D3.3 Adaptive case studies description

Deliverable 3.3

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:	17 December 2019
Edition:	01.01.00

Founding Members

EUROPEAN UNION EUROCONTROL

SESAR JOINT UNDERTAKING

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Luis Delgado / University of Westminster	Project member	10 December 2019
Gérald Gurtner / University of Westminster	Project member	10 December 2019
Silvia Zaoli / Università di Bologna	Project member	10 December 2019
Piero Mazzarisi / Università di Bologna	Project member	10 December 2019
Damir Valput / Innaxis	Project member	10 December 2019

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Graham Tanner / University of Westminster	Project member	17 December 2019
Andrew Cook / University of Westminster	Project coordinator	17 December 2019
Fabrizio Lillo / Università di Bologna	Project member	17 December 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	17 December 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	11 October 2019	Release	Domino Consortium	New document for review by the SJU
01.01.00	17 December 2019	Release	Domino Consortium	Incorporating SJU comments

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 improvement planned to be performed until the end of the project regarding the model (implementation changes, recalibration and the simulation outputs), plus the metrics and scenarios that will be re-run with the model. These changes are based on the insights gathered through the analysis activities performed in the scope of investigative case studies (see D3.2 Investigative case studies description and D5.2 Investigative case studies results) and the feedback obtained from experts and stakeholders on the different workshops activities performed (see D6.3 Workshop results summary). These insights highlighted missing features of the model and potential improvements, as well as some gaps and shortcomings.

The scenarios for this analysis have been chosen highly selectively in order to prioritise the depth of the analysis and methodology development over a large number of scenarios, as these have already been analysed in the scope of the investigative case studies.

Table of Contents

List of tables

Executive summary

Based on the insights gathered through the analysis activities performed in the scope of investigative case studies (see D3.2 Investigative case studies description and D5.2 Investigative case studies results) and the feedback obtained from experts and stakeholders on the different workshops activities performed (see D6.3 Workshop results summary) a set of modifications that are required by the model, the metrics and the selection of the final scenarios have been identified.

The model needs to be upgraded on three different aspects:

- Changes to the model to improve the model capabilities and reliability,
- Changes to the calibration of the model, in particular to the cancellation rate and the performance of some mechanisms,
- Improvements to the output generated by the model to facilitate its post-analysis.

None of these changes are critical and mainly focus on providing a more robust, reliable output which can then be analysed in more detail.

Changes are expected to the metrics computed in Domino. These are driven by the feedback obtained from the consultation with stakeholders and experts (reported in D6.3). Modifications to classical metrics will focus on improving passenger indicators, further work will be performed on the complex network metrics (centrality and causality) to improve their operability.

Finally, case studies have been defined in order to reduce the number of scenarios but increasing the depth of the analysis. The objective is to target relevant operational questions that stakeholders could consider when analysing the impact of introducing modifications to the ATM system.

1 Introduction

1.1 Final stages in Domino

Deliverable D3.2 "Investigative case studies description" [5], defined the full set of scenarios that could be analysed within Domino. A selection of these scenarios (based on consultation with stakeholders reported in D6.2 "Stakeholders consultation on system and investigative case studies" [7]) were modelled (based on the model designed and presented in D3.1. "Architecture definition" [1]) and D4.1 "Initial model design" [6], and analysed (see D5.2 "Investigative case studies results" [9]). The results of these analysis were presented back to experts and stakeholders (see D6.3 "Workshop results summary" [10]). From these consultation activities a set of priorities were drawn. In particular, the fact that network metrics are considered of interest by the community but further research is needed in order to identify how they could be used operationally. The simulations performed with the first version of the model, along with the calibration, helped us to identify a set of shortcomings that should be addressed for the final version of the model. This deliverable thus presents the planned modifications that will be made to the ABM model and the metrics, and which scenarios will be analysed for the final version of the project.

Several shortcomings and missing features of the model have been observed, and bottlenecks for the analysis were identified. In this deliverable, we start by presenting those shortcomings and propose solutions for each one, additionally assigning to each shortcoming a priority level (high, medium, low). We classify them into one the three observed categories:

- changes that need to be performed in the model (i.e., in the code) in order to enhance its capabilities and cover noticed missing features;
- recalibration of certain parameters that will ensure more realistic behaviour of some agents; and,
- recoding additional metrics as part of the model output that will facilitate the process of the analysis of the results and yield more complete and certain insights.

Additionally, we present the improvements to the metrics that we are planning to make, motivated by the goal of making the advanced network metrics more transparent and relating them more closely with some of the commonly used operational metrics.

The set of scenarios, which the updated model will run and the results of that are going to be analysed using the evolved metrics, is presented. The capability of the model to represent complex interactions at the ECAC level has already been presented in the analysis presented for the investigative case studies. Therefore, for this final modelling, we prioritise the detail of the analysis of the output of the model over the number of different scenarios. Based on the output from D5.2, the most interesting scenarios have been selected and focus will be put on the capabilities of Domino's metrics (both classical and those drawing on complex network theory) to answer specific questions about the system performance that stakeholders can consider.

1.2 Structure and contents of this deliverable

Section 2 presents the planned Domino model evolution (including changes to the model, calibration and output). The modifications of the metrics planned for the final version of the project are collected in Section 3. Section 4 presents the scenarios that will be analysed. The deliverable closes with next steps and a look ahead in Section 5.

2 Model evolution

The Domino model implemented and used to generate the results for D5.2 "Investigative case studies" [9] already introduced most of the functionalities needed to fully model the ATM system. However, a set of shortcomings were identified during the analysis of the results. In particular these can be grouped as per:

- **Model changes**: Actual modifications needed to the model capabilities to better capture some of the system performances and capabilities;
- **Model calibration**: Need to recalibrate some parameters to ensure that the model is validated;
- **Model output**: The analysis of the results highlighted the need to improve some of the output variables of the model in order to facilitate and expedite the analysis of the scenarios.

Table 1 summarises the main changes needed for the different groups with an impact, criticality and priority associated to them. The impact is defined as follows:

- Low: No expected significant variation on results from these functionalities; they add further capabilities to the model.
- Medium: Some results may be affected by these functionalities; the main findings are expected to be stable without them.
- High: Significant expected variation on the results related to these functionalities.

The criticality is defined as:

- Low: Few scenarios affected by these functionalities.
- Medium: Baseline scenarios affected by these functionalities.
- High: Baseline and case study scenarios affected by these functionalities.

The priority has been defined considering these impacts and criticality, to ensure the validity of the results and the added value to the Domino capabilities. Note that for the outputs, the priority in some cases is defined considering the simplification of the analysis of the outputs produced, rather than the impact on the validity of such results. Some of the changes required are the result of the feedback obtained from the consultation activities reported in D6.3 [10].

Table 1. Summary of the evolution of the Domino model

Part	Change	Brief description	Impact	Criticality	Priority
Model changes	Curfews	• Introduction of 'hard' and 'soft' curfews[1] • Consideration of curfew expected costs by airlines during their decision process • Better information on which airports apply curfews	High	High	High
	Route clustering	• Increase number of alternative routes between origin and destination pairs	Low	Medium	Low
	Cruise wind uncertainty	• Add uncertainty to the wind encountered by flights during the cruise phase	Medium	High	Medium
	Other miscellaneous changes	• ATFM probabilities for airspace issued regulations (not explicitly modelled) in stressed scenarios • Ensure flights are within flying envelope on all points • General code optimisation	Low	High	Low
Model calibration	Cancellation rate	• Adjust the cancellation rate to ensure it is aligned with historical reported values	High	High	High
	Flight time	• Adjust uncertainty parameters managing the flight (including airport capacity/holding times, cruise wind)	Medium	High	Medium
	ATFM delay maximum values	• Recalibrate ATFM distributions to avoid too high ATFM delays being issued	Medium	Medium	Medium
	4DTA mechanism calibration	• Adjust number of flights and velocities used to recover delay in Level 0 • Adjust number of flights that decide to slow down in Level 2 (i.e., the criteria by which flights make that decision)	High	Medium	High
Model output	Output formats	• Generate the output (simulation results) in a format which is suitable to be directly used in the post-analysis	Low	High	High
	Flights output	• Add explicit information on the reactionary delay of the flights	Medium	High	High

[1] See Section 2.1.1

Part	Change	Brief description	Impact	Criticality	Priority
	Passengers output	• Directly provide information on planned vs. executed passenger itineraries • Incorporate information on missed connections and waiting times	Medium	High	High
	4DTA output	• Ensure decision making points are recorded for all flights to facilitate comparison across mechanism levels • Record information of the flights that perform wait for passengers and wait times • At Level 2, separate the delay information regarding WFP and DCI (although observed conjointly)	Medium	High	High

2.1 Model changes

The impact of the changes will be carefully tested to avoid destabilising other parts of the model.

2.1.1 Curfews

Airport curfews may have a high impact on the airlines' business model. Indeed, the curfew may place an additional constraint on the cost function by impelling an airline to try to get all the aircraft to 'position' (i.e., where they were planned to be). As reported in [1], the application of curfews can be very complex. For example, the curfew may be active only for a certain type of aircraft, for flights coming from a certain direction, for departures, for arrivals, etc. This complexity is driven by the fact that some of these limitations are related with environmental practices (e.g., noise pollution). We can summarise the impact of infringing curfews, i.e., landing after a certain time of the day, according to two categories:

- a penalty or fee to pay ('soft' curfew),
- a rejection of the flight plan, i.e., flights cannot be planned to land after a certain time in some airports ('hard' curfew).

Note that these terminologies are not official, standard terms, and the Domino team is currently investigating the extent of their formal status, and how they are applied at various airports, through consultation with industry experts and EUROCONTROL [10]. This information will be incorporated into the model so that airlines can estimate the expected cost at the end of the day, by the propagation of delay from earlier flights. Further work across this research area is needed, and curfew data needs to be made available, to enable comparisons between projects.

In the current implementation, all curfews are considered 'hard', i.e., flights are not allowed to submit a flight plan which is expected to arrive at the destination after the curfew time. The delay costs need to take into account the costs associated with cancelling a flight overnight (including associated passenger costs such as re-booking and duty of care (including accommodation for passengers)) and ending the day at a different location than planned, resulting in aircraft 'out of position'.

2.1.2 Route clustering

The model uses a restricted number of potential horizontal trajectories between each origin-destination pair, as explained in D5.2. These flight plans are chosen based on a clustering method, used on empirical flight plans in order to reduce the high variability of these trajectories to keep only a few typical ones.

As described in D5.2, the clustering has been performed first by putting all the trajectories following the same sequences of ANSP airspaces together. Then, using a simple distance function based on the total trajectory length, each bundle was run through a clustering algorithm (kernel density estimation). This function gives some good results, but suffers from some shortcomings. For instance, different examples of asymmetric trajectories were considered to be the same, because they had roughly the same total distance. As a consequence, a better clustering method could be used. The project ER3 ADAPT, in which UNITS and UoW both participated, has dealt with similar issues and performed a similar analysis. Domino will thus adapt some of these procedures for its own model.

This is however, a low priority on the model development as the results provided are already in line with expected route choice. More alternatives will be beneficial but we don't expect to see changes on the overall system performances as flight indicators (e.g., time and fuel) will be equivalent.

2.1.3 Cruise wind uncertainty

As reported in D5.2, the trajectories are operated assuming an average wind between each of the ANSPs of the origin-destination airport pairs. The model provides the possibility to generate some uncertainty around the actual cruise wind encountered by the flights. However, this functionality has not been activated, i.e., the simulated, executed average wind is the same as the planned one. This average wind is based on historical, estimated average cruise winds between origin and destination ANSPs. The calibration carried out in the final version of the model will select which percentile of the wind to select to ensure that the airborne time is properly calibrated with historical datasets.

2.1.4 Other miscellaneous changes

Other changes that are required in the code include:

- To adjust the probability of having ATFM delay when it is not explicitly modelled as a regulation (i.e., when the delay is issue for regulations in the airspace) for the stressed scenarios.
- General optimisation of the code. This will be done considering the procedures which have the higher computational execution time and considering their optimisation.
- Ensure that flight parameters are within the flight performances as defined by the BADA 4 models.

2.2 Model calibration

Once the model has been updated, the calibration processes will be redone. This will be performed considering historical data from EUROCONTROL CODA [1, 3] the main reference. In the analysis of the results produced for D5.2 a few specific points were already highlighted as requiring further calibration.

2.2.1 Cancellation rate

The model incorporates a probability of cancellation which should be calibrated considering historical reported data. However, as some cancellations were explicitly modelled (i.e., due to curfew), the cancellation rate that the model generated was too high for D5.2. In this final version, once the implementation of the curfew has been adjusted as described in Section 2.1.1, the cancellation rate will be adjusted to ensure that the model is properly calibrated.

2.2.2 Flight time

The flying time will be re-calibrated considering the actual reported times between take-off to landing from DDR2 data and reported delay reasons. This is relevant to consider adjustments if needed on airport arrival capacity and flight uncertainty. This will be done once the introduction and calibration of wind have been performed (see Section 2.1.3).

2.2.3 ATFM delay

The distributions of ATFM delay will be recalibrated to ensure that unrealistic high delays are not issued to flights or maintained below the expected threshold based on historical data (i.e., analysing ATFM delay from DDR2 data).

2.2.4 4DTA mechanism

The results analysed in D5.2 highlighted that for the 4DTA mechanism at Level 0, when DCI is performed, the speeds selected tend to be unrealistic high. The decision on Level 0 is not based on the cost but just on trying to recover the delay up to a given threshold. This led to delay recovery and fuel consumptions abnormally high in the baseline scenarios. For this reason, 4DTA mechanism should be recalibrated to consider, up to a given point, the expected amount of fuel required to recover the delay so that the maximum speeds are limited. This was validated with feedback obtained from stakeholders, as reported in D6.3 [10].

Similar to Level 0, 4DTA mechanism at Level 2 that allows a flight to slow down at the top of the climb in case it expects to arrive 15 minutes before the expected arrival time should be recalibrated, according to the results analysed in D5.2. As it turned out, a large number of flights that decide to do so end up arriving late, and as this decision is not based on the cost function optimisation as the other ones, this sub-mechanism needs to be recalibrated and potentially even implemented in a more intelligent way.

2.3 Model output changes

The results of the model are stored in a SQL database (on a MySQL sever). This has certain beneficial properties, such as data consistency, the selection of data can be filtered, it minimises data duplicity and facilitates the joining of output and input data. However, when the results produced were analysed in D5.2, a set of drawbacks were identified. These can be summarised as shown below.

- Execution time required to store the results in the database: As the model produces many low level indicators for the output of the mechanisms, flights and passengers, the amount of output data is significantly large. In particular for passengers' output, the model produces a registry per passenger itinerary. This mean that the output of the passengers was very large and required a large amount of time to be stored in the adequate output table. As several simulations were run in parallel, this led to concurrence issues which affected the overall execution times.
- Many low level output results: Many of the output metrics generated by the model are very low-level (e.g., time, fuel, distance used for climb, cruise, descent of planned and executed trajectories) and too detailed for the system performance analysis carried out in Domino. They were required for the verification and validation activities and in some cases can be relevant to understand the dynamics of the model. However, they increase the output size significantly and when focus is put at stakeholders' KPIs and network metrics, they might not be required.
- Difficulty to retrieve the output data for post-processing: Some of the output data produced by the model (in particular the passenger output) is very large. The difficulty accessing the data as queries requires a long time to execute, and data acquired.
- Difficulty to join executed (output) data with planned (input) data: Each scenario is executed several times to allow us to perform statistical analysis of the results. This leads to large datasets which are difficult to join back with the input data. This in its turn difficult the assessment of variations between planned and executed flights and passenger itineraries.
- Analysis of output data with Matlab: Matlab has been used to compute the network metrics. To simplify the access to the data, the output which was stored in the SQL database, was extracted into csv files which were then loaded into Matlab.

Considering these limitations, some changes will be performed to the output produced by the model.

2.3.1 Output format

The model will be modified to enable the selection of which format to use as output. We consider that having the input of the model in a relational database (SQL) is adequate as it provides functionalities such as data consistency. However, the output could be stored in a NoSQL database which will provide faster writing capabilities, or even directly in the format that will be used during the post-analysis of the simulations. This means saving the results directly in csv or Matlab binary (.mat) files. Due to the very high storage capacity that the simulation results require, when saving them to .mat files saving with compression will be performed, which will make the handling and exchanging the files much easier (especially useful until NoSQL database is established).

This approach will reduce the problems of degraded performance due to parallel writing and expedite the gathering of the output of the model to be processed to compute system and network performance metrics.

It has the drawback that filtering capabilities will be limited and it would be difficult to join the data to the original planned one. To solve this, all the required information to compute the system's metrics will be stored in the output. This means that the planned information (e.g., planned flight plans and passenger itineraries) will be stored as part of the output. In order to keep the size of the output manageable, only metrics that are used to compute system and network indicators will be saved. Finally, this has the additional advantage that the output files will be self-contained and they won't require a link to the input data to be analysed, facilitating the sharing of model results.

2.3.2 Flights output

The main addition to the flight output considered for the final version of the model is adding explicit information on the reactionary delay and on the number of passengers in the flight (planned and actual). The flight output will contain all the required information to estimate the flight-centred metrics of the system.

Besides detailed flight metrics, the following fields (note that these are variable names, so some acronyms appear in lowercase) will be stored per flight to be used for the system and network performance analysis:

- Identification of simulation and flight
    - model version
    - scenario id
    - number iteration
    - flight id
- Information on scheduled
    - origin airport
    - destination airport
    - airline operator
    - aircraft model
    - aircraft registration
    - sobt: scheduled off-block time
    - sibt: scheduled in-block time
- Information on flight execution
    - pbrt: push back ready time
    - aobt: actual off-block time
    - aibt: actual in-block time

- Information on delay
    - atfm delay
    - atfm reason
    - reactionary delay
    - departure delay
    - arrival delay
    - main reason of delay
- Information on flight trajectory planned
    - flight plan distance
    - flight plan time
    - exot: expected taxi out time
    - exit: expected taxi in time
    - planned fuel
- Information of flight trajectory execution
    - atot: actual take off time
    - alt: actual landing time
    - actual flight distance
    - actual flight time
    - holding time
    - actual fuel
    - holding fuel
- Information on costs
    - planned fuel cost
    - actual fuel cost
    - duty of care
    - compensation cost (R261)
    - transfer cost
    - soft cost
    - non-pax cost
    - CRCO charges

2.3.3 Passenger output

For each passenger itinerary, information on which flights were planned to be used and which ones were actually used, will be stored. We will ensure that it is possible to compute the arrival delay experienced by passengers and information on their connections (waiting times at the airport, if they make all the connections or were rebooked).

The following fields will be stored per passenger itinerary group:

- Identification of simulation and passenger itinerary group
  - model version
  - scenario id
  - number iteration
  - passenger type
  - flag to indicate if connecting passengers
  - pax group id planned
  - number passengers planned
  - passengers fare
  - pax group id executed
  - number passenger executed
- Planned itinerary
  - initial leg sobt: initial scheduled off block time
  - final leg sibt: final scheduled in block time
  - planned gate-to-gate time
  - leg n origin: airport of origin of leg n
  - leg n destination: airport of destination of leg n
  - flight id n: planned flight for leg n
  - airline operator n: planned airline operator for leg n
  - connection time n: connection time between n and n+1 leg planned

- Executed itinerary
    - initial leg aobt: for the first leg the actual off-block time
    - final leg aibt: for the final leg the actual in-block time
    - reach final destination: flag to indicate if itinerary reach final destination or not
    - itinerary disrupted: flag to indicate if itinerary is the original one or disrupted. If disrupted, it contains the number of the leg that was missed.
    - actual gate-to-gate time
    - delay at final destination
- Compensation information
    - compensation cost: Reg261
    - duty of care

2.3.4 4DTA output

For the 4DTA mechanism, we will add information on the wait for passenger sub-mechanism. We'll record for each flight their decision for waiting for passengers (if waiting or not, for how many passengers and how long), and for modifying their cost index.

The following fields will be stored per decision:

- Identification of simulation, flight and decision point
    - model version
    - scenario id
    - number iteration
    - flight id
    - time stamp decision: pushback_ready, TOC, pax_check
- Flight information
    - airline flight
    - airport origin
    - airport destination
    - planned selected speed
    - estimated departure delay
    - estimated arrival delay

- DCI decision
  - speeding up flag: 0 - no speeding up, 1 - speeding up, 2 - reducing speed, None - no making decision on speeding up
  - selected speed
  - amount of delay recovering
  - expected extra fuel usage
  - extra fuel available
  - recoverable delay
  - pax related delay (due to WFP)*
  - non pax related delay*
  - additional cost expected if no change
  - expected cost reduction (with the taken DCI decision)
- WFP decision
  - waiting for passenger flag: 0 - no waiting, 1 - waiting, None - no making decision on waiting-for-pax
  - (total) waiting time
  - number of passengers and passenger groups waited for
  - number of passengers and passenger groups not waited for

* These two values are going to differ only at Level 2, pushback_ready time stamp.

2.3.5 FP output

For the FP mechanism the following fields will be maintained from the previous model version:

- Identification of simulation and swapping
  - model version
  - scenario id
  - number iteration
  - flight id 1 swap
  - flight id 2 swap
- Flights information
  - airline flight 1
  - airline flight 2
  - airport destination
  - order of swap

- Expected costs of swap
    - cost swap

2.3.6 FAC output

For the FAC mechanism the following fields will be kept from the previous model version:

- Identification of simulation, flight and E-AMAN
    - model version
    - scenario id
    - number iteration
    - flight id
    - airport of destination
- Planning horizon information
    - planned clt: planned landing slot time
    - planned assigned delay: delay assigned at planning horizon
    - planned absorbed air: delay will be absorbed during the flight by modifying speed/trajectory
    - planned speed selected: speed selected by flight
    - planned fuel usage: fuel used/saved during the E-AMAN scope
- Tactical horizon information
    - tactical clt: tactical landing slot time
    - tactical assigned delay: delay assigned by tactical horizon (which will be done as holding)

3 Metrics evolution

3.1 Classical metrics evolution

Upon the result analysis presented in D5.2 [9], the conclusion obtained was that classical metrics designed and measured in the simulation result set were fairly exhaustive and complete, especially regarding flight delay and cost metrics. However, we noticed the several passenger-centred metrics missing in the result set that would enable easier and better tracking of the passenger movements, and facilitate the analysis via simpler and faster queries of the data. For example, the analysis process showed the need to explicitly indicate which passengers missed their connection. Similarly, we are going to explicitly measure (and save in the output) what passengers were waited for by their connecting flights and for how long. The changes on classical metrics will therefore focus on passenger metrics and to better understand the trade-offs between connecting and non-connecting passengers.

For the full list of additional metrics that are going to be added to the output of the model and recorded, refer to the Section 0 Model output changes of this deliverable.

3.2 Network metrics evolution

3.2.1 Improvement in passenger centrality

Drawing on complex network theory, in Deliverable 5.2, we introduced a centrality metric based on passengers' itineraries, termed "Passenger centrality". In principle, the realised passenger centrality of an airport should be computed using the realised passenger itineraries for each model iteration, showing which passengers did not manage to make their scheduled connections. Since retrieving this information from the model outputs takes a considerable amount of time, given the organisation of the output in the database, the realised centrality was instead computed assuming that a passenger misses a connection if the connecting time is shorter than a lower bound of 20 minutes. In reality, the minimum connecting time to get a connection depends on the airport, on the flights and on the stochastic behaviour of the passenger. In Deliverable 5.3, we will use the real missed connections for this computation, which will be more easily accessible in the database.

3.2.2 Trip betweenness centrality

As remarked in Deliverable 5.2, while the loss of outgoing passenger centrality of an airport reflects also the missed connections at that airport (on top of downstream missed connections and cancellations), the loss of outgoing trip centrality only reflects the missed potential connections downstream, and is therefore not useful to measure the performance of a hub. Therefore, we are going to develop a new centrality, also based on potential connections (i.e., not only the ones actually used by passengers on that day), but focusing on the itineraries passing through the considered airport. This new metric, closer to the concept of "betweenness centrality", will focus on the itineraries having the considered airport as an intermediate node and being realistically used by passengers. In the standard betweenness centrality, only the shortest paths are considered, i.e., if a path from node i to node j which passes from node k is not the shortest path joining i and j it would not contribute to the centrality of node k. Note that in a temporal network the shortest paths could be shortest in number of legs, or in duration, or a combination of the two. For ATM applications, however, we deem it more realistic that passengers would not only use the shortest paths, as these could only be available at certain times of the day or could be more expensive than a slightly longer path. Therefore, we plan to consider all the paths that satisfy a constraint on the number of legs (e.g., less than three) and on the duration/connecting time (e.g., upper bound on total duration, and/or lower and upper bound on connecting time). As in standard betweenness centrality, each path will contribute less to the centrality of a node it traverses if there are other considered paths joining the same origin-destination pair. While on a day in which we know exactly all passengers' itineraries we can compute exactly the passenger fluxes connecting in each hub, the proposed version of betweenness centrality would estimate the potential fluxes on a generic day based on all the potential itineraries connecting in that hub. It could be made more precise by weighting paths by the average demand for each origin-destination pair. We will check if the proposed metrics correlates, as expected, with the actual known fluxes of connecting passengers in each airport on the day considered.

In the realised network, an airport loses the contributions to its betweenness given by paths that are not feasible any more due to delays or cancellations. Therefore, the change of this betweenness metric between the scheduled and the realised network measures how much of the potential connecting flux in an airport is disrupted.

With such centrality metric, as with trip centrality, there is the possibility to weight differently itineraries that use flights of the same airline or alliance and itineraries that use flights of airlines belonging to different alliances. This will allow us to show an application of the proposed betweenness metric to assess the effect of the introduction of insurance for inter-airline connections, such as with the GatwickConnects. As a result of such insurance, inter-airline itineraries would be more probably used by passengers, and should therefore be considered in computing the centrality. Compared to the case in which no inter-airline/alliance connections are considered, the betweenness of an airport will increase as more itineraries contribute. Such an increase represents an estimate of the increase of the connecting flux in an airport with the introduction of such insurance. If itineraries are weighted by demand, this is an estimate of the passenger flux increase. We could compare the hypothesis of the introduction of insurance in different airports by comparing the relative increase of betweenness of such airports.

3.2.3 Improvement in causality metrics

In Deliverable 5.1 "Metrics and analysis approach" [8], we introduced causality metrics to capture the propagation of delays in the ATM networked system, in particular by focusing on the network of airports and flights. By defining the state of delay of one airport as the average departure delay of flights within a given time window, we studied the dependence structure between each couple of states, in particular by testing if the information about the previous state of an airport helps in forecasting the future state of another airport, thus assessing the presence of a directional causal relation between the two airports. It is the well-known metric of causality first proposed by Granger C. W. [13]. Furthermore, sometimes it should be preferable to study only the propagation of extreme delays, since small delays are unimportant in terms of ATM performance because easily absorbed during the en-route phase. Thus, we proposed to use Granger causality in tail metric, to test the presence of a causal relation between two airports, but focusing only on the propagation of extreme positive events, which can be defined as the states of delay falling in the right tail of the distribution and, as a consequence, describe the states of congestion of an airport.

When defining the state of delay of an airport as the average of flight departure delays, some issues may arise because of the airport size: (i) for small airports, one delayed flight may increase significantly the state of delay of that airport, also in the presence of normal functioning; (ii) for large airports, the presence of delayed flights which may propagate delay within the system could be averaged away because of the large number of departing flights. Since the causality metrics evaluate how much the states of delay of airports 'correlate' each other, this definition issues may result in assessing the peripheral airports as more central than the large airports, as highlighted in the analysis presented in D5.2. In order to test the robustness of this finding, we propose to explore different definitions for the state of delay of an airport, in particular by considering the quantiles of the distribution of flight departure delays, and not only the average. Then, the analysis presented in D5.2 suggested that the presence of a causal relation between two airports can be related to the buffers adopted by airlines and the airports, large buffers being more functional in absorbing reactionary delays, thus preventing the propagation. For this reason, we propose to repeat a similar causality analysis, but using only reactionary delays of flight in the definition of the state of delay of an airport.

The analysis of the network of causal relations for the ATM system presented in D5.2 have pointed out the presence of some positive feedback subsystems working as amplifying channels for delay propagation, namely reciprocal causal links and feedback triplets or, equivalently, couples of airports which propagate delays each other and triplets of airports where delay propagates in a circle. Recently we showed that the adopted method [14] is sensitive to autocorrelation of time series leading to an overestimation of reciprocated causal links. For this reason, we will develop a more robust method which is able to mitigate such an effect and identify more precisely these loops.

Finally, all the causal analyses developed so far adopt a pairwise causality analysis, i.e., we test the presence of a causal relation by considering one couple of airports at a time, thus neglecting the influence of a third airport which could 'cause' both airports. As a consequence, these network effects may result in a number of feedback subsystems, which are spurious. The difficulty is of course the high dimensionality of a causality test taking into account simultaneously more than two time series. Hence, we propose to test the robustness of this result by generalising the causality analysis to the multivariate case, i.e., when all states of delay are considered at once in assessing the presence of causal relations between airports. To this end we will make use of the inference of the kinetic Ising model which allows us to identify the network of interaction of congestion states of many airports simultaneously.

3.2.4 Relation with commonly used metrics

In order to clarify the potential operational use of the centrality and causality metrics, we will investigate how they are related to more commonly used ('classical') metrics of interest for different stakeholders. The need to explore this link to classical metrics was highlighted as part of the feedback obtained in the dedicated Domino workshop (see D6.3 [10]). For example, from the airport point of view it could be interesting to be able to estimate the change in the expected flux of connecting passengers if there is a change either in the demand or in the scheduled flights. Trip betweenness, when weighted by demand, provides an estimate of the flux of connecting passengers. It is just an estimate because it assumes that the demand is divided equally among the possible itineraries connecting that origin-destination pair that are considered by trip betweenness according to the defined constraints. We are going to check if the estimate correlates well with the actual fluxes on the analysed day.

From the passengers' point of view, interesting information would be the probability to miss a connection (or have one's itinerary disrupted) when connecting in a particular airport, and the average delay at the final destination incurred when connecting there. This information could guide the choice of the itinerary, however to compute it exactly one would need to know the scheduled and realised itineraries of passengers in the past. An estimate of the probability to have one's itinerary disrupted is however given by the relative trip betweenness loss of the airport, to compute which only the DDR data and the average demand is needed. We will investigate if the relative loss of trip betweenness of an airport correlates with the fraction of passengers connecting in that airport having their itinerary disrupted and with their average delay at the final destination. Similarly, we will also look at the correlation between the relative loss of incoming trip centrality and the fraction of passengers having that airport as their final destination that have their itinerary disrupted, and with their average delay.

Both centrality metrics (trip centrality and trip betweenness), aim to evaluate the effects of delay in terms of itinerary disruption. Given that costs arise from these effects, we will further investigate whether centrality losses are related to 'likely real' costs at different levels: over the entire network, at one airport, or for one airline. By 'likely real' costs, we refer to the fact that we need to differentiate further between likely and less likely passenger connections in the model, as will be detailed in subsequent reporting.

Causality metrics aim to capture the propagation channels for both delays and costs within the network of airports, thus revealing the importance of the nodes themselves in the process. In other words, causality metrics allow a network-wide assessment of the role of the airports in propagating distresses within the system. The commonly used metrics of interest are built with the information on either delays or costs, but long-range correlations are averaged away, in particular the dependences at the network level. Hence, we are going to investigate more explicitly what are the extra features described by causality, but not captured by the commonly used metrics. For instance, the average delay of flights at one airport is a common measure of congestion for that airport, however no standards exist to describe how congestions cluster in time during the day of operations. Then, we are going to define an operational framework to integrate common metrics with temporal dependences between either delays or costs, but exploiting the network of causal relations in order to restrict the time dependence structure to the significant propagation channels. Such tool can be of interest for the Network Manager to assess in real-time the state of the ATM system at some level of aggregation, from a single airport to the whole network, but also for short term prediction.

From the point of view of regulators, an interesting point could be reconstructing the process of cost propagation, but exploiting only the information on delays. Since the cost of delay depend non trivially on the delay itself because of network effects, e.g., missing connections, standard metrics are not able to correlate properly delays with costs. Hence, we are going to investigate this dependence structure within the framework given by the agent-based model developed in Domino in order to give some insights on how using DDR data to infer, at least partially, both the most vulnerable nodes and the channels associated with the process of cost propagation.

4 Scenarios evolution

As previously described, the scenarios that will be modelled in the final version of Domino will focus on the analysis of the output from the perspective of different stakeholders. Therefore, the number of different scenarios to be tested will be limited in order to increase the depth and detail of the analysis of the results. The objective is to produce meaningful case studies rather than testing the capabilities of the platform. In D5.2 “Investigative case studies results” [8], the individual implementations of the different mechanisms were analysed and presented. This produced a broad perspective of the model, metrics and mechanisms characteristics. The following scenarios were modelled and tested:

- Default scenario (with all mechanisms at level 0);
- Default scenario + 4DTA level 1;
- Default scenario + 4DTA level 2;
- Default scenario + FP level 1;
- Default scenario + FP level 2;
- Default scenario + FAC level 1;
- Default scenario + FAC level 2.

These scenarios were executed considering the system to be operated under nominal conditions (default delay) and in a congested environment (stressed scenarios), leading to a total of 14 scenarios.

The Domino platform has proven to be capable of capturing the complex interactions of the ATM elements at the ECAC level, producing metrics from a flight, passenger, delay and cost perspective. The number of potential scenarios that could be tested is large but due to the time limitations of the project, we prefer to focus on two case studies: the management of delay at hub and effect of E-AMAN scope on the management of arrivals. This will lead to a total of 6 scenarios and a baseline used for calibration purposes (totalling to 7 scenarios). The output produced by these scenarios will be analysed in depth considering the operational questions that different stakeholders could consider. This is in line with the feedback obtained from the consultation with ATM experts, where research effort should be devoted to better understand the meaning and operational capabilities of the metrics developed in Domino (see D6.3 [10]). The case study of the Hub delay management will be prioritised and reported in D5.3 “Final tool and model description and case studies results”, whilst other scenarios may be done as further work focused on contributions to papers and dissemination.

4.1 Case studies

A baseline scenario with all the mechanisms at Level 0 and with nominal conditions (default amount of delay) will be reproduced and used for the calibration of the model and as reference. As reported in D5.2 [9], more interesting results are observed when the system is under stressed conditions. Therefore, the different case studies will focus on that environment. The full details of the mechanism implementation and the scenarios will be provided in D5.3.

4.1.1 Hub delay management

In this case study, we want to understand the impact of introducing mechanisms to deal with ATFM delay when main hubs are affected by disruptions. Airport capacity has been identified as one of the largest challenges in term of capacity demand in the future and hence, it will be more likely to suffer these type of regulations [11]. ATFM regulations will be explicitly modelled at hubs (LFPG, EGLL and EHAM[2]) to generate the reference scenario.

The disruptions will be created with the following assumptions:

- Starting and finishing in the morning: 06:00 - 14:00 local time
- Reducing the capacity at the airport to half their nominal capacity: LFPG with 44 arrivals/hour, EGLL with 54 arrivals/hour and HEAM with 45 arrivals/hour.

Then the 4D Trajectory Management (4DTA) and the Flight Prioritisation (FP) mechanisms will be applied, as per:

- **Hub delay management baseline**: Disruptions at the three hubs without any mitigation action to be used as baseline. All mechanisms are implemented at Level 0.
- **Hub delay management FP Level 2**: Disruptions at the three hubs with FP at level 2 (flight swapping between airlines allowed).
- **Hub delay management 4DTA Level 2**: Disruptions at the three hubs with 4DTA at level 2 (full DCI and WfP with conjoint decision).

Note that even if the manually set disruptions are only defined for the three hubs, the mechanisms are implemented everywhere, and the whole network is simulated in each case.

4.1.2 Effect of E-AMAN scope on arrival management

This case study will focus on understanding the impact of having the Flight Arrival Coordination (FAC) mechanism implemented with different planning scopes. It is expected that as the scope increases, the benefits in terms of cost savings will increase. This will focus on understanding if the management of arrival capacities tactically (by assigning slots earlier) produces benefits for the

[2] These airports are large hubs which are in the top 20 airports with higher capacity/demand issues forecasted by the EUROCONTROL Challenges of Growth 2018 report [11].

different stakeholders. In this case study four scenarios will be modelled (while focus will be given to the comparison of the baseline with the two Level 2 scenarios):

- **E-AMAN scope on arrival baseline**: 'Stressed' baseline scenario, with all mechanisms at level 0 and high delay across the system.
- **E-AMAN scope on arrival FAC Level 0 extended range**: 'Stressed' scenario with FAC at level 0 and 600 NM as planning horizon.
- **E-AMAN scope on arrival FAC Level 2 nominal range**: 'Stressed' scenario with FAC at level 2 (full cost minimisation) and 200 NM as planning horizon.
- **E-AMAN scope on arrival FAC Level 2 extended range**: 'Stressed' scenario with FAC at level 2 again, but 600 NM as planning horizon.

Note that the Pilot Common Project identifies that E-AMANs will operate in the range of from 100-120 NM to 180-200 NM [12], and that currently some airports such as Heathrow their AMAN Planning Horizon can be as large as 500 NM.

We will focus on the scenario where the mechanism is modelled at Level 2. The extended range at Level 0 is only computed as another reference to better understand the effect of extending this horizon.

4.2 Stakeholders perspective

In addition to the computation of delay and cost metrics, Domino will produce indicators that are relevant for different stakeholders. Examples of the analysis that might be performed are:

- **Airports perspective**
  - How does the introduction of a given mechanism affect passenger itineraries connecting at a specific airport? Does it render more or less probable that such itineraries are disrupted?
  - How does it affect passenger itineraries starting at that airport?
  - How critical is one airport, with respect to other airports, in propagating delays and costs?
  - Which airports are propagating delay and costs towards a given ('my') airport? What is the impact of introducing a certain mechanism?
- **Airlines perspective**
  - How is the connectivity impacted for one airline, when mechanisms are implemented with respect to other airlines?
  - Which elements in the system are generating costs for flights?
  - How do hub-based airlines (e.g., network carriers) perform compared to non-hub-based airlines (e.g., point-to-point carriers), at different airports?

- **Passengers' perspective**
  - What trade-offs exist between different types of passengers (i.e., connecting and non-connecting)?
  - How does the likelihood of having a disrupted itinerary vary when mechanisms are introduced, if an itinerary starts/ passes through / ends at a given airport?

5 Next steps and look ahead

The changes highlighted in Section 2 (Model evolution), will be implemented in the model. This final version of the model will be reported in D4.2 “Model source code”.

The priority is to resolve some shortcomings of the model, recalibrate parts of the model, and focus on producing outputs that are suitable for the analysis of the system, and useful to stakeholders.

The scenarios identified in Section 4 (Scenarios evolution), will be implemented and executed. The metrics presented in D5.1, considering the changes indicated in Section 3 (Metrics evolution), will be computed producing the final results of the model. These will be presented in D5.3 “Final tool and model description and case studies results” (due NOV19). D5.3 will, therefore, present the methodology developed in Domino along with the model, metrics and their applicability.

6 References

1. Boeing, 2019. Airports with Noise and Emissions Restrictions, https://www.boeing.com/commercial/noise/list.page
2. CODA, 2015. CODA Digest 2014. All-causes delay and cancellations to air transport in Europe - 2014.
3. CODA, 2018. CODA Digest 2017. All-causes delay and cancellations to air transport in Europe - 2017.
4. Domino Consortium, 2018a, D3.1 - Architecture definition
5. Domino Consortium, 2018b, D3.2 - Investigative case studies description
6. Domino Consortium, 2018c, D4.1 - Initial model design
7. Domino Consortium, 2018d, D6.2 - Stakeholders consultation on system and investigative case studies
8. Domino Consortium, 2019a, D5.1 - Metrics and analysis approach
9. Domino Consortium, 2019b, D5.2 - Investigative case studies results
10. Domino Consortium, 2019c, D6.3 - Workshop results summary
11. EUROCONTROL, 2018. European aviation in 2040 - Challenges of growth – Edition 2
12. European Commission, 2014. 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.
13. Granger, Clive WJ. “Investigating causal relations by econometric models and cross-spectral methods.” Econometrica: Journal of the Econometric Society (1969): 424-438.
14. Hong, Yongmiao, Yanhui Liu, and Shouyang Wang. “Granger causality in risk and detection of extreme risk spillover between financial markets.” Journal of Econometrics 150.2 (2009): 271-287.

7 Acronyms

4DTA: 4D Trajectory Adjustment

AMAN: Arrival Manager

ANSP: Air Navigation Service Provider

ATFM: Air Traffic Flow Management

ATM: Air Traffic Management

DCI: Dynamic Cost Indexing

E-AMAN: Extended Arrival Manager

ECAC: European Common Aviation Area

FAC: Flight Arrival Coordination

FP: Flight prioritisation

H2020: Horizon 2020 research programme

SESAR: Single European Sky ATM Research

SJU: SESAR Joint Undertaking

TBO: Trajectory-Based Operations

TOC: Top of Climb

UDPP: User-Driven Prioritisation Process

WFP: Wait For Passengers

-END OF DOCUMENT-