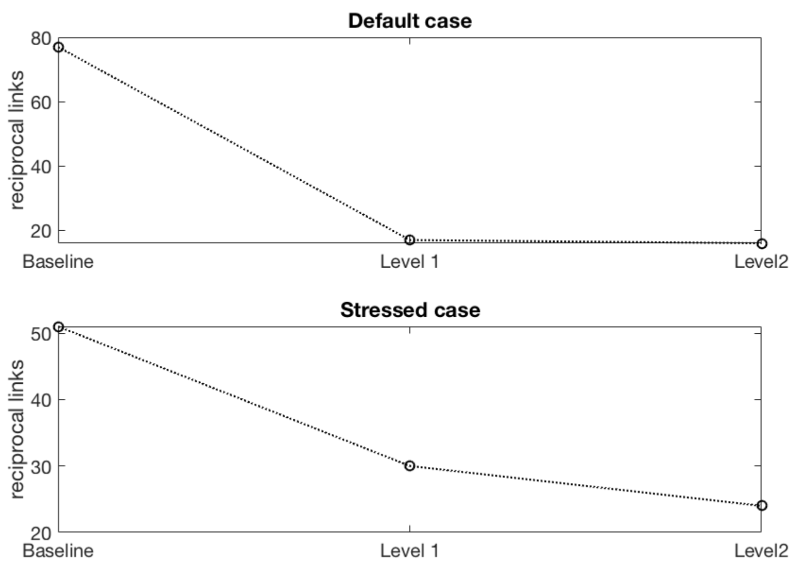


Figure 65. Number of reciprocal links (GC in mean network built with the state of delay) at the different levels of implementation for both the default and stressed case. FAC

We measure the value of the same network metrics in Table 58 for the causality networks obtained for the scenarios where the FAC mechanism is implemented and show the percentage variations with respect to the corresponding baseline in Figure 62 and Figure 63. Let us notice the overall decrease of link density at any level of implementations for both the whole network (Figure 62) and the subgraph obtained for nodes representing airports where E-AMAN is not implemented (Figure 63). The decrease of link density can be interpreted as a decrease of the level of causality, i.e., a smaller number of propagation channels for the delay. With the exception of the number of feedback triplets in the stressed scenario with FAC at Level 1, we observe a decrease of the over-expression of the clustering coefficient, the number of feedback triplets, and the reciprocity. Finally, in Figure 62 and Figure 63, we show that reciprocity is inversely correlated to the variation of link density. Similarly to the 4DTA mechanism, in order to better characterise this effect, we consider this simple Monte Carlo simulation: (i) given the network for the baseline scenario, and (ii) observed a given decrease of link density for one scenario with implemented FAC, (iii) we can randomly erase some links from the baseline network in order to target the same number of links of the network for the FAC scenario, and, finally, (iv) measure the value of reciprocity. The result of this simple Monte Carlo experiment is shown in Figure 64 for the Granger causality in mean networks built with the states of delay. Note that the expected measure of reciprocity according to Monte Carlo simulations is always smaller than the observed one. This suggests that: (i) the overall number of reciprocal links decreases by following the negative variation in the total number of links (see Figure 65), but (ii) the over-expression of reciprocity is inversely correlated with link density because (iii) some reciprocal subsystems are unaffected from FAC.

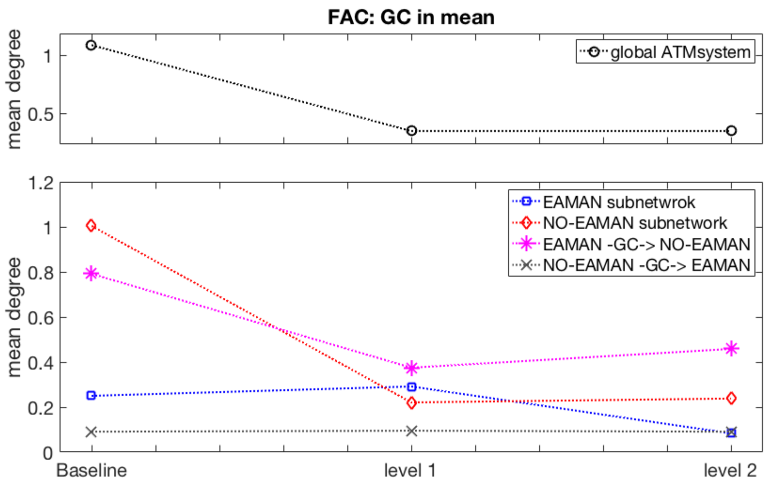


Figure 66. Change of the mean degree (for GC in mean) when we move from the baseline to the scenario with implemented FAC mechanism at Level 1 and Level 2 (default case). We obtain qualitatively similar results for the stressed case

When applying GC in mean to the state of delay of airports, we can notice an overall decreasing of the level of causality (in mean) when we move from the baseline to any scenario with FAC implemented. In particular, airports where E-AMAN is implemented (Granger) cause less the airports without the FAC mechanism (see Figure 66). A smaller level of causality is observed also in the subnetwork of airports where E-AMAN is not implemented, thus highlighting an improvement for the global system as a secondary effect of applying E-AMAN in the subset of the 24 airports (characterised by high traffic). This externality represents an overall improvement in the performance of the aggregate in terms of delay propagation.

Causality in tail

Table 59. Metrics for the Granger causality in tail network FAC (delay), for the baseline scenario (both default and stressed).

Delay	Baseline default scenario	Baseline stressed scenario
Link density	0.1594	0.3460
Mean degree	40.5	87.9
Clustering coefficient	0.43	0.56
Over-expression of feedback triplets	3.7	1.9
Reciprocity	0.21	0.28

In this paragraph we show the results obtained for the Granger causality in tail networks, i.e., we study the channels of propagation of extreme delay events (Table 59).

The selected network metrics measured on the Granger causality in tail network built with simulated data for the baseline scenario, in both the default and stressed case, are shown in Table 2. We can note that link density in the stressed scenario is about twice with respect to the default case. We argue that this difference is because of one leg effects which strongly correlate in the subnetwork of low traffic airports, see Figures 6 and 7. The largest number of causal links is among NO-E-AMAN airports which are largely the ones with low traffic. Here, a highly delayed flight affects importantly the state of delay of the airport, thus its state of congestion, because of the small number of flights departing from it within the 1 hour time window. Then, it is likely that this extreme delay is propagated by that aircraft to the airport where it will land and depart again. In the stressed case, when we observe more extreme delays, the described behaviour tends to increase the number of causal relations between low traffic airports.

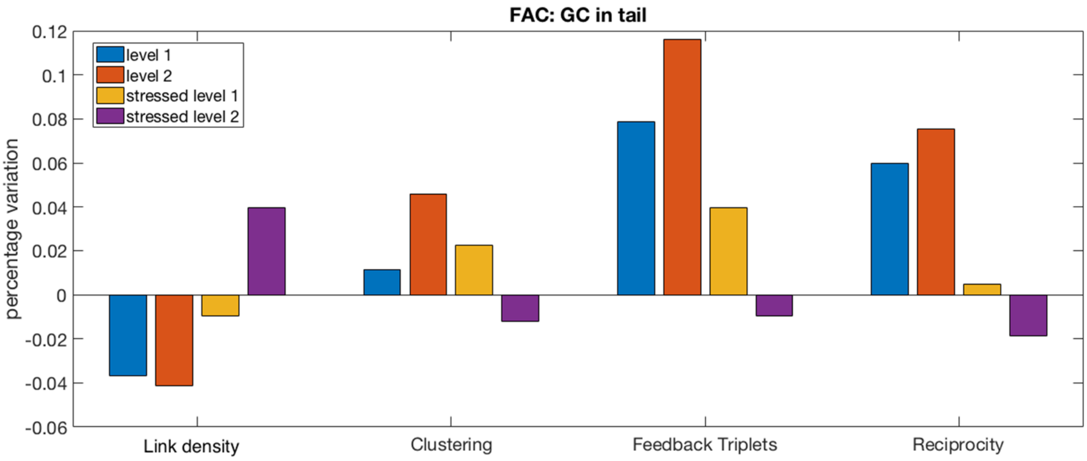


Figure 67. Percentage change of Granger causality in tail network, i.e., mean degree, clustering coefficient, number of feedback triplets, and reciprocity coefficient, when we compare the scenario with FAC mechanism with the corresponding baseline (for both default and stressed case).

We measure the network metrics (clustering, feedback triplets and reciprocity) for the Granger causality in tail networks obtained for the scenarios where the FAC mechanism is implemented and, in Figure 67, show the percentage variations with respect to the corresponding baseline. Again, it is evident the inverse dependence between changes in link density and variations of the over-expressions of the network metrics measured on the FAC scenarios at any level of implementations. As before, this is a signal of the over-expression (w.r.t. to the random case) of the amplifying subsystems which are preserved on average by the implementation of the FAC mechanism, both at Level 1 and 2 and for the default and stressed cases.

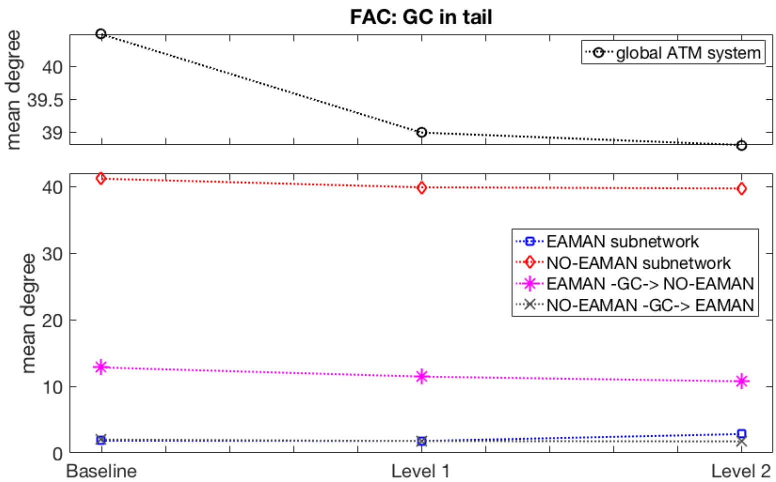


Figure 68. Change of the mean degree (for GC in tail) FAC when we move from the baseline to the scenario with implemented FAC mechanism at Level 1 and Level 2 (default case)

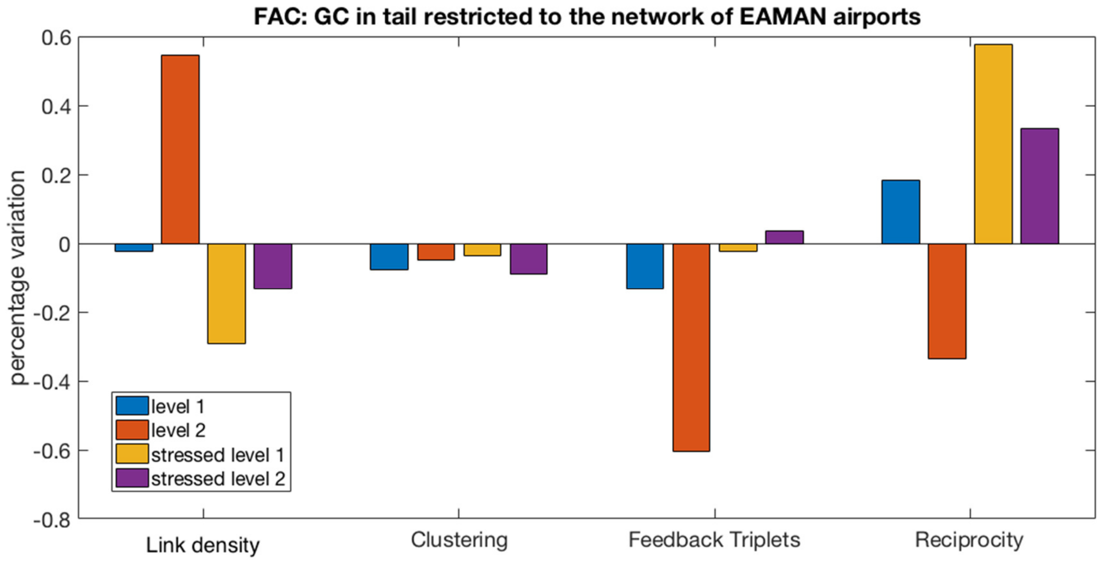


Figure 69. Percentage change of Granger causality in tai network only airports with FACl, i.e., mean degree, clustering coefficient, number of feedback triplets, and reciprocity coefficient, when we compare the scenario with FAC mechanism with the corresponding baseline (for both default and stressed case).

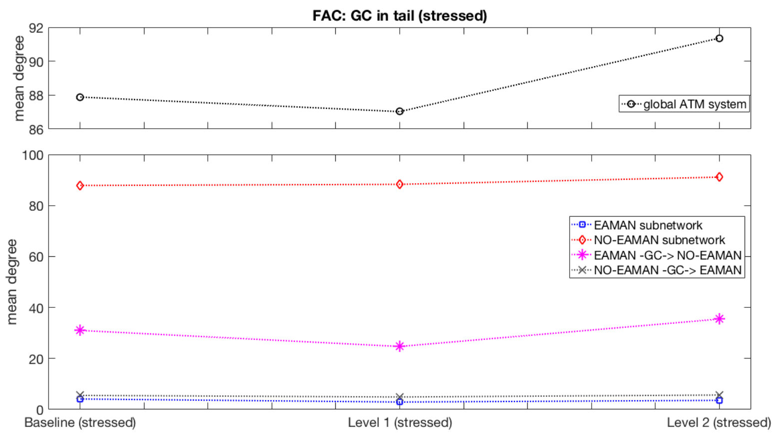


Figure 70. Change of the mean degree (for GC in tail) stressed case FAC when we move from the baseline to the scenario with implemented FAC mechanism at Level 1 and Level 2 (default case)

We observe a similar behaviour of GC in mean in relation to the variation of the mean degree of the network when moving from the baseline to FAC scenarios. Let us focus on the default case in Figure 68. We notice a decrease for the level of causality, which is measured as a reduction of the mean degree at the global level. In particular, this is due to less causal links from the E-AMAN airports to the NO-E-AMAN ones (purple stars in Figure 68) and in the subnetwork of airports where E-AMAN is not implemented. Interestingly, this improvement for the system (i.e., less channels of extreme delay propagation) is associated with an increase in the over-expression (w.r.t. the random case) of triangles, feedback triplets, and mutual links. However, this behaviour is not observed in the subnetwork of airports where E-AMAN is implemented, see Figure 69. Here, we measure (with the exception of reciprocity at Level 1) that the subsystem of E-AMAN airports perform better than the corresponding default baseline, and there are less subsystems amplifying delay propagation. Hence, in the subnetwork of E-AMAN airports, FAC mechanism has a positive impact overall from the point of view of delay propagation. Finally, when we consider the stressed scenarios, the implementation of FAC at Level 1 exhibit results similar to the default case but the implementation at Level 2 show quite the opposite (as we can expect because the link density increases), see Figure 70.

Comments

The FAC mechanism tends to increase the overall stability to the propagation of both delays and congestions among airports (because the overall link density decreases). By looking at the whole causality networks, FAC tend to preserve the subsystems amplifying the (extreme) delay, whereas, in absolute terms, the total number of mutual causal links decreases. Interestingly, by restricting to the subgraph of E-AMAN airports, the over-expression of clustering and feedback triplets decrease (on average) too, for both GC in mean and GC in tail. Hence, FAC represents an overall improvement for the restricted subsystem of airports where E-AMAN is implemented.

5 Conclusions

This deliverable presented the results obtained with the first investigative case studies. The model is running smoothly on a European scale and is able to simulate 14 scenarios already. Its output can be analysed with powerful tools developed by the project, in particular with the new metrics of causality and centrality proposed in D5.1 [3].

The scenarios simulated for this deliverable are focused on trying to understand the impact of each of the mechanism (4D Trajectory Adjustments, Flight Prioritisation and Flight Arrival Coordination). These mechanisms are implemented at different levels, which gives some insight on the potential benefits on corresponding SESAR solutions. Moreover, the mechanisms are tested on two different baselines: a standard (default) one corresponding to a normal traffic volume and a stressed one, with high delays. This allows us to understand the benefits of the mechanisms in different operational environments. This is particularly important since higher delays are expected in Europe in the future due to increasing traffic volumes.

The calibration of the model has been performed only approximatively for this deliverable, due to foreseen changes in the code. It has been calibrated using averages on a few key metrics, including departure delay. The point of comparison is the historical operational day for which the traffic is set (12th September 2014) and 2017 CODA reports [1]. When studying the output of the calibrated models, a few key points can be noted, including the fact that flights plan trajectories slightly too long and operate them too quickly, probably due to small taxi times and congestion at arrival. Some distributions of delays are also notably different, in particular the departure delay, where negative delays are absent from the model. Finally, the cancellation rate on this deliverable is significantly higher than it should be due to the introduction of explicit cancellation for missing curfews. This will be revised in the final version of the model.

Many metrics can be computed from the output of the model due to its low-level detail. This includes standard metrics, such as average delays, but also other advanced metrics developed in this project specifically to capture subtle network effects in the system. The first one is a centrality measure able to indicate how a node in the network (an airport in this study) is important for the connectivity of the network. The second one is able to reflect indirect causality relationships between nodes in the network, in particular helping at understanding delay propagation. These metrics were defined in D5.1 [3] and applied now to the full outcome of the European wide model. Some of the new advanced network metrics have already show their potential to be used in a more operational manner. This will be explored further in the next activities of Domino, for example by relating them to other indicators, such as those that are cost-based.

The analysis of the result has been structured by contrasting the different mechanisms, their different levels of implementation, and their effect in default and stressed baselines. Although lots of detailed results are interesting *per se*, one can extract a few key ones. First, it seems that in the stressed case, all levels and mechanisms improve the cost impact on the airlines. For some of them, this is counter-balanced by a deterioration of other KPIs, like average arrival delay. It happens also that passengers might experience a worsening on their trip experience as arrival delays might increase in some cases. The centrality of airports usually improves with the introduction of the mechanisms, even if sometimes the airports not concerned directly by the mechanisms see a deterioration of their situation. The results with causality are less unanimous, with some improvement or deterioration depending on the exact metrics and the airports considered. Overall, the mechanisms seem to have a lot more positive impact in a stressed environment, which indicates that they would be beneficial to implement in the future, where the traffic volumes are much higher. All these results indicate that, even if the model needs finer calibration and improvements, it is already able to capture the intricate effects coming from the massive number of interactions and the tight connection of the system elements, and to reflect the impact of changes in the ATM environment.

Finally, it is important to understand that the results presented in this deliverable have been useful as part of the validation of the model and in the process of obtaining feedback from experts and stakeholders. This will be used to improve the model and target the final scenarios in Domino (see Section 6, Next steps and look ahead, for more details).

6 Next steps and look ahead

D5.2 presents the results of the first investigative case studies focusing on the effect of the mechanism independently implemented. These results have been shared with experts and stakeholders in dedicated workshops (see D6.3 Workshop results summary). The feedback obtained from these activities along with the outcome of the validation of the model and the analysis of the mechanisms presented in this deliverable are the base of the changes in the model and the definition of new scenarios that will be presented in D3.3 - Adaptive case studies description (expected August 2019). D3.3 will become the blueprint of the activities that will be done until the end of the project including: the evolution of the model (with its validation), modifications on the metrics (with special focus on their operability) and selection of scenarios to be analysed to highlight the capabilities of the model and metrics.

The database structure that has been used to store the input and output of the model (including the needed adaptation to incorporate the changes considered for the final model implementation) will be reported in D2.2 - Database structure (due July 2019).

D3.3 will be the starting point of the changes that will be performed to the model which will be released in D4.2 - Model source code (due September 2019). This final model version will be executed to generate the output that will be analysed and presented in D5.3 - Final tool and model description and case studies results (due October 2019). The final project results will be summarised in D1.2 - Final Project Results Report and delivered with the Project Dissemination Report (D6.4) closing the project by November 2019.

7 References

[1] CODA, “CODA Digest 2017. All-causes delay and cancellations to air transport in Europe - 2017,” 2018.

[2] Domino Project Consortium, “D3.1 Architecture definition,” 2018.

[3] Domino Project Consortium, “D5.1 Metrics and analysis approach,” 2019.

[4] Domino Project Consortium, “D6.3 Workshop results summary,” 2019.

[5] Domino Project Consortium, “D3.2 Investigative case studies description,” 2018.

[6] Domino Project Consortium, “D4.1 Initial model design”.

[7] European Commission, “Commission implementing regulation (EU) No 716/2014 of 27 June 2014 on the establishment of the Pilot Common Project supporting the implementation of the European Air Traffic Management Master Plan,” 2014.

[8] CODA, “IATA Summer Season 2010. Taxi-In (TXI) times in minutes,” 2011.

[9] CODA, “IATA Summer Season 2010. Taxi-Out (TXO) times in minutes,” 2011.

[10] SESAR Joint Undertaking, “E.02.14 - DCI-4HD2D D3.2 Final Technical Report - 00.00.01,” 2016.

[11] EUROCONTROL, “DDR2-Webportal,” 8 June 2018. [Online]. Available: https://www.eurocontrol.int/ddr.

[12] SESAR Joint Undertaking, “E.02.06 - POEM-D6.2 Final Technical Report,” 2013.

[13] SESAR Joint Undertaking, “E.02.06 - POEM-D4.2 Consolidated design review an models refinements,” 2012.

[14] Airbus, “Performance Engineering Programme,” 2014.

[15] CODA, “CODA Digest 2014. All-causes delay and cancellations to air transport in Europe - 2014,” 2015.

8 Acronyms

4DTA: 4D Trajectory Adjustment mechanism

ABM: Agent-based model

AIRAC: Aeronautical Information Regulation and Control

AMAN: Arrival Manager

ANSP: Air Navigation Service Provider

ATC: Air Traffic Control

ATFM: Air Traffic Flow Management

ATM: Air traffic management

CI: Cost Index

COBT: Calculated Off-Block Time

CODA: Central Office for Delay Analysis

DCI: Dynamic cost indexing

DDR2: Demand Data Repository

E-AMAN: Extended Arrival Manager

FAC: Flight Arrival Coordination mechanism

FP: Flight Prioritisation mechanism

G2G: Gate to Gate

GC: Granger Causality

KPI: Key Performance Indicator

Q-Q: Quantile-Quantile

TOC: Top of Climb

TOD: Top of Descend

-END OF DOCUMENT-