Resilience 2050

Resilience2050.eu

New design principles fostering safety, agility and resilience for ATM

Grant agreement no: 314087

Theme AAT.2012.6.2-4. Building agility and resilience of the ATM system beyond SESAR

Funding Scheme: Collaborative Project (small or medium-scale focused research project)

D3.2 The resilience metrics for the European ATM system

Revision: Issue 1 Draft 1

Date: 22/12/2013

Status: Final version

Dissemination Level: PU

	Organisation
Prepared	The Innaxis Foundation and Research Institute
Reviewed	Resilience2050 consortium
Approved	The Innaxis Foundation and Research Institute

RECORD OF REVISIONS

Revision	Date	Reason for Revision	Modified sections/page
v0.01	*11*/2013	Creation of the document	All
v0.02	*12*/2013	Version for review by partners	All
v0.03	*12*/2013	Version reviewed, ready for final approval	All
v0.04	*12*/2013	Final version	All

Abstract:

This document is deliverable D3.2 "The resilience metrics for the European ATM system", which as described in the DoW is the third deliverable of WP3, with 11 person-months involved in its corresponding tasks. The original report delivery date was T0+9 which corresponded to September 2013, however, as agreed with the project officer, D3.2 was delayed until December 2013 in order to wait until D3.1 Multilayer representation was concluded, which delivery date was 1st September 2013. The following deliverable D3.2 describes the Resilience Metrics of the ATM system within Resilience2050 project.

TABLE OF CONTENTS

TABLE OF FIGURES

1 INTRODUCTION

1.1 General Introduction

Resilience2050.eu is a collaborative project funded through the FP7 AAT Call 5, topic AAT.2012.6.2-4: Identifying new design principles fostering safety, agility and resilience for ATM.

The project aims to:

· Develop adequate mathematical modelling and analysis approaches to support systematic analysis of resilience in ATM scenarios, taking into account the different ATM disturbances that can take place in the European airspace.

· Develop metrics to systematically define resilience, addressing the concepts of "Responding", "Monitoring", "Learning" and "Anticipating". This work will result in a Resilience Analysis Framework (RAF 2050), to enable the definition of new ATM design principles fostering safety, agility and especially resilience.

· Provide an extensive overview of human contributions to resilience in current ATM.

The project is carried out by an international consortium composed of: The Innaxis Research Institute, (Project Coordinator, Spain), Deutsches Zentrum für Luft- und Raumfahrt e.V (DLR, Germany), Universidad Politécnica de Madrid (UPM, Spain), Nationaal Lucht- en Ruimtevaartlaboratorium (NLR, Netherlands), Istanbul Teknik Üniversitesi (ITU, Turkey), Devlet Hava Meydanlari Isletmesi Genel Müdürlügü (DHMI, Turkey) and King's College London (KCL, UK).

The project was launched on the 1st of June, 2012 and will last 36 months.

The document structure presents, through a top-down approach, the top level Resilience2050 project information, the role of Work Package 3 within the project and, finally, the work performed under task D3.2, regarding the resilience metrics in the European ATM system. Concretely:

- **Section 0** consists of the title, cover page, record of revisions, abstract and table of contents.
- **Section 1** is the overall introduction of deliverable D3.2, including general information about the Resilience2050 project and the current structure of the deliverable with brief explanations of each section.
- Taking into account that D3.2 is the core report of WP3, **section 2** introduces the main goals of WP3 in the Resilience2050 project. It also includes a description of D3.2 in relation to the WP3 framework and the approach and scope within this work package, in addition to the explanation of the links between the rest of the deliverables in WP3 (D3.1 and D3.3). The information flow in the WPs and deliverables is also depicted in this section 2.
- **Section 3** describes the data preparation process that was required prior to the resilience metrics building. Firstly, the different inputs from D2.2 and its influence regarding the disturbances are described. Different sub-sections for each one of the process steps are included. The final output of this data preparation process were the CSV files named Analysis 4, Analysis 5 and Analysis 6. It has been paid particular attention to a full alignment with the Resilience mulitlayer representation (D3.1), the results from disturbances identification in WP2 and the future research in WP4 and WP5
- **Section 4** consists of the description of the Resilience Metrics. Hence, it is the crucial section of the deliverable and merge quantitative and qualitative analysis of this new property of the ATM system. It covers the mathematical-technical tasks that were required in order to achieve the Resilience metrics , and secondly provides and describes the Resilience Metrics by themselves per airport(s) and per kind of disturbance.
- **Section 5** describes the causality perspective analysis that has been done in a parallel fashion to the previous sections. Meanwhile the resilience metrics proposed in section 4 are centred on the analysis of disturbances and their propagation through the network, this section goes one step ahead building the

European airport network structure from a causality perspective. A Granger causality test with the top 50 airports and their relationships is included.
- **Annex I** provides the acronym list.
- **Annex II** consists of the references compilation.

1.2 Resilience Metrics introduction (and comparison with current reports and metrics etc)

The Single European Sky legislation and research programme (SESAR) is based on providing solutions which performance enhancement can be measure through the right performance indicators.

Although those performance indicators can be defined conceptually, the difficulties of gathering the data, can easily lead to a measurement paradigm that is less than optimal. For instance, data might not be available or complete, the data needed to develop the metric might not arrive concurrently, be in the same format or event represent exactly the same phenomena. This leads to situations in which the accuracy of the data could be less than ideal and, therefore, requires that statistical error and uncertainty in the measurement are correctly tracked to ensure that, whenever a phenomena is observed, a correct understanding of the limitation of the observation is conceived. A performance-based air traffic management system is something to aim for and a powerful Data Science practice is what will enable it.

This document presents a performance metric for the resilience of the system against disturbances, as defined in previous work in this project. This is the first study aimed at defining a performance metric that measures how the system reacts against those disturbances. The document presents a proof-of-concept for this metric, including not only the formal definition of the metric, but also a complete methodology on how to derive this metric for different disturbances across the air transport system in Europe. The methodology presented tracks the statistical error and the uncertainty in the performance measurement, which is key to ensure a scientific foundation of the work.

Additionally, this document reports on the first figures for this metric for the available data. It is worth taken into account that the datasets used are available to Eurocontrol in a very stable format on a day-to-day basis, which would allow for the total automation of the performance system, increasing the usability of this metric. Some of the basic metrics used today to monitor the performance of the air transport system require, for instance, reliable input from airlines, which is frequently not obtained with the detailed needed. The "resilience" metric presented in this document could be perfectly computed as a close-to-real time paradigm, which will enable a permanent monitoring of the performance of the air transport system without additional input from stakeholders.

This document focusses on reporting the methodology without looking into the scalability of the computation mad. In future deliverables, the team will report on a full solution on how to automate the computation of the "Resilience of Air Transport against disturbances" metric.

2 D3.2 in Resilience2050 project

2.1 Resilience2050 and deliverable 3.2

The key objective of the Resilience2050.eu project is to define, analytically, the concept of "resilience" meaning, within the context of Air Traffic Management. The sensible steps, as described in the DoW and some of them already achieved, are the following:

- The first WP provided the theoretical framework, that leaded to a Resilience definition within the ATM domain. They were also explored other novel ideas such as the human role factor in the ATM resilience. Some learning about the proper terminology (Resilience, Robustness, Disturbances, Perturbations) were taken from other socio-technical domains.
- WP2 tackled the data sets and data mining analyses that enabled, together with WP1, a deep study of which is the "Resilience level" in the current European ATM system for a each particular disturbance. They were also included some insight of the delay propagation patterns in the European ATM system.
- The third WP and current should built a sensible structure where the Resilience concept could be represented, enabling afterwards to a full list of Resilience Metrics of the current ATM system

2.2 Research steps and approach

In the context previously explained, the current Deliverable 3.2 'The resilience metrics for the European ATM system', have firstly required input from WP2 D2.1: the specific information of the data sources involved in each layer. In order to provide the proper datasets, it was done the selection of the scenarios that would be significant for the resilience study. Finally, taking into account operational experience of the consortium partners and the most common delay causes (from Eurocontrol CODA and NOP), the scenarios depicted were:

- Weather hazards (thunderstorm, rain, hail, snow, tornado, fog etc)
- Bad visibility issues
- Runway operations: runway configuration changes
- Staffing problems: ATC strikes, illnesses etc
- Capacity issues -in the macromodel: sectors, regulations etc-

Once selected the scenarios and data sources, the input regarding the specific layers connections were taken from the resilience data mining tasks done in WP2 D2.2. In addition partial results of D2.3 achieved at the time of delivering this report have also been interesting as they enrich the multilayer connection perspective.

WP3 "Development of new design principles" targets the analysis of the resilience of the current system. WP3 started with the already delivered D3.1: By means of the development of a "layered resilience assessment" metaphor, different elements (for instance, airports and their performance, disturbances) were organized in different logical “layers” or "views" of the system, and where relations between them were evaluated making use of WP2 outcomes. As a result, the abstract concept of resilience could be represented, measured and communicated, and different strategies to achieve an increased resilience in air transport operations could be developed.

Due to the complexity of the Air Transport with hundreds of European stakeholders involved (ATC, airports, airlines), and the complex relations between them, Resilience2050 consortium made a huge effort in creating a multilayer representation in D3.1 simple enough to understand the delay propagation, and complex enough to provide interesting metrics of resilience as an ATM characteristic. In this context, D3.2 only differs to what was planned in D3.1 in two points:

- The multiple disturbance matrix have not been delivered in D3.2, due to the low number of instances per airport and per disturbances' combinations, that matrix would lead to a lack of significant results/metrics.

- On the other hand, an extra analysis -not expected at D3.1 delivery but of a high scientifical interest- has been provided in a parallel fashion to the Resilience Metrics: the consortium have built the European airport network structure from a causality perspective. A Granger causality test with the top 50 airports and their relationships has been included in section 5 of the current deliverable.

Hence, visually summing up the general information flow that was required for prior to the creation of the current deliverable:

WP1	WP2	WP3
Resilience definition: Other socio-technical domains (D1.1) Human factor, disturbances (D1.2) ATM resilience, basic modelling approach (D1.3). (The full modelling approach connects D1.3, D3.1, D4.1 and D4.5 as explained above)	**Data mining exercises:** Data sources and scenarios definition (D2.1) Data mining activities (D2.2)	**Resilience metrics:** Multilayer representation (D3.1) Resilience Metrics - current document-

Looking at it the other way around, the current deliverable will be an input for the following research activities and deliverables:

WP3	WP4	WP5
New design principles (D3.3)	Developing the model D4.1	Stress testing of new concept D5.i

Further interrelations between other deliverables and WPs (2, 3, 4 and 5) has also been pointed out in the following figure. Ticks represent finished deliverables, and the ones with the figure highlighted in red represent those in which the consortium is currently focused on:

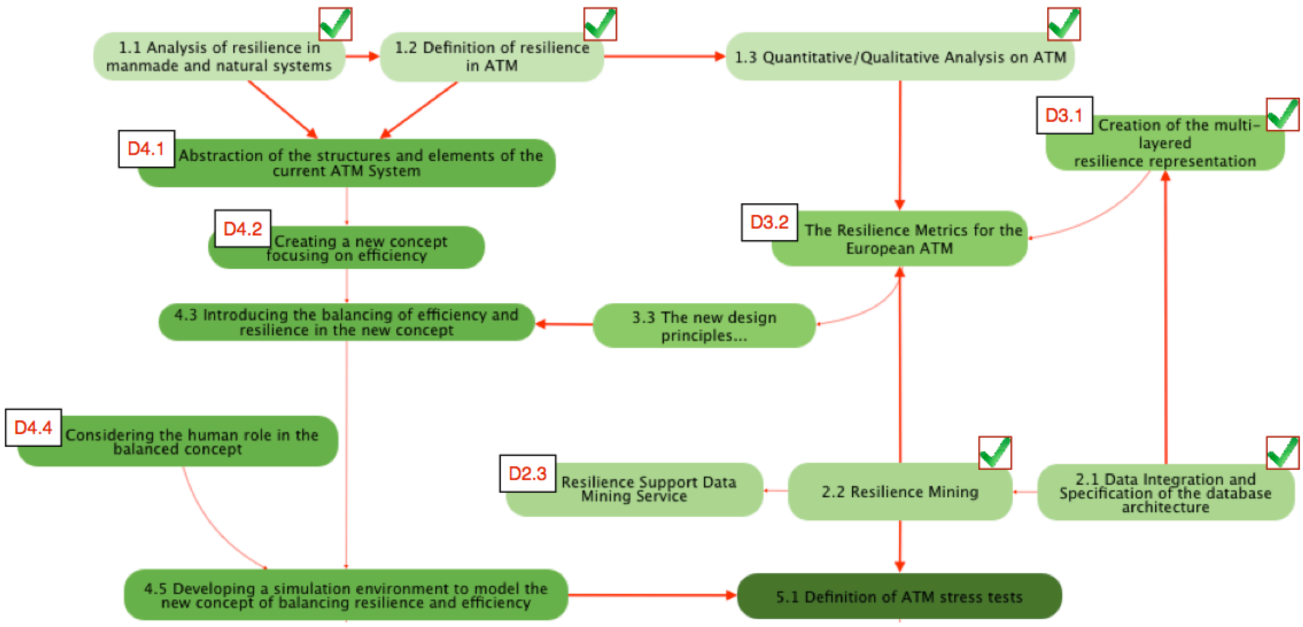

Figure 1 Interrelations between deliverables

In conclusion, the data mining exercises run in the project provided strong guidance to build D3.2 in a way that this Resilience model and the corresponding metrics will be purely data-driven. The different data mining exercises exploited the different data sources available and this deliverable provides a set of resilience metrics that encompass those data mining activities and provide guidance which is the ATM performance of different stakeholders in terms of this "new" property. For the resilience metrics and the current deliverable it has also been taken into account future steps and deliverables, especially those involving the modelling.

3 Data preparation process

In the current third section it is provided the information of the data preparation process that was required (as a preliminary task) to reach the Resilience Metrics. It has been covered and analysed which has been the inputs and outputs in each of the process stages, in addition to any relevant information that was considered significant. It has not been included the Python code used, however the pseudo-code and key ideas can be easily followed as the overall process has been documented in detail in the current deliverable.

Meanwhile the initial input were the historical databases (flights and disturbances), the final outputs have been the CSV files, used afterwards in the different analyses and in the resilience metrics,

As intermediate steps in the process 6 queries and 6 tables have been used. A specific chapter for each of them has been prepared covering: a full explanation of the material contained (tables, raws, files etc) and the data dealing tasks involved (filtering, merging, comparing etc)

In brief, the data flow process has been the following:

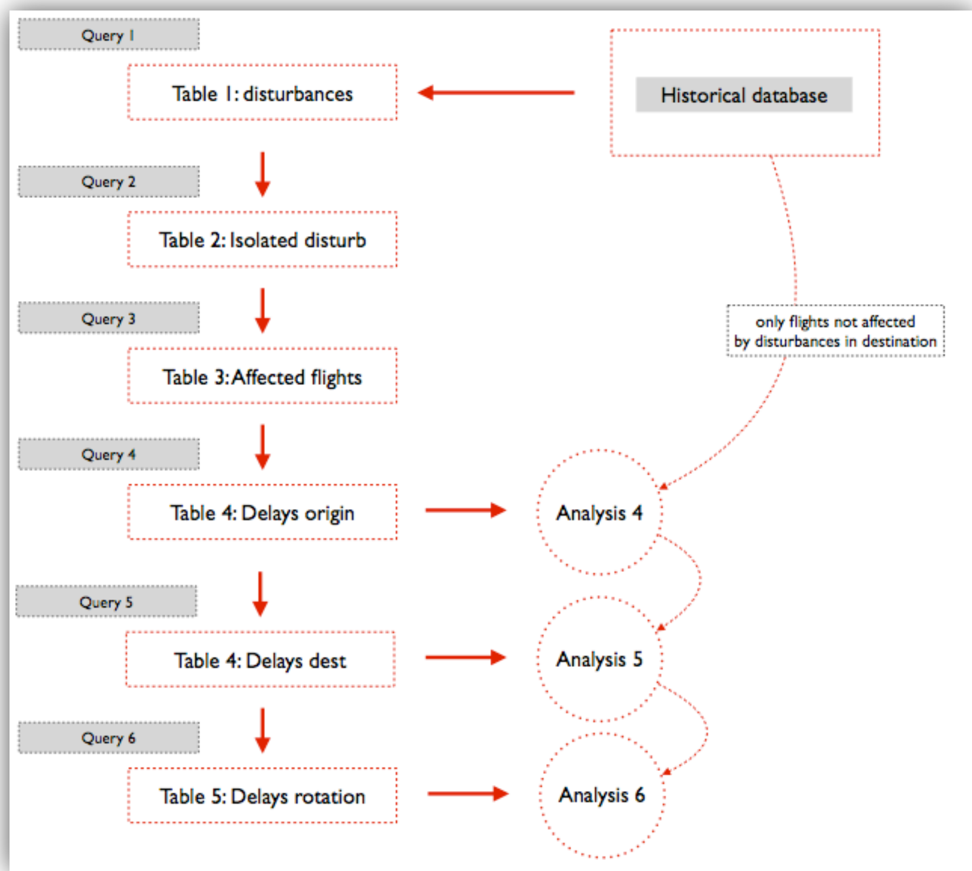

Figure 2: Data flow process

3.1 Query 1 and Table 1

Input: ALLFT+, METAR and AAC headlines databases

Output: A table with the detected windows for each type of disturbance with the structure detailed below

Description of the process: The objective for this table is to represent the periods where a disturbance took place in an airport. For this task, the implemented process analyses the appropriate database and extracts the starting and ending times where a disturbance occurred, as well as the affected airport by the disturbance. Since the multilayered model is only going to be applied to the selected airports, this table is going to be constructed for the selected airports described in WP 2.2. Depending on the type of the disturbance a different database has been selected and analysed, concretely:

To analyse the change of configuration disturbance, the ALLFT+ is used to obtain the list of flights that took of for each of the selected airports and to detect when a change of configuration occurred (following the same process that was used in WP 2.2). To analyse the impact of the changes of configuration, it has been decided to select a one hour time window around each change of configuration detected.

For the staffing disturbance, the AAC headlines is used to identify the days for each airport where a staffing problem took place (as was conducted in WP 2.2). Therefore, for each detected staffing problem, a time window of 24h is generated in the table identifying not only the time window but also the affected airport.

For all the weather-related disturbances the METAR dataset is analysed to extract the starting and ending time, i.e. the time window, where each weather phenomenon affected each airport.

Finally, for the visibility disturbance, a similar step to the previous one has been followed with the exception that, in this case, the visibility information is analysed to detect the time windows where this value fell below 400m.

The following table represents the structure that Table 1 contains.

Airport	Disturbance kind	Start time stamp	Ending time stamp
LEMD	1 [(1=rain or directly rain)	13:25	13:55

3.2 Query 2 and Table 2

Input: Table1

Output: Table 2, a filtered version of table1

Description of the process: This table represents a filtered version of Table 1. Therefore, it presents the exactly same structure that was described in the previous section. The process starts by reading all the time windows of Table 1 and grouping them by the disturbance type. Then, for each time window of each disturbance, it verifies if there are no other time windows of another disturbance that overlap with the analysed time window. If no overlap is found, the record is stored in Table 2. Thus, this table represents the time periods where a disturbance affected an airport and no other disturbance occurred at the same time so we can be sure that the corresponding perturbance measures that we are going to analyse afterwards are only being affected by a single disturbance. There is only one exception to this filter and that is the case of the visibility and weather disturbances. Since these two disturbance are closely related (the visibility disturbance is caused by a weather phenomenon) we are not going to apply the filter whenever the time windows involved in an overlap belong to only these two types of disturbances.

Table structure:

Airport	Disturbance kind	Start time stamp	Ending time stamp
LEMD	1	10:14	10:25

3.3 Query 3 and Table 3

Input: Table 2 and the ALLFT+ database

Output: Table 3

Description of the process: The objective of this table is to identify the flights that were affected by each disturbance included in Table 2 as well as some useful information for conducting the analysis of the multilayered model. Therefore, for constructing this table, the process reads each entry of table 2 that identifies a time window where a disturbance affected an airport, and uses this information to access the ALLFT+ database to extract the flights that were departing when the disturbance occurred. With this set of flights, the next step is to filter the flights that were not arriving at one of the selected airports (listed in WP 2.2). Then, for each flight, several fields that will help in the posterior analysis are retrieved, like, for example, the aircraftID, the registration number, the departure and arrival airports and the severity measure. The severity measure is only applied to the configuration change disturbance and the visibility disturbance which represents a quantitative value of the impact of the disturbance. In the case of visibility, this measure represents the visibility value (measured in meters) that occurred when the flight took place, whereas with the change of configuration disturbance, this measure represents how close the departure took place to the moment where the change happened. The structure of this table is as follows:

Flight number	Registration number	departuring airport	arriving airport	Disturbance affecting	Severity of disturbance affecting
IBE1234	XXX	LEMD	LFCG	disturbance identification	X

3.4 Query 4 and Table 4

Input: Table 3 and the ALLFT+ database

Output: Table 4

Description of the process: This table extends Table 3 to include a perturbance measure that is going to be used to analyse the impact of each disturbance. Concretely, for this table, the measure that is calculated is the delay at the departure airports, computed as described in WP2.2, i.e., the difference between the estimated departure time included in the flight FTFM plan and the real time stamp stored in the radar points set, being both fields obtained from the ALLFT+ database. The new departure delay values together with the information contained in Table 3 constitute the data of Table 4 represented in the following table structure.

Flight number	Registration number	departuring airport	arriving airport	Disturbance affecting	Severity of disturbance affecting	Departuring Delay
IBE1234	XXX	LEMD	LFCG	disturbance identification	X	140 (seconds)

3.5 Analysis 4

Input: CSV file from Table 4

Output: Resilience Metrics of the European ATM system

The information of the analysis done is covered under the next section "Resilience Metrics"

3.6 Query 5 and Table 5

Input: Table 3 and the ALLFT+ database
Output: Table 5

Description of the process: Similarly to Table 4, Table 5 extends Table 3 to include a perturbance measure: the arrival delay. This measure is computed from the data contained in the ALLFT+ database as detailed in WP2.2, i.e., the difference between the estimated arrival time and the real arrival time stored in the radar points set. Therefore, the new table is constructed by adding the arrival delay measure to the fields contained in each record of Table 3. Since this table is going to be used to analyse the impact of each disturbance in the arrivals of the flights, a similar filtering process to the one followed with Table 2 is conducted but, in this case, it is centred around the arrival airports of the flights, i.e., the flights that arrive at an airport that was also being affected by a disturbance are discarded from the final dataset so that it only contains flights that were being affected by a disturbance at its departure. The following table represents the structure of Table 5.

Flight number	Registration number	departuring airport	arriving airport	Original Disturbance affecting	Severity of disturbance affecting	Departuring Delay (at origin)	Arrival Delay (At destination)
IBE1234	XXX	LEMD	LFCG	disturbance identification	X	1400 (seconds)	800 (seconds)

3.7 Analysis 5

Input: CSV file from Table 5

Output: Resilience Metrics of the European ATM system

The information of the analysis done is covered under the next section "Resilience Metrics"

3.8 Query 6 and Table 6

Input: Table 3 and the ALLFT+ database
Output: Table 6

Description of the process: The objective of this table is to contain the appropriate information to analyse how the impact of a disturbance affecting the flights departing from an airport propagates to the next leg that conduct the aircraft of the flights. Therefore, to construct this table, Table 3 is read (which contains the set of flights affected by all the proposed disturbances) and, for each flight, the next leg flight is obtained from the ALLFT+ database by obtaining the first of all the flights with the same registration mark field that departed from the same airport and after the Table 3 flight arrived.

Similarly to the previous tables, these next leg flights are filtered to contain only the flights that arrived at one of the selected airports of WP 2.2 and that they were not being affected by any disturbance at their departure (since we are analysing the influence of a disturbance in the departure of the next leg flight we do not want to include next leg flights that were also being affected by another disturbance). As happened with Table 3, several fields that describe the flight are gathered from the ALLFT+ database as well as a perturbance measure: the delay at the departure of the next leg flights (computed as described in Table 4) to construct the data of Table 6 which structure is represented in the following table.

Flight number	Registration number	departuring airport	arriving airport	Disturbance affecting	Severity of disturbance affecting	Departuring Delay (at origin)	Arrival delay (at destination)	Departuring Delay (At destination)
IBE1234	XXX	LEMD	LFCG	disturbance identification	X	1400 (seconds)	1000 (seconds)	800 (seconds)

3.9 Analysis 6

Input: CSV file from Table 6

Output: Resilience Metrics of the European ATM system

The information of the analysis done is covered under the next section "Resilience Metrics"

4 Resilience Metrics

The current section includes detail information on:

- Data sources required from the data process preparation and basic explanation
- Data representation guidelines
- Data cleansing
- Data fitting
- Reference delay rate
- Disturbance effect
- The Resilience graph
- The error and confidence rate
- The global resilience picture
- The resilience matrix

4.1 Data source and basic explanation:

The data process preparation outputs included different CSVs containing:

- from the CSV in Analysis 4 it has been extracted which is the departure delay of the flights affected by each of the disturbances in each particular airport.
- from the CSV in Analysis 5 it has been extracted which is the arrival delay (next leg, hence destination airport) of those flights.

The combination of both tables, taking into account the reference state consideration explained in the following paragraph, enabled determining the "en route" resilience. The information included is classified per pair of airports and per kind of disturbance.

On the other hand, the "turn around" resilience was extracted from:

- the CSV in Analysis 5 contained which is the arrival delay of the flights affected by each of the disturbances (at the previous leg)
- the CSV in Analysis 6 contained which is the departure delay (after turn around) of the flights affected by each of the disturbances (at the previous leg)

For both "en-route" and "turn around" resilience the data included two different kinds of information:

- When no disturbance (of those identified in the project) was taking place, the delay has been named; "reference state delay", representing the nominal condition of the system, or in other words: average delay when system not presumedly disturbed.
- When an isolated disturbance occurs, the delay is named and linked to that disturbance: "disturbance X delay"
- Multiple disturbances taking place at the same time have not been analysed for the reasons depicted in section 2 of the current document

Original data sets names were:

	Pair of airports	Single airport
No disturbance	(R1) Reference en-route delay	(R2) Reference turnaround delay
Isolated disturbance	(Dx.1) En-route delay under disruption	(Dx.2) Turnaround delay under disruption

Where x ranges across all considered disturbances.

4.2 Data representation

The following graph is just an introductory example of the representations of the data. The idea is to include this graph explanation previously to any further data tasks applied that would make the graph and metrics understanding more complex.

Each flight data is represented as a dot. The abscissa and ordinate values represent different delays values (minutes) at different aircraft flight stages/legs. The stage considered varies depending on if it is analysed enroute or turn around resilience. There is one graph per disturbance and per airport considered. Colours represent different destination airports in the case of en-route resilience. The colour is the same for turn around resilience. The abcissa and ordinate units were originally seconds. For a better undertanding in the representation have been transformed into minutes (') or hours (h)

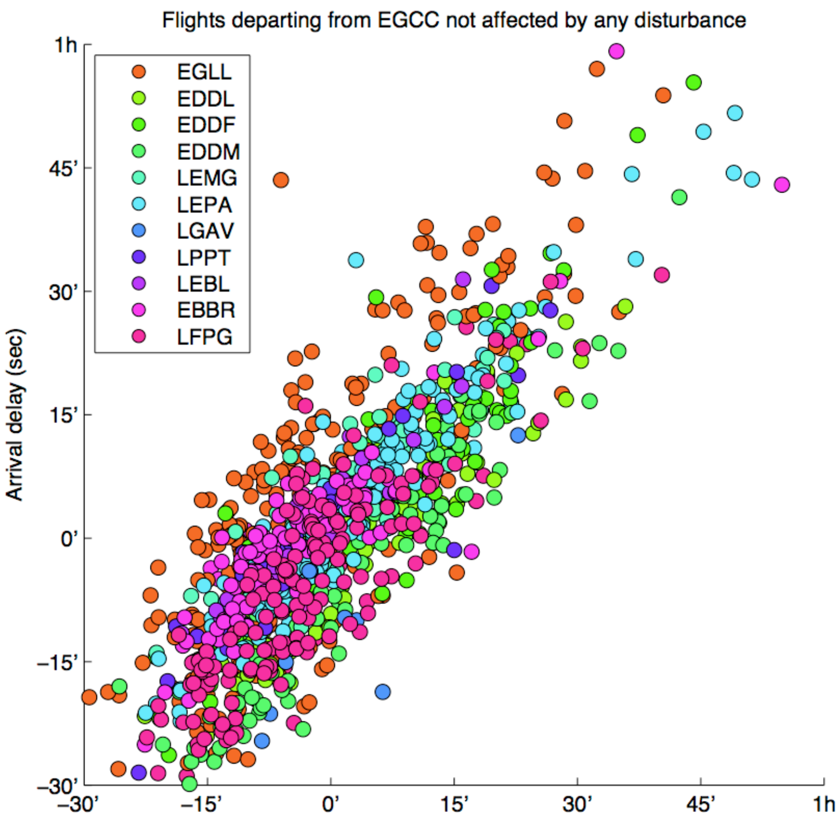

Figure 3 Data representation example

4.3 Data cleansing

Regardless of the source no data set can ever be considered errorless. Undesirable data errors can derived from many different sources, including: system failures, such as wrong or noisy lectures, concurrent database writing or even due to human factors. Recently, in the emergence of Data Science, a new term have been coined: "data artifacts" (J. Van den Broeck, S. A. Cunningham, R. Eeckels, and K. Herbst. Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities. PLoS Med. 2(10):e267, 2005.) In general artifacts refer to errors, discrepancies, redundancies, ambiguities and incompleteness that hamper the efficacy of analysis or data mining. Currently, the detection and correction of such artifacts is a very alive research area with many different fields of application.

The volume and data throttle in Big Data require any artifact detection technique to be both: completely automated and fast enough to not to interfere with the data collection process. In Resilience2050 two methods for data cleansing and artifacts detection are proposed: Extreme values analysis (EVA) and Outlier Detection from Data Subspaces (ODDS).

The rationale behind the EVA analysis is as follows; Flights with excessive accumulated delay are most likely to be handled differently by operations and, therefore, their dynamics are expected to be substantially different as for the rest of the flights. In terms of Data Mining most of this events could be considered as "Dragon Kings" (D. Sornette. Dragon-Kings, Black Swans and the Prediction of Crises. 2009arXiv0907.4290S, 2009) and therefore should be excluded from any analysis of the system itself. The general approach for a EVA analysis consist on estimate the distribution of the maximum using the "Fisher-Tippet–Gnedenko Theorem" for asymptotic statistics. This asymptotic result however depends on the shape of the tail of the original distribution, which has to be estimated from the available data. Once the distribution of the maximum is known (or at least approximated) it is fairly easy to estimate the probability of events over a given maximum. Almost any statistical package contains an implementation of EVA, more information can be found in (S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, London. 2001).

The ODDS method detects deviation patterns for common attributes. First one has to define one of two metrics, usually referred to as the Q-measure and the O-measure, using either of them will produce a cleaner data set, however the Q-measure requires less computational power at the expense of accuracy performance. Using either of them a CA-outlier is defined for each subset of the sample, the CA-outlier defines a score rank over the enumeration of all the sample's subspaces, this enumeration allows the subspaces to be arranged as a lattice. This lattice is pruned by a cut-off threshold given by the rate-of-change of the measure selected previously, dividing the sample set between outliers and non-outliers. The theoretical details of the previous algorithm can be found in (C.C. Aggarwal, P.S. Yu. An Effective and Efficient Algorithm for High- dimensional Outlier Detection. VLDB Journal, 14(2):211-221, 2005.) and many common statistical packages include a more or less efficient implementation of the ODDS algorithm.

Each of the previously classified data sets (R1, R2, D1.1, D1.2, ..., Dn.1,Dn.2) are cleansed individually following the previous method, and so producing a new set of cleansed data sets (R1*, R2*, D1.1*, D1.2*, ..., Dn.1*,Dn.2*)

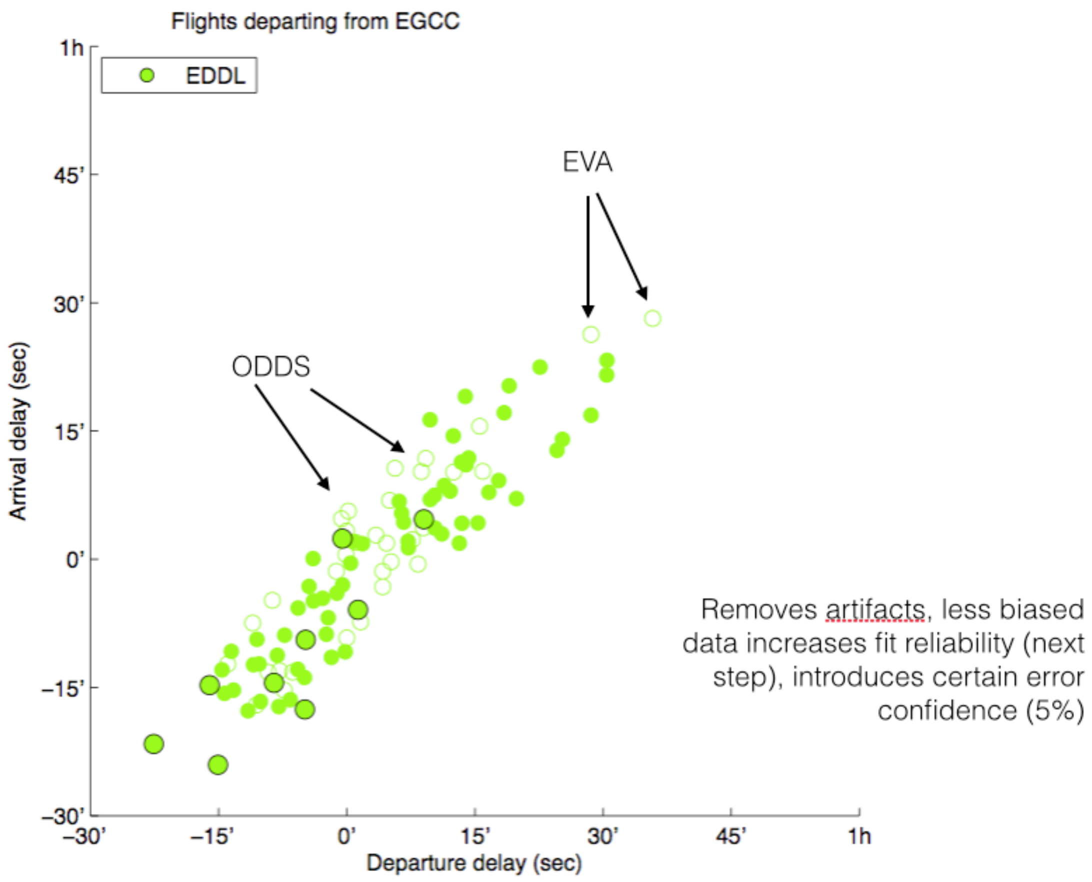

Figure 4 Data cleansing

Data sets names references after data cleansing:

	Pair of airports	Single airport
No disturbance	R1*	R2*
Isolated disturbance	Dx.1*	Dx.2*

Where x ranges across all considered disturbances.

4.4 Data fitting

Once the data set has been cleansed, extreme values and outliers removed, the data is fitted using a linear regression. The linear regression models establishes a linear relation between a dependent variable (e.g. arrival delay) and a explanatory variable (e.g. departure delay). Usually, this relation is obtained by the less squares approach; for instance the Ordinary Leasts Squares or OLS which minimizes the distance between the linear function and the known data set. This method always produces the best possible linear fits in terms of euclidean distance. The error of the linear fitting is represented by the Sum of Squares Error or SSE, or the Root Mean Square Error or RMSE. (A. Björck. *Numerical methods for least squares problems*. Philadelphia: SIAM.ISBN 0-89871-360-9. 1995)

However, the relation between the dependent and explanatory variables may not be necessarily linear. In order to assess the "linearity" of the data set two measures can be used, the Pearson's correlation coefficient and the Spearman's rank correlation. The former is related to linear relations whist the later is refer to monotonic relationships. In general we will use Pearson's unless results are not clear in which case we will use Spearman's and give them both.

We now apply the linear regression model explained previously to each of the cleansed data sets (R1*, R2*, D1.1*, D1.2*, ..., Dn.1*,Dn.2*) producing a new set of linear fits (r1, r2, d1.1, d1.2, ..., dn.1,dn.2). Note that each linear fits contains two parameters (e.g. r1 $=(a_{r1},b_{r1})$, dn.1$=(a_{dn.1},b_{dn.1})$, ...)

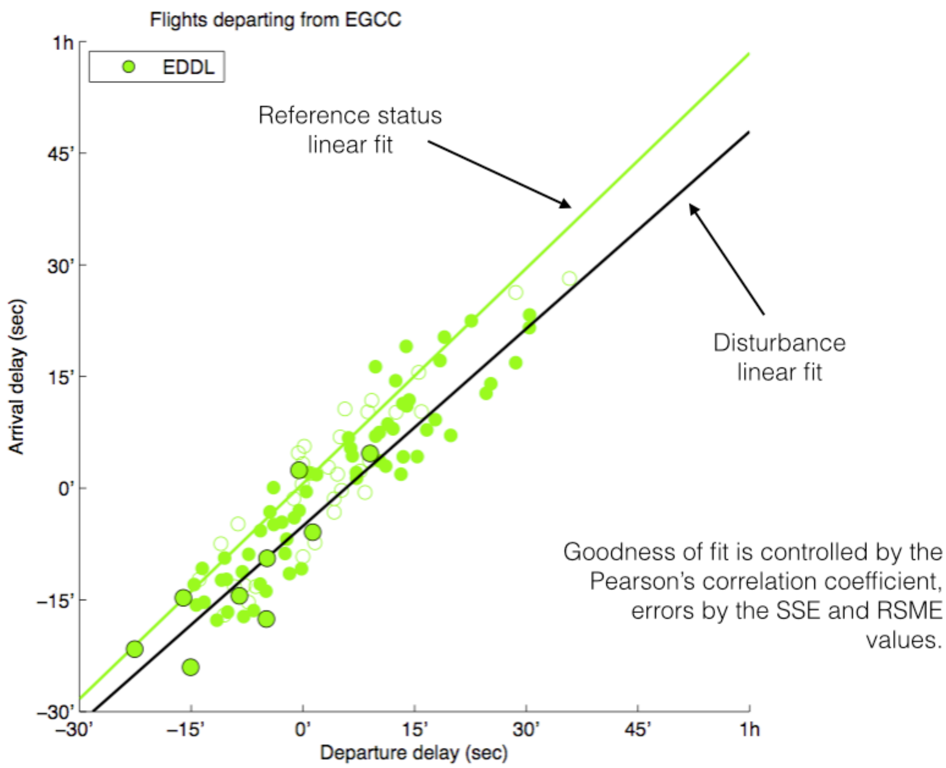

Figure 5 Data fitting

Data sets names references after data fitting:

	Pair of airports	Single airport
No disturbance	r1 $=(a_{r1},b_{r1})$	r2 $=(a_{r2},b_{r2})$
Isolated disturbance	dx.1$=(a_{dx.1},b_{dx.1})$	dx.2$=(a_{dx.2},b_{dx.2})$

Where x ranges across all considered disturbances.

4.5 Reference delay rate

The linear regression previously explained determines a linear dependence between the a dependent variable (arrival delay in the case of en-route delay or departure delay in the case of turnaround delay) and the explanatory variable (departure delay in the case of en-route delay or arrival delay in the case of the turnaround delay). Any linear relation is defined by two parameters; a slope and a constant term. The slope measures how delay is amplified (if greater than one) or adsorbed (if smaller than one), when considering the en-route delay the slope represents the system's capability to reduce delay by means of ATFM whilst when considering the turnaround it represents how the airports are capable of absorb delay.

The constant term is a residual delay which is intrinsic to the system (the value of the dependant variable does not really depends on the explanatory variable). In theory both parameters are useful, however, data analysis have proven the constant terms to be negligible (smaller than one minute for the 90% of the data) and therefore for the sake of simplification will be dropped from the resilient metrics.

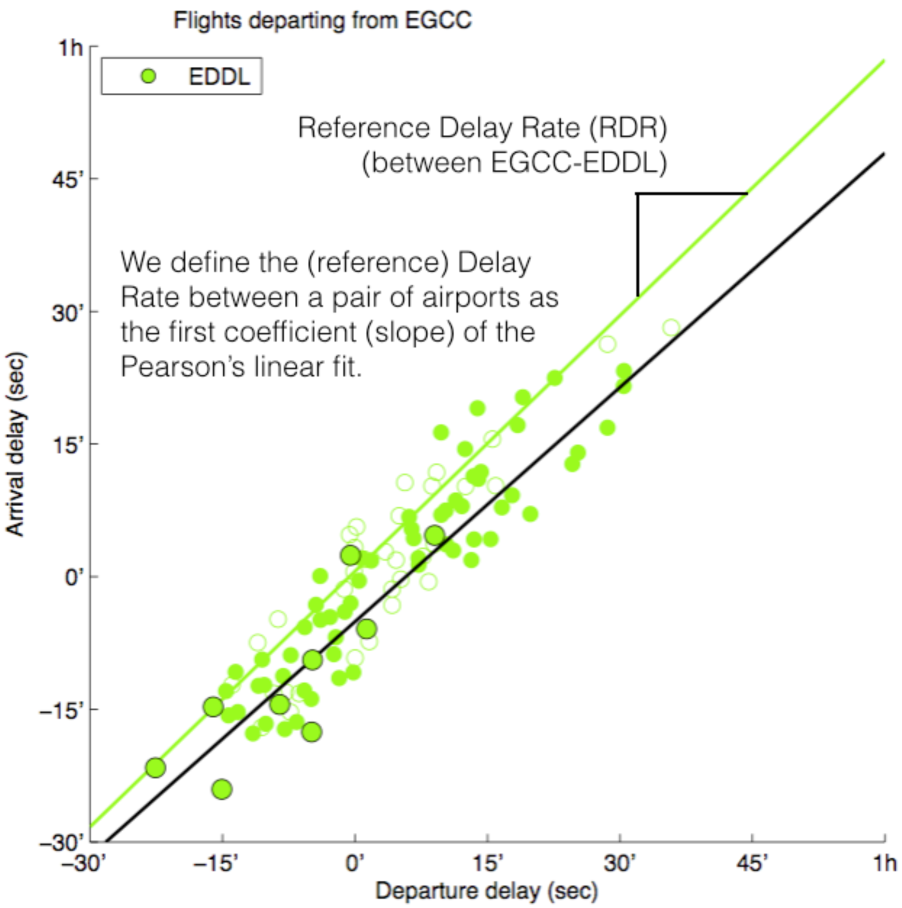

Figure 6 Reference delay rate

We define the performance of an airport pair or a single airport turnaround by the slope of the linear regression. In previous notation and when no disturbance is affecting the system the Delay Rate would be a_{r1} when considering en-route delay and a_{r2} when considering turnaround delay. Similarly when the system is affected by the (isolated) disturbance x, the Delay Rate would be $a_{dx.1}$ when considering en-route delay and $a_{dx.2}$ when considering turnaround delay:

	Pair of airports	Single airport
No disturbance	a_{r1} (En-route Reference Delay Rate)	a_{r2} (Turnaround Reference Delay Rate)
Isolated disturbance	$a_{dx.1}$ (En-route Disrupted Delay Rate)	$a_{dx.2}$ (Turnaround Disrupted Delay Rate)

Where x ranges across all considered disturbances.

4.6 Disturbance Effect

The advantage of considering the slope of the linear regression as the Delay Rate is that it does not depend on a particular level of delay (e.g. produced for instance by congestion), it is a "relative" value resembling the performance of the system (en-route or turnaround). When no disturbance occurs this Delay Rate represents the "normal" behaviour of the system. This "normal" behaviour is call Reference Status and must not be emroiled with the Desired Status. The Reference Status represents the actual performance of the system, whilst the Desired Status is an entelechy representing more of an aspiration than a reality. In the same way disturbances should be measured with respect to the Reference Status, that is; to quantify the impact of a disturbance we have to measure the alteration of the Reference Status while the turmoil is still active.

In order to capture the effect that a disturbance has on the system, the proposed metric is the relative difference (percentage) of the Delay Rate (previously defined) between the Reference Status and the Disrupted Status. A Disturbance Effect close to 0 would be a very low impact on the system, while a (positive) value >>0 would mean a huge (positive) impact on the system, a (negative) value <<0 would represent a vast (negative) impact on the system. It is a signed measure of how far the system is from performing as in the Reference Status when affected by a disturbance.

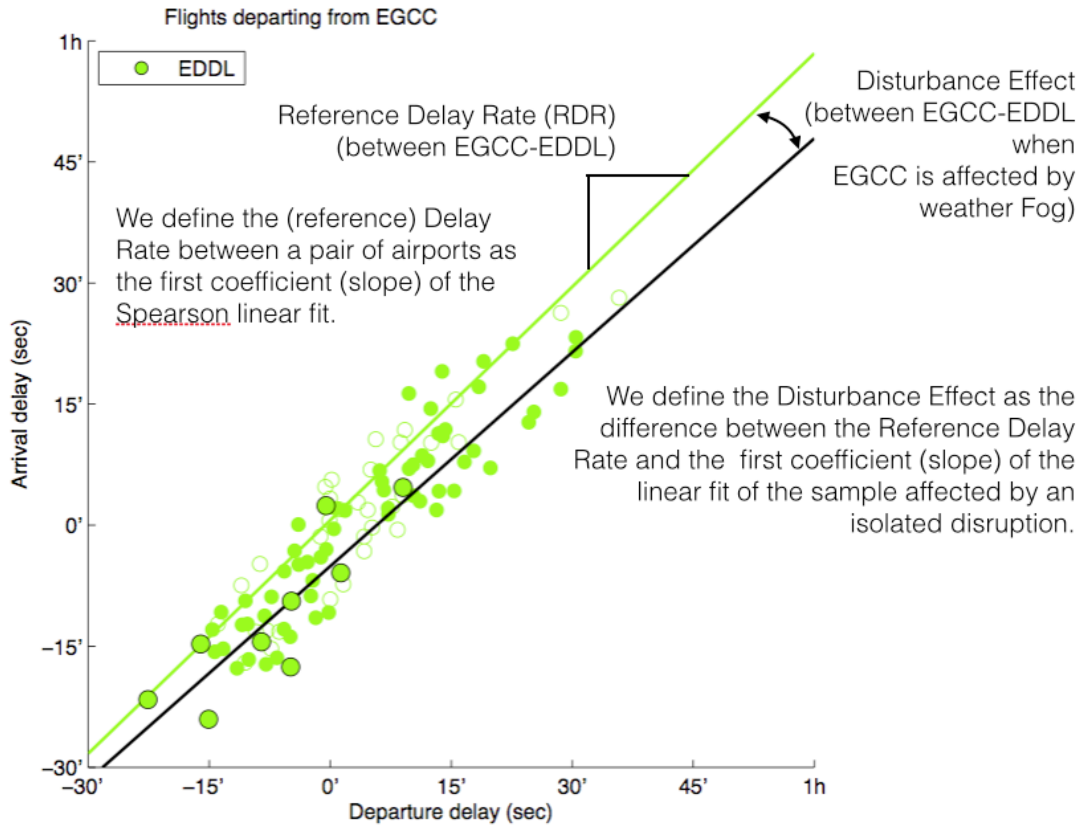

Figure 7 Disturbance effect

In previous notation the Disturbance Effect for the disturbance x would be $e_{x1} = (a_{r1} - a_{dx.1})/a_{r1}$ for en-route delay and $e_{x2} = (a_{r2} - a_{dx.2})/a_{r2}$ for turnaround. The Disturbance Effect for the Reference Status remains undefined, or just set to one for coherence.

	Pair of airports	**Single airport**
No disturbance	N/A (or =1)	N/A (or =1)
Isolated disturbance	e_{x1} (En-route Disturbance Effect)	e_{x2} (Turnaround Disturbance Effect)

Where x ranges across all considered disturbances.

4.7 The Resilience graph

In this section we will provide an algorithm to generate a layered graph model, using the numerical representation of Disruption Effects and Reference Status explained before. A layered graph model consist on several graphs in which nodes are identified across. From a strictly theoretical point of view a layered graph does not add additional information, however, in practice there is a conceptual difference. Since nodes in each layer are identified, it is possible to switch across layers depending on the context. For instance, as we will see later on, one layer would represent the Reference State and there would be also additional layers for each disturbance considered. Then, if we are interested in a particular route with multiple flight legs, we can start in the Reference Graph if there is no disturbance, track the flight in that layer until a disturbance occurs, then we transition to a disrupted stated. Now we switch to the corresponding Disrupted Graph and continue there while the disturbance is active.

The advantages of having a (layered) graph model is that it ultimately allows a broad variety of algorithms and concepts of Graph Theory to be applied. All the common elements such as Critical Paths, Minimum vertex cover, Loops, Cycles, Cliques, Maximal Flows, etc. can now be interpreted in terms of system status and disruption effect.

The layered graph is defined as follows; there is a base layer or Reference Graph in which airports are represented by nodes and two airports are connected by an edge if there are any flights between them each edge is weighed by the Delay Ratio (en-route for two distinct airports, and turnaround for loops on each node) of the Reference State. An additional layer is generated for each disturbance considered, as in the Reference Graph, airports are represented by nodes and two airports are connected by an edge if there are any flights affected by that disturbance between them. But, in contras to the Reference Graph, edges now are weighted by the Disturbance Effect (en-route for two different airports and turnaround for the same airport).

	Pair of airports	**Single airport**
No disturbance	Reference state graph **edges** weighted by $\mathbf{a_{r1}}$	Reference state graph **node loops** weighted by $\mathbf{a_{r2}}$
Isolated disturbance	Disrupted state graph **edges** weighted by $\mathbf{e_{n1}}$	Disrupted state graph **node loops** weighted by $\mathbf{e_{n2}}$

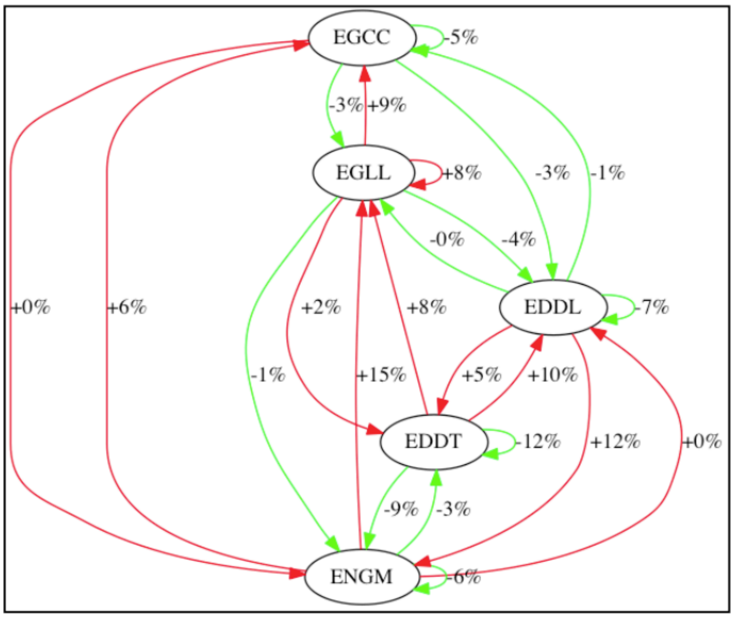

Figure 8 Resilience graph (single disturbance)

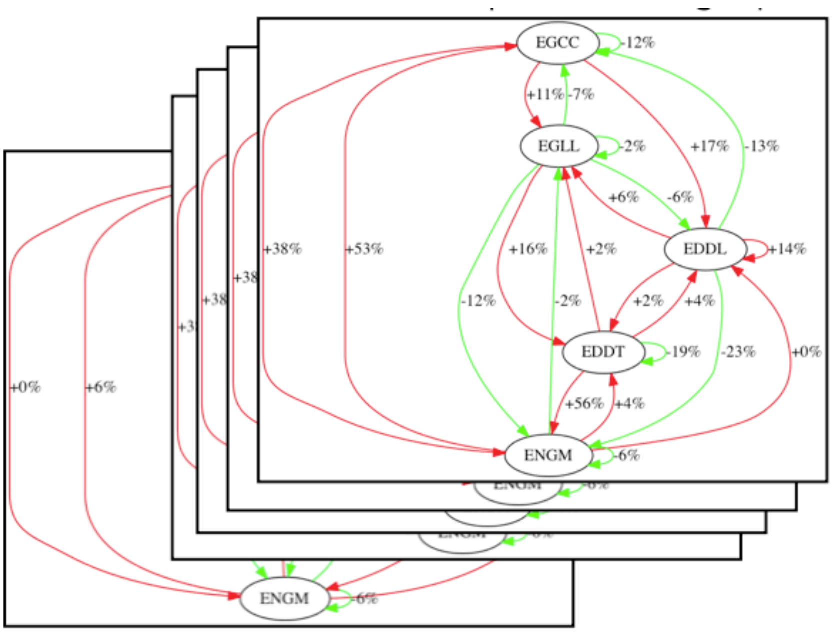

Figure 9 Resilience graph: multiple disturbances

4.8 The error and confidence layer

The previous layered graph model provides an overall picture of the resilience of the system under disturbances. For many uses this picture would be enough to consider, however, in practice it is also of importance to consider not only the values of the parameters and metrics, but also a set of "meta-information" bounding the error and confidence of those. For this reason attached to each of the previous graph layers (Reference Graph and Disrupted state Graphs) there is an additional graph layer with the same nodes and edges but weighted by two parameters, a statistical significance or p-value (how sure one can be about the statistical approximation, typically 95% although it could go down depending on the sample distribution) and a confidence interval (a proxy of the most likely error in the parameter estimation within the statistical significance), While doing the data cleansing both methods EVA and ODDS provide a statistical significance, the linear fitting also provides a confidence interval for the parameter estimation. Assuming the processes are independent it is straightforward to approximate the global significance and confidence interval.

The following is just an example of what the error and confidence layer looks like:

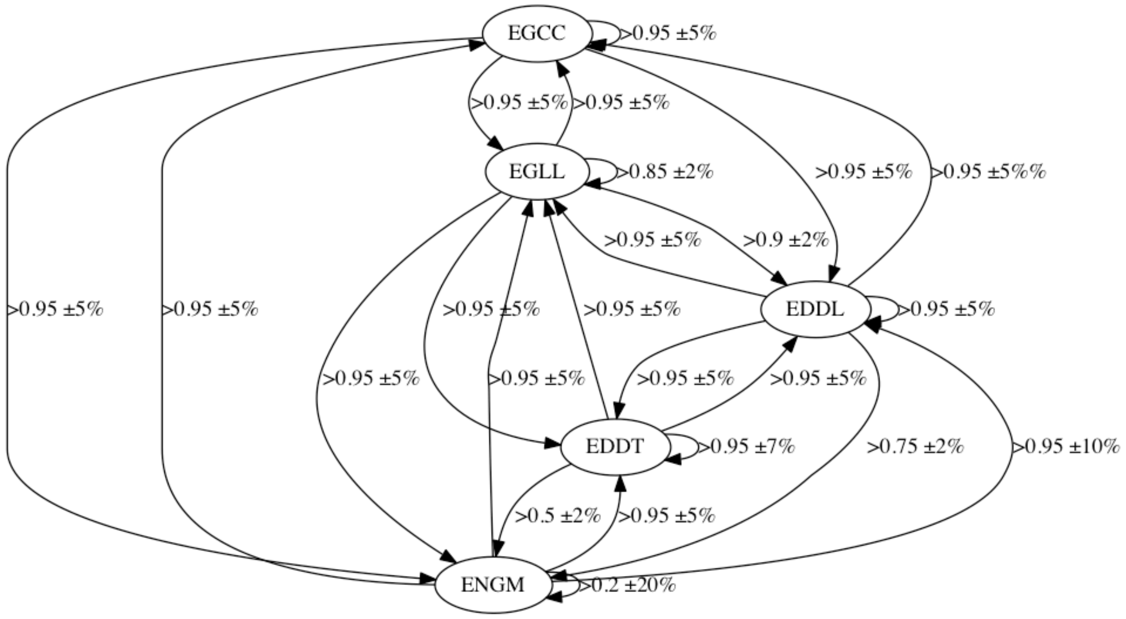

Figure 10 Error and confidence

4.9 The global resilience picture

All the elements described in this section determine the global resilience picture; A ground two-graphs layer, containing the information regarding to the Reference State of the system and, over it, several two-graphs layers (one for each disturbance considered) containing relative information of the system under certain isolated disturbance or Disturbed State with respect to the Reference State. Each layer is composed of two graphs: one containing the Delay Rates, in case of the Reference State, and the Disturbance Effect, for the disruption layers, and another layer containing the statistical significance and confidence intervals of those.

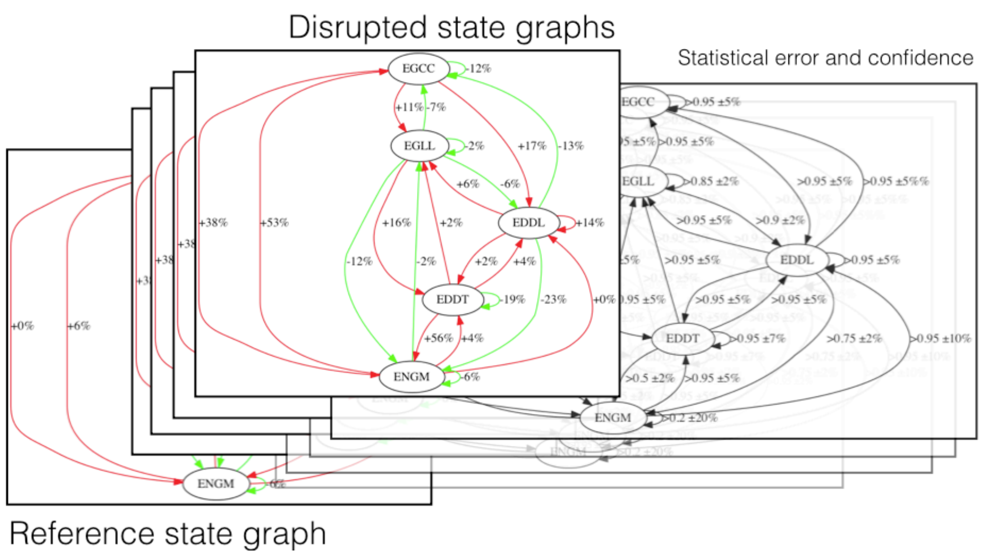

Figure 11 Global resilience picture

4.10 The Resilience Matrix

Alternatively to the graph representation one can consider the adjacency matrix. Adjacency matrices comprises exactly the same information as graphs, but they could be a better presentation when the graphs are highly connected specially if there is a considerable number of nodes involved. In addition matrix representation allows a better communication in a digital world providing a more suitable portrayal for exporting and integration with current systems. In the Resilience Matrix (or matrices) values are stored in several csv files, two for the Reference Status plus two additional files for each disturbance studied (or Disturbed Status). Each file contains as many columns and rows as the number of nodes (airports) considered, in each position the weight of the respective graph is stored.

	EGCC	**EGLL**	**EDDL**	**EDDT**	**ENGM**	**...**
EGCC	-5%	-3%	-3%	n/a	0%	...
EGLL	+9%	+8%	-4%	+2%	-1%	...
EDDL	-1%	0%	-7%	+5%	+12%	...
EDDT	0%	+8%	+10%	-12%	-9%	...
ENGM	+6%	n/a	0%	-3%	+6%	...
...	...	...	...	...	...	...

5 A causality perspective

5.1.1 Rationale for a causality delay analysis

As described in Section 4, the methodology previously proposed is centred on the analysis of single disturbances, and on how these disturbances propagate through the network after one or two flights of an affected aircraft. This leaves several questions unanswered, as for instance: how delays propagate after these first two flights? Is there any global pattern of delay propagation, i.e. independent of the originating disturbance? Within this propagation process, are there some airports more important than others? In this Section, we try to answer this questions by analysing the propagation of any delay through the network by means of causality measures.

The rationale behind this analysis is that, when delays appear in one airport, these may be propagated to other airports by connecting flights, thus creating a *cascade effect*. By analysing how delays evolve in pairs of airports, such effect should be detectable; yet, a simple correlation analysis may provide unreliable results, as for instance two airports may have delays because of bad weather, and not because a real delay propagation has occurred. The solution to this problem has been already proposed 40 years ago in economy, as involves the use of *causality* metrics, *i.e.* metrics able to detect the presence of a "forcing" between two time series, discarding any co-effect. In what follows, such causality metric is described, and then applied to the evolution of delays in 50 european airports.

5.1.2 Granger causality

The Granger causality test is an extremely powerful tool for assessing information exchange between different elements of a system, and understanding whether the dynamics of one of them is led by the other(s). Firstly introduced by the Nobel Prize winner Clive Granger (Granger, 1969), its main applications have been inside the field of economics (Hoover, 2001); yet, recently has been successfully applied to other fields of research, as for instance the analysis of biomedical data (Brovelli *et al.*, 2004; Kamiski *et al.*, 2000; Roebroeck *et al.*, 2005).

Classical statistical instruments, like, for instance, correlation analysis, are only able to assess the presence of some common (equivalent) dynamics between two or more systems. However, correlation does not imply causality. Granger causality, on the other hand, is held to be one of the few tests able to detect the presence of causal relationships between different time series. The two axioms, on which this test are based, are as follows: firstly, causes must precede their effects in time, and secondly, information relating to a cause's past must improve the prediction of the effect above and beyond information contained in the collective past of all other measured variables (including the effect).

Following the previous ideas, a time series p is considered to Grange-cause another time series q if the inclusion of past values of the series q can improve the process of forecasting the values of the time series p. In this case, the future evolution of p also depends on the past values of q. Also, it should be noted that two time series presenting a high correlation, or two time series that are 'forced' by a third system, do not pass the Granger causality test: as they have similar values, one of them cannot convey useful information for the forecast of the other. Yet, claims of causality from (multiple) bivariate time series should always be taken with caution, as true causality can only be assessed if the set of two time series contains all possible relevant information and sources of activities for the problem (Granger, 1980), a condition that real-world experiments can only rarely comply with (Zanin and Papo, 2013).

5.1.3 Network reconstruction

The network reconstruction process starts with time series representing average landing delays across European airports. This information has been extracted from the Flight Trajectory (ALL-FT+) data set as provided by the EUROCONTROL PRISME group, as described in Resilience2050 D2.1 and D2.2.

For the top-50 European airports (in terms of movements, as recorded in 2011), a time series has been extracted, representing the average hourly delay of arriving flights. Due to some missing days, each time series comprises 7440 values. These delay time series presented several trends, as strong delays are expected mainly at peak hours, during week days, and during summer, *i.e.* those periods in which the traffic is higher. A detrend process has then be performed using delay values corresponding to one week (168 hours) before and after each value.

Starting from this information, a network is reconstructed, where each one of the 50 nodes represents an airport of the set. For each pair of nodes (airports), the Granger Causality is calculated. This involves, at each available time step, the forecast of the next value of the time series by means of a multi-linear regression, using the information of the last 24 hours; two errors are then compared: the one corresponding to the forecast obtained using only information about the first node, and the one corresponding to the forecast including information extracted from the time series corresponding to the second node. The result is then expressed as a F-Statistics significance level, assessing whether the two forecast errors are significantly different, and thus whether some causality has been detected in the data. A link between two nodes, A and B, is then created when two conditions are simultaneously met: there is a significant causality between A and B (significance level greater than 0.99), and no causality is detected between B and A (significance level lower than 0.99). This reduces the effects of confounding factors, *e.g.* the presence of a third airport forcing both A and B, that would result in bidirectional causalities. The result is an unweighted directed network, where bidirectional links are forbidden.

The following image depicts the resulting network, with arrows indicating the presence of causality between the delays of pairs of airports. The size of the nodes represents their *out-degree*, *i.e.* the number of airports they are "driving", while the color represents their *in-degree*, *i.e.* the number of airports forcing their dynamics.

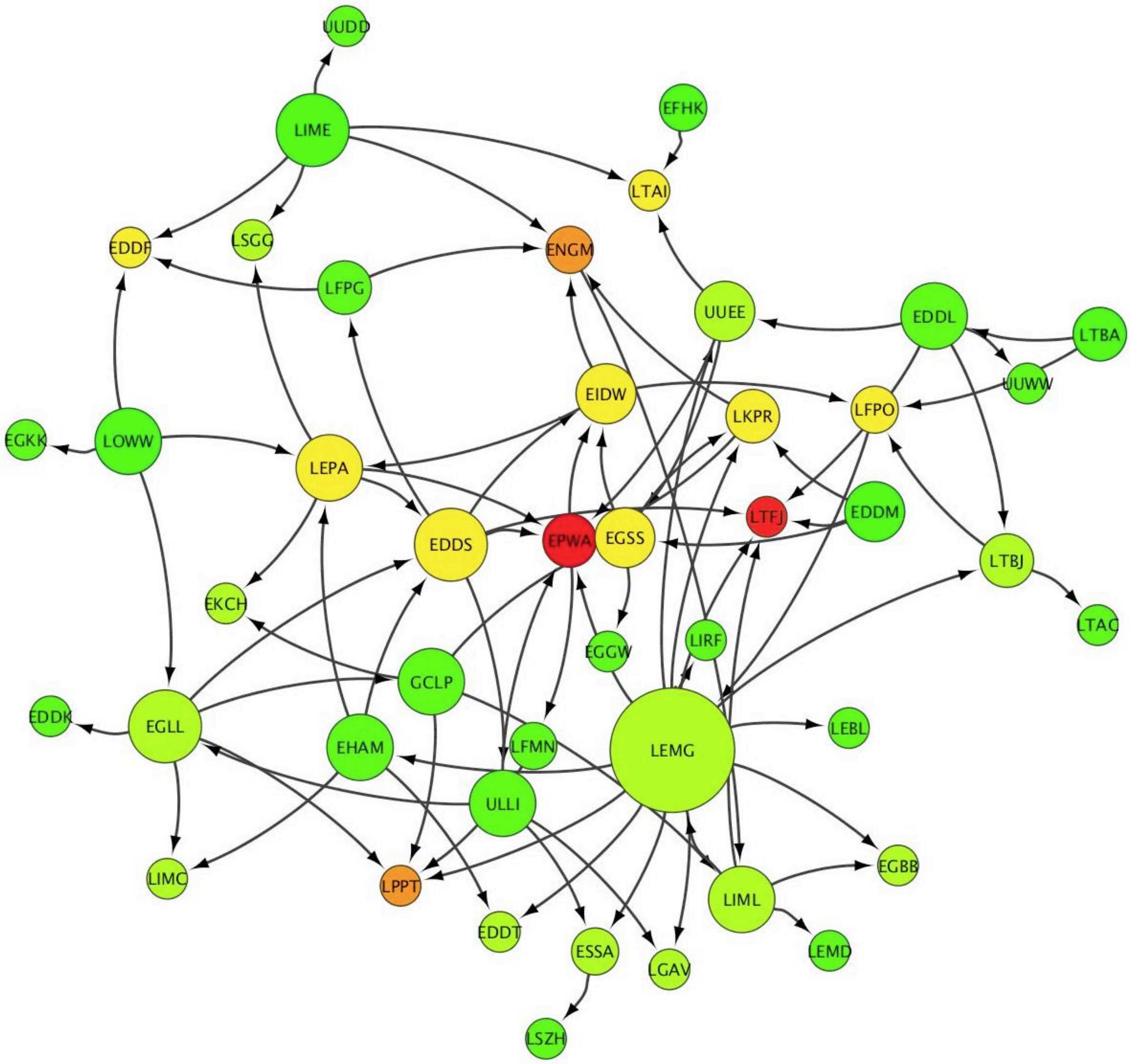

Figure 12 Causality analysis

As it can be noticed, the most important airport of the network, in terms of spreading delays through the system, is LEMG - Malaga, Spain. Due to its high connectivity with different part of Europe, delays here generated can easily propagate to Germany (EDDT), Nederland (EHAM), Greece (LGAV), Russia (UUEE) or Italy (LIRF). On the other hand, the two airports that are mostly being driven by other airports are EPWA (Frederic Chopin Airport, Warsaw) and LTFJ (Sabiha Gökçen, Istanbul).

6 Annex 1 - Acronyms

Table 6-1 List of acronyms

Term	Definition
AAT	Aeronautics and Air Transport
ALL_FT+	All Flights Trajectories
ANSP	Air Navigation Service Provider
ATFCM	Air Traffic Flow Capacity Management
ATFM	Air Traffic Flow Management
ATM	Air Traffic Management
CDM	Collaborative Decision Making (Airports)
CFMU	Central Flow Management Unit
CORDIS	Community Research and Development Information Service
CRCO	Central Route Charges Office
CSV	Comma Separated Value
D.X.Y	Deliverable X.Y
DDR	Demand Data Repository
DHMI	Devlet Hava Meydanlari Isletmesi (The General Directorate of State Airports Authority in Turkey)
DLR	Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Centre)
DoW	Description of Work

EC	European Commission
FP	Framework Programme
ICAO	International Civil Aviation Organization
INX	The Innaxis Foundation and Research Institute
ITU	Istanbul Teknik Universitesi (Istanbul Technical University)
KCL	King's College London
METAR	Meteorological Aerodrome Report
NLR	Nationaal Lucht en Ruimtevaartlaboratorium (The National Aerospace Laboratory – The Netherlands)
NOP	Network Operation Plan
PRISME	Pan-European Repository of Information Supporting the Management of EATM
QNH	Atmospheric Pressure (Q) at Nautical Height
R&D	Research and Development
RAF2050	Resilience Analysis Framework
SES	Single European Sky
SESAR	Single European Sky ATM Research Programme
TAF	Terminal Area Forecast
UPM	Universidad Politécnica de Madrid (Madrid Technical University)
WP	Work Package

7 Annex 2 - References

[1] J. Van den Broeck, S. A. Cunningham, R. Eeckels, and K. Herbst. Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities. PLoS Med. 2(10):e267, 2005.

[2] D. Sornette. Dragon-Kings, Black Swans and the Prediction of Crises. 2009arXiv0907.4290S, 2009

[3] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, London. 2001.

[4] C.C. Aggarwal, P.S. Yu. An Effective and Efficient Algorithm for High- dimensional Outlier Detection. VLDB Journal, 14(2):211-221, 2005.)

[5] Björck. *Numerical methods for least squares problems*. Philadelphia: SIAM.ISBN 0-89871-360-9. 1995

[6] Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., Bressler, S. L., 2004. Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. Proceedings of the National Academy of Science USA, June 29, 101 (26), 9849-9854.

[7] Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37 (3), 424–438.

[8] Granger, C. W. J., 1980. Testing for causality: a personal viewpoint. J. Econ. Dyn. Control, 2, 329352.

[9] Hoover, K.D., 2001. Causality in macroeconomics. Cambridge University Press.

[10] Kamiski, M., Ding, M., Truccolo, W. A., Bressler, S. L., 2000. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. Biological cybernetics 85 (2), 145-157.

[11] Roebroeck, A., Formisano, Elia., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. NeuroImage 25 (1), 230-242.

[12] Zanin M. and Papo D., 2013. Efficient neural codes can lead to spurious synchronization, Frontiers in Computational Neuroscience 7, 125.