D4.3 Final Validation Report

Document information	
Project Title	Empirically grounded agent based models for the future ATM scenario
Project Number	E.02.18
Project Manager	Deep Blue
Deliverable Name	D4.3 Final Validation Report
Deliverable ID	4.3
Edition	00.00.05
Template Version	03.00.00
Task contributors	
XXXXXXXXXXXXXX	

Abstract

This deliverable contains the ELSA final validation report, detailing the objectives, the exercises carried out and their outcomes.

Authoring & Approval

Prepared By - *Authors of the document.*		
Name & Company	Position & Title	Date
XXXXXXXXXXXXXX / Deep Blue	Validation Expert	15/02/2014
XXXXXXXXXXXXXX / Deep Blue-UniRoma	WpE PhD Candidate	15/02/2014
XXXXXXXXXXXXXX / UniPa	Grantholder	15/02/2014
XXXXXXXXXXXXXX / SNS	Post-Doc	15/02/2014

Reviewed By - *Reviewers internal to the project.*		
Name & Company	Position & Title	Date
XXXXXXXXXXXXXX / Deep Blue	Validation Expert	20/02/2014

Approved for submission to the SJU By - *Representatives of the company involved in the project.*		
Name & Company	Position & Title	Date
XXXXXXXXXXXXXX / Deep Blue	Project Coordinator	24/02/2014

Document History

Edition	Date	Status	Author	Justification
00.00.01	15/01/2014	First Outline	XXXXXXXXXXXXXX, XXXXXXXXXXXXXX	New Document
00.00.02	25/01/2014	First Complete Draft	XXXXXXXXXXXXXX, XXXXXXXXXXXXXX i	First version drafted
00.00.03	20/02/2014	Second Draft	XXXXXXXXXXXXXX, XXXXXXXXXXXXXX, XXXXXXXXXXXXXX	Version updated
00.00.04	24/02/2014	Proposed Version	XXXXXXXXXXXXXX	Version updated
00.00.05	20/03/2014	Second Proposed Version	XXXXXXXXXXXXXX	Version updated

Intellectual Property Rights (foreground)

This deliverable consists of Foreground owned by one or several Members or their Affiliates.

Table of Contents

EXECUTIVE SUMMARY ... **6**

1 INTRODUCTION ... **7**

1.1 PURPOSE OF THE DOCUMENT ... 7
1.2 INTENDED READERSHIP ... 7
1.3 STRUCTURE OF THE DOCUMENT ... 7
1.4 ACRONYMS AND TERMINOLOGY ... 7

2 VALIDATION OBJECTIVES ... **9**

2.1 V0 ATM NEEDS ... 9
2.2 V1 SCOPE ... 10
2.2.1 WP1 ... *10*
2.2.2 WP2 ... *11*
2.2.3 WP3 ... *11*
2.3 STAKEHOLDERS INVOLVED IN ELSA VALIDATION ACTIVITIES ... 12
2.4 SUCCESS CRITERIA ... 13

3 VALIDATION METHODOLOGY AND EXERCISES ... **14**

3.1 WP1 ... 14
3.2 WP2 ... 15
3.3 WP3 ... 16

4 RESULTS OF THE VALIDATION ... **18**

4.1 WP1 ... 18
4.1.1 Best analyses: feedbacks collected from ATM experts ... *18*
4.2 WP2 ... 19
4.2.1 Tactical layer ... *19*
4.2.2 Strategic layer ... *22*
4.3 WP3 ... 24
4.3.1 Scope ... *24*
4.3.2 Accuracy ... *25*
4.3.3 Expected Operational Benefits ... *25*
4.4 SUCCESS CRITERIA MATCHING ... 25

5 CONCLUSIONS ... **27**

6 REFERENCES ... **28**

A.1 COMMUNITY ANALYSIS: SEASONAL VARIATIONS AND IMPACT OF THE ASH CLOUD ... 29
A.2 ATCO MANAGEMENT STRATEGIES: DIFFERENCES AMONG COUNTRIES AND SECTORS ... 35
A.3 ANALYSIS OF THE MOST CRITICAL NAVPOINTS OF THE ITALIAN AIRSPACE ... 41
A.4 EXPLORATORY ANALYSIS OF SAFETY DATA AND THEIR INTERRELATION WITH FLIGHT TRAJECTORIES AND NETWORK METRICS ... 47

List of tables

Table 1. Stakeholder involved in Validation activities 12
Table 2. Mapping between validation objectives and success criteria. 13
Table 3: WP1 Validation Exercises 14
Table 4: WP2 Validation Exercises 16
Table 5: WP3 Validation Exercises 17
Table 6: Matching of the validation objectives with their related success criteria........ 25
Table 7: Rank of the ten most critical nodes in terms of strength........ 42
Table 8: Rank of the ten most critical nodes in terms of betweenness 43
Table 9: Rank of the ten most critical nodes in terms of Origin-Destination betweenness 43
Table 10: Rank of the ten most critical nodes in terms of Random-Walk Betweenness 44
Table 11: Overview of navpoints ranking with respect to the different metrics........ 45
Table 12: Rank of the ten most critical nodes in terms of strength and betweenness 48

List of figures

Figure 1: Cumulative distribution of Delay in Simulation compared with Real data 20
Figure 2: Comparison of FRAC values of simulated and real data when varying the maximum re-routing angle (all Airac 334). 21
Figure 3: Comparison of FRAC values of simulated and real data when varying the maximum re-routing angle (6 May 2010). 21
Figure 4: Correlation between frac on the days of Airac334........ 22
Figure 5: Real and simulated distribution of degree for N_{sp} =1 (left) and N_{sp} = 10 (right)........ 23
Figure 6: Real and simulated distribution of topological length of flights for Nsp =1 (left) and Nsp = 10 (right) 23
Figure 7: Distance between real and simulated distributions for the four targets as a function of the parameter of control N_{sp} 24
Figure 8: Map of airport network in day 11/04/2010 of AIRAC 333 (before ash cloud crisis) portioned in communities according to the Infomap algorithm. The points correspond to the airports, while their color identify the communities they belong to. For readability purposes, the communities with less than 10 airports are all in black. 30
Figure 9: : Map of airport network in day 18/04/2010 of AIRAC 333 (during ash cloud crisis) portioned in communities according to the Infomap algorithm. The points correspond to the airports, while their color identify the communities they belong to. For readability purposes, the communities with less than 10 airports are all in black. 31
Figure 10: Map of airport network during the week starting on July 15th 2010 portioned in communities according to the Infomap (top) and Blondel (bottom) algorithms. The points correspond to the airports, while their color identify the communities they belong to. For readability purposes, the communities with less than 10 airports are all in black. 32
Figure 11: Map of airport network during the week starting on November 11th 2010 portioned in communities according to the Infomap (top) and Blondel (bottom) algorithms. The points correspond to the airports, while their color identify the communities they belong to. For readability purposes, the communities with less than 10 airports are all in black. 33
Figure 12: Spain ATCO management strategy - May 2010........ 36
Figure 13: Germany ATCO management strategy - May 2010 36
Figure 14: UK ATCO management strategy - May 2010 37
Figure 15: UK ATCO management strategy - September 2010 37
Figure 16: Germany ATCO management strategy - September 2010 38
Figure 17: Spain ATCO management strategy - September 2010........ 38
Figure 18: France ATCO management strategy - September 2010........ 39
Figure 19: Italy ATCO management strategy - September 2010 39
Figure 20: ATCO management strategy in sector MIOW of Rome FIR 40
Figure 21: ATCO management strategy in sector OV of Rome FIR 40
Figure 22: Map of the strength of the nodes in the Italian airspace........ 42
Figure 23: Map of the betweenness of the nodes in the Italian airspace 43
Figure 24: Map of the Origin-Destination betweenness of the nodes in the Italian airspace 44

Figure 25: Map of the Random-Walk betweenness of the nodes in the Italian airspace 45
Figure 26: Map of the strength of the nodes in the Italian airspace... 48
Figure 27: Map of the betweenness of the nodes in the Italian airspace 49
Figure 28: Navpoints in the Rome FIR with their values of strength and betweennes related to the number of STCA assigned to them. The vertical and horizontal dashed lines correspond to the average value of strength and betweenness respectively 51
Figure 29: Navpoint Network in the Rome FIR. Red nodes are nodes correctly identified as critical crossing the strength and betweenness centrality metrics, blue nodes are critical nodes with small metrics. The size of the nodes is proportional to their value of strength. The red dashed circles highlight the presence of outliers in some boundary areas. 52
Figure 30: Critical Nodes across airways in the central part of the network. The red circle highlights outliers close to the boundary of a sector. 53
Figure 31: Navpoints in the Rome FIR with their values of strength and Alt related to the number of STCA assigned to them. The vertical and horizontal dashed lines correspond to the average value of Alt and strength respectively 54

Executive summary

This document presents the validation activities carried out for the ELSA project. The validation activities were dedicated to three main areas: refinement of the results of the first ELSA work package, steering the development of the Agent Based Model and support the definition of the proof of concept of the decision support tool.

WP1 proposed a set of analyses to measure the propagation of disturbances in the ATM system. The validation objectives for the analyses were related to the identification of a set of “best analyses” that were progressively short-listed from the initial WP1 set to select the ones which seemed more relevant from the operational point of view. Discussions with ATM experts were carried out to collect their feedback and steer further improvements of the proposed solutions. For each analysis it was produced a brief factsheet containing all the relevant information and in particular the solution expected benefits and limitations. In this document we focus on the main feedbacks gathered from the experts to underline the potential applications of the proposed solutions.

The second area of validation was related to the development of the Agent-Based Model. The contribution of the validation activities was to make sure that both layers remain connected to the WP1 results, both as an input and as a comparison of output, and to inform all the design decisions with operational considerations, to ensure a high level of consistency between the operational world and the models. For this purpose, there was a constant interaction among the ELSA partners and with external experts from the ATM world and from a major EU airline. In this document we report the analyses specifically carried out to validate the model outputs. We show that both layers are capable of reproducing the real ATM scenario with a good degree of accuracy.

Finally, the third area of validation involved the design of a proof-for-concept of the ELSA decision support tool. The validation activities for WP3 lead to the refinement of four aspects: roles, timeframe, scenarios, supported decisions.

1 Introduction

1.1 Purpose of the document

This document describes the validation activity and its results. The purpose is to provide a detailed description of the validation process carried out for the ELSA project, including the methodology adopted and the exercises performed. The aim is to present how the research work benefited from the validation activities and the different level of achievement of the validation objectives. Thus the document will provide a detailed discussion of the limitations and problems encountered.

1.2 Intended readership

This document is relevant for the ATM research community at large and more specifically for the WP-E research community.

External experts may also be part of the readership, in all those cases where their expected contribution to ELSA could be clarified by extracting text from this document.

1.3 Structure of the document

The Validation report structure has been derived from the SESAR Validation Plan Template [1] and adapted for a long-term research initiative, considering the E-OCVM Methodology [2]. The present document includes the following sections:

- Introduction,
- Validation approach, with objectives, methods and techniques, success criteria,
- Description of the validation results
- References.

1.4 Acronyms and Terminology

Term	Definition
ABM	Agent-Based Model
ACC	Area Control Centre
AIRAC	Aeronautical Information Regulation and Control
ANSP	Air Navigation Service Provider
AOC	Aircraft Operations Centre
ATCO	Air Traffic Controller
ATM	Air Traffic Management
DBL	Deep Blue
ECAC	European Civil Aviation Conference
EEC	EUROCONTROL Experimental Centre

Term	Definition
ELSA	Empirically grounded agent based models for the future ATM scenario
FIR	Flight Information Region
NM	Network Management
OCC	Operation Control Centre
SESAR	Single European Sky ATM Research Programme
SJU	SESAR Joint Undertaking (Agency of the European Commission)
SNS	Scuola Normale Superiore of Pisa
STCA	Short Term Conflict Alert
TMA	Terminal Manoeuvring Area
UniPa	University of Palermo
WP	Work Package

2 Validation objectives

According to the principles set up in the European Operation Concept Validation Methodology (E-OCVM) the objectives of the validation work for ELSA were specified according to the Maturity Level of the proposed solutions. Therefore main objectives for the V0 Maturity level were related to the identification, definition, and refinement of the following aspects: ATM Needs, Scope, Target Users, and identification of R&D Areas.

The second set of validation objectives was related to the V1 maturity level, concerning the identification of solutions for the target ATM needs. As the project progressed through the WPs, these aspects were subject to a progressive refinement due to an increasing understanding of the domain and its related needs in a long term timeframe.

2.1 V0 ATM Needs

The identification and definition of the ATM needs was the main objective at this level. The initial understanding of the main ATM need was to "Understand propagation of disturbances, like delays, or safety events".

However, while progressing during the project this need had to be refined to take into account the set of information that could be retrieved from the available data set. In fact since only data after the take-off until the landing were available, the understanding of the propagation of disturbances and delays had to be limited to this phase. The refined ATM needs were then the following:

- analyse changes and variation in predictability during execution phase of the flight,
- analyse the sensitivity of predictability to external disturbances, like weather and traffic load,
- analyse the sensitivity of predictability to change of strategies by ATM actors, i.e. NM and Airline Operators.

The main R&D need was to deepen the understanding of the ATM problem and its contributing factors in order to develop or further elaborate the proposed analyses.

These needs were mainly addressed by WP1 which focused on identifying a set of best analyses and assessing their fitness-for-purpose. WP2 was meant to address the same needs also in a SESAR scenario. However, as it will be discussed in section 2.2.2, its contribution was limited to the current ATM scenario.

The results obtained in WP1 and the initiation of WP2-3, moved the R&D focus to more applied areas, less research-oriented than the ones just mentioned. The target ATM needs for this phase were the following:

- improve the strategic planning phase by using network analysis,
- improve the monitoring phase of the flight execution by analysing its predictability,
- improve the management of standard (e.g. seasonality) and non-standard (e.g. big shocks) fluctuations.

Such a formulation resulted in the following WP3-specific ATM Needs:

- identify how to manage ATM as a complex system,
- detail the work of ATM operators in a future SESAR scenario, including their tools, tasks, and communication flows.

The **scope** was limited to analyses at the level of the ECAC area, analysing only en-route delays (i.e no reactionary delays) since only data on the execution phase of the flights were available. Outlier cases, like the volcano ash cloud or less extreme events like bad weather, were also included in the analysis.

The identification of **target users** proved to be a very difficult task. We were able to identify target users in sufficient detail, but it was more difficult to gather their expectations in terms of validation. The methodological problem we faced was that target users tended to provide their inputs as if ELSA was supposed to deliver solutions for the near future, while we needed them to position in the far future, in an operative context which was often beyond their imaginative capacity. Nevertheless **Target users** have been progressively identified, starting from an initial target of generic stakeholders, interested in network management, moving to target users like the Central Office for Delay Analysis and the Network Management at the ECAC level. At the national level, the target users included the ACC supervisor, the ANSP Performance Review Unit and the Safety Department.

2.2 V1 Scope

The main objective of this phase was to identify the operational and technical solutions for meeting the target performances identified in phase V0 and addressing the related needs. The main R&D need in this phase was the identification of the context of use of the proposed solutions. However, this need coexisted with that of continuously increasing the understanding of the ATM problems. For example, what is clarified and understood in today's scenario may require additional fine-tuning in a SESAR one. The validation activities that led to the definition of the proposed solution were carried out in an iterative way in order to progressively refine the proposed solutions to better match the ATM needs that were simultaneously clarified.

The objectives at this level were mapped per WP, as V1 objectives were addressed in all the three technical WPs, with some differences depending on the technical work carried out.

2.2.1 WP1

This WP focused on delivering a set of analyses to address the target ATM needs. The validation objectives for the analyses were related to the identification of a set of "best analyses" that should be able to address the identified ATM needs. For each of the identified analysis we set the following objectives:

- describe the proposed solution,
- describe the expected benefits,
- match the analysis results with KPA
- describe the target operating scenario,
- identify the enablers and constraints,
- expose any hidden assumption,
- describe the limitations of the proposed solution.
- anticipate new problems that could be created if the new solution becomes operational.

Each analysis was also presented according to a standard factsheet that contains the following fields:

- Title of the proposed analysis,
- What: brief description of what the proposed analysis aims to measure,
- Why: rationale for selecting this analysis, including ATM relevance,
- Results: summary of main results obtained,
- How: details on how the analysis is carried out,
- References: list of relevant references if any.

The validation activities have enabled us to show that all analyses have potential relevance for ATM, but while some of the outcomes may be relevant in a future SESAR scenario, in the present one their applicability is limited.

With respect to the specific WP1 validation objectives we were able to clearly achieve the following:

- describe the proposed solution,
- describe the expected benefits,
- describe the limitations of the proposed solution,

- identify the enablers and constraints.

The standard factsheets used for presenting the analyses to the ATM stakeholders proved to be very effective in describing the proposed solutions and their expected benefits and limitations. All the interviewed stakeholders were able to clearly understand the main points of the proposed analyses. Specific validation activities contributed in clarifying the possible application of the proposed analyses along with the enablers and constraints that could facilitate or make harder a successful application.

The other validation objectives were instead more difficult to achieve. In particular the identification of the target operating scenario was one of the major issues. In fact the interviewed stakeholders found it difficult to position themselves in the future and to imagine a possible scenario of application. Their understanding of how the future ATM will be structured was closely linked to the present scenario with a limited perception of the innovations that will take place with SESAR and beyond.

Details of the feedbacks collected from the ATM experts are reported in section 4.1.

2.2.2 WP2

The main validation objectives for the ABM development were related to its ability in simulating the SESAR scenario, thus allowing a better analysis of the proposed solutions expected benefits and the possibility of anticipating new problems that could be created if the new solutions becomes operational. However, the development of the Agent-Based Model led to a redefinition of the main validation objectives related to WP2. In fact the ABM did not reach a sufficient maturity in the current scenario to be extended to the SESAR one. The achievement of the following objectives required in fact more effort than expected:

- inform the ABM characteristics from a set of regularities identified by WP1. The identification of these regularities and its application to the model required a lot of effort both in the programming and the validation part. In fact some of the regularities identified in WP1 were not applicable to model while others have to be refined with the help of ATM experts to clarify how to implement them correctly. Not all the WP1 results could be directly fed into the model and some aspects required additional work to achieve the level of detail required by WP2.
- inform the ABM outcome metrics from a set of WP1 regularities.

In addition WP2 activities suffered from a cascading effect due to WP1 delay. In fact we had to invest 8 months in understanding the data and finding the right data set before we were ready to start the analysis process. Moreover, WP1 had to be left open in order to feed the “right” regularities to WP2, because most of those emerged in WP1 were not directly useful for the ABM development. Finally, during the interviews with the ENAV experts, we noticed that multi-sector directs were already a common practice in the controllers’ behaviour. Therefore our idea of simulating the SESAR scenario by replicating the first sector was not enough and we would have needed to further extend the model to the size of at least a complete FIR.

For the above reasons the following objectives were discarded:

- build a multi-layer ABM,
- build a scalable ABM, that can simulate from one sector to multiple sectors,
- simulate a realistic SESAR scenario,

Besides these we also avoided to integrate the Strategic and the Tactical layer because it was considered not a priority since the M1 trajectories used as input for the Tactical layer could be retrieved from our dataset in a more simple and realistic way especially when simulating a single sector scenario.

2.2.3 WP3

The validation objectives for WP3 were mostly related to the R&D areas of:

- context of use: definition of the target SESAR scenario,
- human-technology interaction and to the definition of the needs of new roles, or modification of existing ones.

WP3 defined a solution to the target ATM needs by proposing a proof-for-concept of decision-support tools, including details on the target user(s), their role, the application scenario, and on the human-technology interaction modalities.

As such, WP3 especially addressed the following validation objectives:

- describe the proposed solution,
- describe the expected benefits,
- describe the target operating scenario.

Specific WP3 validation objectives were also identified, directly concerning the decision-support tool. These objectives were mostly related to the DST domain suitability and include:

- proposed visualisations for DST are suitable for the ATM tasks to be supported,
- proposed interaction modalities are suitable for the ATM tasks to be supported.

The interactions between DST and the ABM were initially included and later excluded in order to maximise the effort invested in the two WPs, minimising the mutual constraints.

2.3 Stakeholders involved in ELSA Validation Activities

The ELSA stakeholders include four major groups: the research community, SESAR/EUROCONTROL, ANSPs, Airlines. Their involvement in the Validation activities contributed, at the beginning of the project, to the definition and refinement of their high level needs. Subsequently they also effectively contributed in the validation of the proposed solutions. Table below list the main stakeholder groups, specific names attached to them, and their involvement in the ELSA validation activities.

Table 1. Stakeholder involved in Validation activities

Stakeholder	Internal/ External	Objectives	Relevant WP/ Activities
Research community	External	Application of complex science to ATM.	WP1 analyses and application in WP2.
SESAR JU	External	Realistic SESAR ABM. Identification of perturbation problems in SESAR scenario.	Multi-layer ABM – WP2. Application of analyses in WP2.
ANSP **- ENAV**	External	Applied analyses for monitoring of disturbances and of system performance.	Sub-set of WP1 analyses. ABM development
Airspace Users/Airlines **- AOC Alitalia**	External	Design of airline-oriented ATM services.	Concept scenarios delivered by WP3. ABM development
EUROCONTROL - **CODA** - **NOP** - **NOP/DDR/route network sub-group** - **Safety Team** - **MUAC Safety**	External	Applied analyses for delay monitoring. Applied analysis for network management. Applied analyses for safety monitoring.	Sub-set of WP1 analyses.

2.4 Success Criteria

In this section we report the success criteria for each validation objective. Some of them were already defined in D4.1 while others were added subsequently. In any case this approach will make easier to identify if the validation objectives were partially or totally achieved.

Table 2. Mapping between validation objectives and success criteria.

Validation Objectives	Success Criteria
Correct identification of ATM need(s), scope, and target users.	Positive feedback by the identified stakeholders.
Definition of a set of proposed analyses to address the ATM needs, in current and in the SESAR scenario.	Analysis factsheets completed in all the subfields for a selected set of analyses, including definition of benefits and link with ATM needs.
Definition of a set of proposed analyses to address the ATM needs, in current and in the SESAR scenario.	Positive feedback on each of the proposed analysis by at least one stakeholder.
ABM Calibration: • build a multi-layer ABM, • model the interactions between the layers, • build a scalable ABM, that can simulate from one sector to multiple sectors, • simulate a realistic SESAR scenario,	• Positive feedback on the model inputs by the relevant stakeholders (Alitalia dispatcher, ENAV supervisor) • Use the strategic model outputs as inputs for the tactical layer • ABM working at multi-sector level • Positive feedback on the reliability of the model results in the SESAR scenario by at least one stakeholder
ABM Validation: • inform the ABM characteristics from a set of regularities identified by WP1. • inform the ABM outcome metrics from a set of WP1 regularities	• Use selected WP1 regularities to develop the model • Positive comparison of the model results with some of the regularities observed in WP1
Definition of the proof-for-concept of a decision-support tool: target SESAR scenario, human-technology interaction, new roles, or modification of existing ones.	Positive feedback on the proposed proof-for-concept by at least one of the following stakeholders: - Experts on Network and Flow Management, - Experts on Delay, - EUROCONTROL and SESAR JU.
Definition of the proof-for-concept of a decision-support tool: target SESAR scenario, human-technology interaction, new roles, or modification of existing ones.	New roles, or modification of existing ones clearly identified in WP3 and included in proof-for-concept.

3 Validation methodology and exercises

3.1 WP1

The methodology employed to reach the WP1 validation objectives was based on internal reviews and face-to-face and email interviews with ATM experts.

The interviews objectives were:

- assessing the ATM relevance of the proposed analyses;
- supporting the interpretation of analysis results;
- supporting the definition of follow-up analyses.

Internal reviews had instead the role of guiding the refinement of the analyses in order to address the suggestions retrieved from the stakeholders to better match the ATM needs they identified.

The validation exercises carried out for WP1 are listed in Table 3.

Table 3: WP1 Validation Exercises

Validation Exercise	Objective	Date
Visit to CODA (with email follow up on data quality)	Understand the dataset	4th April 2012
Workshop	Check consistency between M1 and M3	9th May 2012
Bilateral meeting between SNS and DBL	Comparison between AIRACs	5th June 2012
Bilateral meeting between UniPa and DBL	Consistency of flight trajectories characteristics, in terms of total number space/time evolution.	28th June 2012
Workshop with CODA & Webex with NOP	Cross-check of 1 day btw ELSA and CODA data	11th-12th July 2012
Internal review	Cross-check of airspace data: sectors, FIRs, TMAs.	6th November 2012
Interview with ENAV Supervisor	First ABM development	21st November 2012
Interview with ENAV Supervisor and Flow Manager	Progressive refinement of proposed analyses	6th March 2013, 22nd May 2013, 23rd June 2013 (by email)
Interview with ENAV Safety	Collect feedbacks on best analyses	November 2013

Validation Exercise	Objective	Date
Interview with SESAR Expert (ENAV Controller)	Feedback on best analyses	27th January 2014
Interview with MUAC Safety Experts	Feedback on a subset of best analyses	13th February 2014
Interview with Deep Blue Safety Experts	Feedback on a subset of best analyses	20th February 2014
Interview with CODA	Feedback on best analyses	18th March 2014

3.2 WP2

WP2 validation was divided in three strands of work: interviews with stakeholders, calibration and output analyses.

The interviews were structured around the following topics:

- Support in defining ABM: scope, scenarios to mimick/analyse, modelling relevant events/dynamics, and modelling relevant agents' behaviours,
- Iteratively assess first proposed ABM versions,
- Support in the interpretation of ABM results.

The ABM calibration instead ensured that:

- the relevant WP1 regularities informed the model
- the processes underlying the operation of the model are analysed and understood.

Finally the output analyses, that were done in the final phases of the ABM development, tried to answer the following questions:

- Which is the minimum set of parameters for the model to replicate selected real features observed in WP1?
- Is the model able to replicate real features of the ATM?

The following analyses were carried out to answer the above questions:

Tactical layer:

- Compare the distribution of en-route delays in the sector with the real one. The comparison was done while varying the de-conflicting strategy. The different de-conflicting strategies employed involved the use of multiple flight levels, the use of directs towards the sector exit point after the first re-routing or just as a global optimization strategy to avoid any conflict.
- Analysis of the metric "frac" defined in D1.3 as the fraction of flights which should have passed by a selected point and have not. We compared real values of "frac" for the points inside the sector LIRROV with the values obtained in the ABM simulation. The comparison was done by varying the maximum deviation angle allowed for re-routing.

Strategic layer:

- Given the network and the trajectories on the simulated network, we defined four distributions of interest for us: the degrees, the strengths, the weights, and the topological lengths of trajectories of the flights. The latter is the number of nodes that the flights cross on their trajectories. The difference between the real distribution and the simulated one gives an idea of how good the model fits the reality.

- Compare the number of rejected flight plans with the real one gathered from the operative knowledge of the Alitalia dispatchers.

A list of the validation exercises carried out for WP2 is presented in Table 4.

Table 4: WP2 Validation Exercises

Validation Exercise	Objective	Date
Bilateral review between SNS and DBL	Review of strategic ABM development status	24th June 2013
Bilateral review between UniPa and DBL	Review of tactical ABM development status	14th March 2013, 14th June 2013
Internal review	Review of ABM development	February 2013
Check if M1 trajectories are conflict free	Inform the ABM with observed features	March 2013
Internal review	Review of ABM development	July 2013
Interview with ENAV Supervisor and Flow Manager	Progressive refinement of tactical ABM	6th March 2013, 22nd May 2013, 23rd June 2013 (by email)
Interview with Alitalia Dispatcher	Progressive refinement of strategic ABM	10th May 2013, 4th June 2013, 8th July 2013
Interview with SESAR Expert (ENAV Controller)	Interpretation of ABM results	27th January 2014

3.3 WP3

The methodology employed to reach the WP3 validation objectives was based on internal reviews and face-to-face and email interviews with ATM experts.

The interview objectives were:

- assess the proposed context of use and scenarios,
- assess the proposed roles.

The validation exercises carried out for WP3 are listed in Table 3.

Table 5: WP3 Validation Exercises

Validation Exercise	Objective	Date
Interview with ENAV Supervisor and Flow Manager	Refinement of target scenarios	6th March 2013, 22nd May 2013, 23rd June 2013 (by email)
Interview with Alitalia Dispatcher	Refinement of target scenarios Assessment of proposed roles	10th May 2013, 4th June 2013, 8th July 2013
SESAR Expert	Refinement of target scenarios	27th January 2014, from 2nd February to 20th February 2014 (several interactions by email)
Interview with ENAV Controller, with previous experience of similar projects	Refinement of target scenarios Assessment of proposed roles	13th September 2013, 30th January 2014,

4 Results of the validation

In this section we present and discuss the results of the validation activities.

4.1 WP1

The validation activities for WP1 lead to the definition of a list of selected "best analyses". The final list is the following:

1. **Seasonal variations of communities:**

   We detected sets (communities) of nodes (airports, or sectors, or navpoints) clustered together. Communities are generically defined as sets of nodes that are more connected among themselves than with the rest of the network. The variations of communities during a period of time (day, week, month, year) can be analysed.

2. **Community analysis of the impact of the ash cloud:**

   This analysis applies the same method of the previous one (community detection), but it focuses on special events, like the volcanic ash cloud in 2010. Compared to the previous analysis a finer time resolution is needed, to detect changes in the community during the day.

3. **Analysis of ATCO management strategies: differences among countries and sectors:**

   Quantitative metrics are used to describe the ATCOs controlling strategies and to compare their behaviour across different countries and in time. For instance we correlated the negative delay (i.e. time gained) and the number of flights for different countries or single sectors.

4. **Analysis of the correlations between STCA events and Complex Network metrics in the Rome FIR:**

   We analysed the correlation between safety events and network metrics, thus matching the number of safety events with specific fixes. A high correlation is found between safety events and the fixes' strength and betweenness. Two thresholds are also found, one below which safety events tend to zero, and one above which there is a high probability of a high number of occurrences.

5. **Analysis of the most critical navpoints of the Italian airspace:**

   We analysed the Italian airspace in order to identify its most critical navpoints in terms of two specific network metrics: Strength and Betweenness. The rankings obtained with these metrics are compared with the operational experts' subjective assessments. Almost all the critical navpoints identified by the ENAV supervisor were among the first positions in our rankings. The end goal of this analysis is to build a quantitative taxonomy of fixes and to match it with the experts' knowledge.

4.1.1 Best analyses: feedbacks collected from ATM experts

Here we report the most relevant feedbacks about the best analyses collected during the interviews with the aviation stakeholders.

<u>Positive feedbacks:</u>

- **Network analysis is a powerful and innovative tool** that may have several applications in the ATM world. Only some of them were explored in ELSA due to the available dataset.
- ATM experts recognized some of the proposed analyses as **promising tools for monitoring the airspace structure especially in the future SESAR scenario**. Community analysis could help in efficiently driving the allocation of resources and improving the coordination between different entities. The analysis of safety occurrences and of the most critical navpoints are instead useful for monitoring the airspace allocation and the associated criticalities.

- The analysis of the ATCO strategies is useful in underlying commonalities in the management strategies that were not foreseen on the basis of the common knowledge of the operating procedure of other ANSPs. Moreover this analysis can be effective in showing different controlling strategies within the same ANSP (younger vs older controllers) to monitor the application of new procedures put in place.
- Some networks metrics (e.g. betweenness) can give useful information for designing the airspace because they are able to predict the location of the most congested areas without basing the assessment on a huge amount of historical flight data, but only on the intrinsic characteristics of the airspace structure.

Negative feedbacks:

- The practical applicability of the proposed concepts is difficult. The analyses are able to monitor interesting characteristics of the ATM structure but how to use these information in the current scenario is not trivial.
- According to the ENAV supervisor it is unlikely that a major redesign of the airspace will happen in the future. Therefore the impact of the analysis on safety occurrences and critical navpoints may be limited to the monitoring phase.
- Some of the network metrics are not easy to understand intuitively. Therefore trusting the proposed results is not immediate especially when the conclusions are not fully in accordance with the operative knowledge.
- The analysis of safety occurrences need to be further explored by enlarging the dataset and by linking the events to different metrics more related to the phase of flight they belong to (descent, ascent, cruise).

4.2 WP2

As stated in section 3.2 several analyses were carried out to validate the outputs of the Agent Based Model. Results of the analyses were compared with selected observed regularities already described in D1.3 when available. When this comparison was not possible we performed new analyses on real data to retrieve the needed information to compare or we asked a qualitative feedback from operative experts. Here we presents the analyses performed and their results.

4.2.1 Tactical layer

The main metric used to compare the results of the ABM simulation with those from WP1 was the en-route delay. This was computed as the difference between the planned sector transit time (derived from M1 trajectories, the last filed flight plans) with the actual one (derived from M3 trajectories, the routes effectively travelled by the aircraft). The distribution of the en-route delays is in fact the most direct measure of the controllers strategies in different operative situations. Therefore a positive matching between simulated and real data on en-route delay can be considered a good validation of the model effectiveness in reproducing the real world. Simulated and real results were compared at different stages of the model development in order to assess which was the minimum set of parameters required to correctly reproduce the real ATM scenario. The model parameters considered were:

- Number of available flight levels
- Use of directs after a re-routing
- Use of directs independently of re-routings as a global optimization strategy

The comparison was done using the Kolmogorov–Smirnov test (K–S test). In statistics, this is a nonparametric test for the equality of continuous, one-dimensional probability distributions that can be used to compare two samples (two-sample K–S test)[3] in order to assess their similarity. At the end of the development process the matching between real and simulated data is satisfactory (see Figure 1).

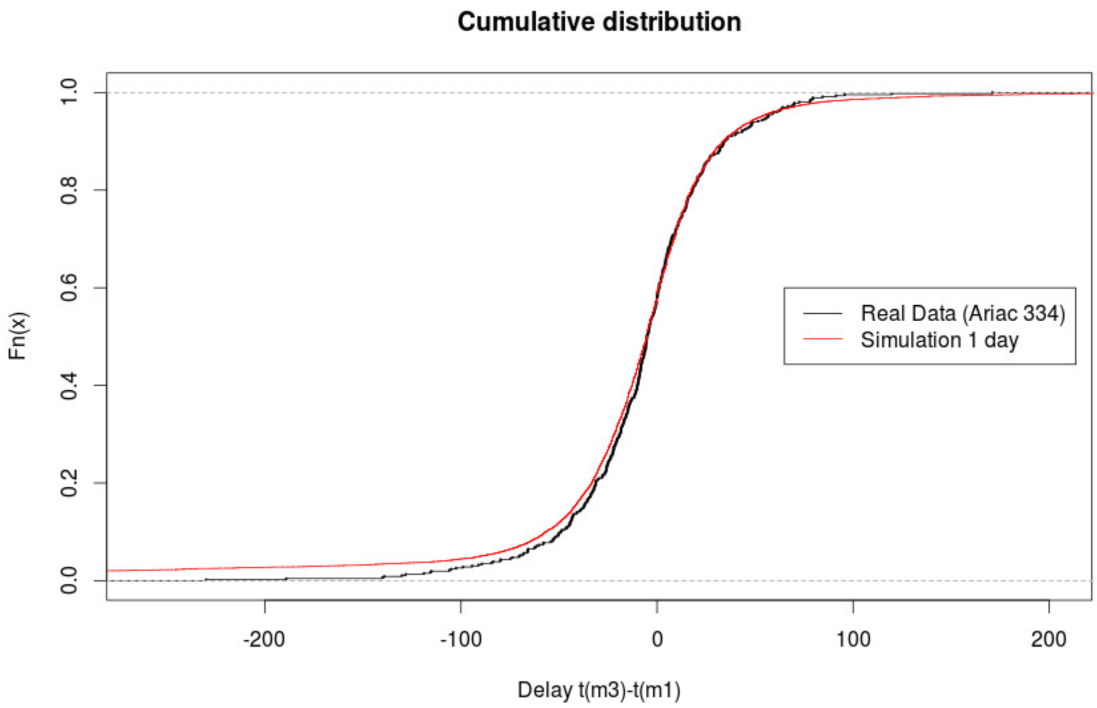


Figure 1: Cumulative distribution of Delay in Simulation compared with Real data

This results has been obtained by using a potentially infinite number of Flight Levels, and directs both after re-routings and as a global optimization strategy (for details see D2.3)

Another set of simulation involved the metric defined in D3.1 as “frac”. It is the fraction of flights which should have passed by a selected navpoint and have not. The “frac” value for each navpoint can be retrieved by comparing the last filed flight plans and the actual trajectories. The average value of “frac” within a sector is a measure of how the available airspace is used. Comparing real and simulated data can therefore help in validating the effectiveness of the model in replicating the real trajectories. This comparison was done in order to select the best value of the maximum allowed angle of deviation between trajectories in rerouting events.

- First we used a set of angles (4; 10; 20; 30 degree) and for each simulation we compared the value of “frac” for the navigation points inside the sector with real data.
- Then we tried to do the comparison by using the average value of “frac” in all Airac334 (Figure 2).
- Finally we compared the simulated results with real data of only the same day of the simulation (Figure 3).

For all cases, the objective comparison was done by computing the standard correlation between real and simulated data. We also computed a weighted correlation in which the standard one was corrected by taking as weights the strength (i.e. the traffic load) of the selected navpoint. Looking at the dynamics of the frac during the month, a particular behaviour across the working week does not seem to emerge. However we calculated the correlation matrix (unweighted) between the frac on each day of the Airac334. Looking at the distribution of the correlation coefficient we can observe that high value of correlation are reached (mean 0.088 std 0.352) (Figure 4). This is an indication of the ability of the model of reproducing the real patterns on a short time-scale which has the length of about one day.

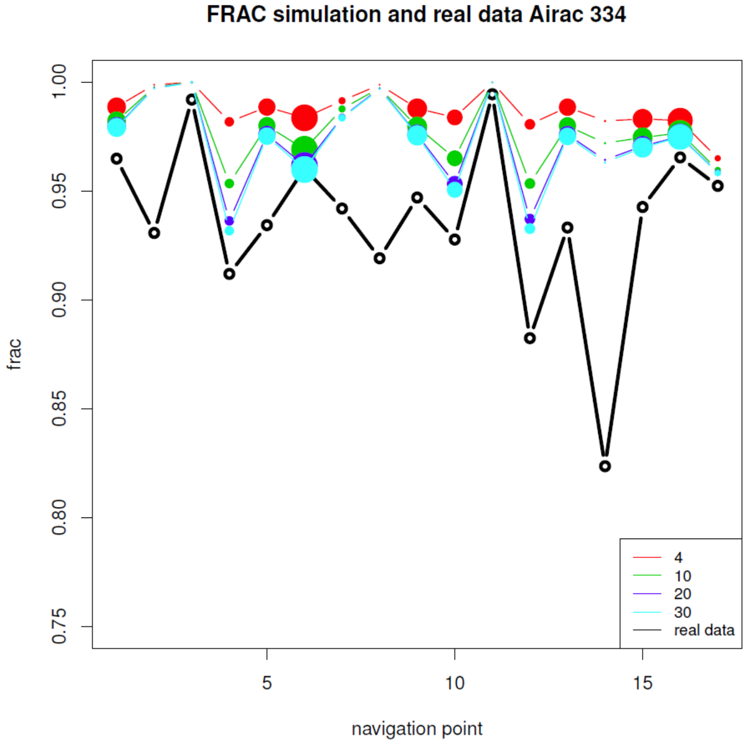


Figure 2: Comparison of FRAC values of simulated and real data when varying the maximum re-routing angle (all Airac 334).

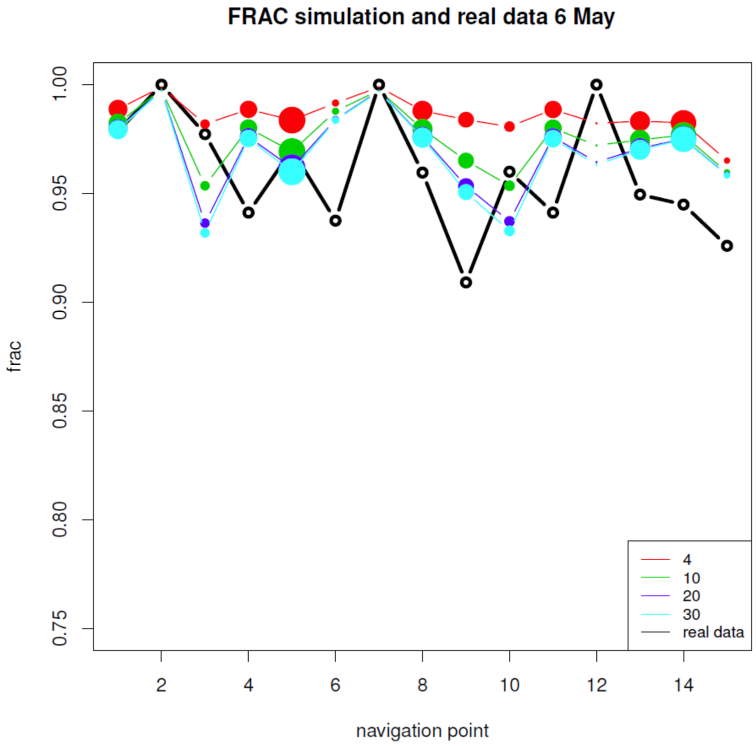


Figure 3: Comparison of FRAC values of simulated and real data when varying the maximum re-routing angle (6 May 2010).

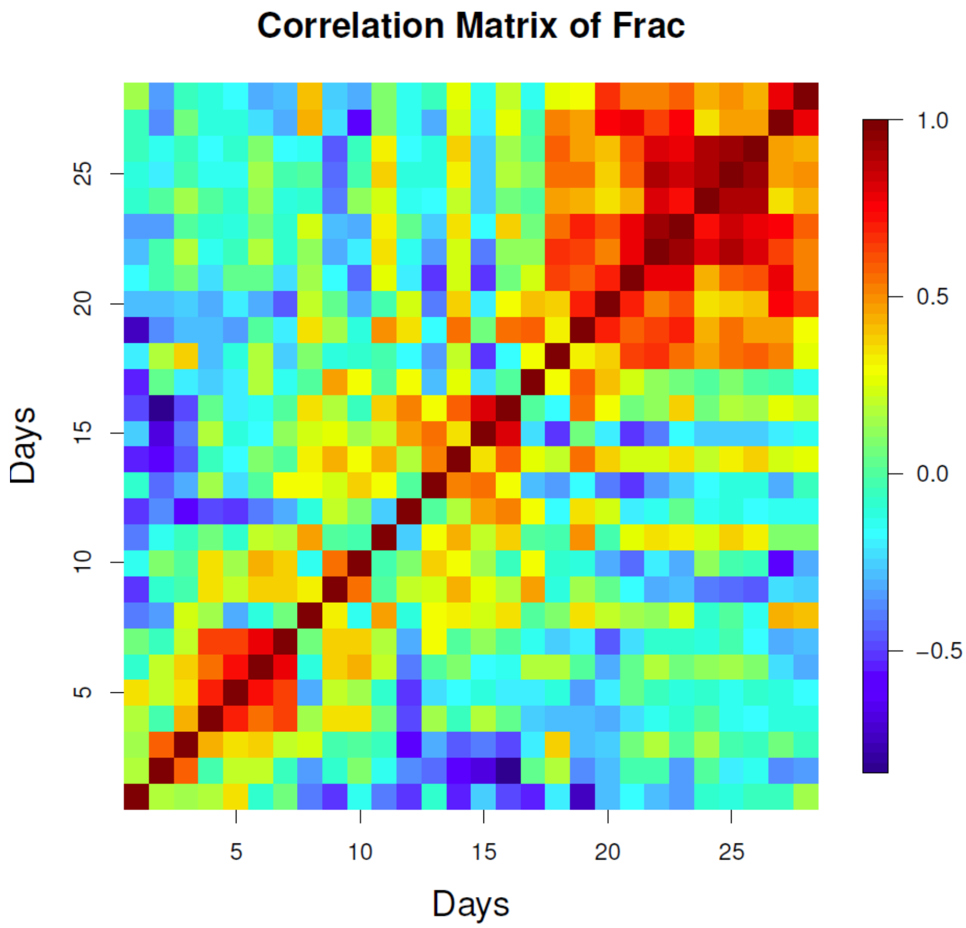


Figure 4: Correlation between frac on the days of Airac334

4.2.2 Strategic layer

The validation of the Strategic ABM relies mainly on the network created by the simulated flights with respect to the one generated by the real flights. Note however that a first, quick validation based on the number of rejected flights plans is possible. By rejected flight plans we mean those that have been submitted to the NM but that have not been accepted for any reason (regulated sector, sector capacity exceeded). We find between 3% and 8% of the flight plans which are rejected with the fully calibrated model, a range which was assessed as realistic by the Alitalia experts.

In order to have a quantitative calibration, we use a realistic setup for the Strategic ABM, as explained in the deliverable D2.3. The idea is to calibrate the ABM on the real network using the real airspace configuration derived from WP1 analyses, generate the simulated trajectories, and compare them to the real ones. Generally speaking, building a network from traffic consists in:

- extract all the nodes from each “step" of the flight,
- putting an edge where at least one flight has flown,
- weight the edges with the traffic between the nodes.

Given the network and the trajectories on this network, we define four distributions of interest for us: the degrees, the strengths, the weights, and the topological lengths of trajectories of the flights. The latter is the number of nodes that the flights cross on their trajectories. The difference between the real distribution and the simulated one gives an idea of how good the model fits the reality. Note that in the Strategic ABM we use these distributions as targets for a calibration on one of the parameter of the model. In practice, the validation comes from how well the targets are hit by the calibration process: a small distance to the target for the calibrated model implies a good consistency with the reality, hence validating the model. On Figure 5 and Figure 6 we show an example of two of these distributions, each of them for two extreme values of the parameter of control, N_{sp}[1]. Clearly, the change of the parameter of control have a strong difference on the distributions. It is quite obvious from these graphs that the choice N_{sp} = 10 gives a better agreement than N_{sp} = 1. In order to have a quantitative criteria with which we can select the best agreement, we use the Kolmogorov-Smirnov distance[3]. It lies between 0 - for perfectly identical distributions - and 1. We compute this distance between the real distribution and the simulated one, for the four targets (degree, strength, weights and topological length) for di different values of the parameter of control. The result is presented on Figure 7. As one can see, some metrics, the strengths and the weights, are quite insensitive to the parameter of control. On the contrary, distances on degrees and topological lengths drop significantly when N_{sp} increases. Hence, this procedure allows us to select N_{sp} = 10 as the best choice and run the simulations with it. Note also that the absolute value of the distance is small compared to 1, especially for the strengths and the topological lengths.

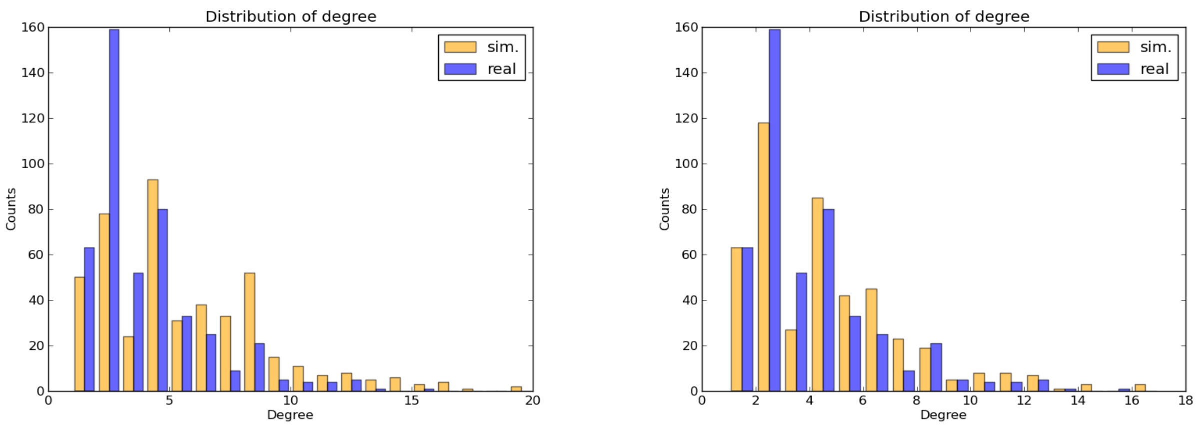


Figure 5: Real and simulated distribution of degree for N_{sp} =1 (left) and N_{sp} = 10 (right)

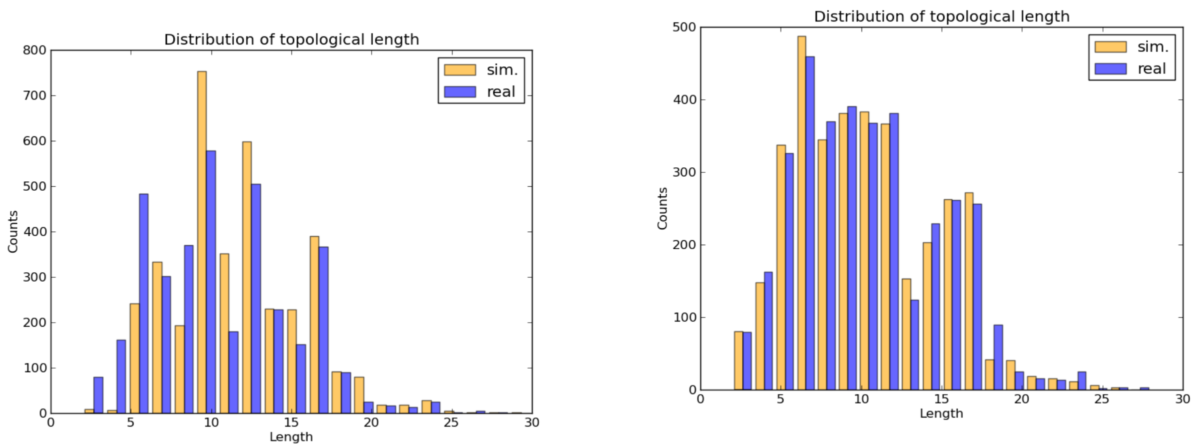


Figure 6: Real and simulated distribution of topological length of flights for Nsp =1 (left) and Nsp = 10 (right)

[1] The ABM being based on two super imposed networks, one of navpoints and one of sectors, a trajectory can be expressed in terms of a series of navpoints or, at a coarser scale, as a series of sectors. When we fix the number of flight plans N_{fp} that a company submits for a given flight, we give them the possibility of choosing between different paths on the network. For this, we select first a certain number of sector paths. Then the parameter N_{sp} gives the number of navpoints paths per sector path. Hence, if N_{fp} =10 and N_{sp} = 1, the air companies will have a choice between 10 distinct sector paths, each of them having one navpoint path. On the contrary, if N_{sp} = 10, there will be only one sector path available, including 10 different navpoint paths.

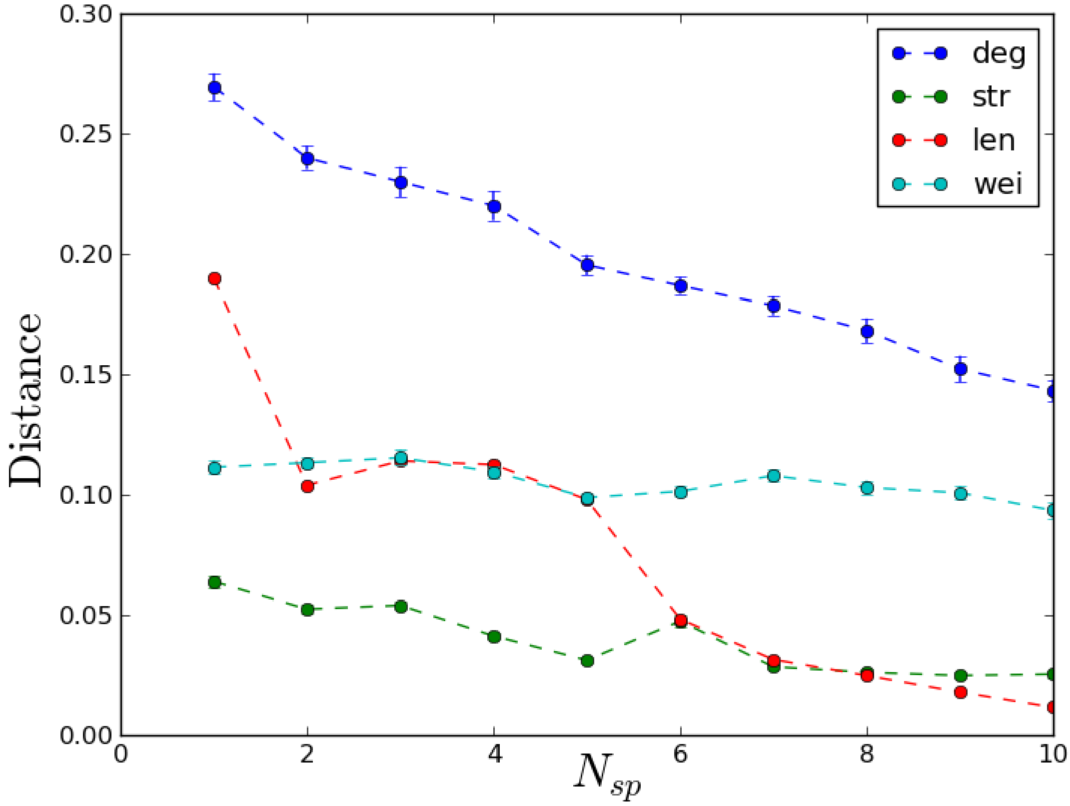


Figure 7: Distance between real and simulated distributions for the four targets as a function of the parameter of control N_{sp}

4.3 WP3

The validation activities for WP3 lead to the refinement of four aspects: roles, timeframe, scenarios, supported decisions. The main results concerned the scope of the decision support tools, the accuracy of the information contained in the prototype, the expected operational benefits.

4.3.1 Scope

The initial list of roles included ATCO, Complexity Manager, Sub-Regional Network Manager, Regional Network Manager, Network Manager, Flight Dispatcher, Flight Crew. The target users were progressively narrowed down to three, including in the decision-support tool prototype the Network Manager, the Complexity Manager, the Flight Dispatcher.

The initial timeframe of application of the decision support tool extended from the long-term scheduling phase (months before the flight), to the strategic phase (days before the flight), to the short-term planning (one day before), to the flight execution, and post-flight data analysis. The selected timeframe was stabilised on the time window from 24 hours before the flight till the monitoring of the flight execution.

The strategic planning phase was excluded as the potential benefits were not clear. All the experts reported that detailed advance planning suffers from too many uncertainties to be carried out with fine details, as it is always subject to adjustments. This consideration was considered to remain true also in a future SESAR scenario. The ELSA decision support tools thus focus on the planning refinement that happen 24 hours before the flight, when more information become available on weather phenomena, crew and aircraft rotation, impact of organisational aspects on airspace effective capacity.

4.3.2 Accuracy

Two scenarios, *24 hours planning* and *execution-monitoring*, were refined in many iterations with the SESAR and ENAV experts to ensure the maximum degree of accuracy.

Elements of accuracy include the following:

- Operational scenario
- Flight details, such as callsigns, routings, flight levels
- Time evolution of the scenario
- Evaluated options and proposed solutions

Flight details were extracted from the ELSA database and altered only to better suit a future SESAR scenario. A final accuracy check was carried out.

4.3.3 Expected Operational Benefits

The typical issue to be debated for this part of the work was to clearly identify which support tool could bring real operational benefit. The key question for this part of the work was “what is the use of this piece of information? Which benefits is it bringing to the *target user*?”.

On the basis of this design question, a series of potential decisions to be supported were discussed with the experts and within the design team. Decisions were the “ELSA approach” did not bring any benefit were excluded. The main reasons for discarding proposed concepts were the following ones.

- Too simplistic solution, relying too much on the principle “the more information, the better”.
- Benefit deriving only from a better information visualisation.
- Unrealistic timeframe for decision making, typically too short.
- Too complex coordination required among too many users.

The concepts that were eventually selected offer a real support to decision making, by easing the information retrieval, the comparison of different options, the simulation of effects and consequences, the sharing of information among involved actors.

4.4 Success criteria matching

The following table reports an overview of the achievement of the validation objectives.

Table 6: Matching of the validation objectives with their related success criteria

Validation Objectives	Success Criteria	Objectives verification
Correct identification of ATM need(s), scope, and target users.	Positive feedback by identified stakeholders.	**Partially achieved**. We gathered a general positive feedbacks, but more detailed feedback was hard to obtain, as it requires the stakeholder to position her/himself very far in the future.

Validation Objectives	Success Criteria	Objectives verification
Definition of a set of proposed analyses to address the ATM needs, in current and in the SESAR scenario.	Analysis factsheets completed in all the subfields for a selected set of analyses, including definition of benefits and link with ATM needs.	**Achieved**. 5 best analyses identified and fully described by completing the factsheets in all the subfields.
Definition of a set of proposed analyses to address the ATM needs, in current and in the SESAR scenario.	Positive feedback on each of the proposed analysis by at least one stakeholder.	**Partially achieved**. We gathered positive feedbacks for all the best analyses but they were followed by some remarks and some limitations were also identified.
ABM Calibration: 1. build a multi-layer ABM, 2. model the interactions between the layers, 3. build a scalable ABM, that can simulate from one sector to multiple sectors, 4. simulate a realistic SESAR scenario,	1. Positive feedbacks on the model inputs by the relevant stakeholders (Alitalia dispatcher, ENAV supervisor) 2. Use the strategic model outputs as inputs for the tactical layer 3. ABM working at multi-sector level 4. Positive feedback on the reliability of the model results in the SESAR scenario by at least one stakeholder	1. **Achieved.** Model inputs were validated in several dedicated interviews. 2. **Not achieved.** The two models were not integrated. 3. **Not achieved.** 4. **Not achieved.**
ABM Validation: 1. inform the ABM characteristics from a set of regularities identified by WP1. 2. inform the ABM outcome metrics from a set of WP1 regularities	1. Use selected WP1 regularities to develop the model 2. Positive comparison of the model results with some of the regularities observed in WP1	1. **Achieved.** Airspace structure was used for the strategic layer, trajectories distributions from WP1 used as inputs for the Tactical layer. 2. **Achieved.** Both layer were validated by comparing the results with selected regularities from WP1.
Definition of the proof-for-concept of a decision-support tool: target SESAR scenario, human-technology interaction, new roles, or modification of existing ones.	Positive feedback on the proposed proof-for-concept by at least one of the following stakeholders: - Experts on Network and Flow Management, - Experts on Delay, - EUROCONTROL and SESAR JU.	Not yet verified, due to late delivery of the prototype.
Definition of the proof-for-concept of a decision-support tool: target SESAR scenario, human-technology interaction, new roles, or modification of existing ones.	New roles, or modification of existing ones clearly identified in WP3 and included in proof-for-concept.	**Achieved.** 3 roles are included in the prototype and detailed in two scenarios.

5 Conclusions

The ELSA project aimed at using methods and techniques from Complex Systems theory to identify and measure disturbances that may affect the ATM system, either in terms of traffic flow perturbations (e.g. delays, deviations from the planned trajectory), or safety-relevant events (e.g. losses of separation and STCA). This objective has been addressed by delivering a set of best analyses that can be applied at the EU level to monitor and characterize geographical areas.

These analyses and the ATM needs they address proved the benefits that Complex System theory can bring to the ATM system, by giving a new way of monitoring the current structure and by providing useful tools to drive major redesigning processes. With this respect validation activities played a crucial role in identifying the most useful applications of the Complex System theories by gathering feedbacks from operational experts on the applicability of the proposed solutions. This approach resulted in the definition of 5 “best analyses” which were identified as the most promising for a real and useful application in the future ATM world. However there are some open questions still pending, the main one regarding the context of application of the proposed analyses. One of the main difficulties found during the validation activities was in fact that the stakeholders found it difficult to position themselves in the future and to imagine a possible scenario of application. Their understanding of how the future ATM will be structured was closely linked to the present scenario with a limited perception of the innovations that will take place with SESAR and beyond.

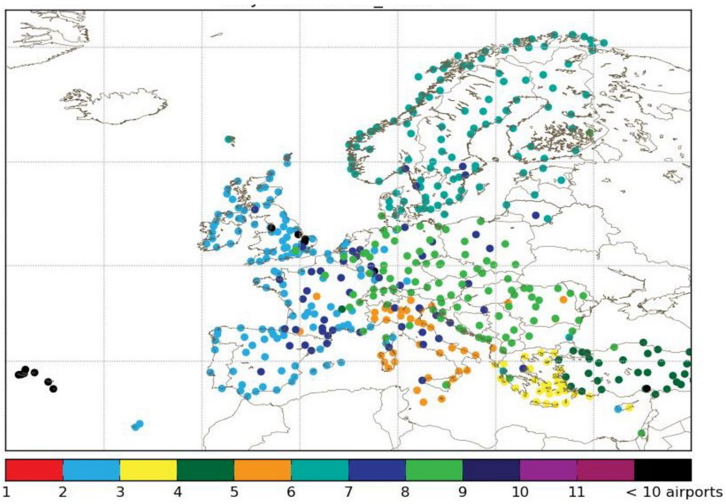
A second outcome of the project has been the construction, calibration, and validation of an Agent Based Model. Its final objective was to simulate realistic Air Traffic scenarios and also to help in analysing the statistical regularities that could not be thoroughly analysed from the empirical data. While the modeling process was completed to an adequate extent during the ELSA project, the model outcomes could not be fully analysed during the project duration. Moreover, thanks to the thoroughly validation activities carried out during the development and the results analysis phases, we now clearly understand two main limitations of the model we built. First, the single sector model is too simplified to capture some relevant phenomena, like the propagation of delays, or multi-sector trajectories, like business trajectories, or long directs. Second, the model took a real sector of Rome ACC as its reference point, while the model outcomes can be significantly improved by making the model more general. The two points are connected. Extending the analysis to different sector types is an enabling condition to assemble a multi-sector ABM, to mimic features such as multi-sector directs or simulate the future SESAR scenario.

The third outcome was the design and implementation of decision support tools, to (i) monitor the current complexity status, (ii) receive a prediction of the likely development, (iii) simulate the effects of changes to the system. The decision support tools were designed at the proof-for-concept level, with the effort focused on defining realistic SESAR scenarios and corresponding visualisations. The proposed scenarios describe the interactions among the Network Manager, the Complexity Manager, and the Flight Operation Centre, using the tools 24 hours prior departure, till during the flight execution. The scenarios were prepared to show how the target roles could visualise and interact with traffic in a SESAR scenario and to design negotiation processes among them.

6 References

[1] SESAR, “Validation Plan Template,” 2012.
[2] EUROCONTROL, “E-OCVM Version 3.0 - European Operational Concept Validation Methodology,” 2010.
[3] Eadie, W.T.; D. Drijard, F.E. James, M. Roos and B. Sadoulet (1971). Statistical Methods in Experimental Physics. Amsterdam: North-Holland.

A.1 Community analysis: seasonal variations and impact of the ash cloud



What

We detected sets (communities) of nodes (airports, or sectors, or navpoints) clustered together, i.e. that share a common profile with respect to the flights that are connecting them. Communities are generically defined as sets of nodes that are more connected among themselves than with the rest of the network. We carried out a qualitative analysis with the objective of highlighting global patterns that can have some operative impacts on the airspace organization. In particular we analyzed seasonal variations of the airport community distribution and how communities reshape in response to an external event like the volcanic ash cloud in 2010. Two main algorithms were employed: a Modularity maximization algorithm (see [1]) and the Infomap (see [2]) algorithm.

Why

The community analysis can be useful to highlight clusters of airports that share common characteristics in terms of flights arriving to and departing from them. Airports that belong to the same community will be more frequently connected with themselves than with airports of other communities. A cluster of airports located in different FABs can be a clear indication of numerous flight connecting these areas (or these airports). Community detection on the sector and navpoint networks could instead inform some of the changes that SESAR plans to introduce in the next years. This may happen either by redesigning the airspace so that the new ACCs are more densely connected inside and have less interface (links) with the adjacent ones, or by identifying those boundaries with high traffic exchange volumes (as measured by the number of links and/or by their strength) in order to devise dedicated coordination tools and procedures.

Results

We obtained two sets of results. The first one is a community detection analysis during the ash cloud crisis in April 2010. The objective of this analysis was to see how communities are reshaped due to external events. We observed that European airspace network is organized in such a way that when some areas are closed for external reasons (like the ash cloud) the others are able to operate independently by just cancelling the connection with the closed areas without impacting the rest of the system. An example of this behavior can be seen in Figures 1 and 2. In figure 1 airports in the southern coast of Portugal and Spain belong to the same community of some English airports. In figure 2, when the ash cloud caused the closure of the northern Europe airspace the communities are reshaped and the Portugal and Spain airports belong all to their national communities. It is clear then that the system has an intrinsic ability to reshape in response to external events that cause closures of part of it by preserving its hierarchical structure.
This result is confirmed when detecting communities on different timescales. The reshaping works in the same way if we detect communities in a single day of the ash crisis or if we do the detection on a weekly basis.

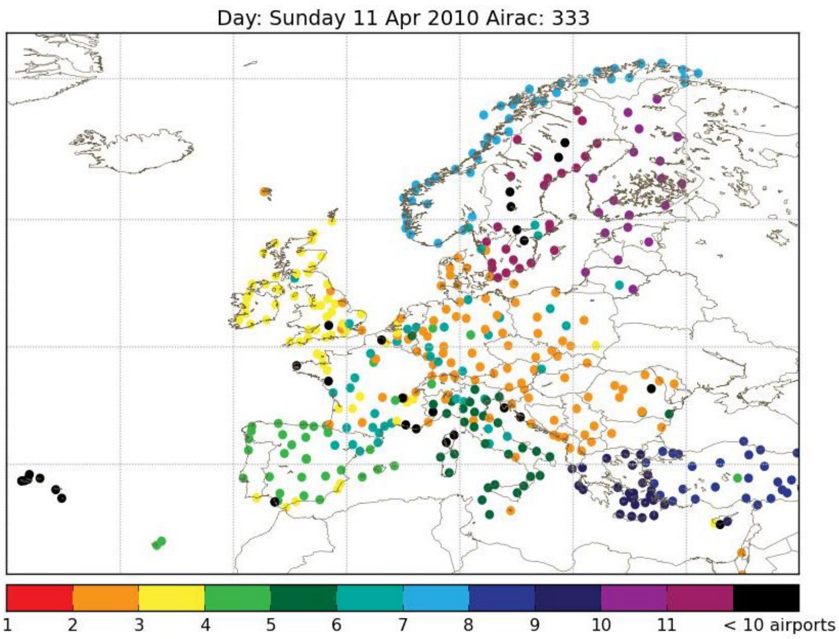


Figure 8: Map of airport network in day 11/04/2010 of AIRAC 333 (before ash cloud crisis) portioned in communities according to the Infomap algorithm. The points correspond to the airports, while their color identify the communities they belong to. For readability purposes, the communities with less than 10 airports are all in black.

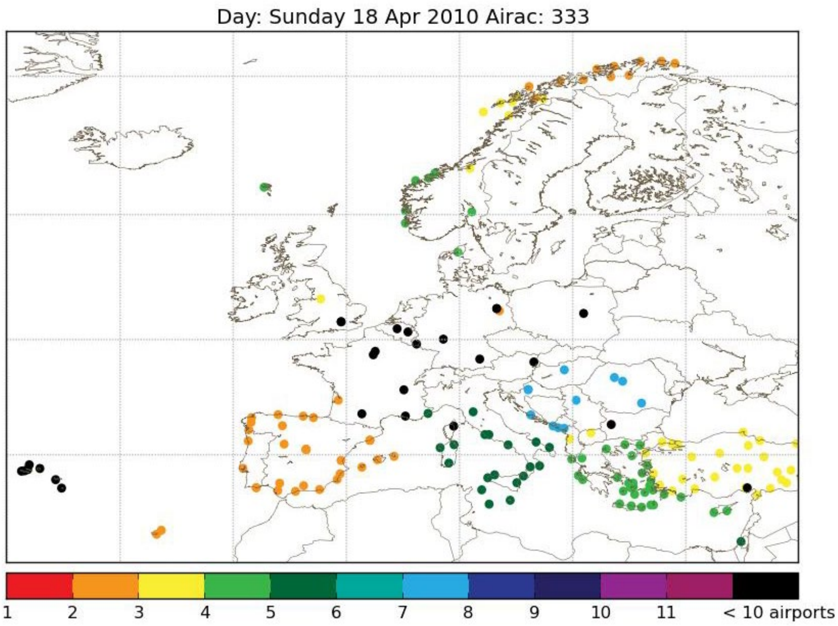
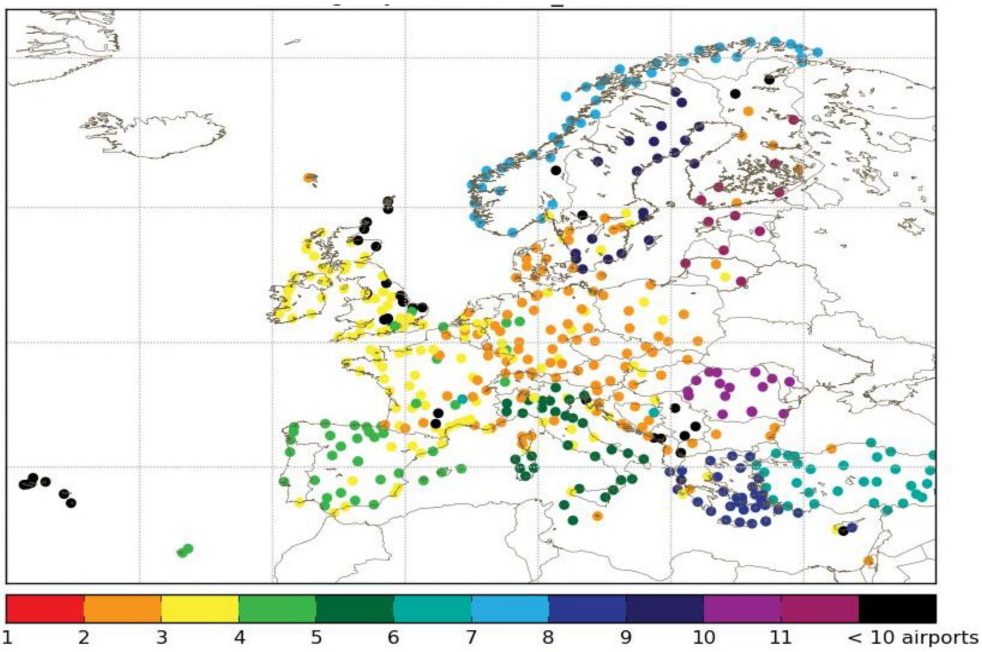


Figure 9: : Map of airport network in day 18/04/2010 of AIRAC 333 (during ash cloud crisis) portioned in communities according to the Infomap algorithm. The points correspond to the airports, while their color identify the communities they belong to. For readability purposes, the communities with less than 10 airports are all in black.

The second set of results is instead related to seasonal variations of communities. Figures 3 and 4 show communities detected in two different weeks, one in July and the other in November. The collocation of Greek airports in this period is an interesting example of communities seasonal variation. In fact during summer all Greek airports or at least some of them are part of a greater community which comprises airports in other parts of Europe. On the contrary in November Greek airports are always part of their own national community. This behaviour is clear by using both community detection algorithms. This result can be positively used to increase coordination between the most linked areas in a particular season by implementing convenient tools like “hot-lines” between airports or ACCs.

Analyses of the networks of sectors and navpoints [4] showed that the detected communities are typically much smaller than the typical size of a country. Depending on the specific community detection algorithm employed communities of navpoints typically have the size of sectors and thus they can be used to conveniently perform a redesign of the airspace at the sector level. Communities of sectors have instead the size of national airspaces or FIRs thus giving information about the airspace at this granularity level.
A final important remark is related to the time resolution of community detection algorithms. In fact the time window on which data are gathered is a key point to detect possible variations of the airspace structure. Our analyses showed that the minimum time interval on which the algorithm is capable of detecting significant variation is of the order of few hours. Below this threshold the number of flights crossing the airspace is not sufficient to obtain a significative differences in the community detection.



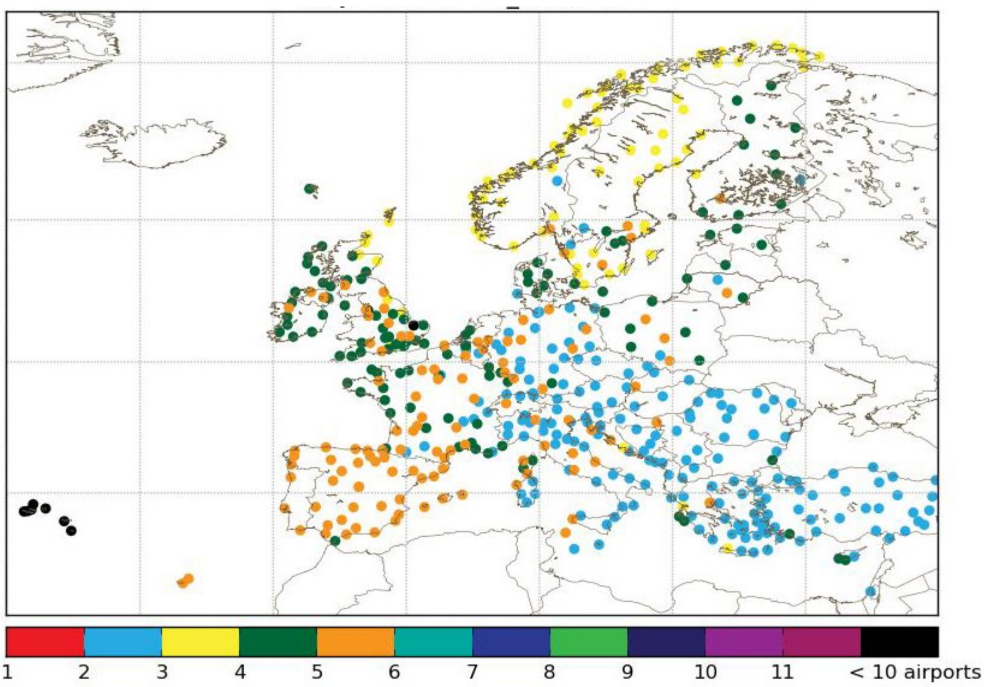


Figure 10: Map of airport network during the week starting on July 15th 2010 portioned in communities according to the Infomap (top) and Blondel (bottom) algorithms. The points correspond to the airports, while their color identify the communities they belong to. For readability purposes, the communities with less than 10 airports are all in black.

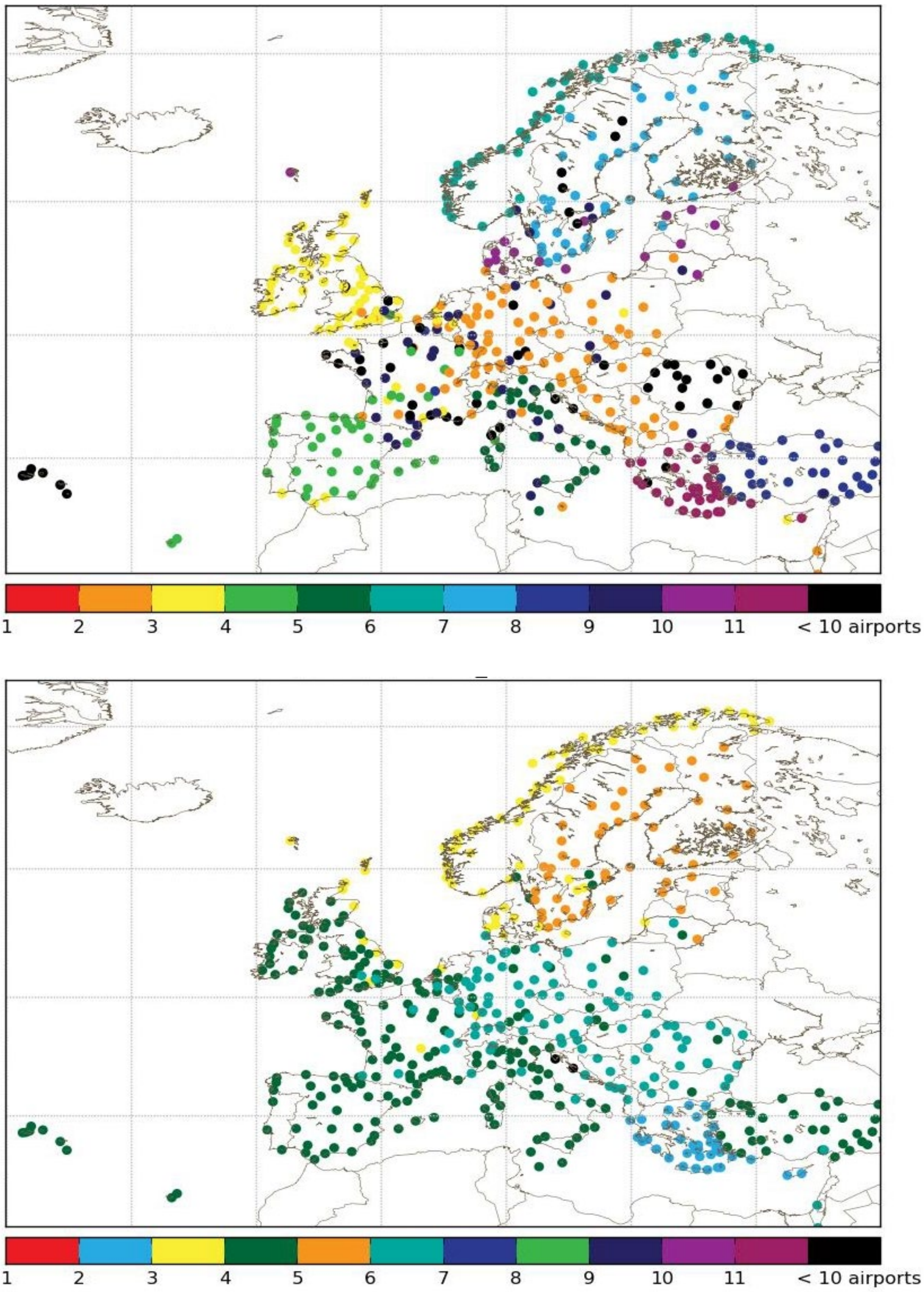


Figure 11: Map of airport network during the week starting on November 11th 2010 portioned in communities according to the Infomap (top) and Blondel (bottom) algorithms. The points correspond to the airports, while their color identify the communities they belong to. For readability purposes, the communities with less than 10 airports are all in black.

Potential applications

The community detection applied to the networks of sectors and navpoints can be helpful in comparing existing airspace partitions and partitions obtained with network community detection in order to improve the design of airspace. This may occur in different ways such as a modification of sector boundaries according to a local difference observed between the boundary of an existing sector and the corresponding community detected one. Moreover network community detection may help in highlighting the boundaries (between sectors, ACCs, NAs, FABs) that usually require intensive coordination to develop dedicated coordination tools and procedures.

Finally, the application of network community detection to the airport network can be used to implement direct communication links between closely-connected distant airports, as they are identified by the community detection algorithm. The implementation of this direct lines can also occur on a seasonal basis to cope with the observed seasonal variations of the airspace configuration.

How

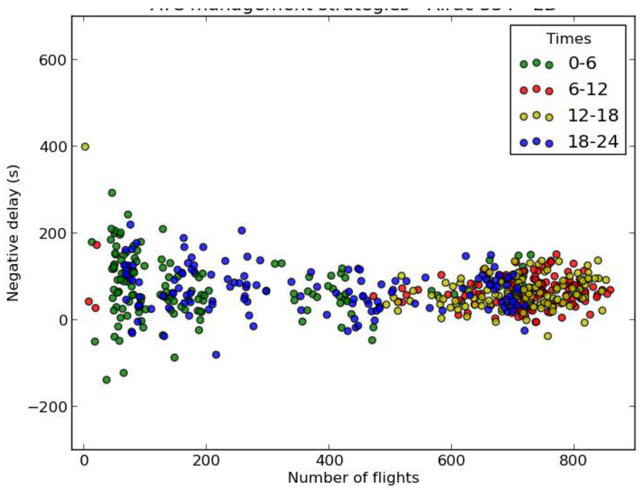
Two main algorithms were employed: a Modularity maximization algorithm (see [1]) and the Infomap (see [2]) algorithm. The first one identifies partitions in the network as the modularity gets its maximal value. Modularity is a network metric computed as the fraction of the edges that fall within given modules minus the expected fraction if edges were distributed at random. Exact modularity optimization is a problematic task from the computational point of view; therefore for large networks approximated algorithms are necessary. Specifically, we use the Blondel version of the Modularity maximization algorithm. This introduces elements of stochasticity in the algorithm search process. We have, therefore, performed 100 iterations of the algorithm for each day in the AIRAC and choose the one with the highest modularity. The Infomap algorithm, by using random walks to analyze the information flow through a network, identifies communities as modules through which information flows quickly and easily. In other words, the communities are formed to minimize the length of a random walk in the graph. As for the Blondel algorithm, for each day of the AIRAC, we have run the algorithm with 100 different seeds and 10 iterations for each of them.

References

[1] V. B. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Fast unfolding of communities in large networks, JSTAT, P10008 (2008)
[2] M. Rosvall and C.T. Bergstrom, Maps of random walks on complex networks reveal community structure, PNAS 105 11181123 (2008)
[3] ELSA deliverable D3.1
[4] G. Gurtner, S. Vitali, M. Cipolla , F. Lillo, R. N. Mantegna, S. Miccichè , S. Pozzi, Multi-scale analysis of the European airspace using network community detection, (unpublished) manuscript submitted for publication to PloS One.

A.2 ATCO management strategies: differences among countries and sectors

ATC management strategies – Airca 334 – Germany



What

We correlated the spatially averaged negative delay and the number of flights for different countries or single sectors. This measure allows to highlight different ATC management strategies among the different countries or sectors. We considered the following countries: Italy (LI), Spain (LE), United Kingdom (EG), France (LF), Germany (ED). The analysis was carried out in May and September 2010. Moreover we did the same analysis on the sectors LIRROV and LIRRMIW of Rome FIR.

Why

This kind of analysis can be useful to show differences in the way controllers handle flights in different operative situations. In particular we can detect if the controllers tend to give directs to the flights when the traffic load is low or instead if their controlling strategies are the same in every situation of traffic load. By comparing results obtained in different periods of the year we are also able to detect if a particular controlling strategy is implemented only once in a while or if it is a permanent habit. Finally, this analysis carried out at sector level in the same FIR can give information on the relationship that may occur between the control strategies and the sector characteristics.

Results

We detected different management strategies of the traffic load. The common behaviour is to give directs whenever it is possible despite of the number of flight that are present in the airspace. Of course, when the traffic load is lower the number of directs has usually a moderate increase (see for instance Figure 12, referred to Spain). Clearly the traffic load is lower during the night and the early morning thus resulting in higher negative delays but there are cases when high negative or positive delays are observed also in other periods of the day. The average en-route delay has small but negative values which means that the controlling strategy is to shorten flight times whenever is possible.

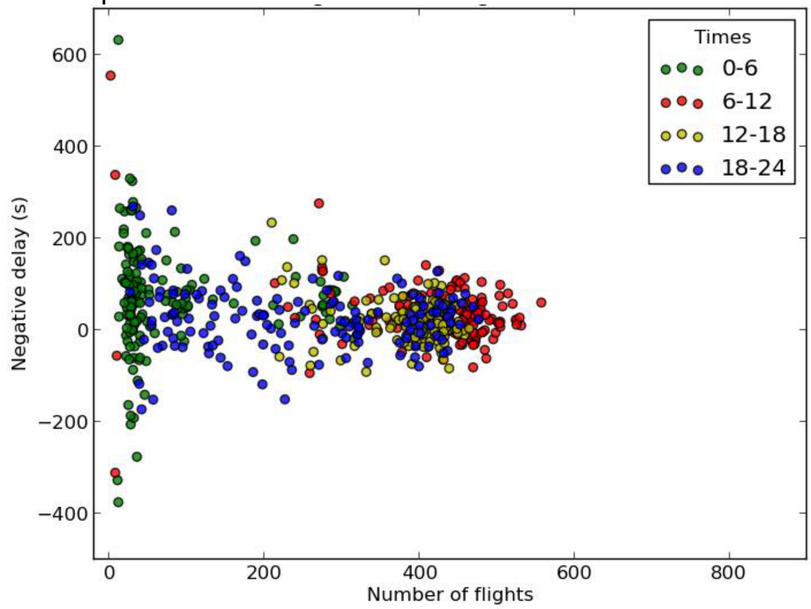


Figure 12: Spain ATCO management strategy - May 2010

In Germany instead the average negative delay is almost constant in every traffic load condition (Figure 13).

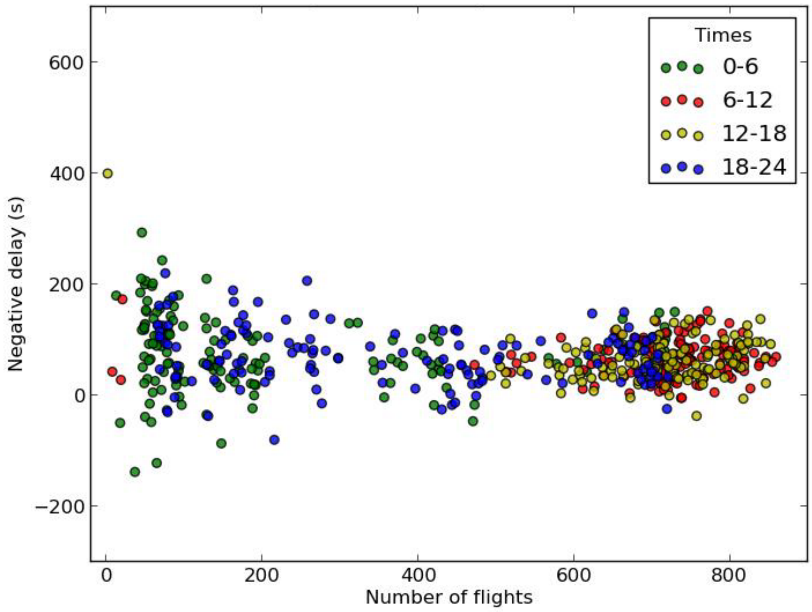


Figure 13: Germany ATCO management strategy - May 2010

This is a clear indication of a management strategy that is kept the same despite of the current number of flights. There is still a moderate increase of the negative delay when in low traffic conditions but it is less significant than what we observed for Spain.
The situation is different in UK where the negative en-route delay has a higher increase when the number of flights decreases (Figure 14). This behaviour is confirmed also when changing the timeframe of the analysis thus suggesting that this

is due to a precise management strategy (see Figure 14 & Figure 15).

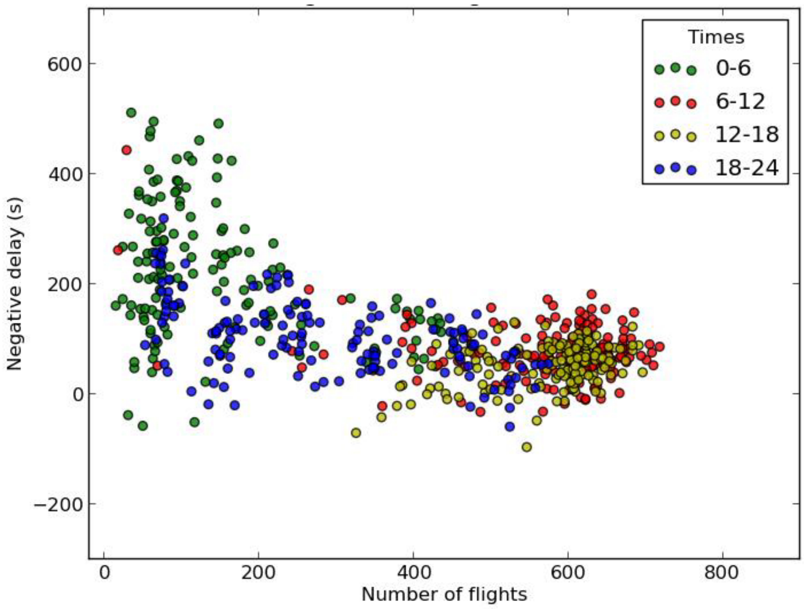


Figure 14: UK ATCO management strategy - May 2010

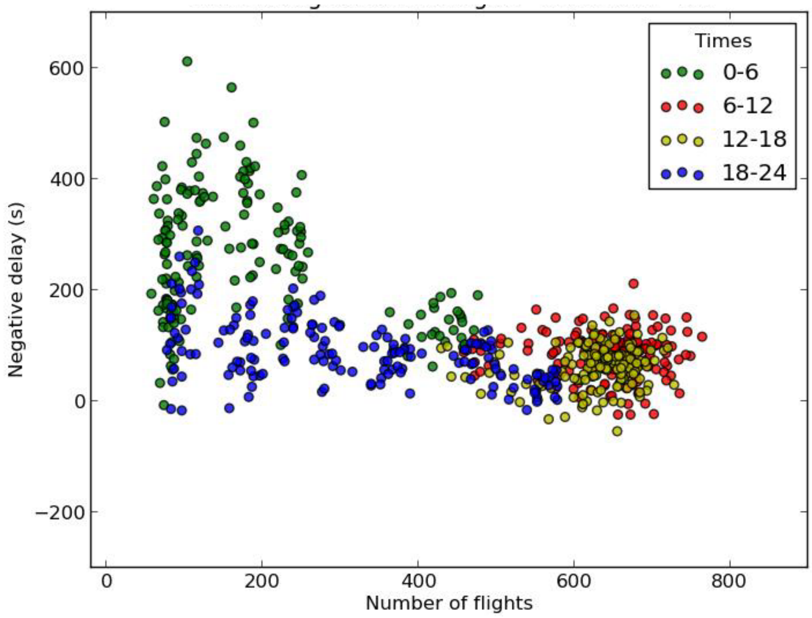


Figure 15: UK ATCO management strategy - September 2010

Moreover in UK the average en-route delay has a higher negative value. This suggests that with respect to planned trajectory, flights can usually take directs to shorten their trajectories. This conclusion has been confirmed by an ENAV controller who added also another explanation. The UK airspace, especially in the London area has very strict airways that flights must follow in high traffic load conditions. When the traffic is lower this constraint is relaxed thus allowing controllers to give more directs. Moreover the airways for approaching London airports are not straight so if a flight gets a direct its en-route negative delay increases more than what we observed in other countries in low traffic conditions.

The management strategies observed in the selected countries in May 2010 are confirmed when looking at a different period of the year thus confirming that they are a permanent habit of the controllers of each country (Figure 16 and Figure 17).

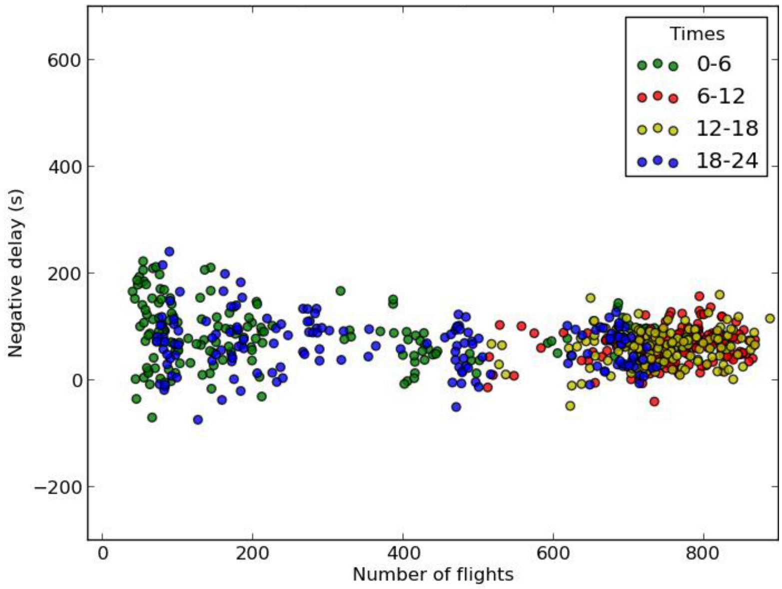


Figure 16: Germany ATCO management strategy - September 2010

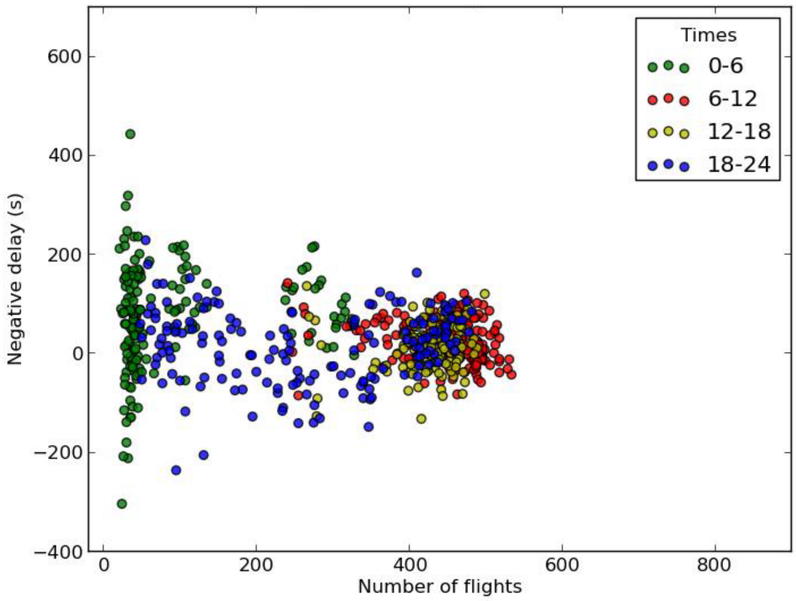


Figure 17: Spain ATCO management strategy - September 2010

The same behaviour can be observed also in France and Italy (Figure 18 and Figure 19).

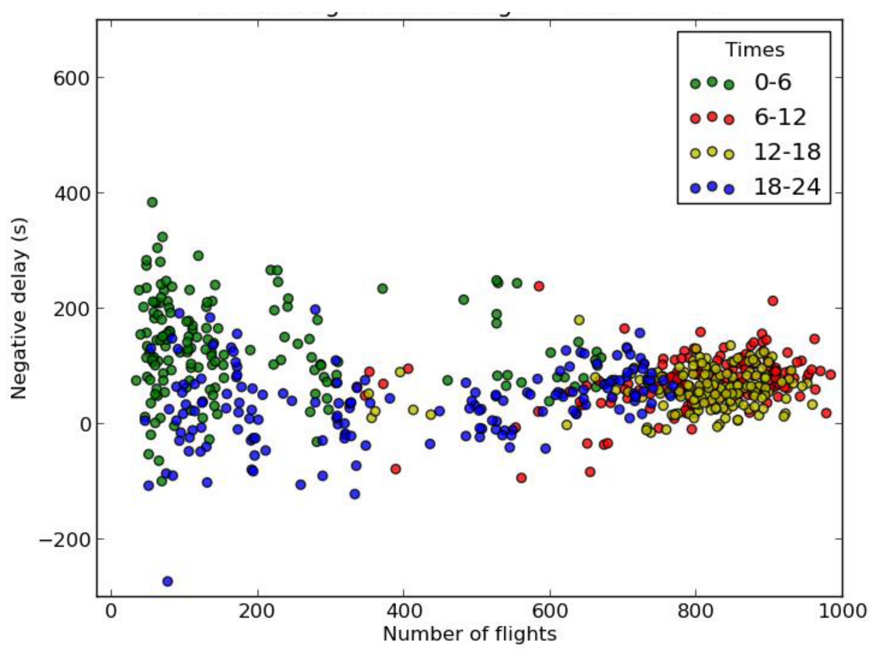


Figure 18: France ATCO management strategy - September 2010

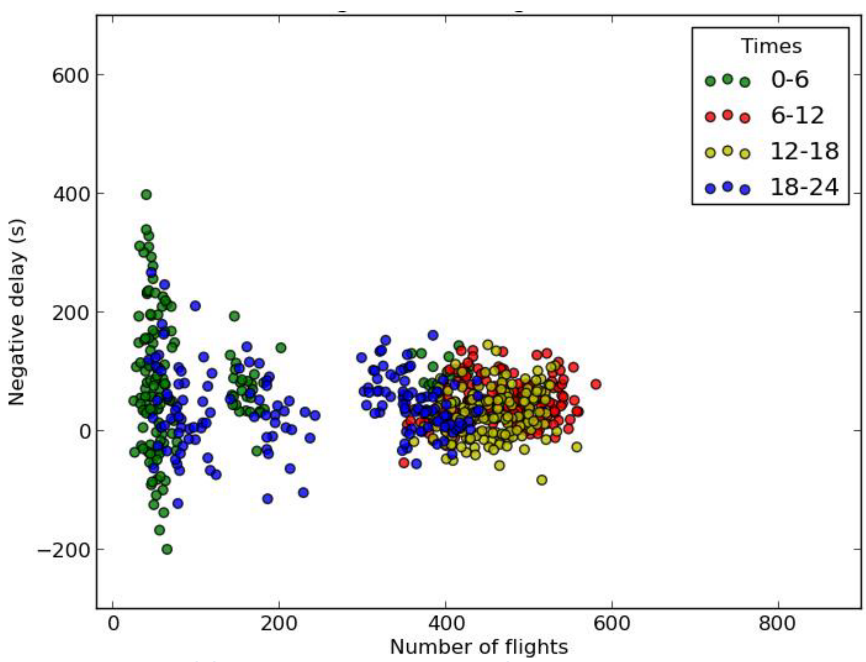


Figure 19: Italy ATCO management strategy - September 2010

The analysis at sector level do not show significant differences with the national cases. Despite of the operative differences of the two sectors we selected for comparison, the management strategy appears to be the same. Sector LIRRMIW is in fact crossed almost exclusively by flight in cruise phase while sector LIRROV has mixed traffic. In both cases the negative en-route delay is almost constant with a slightly increase when the number of flights is lower (Figure 20 and Figure 21). Other metrics are being considered to try and capture the operational differences among sectors.

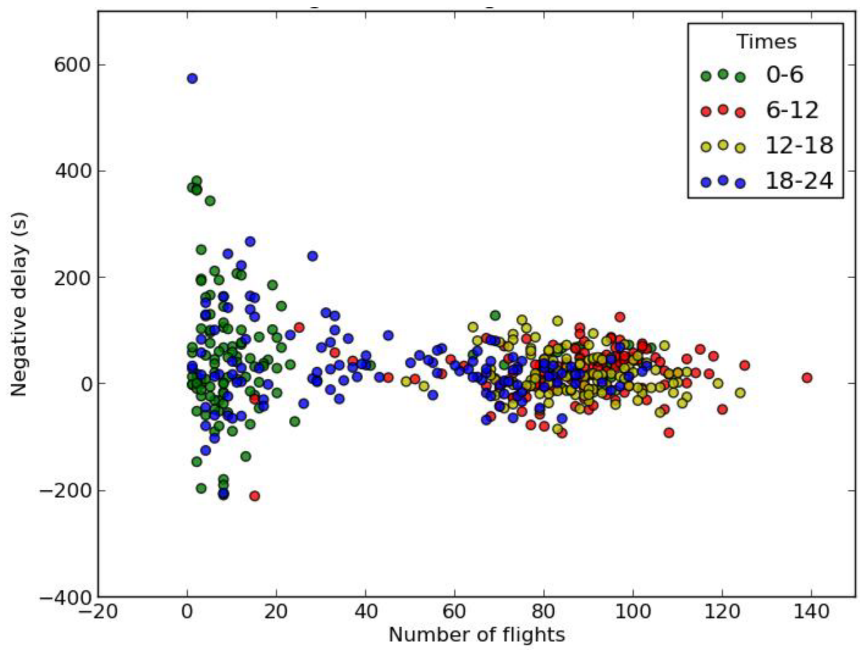


Figure 20: ATCO management strategy in sector MIOW of Rome FIR

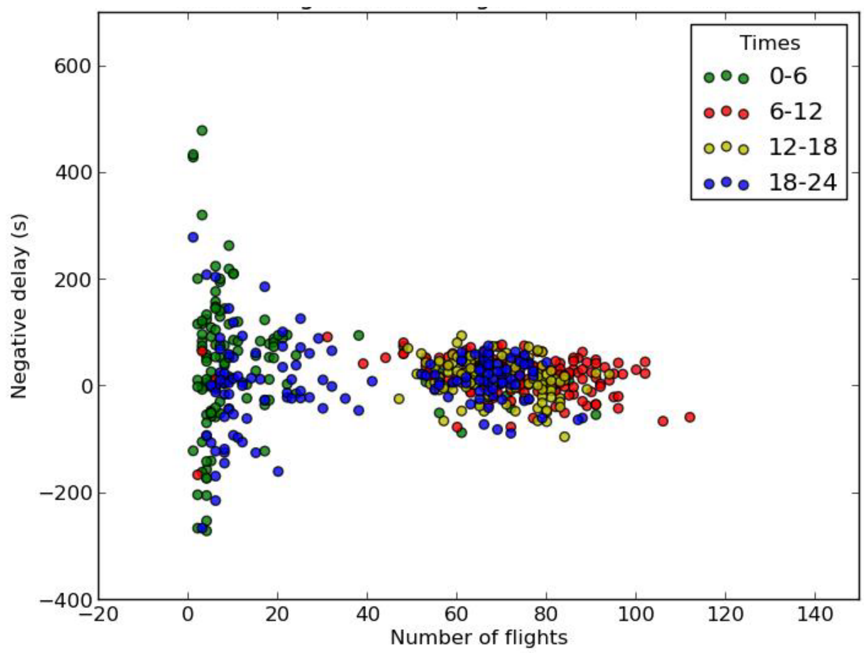
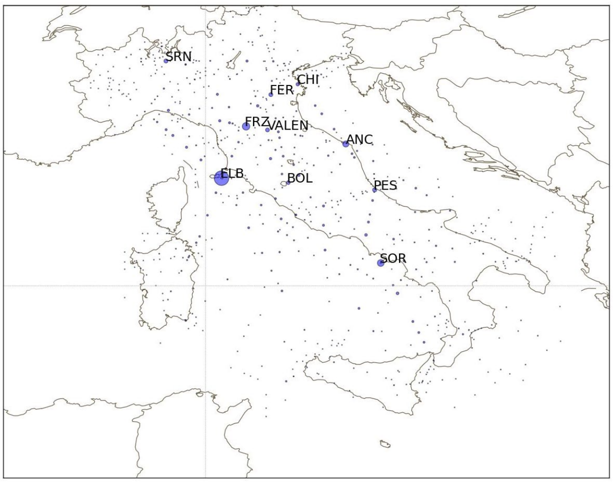


Figure 21: ATCO management strategy in sector OV of Rome FIR

How

We considered all the flights crossing the selected area in a 1-hour period. For each flight we computed its en-route negative delay by comparing the segment of M1 and M3 trajectories that it followed in the selected area during the timeframe of interest. Finally we averaged all the flight negative delays in every hour. Each point on the plots represents one of these averages that were repeated for every hour of the 28 days of the selected AIRAC.

A.3 Analysis of the most critical navpoints of the Italian airspace



What

We analyzed the Italian airspace in order to identify its most critical navpoints in terms of two specific network metrics: Strength and Betweenness. These two metrics are able to describe the airspace network giving information on the traffic load and probability of conflicts around a navpoint [1]. Using appropriate tools derived from the complex network science [2] we measured the Strength and Betweennes of all the navpoints in the Italian airspace and we derived a ranking of the ten most critical ones for each of the selected metrics. The rankings obtained were then validated with air traffic controllers from the Italian ANSP in order to confirm the ability of our analysis to identify correctly the most critical navpoints in the Italian airspace.

Why

This technique could be positively used to assist in the redesign of the airspace or to correlate with ATCOs' workload subjective ratings. In particular, measurements of betweenees, being independent from the traffic load, may help in determining which fixes will be more critical in terms of number of conflict occurring around them. A proper airspace redesign could then be conveniently driven by these useful information on the network topology. The ideal situation would be to design a network in which every node has the same value (preferably small) value of betweennes.

Results

We first present our rankings of the most critical fixes for each of the selected metrics.

<u>Strength</u>
The strength of a node is essentially a measure of the traffic load on its adjacent links [2]. It is the most direct measure of the traffic load on the node, then the nodes with

high strength should also be the most busiest. Table 7 reports the ten nodes of the Italian airspace with the highest value of strength. Their geographical location is shown in Figure 22. According to the air traffic controller we interviewed the first four navpoints are actually the most trafficked in the Italian airspace network and the traffic load around them reflects the rank we found. The others are less busy and their order in the ranking may vary on a seasonal or even day-by-day basis.

Rank	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
Navpoint	BZO	GEN	SRN	ELB	SOR	TOP	BOL	VIC	CHI	FRZ

Table 7: Rank of the ten most critical nodes in terms of strength

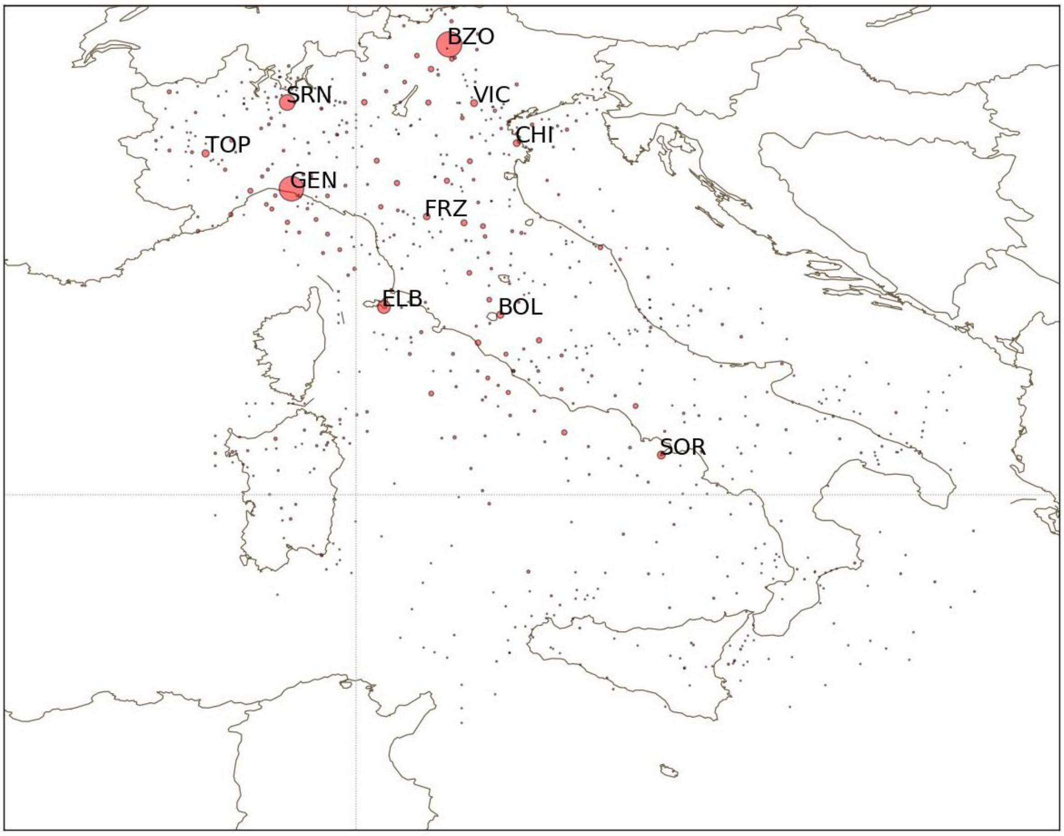


Figure 22: Map of the strength of the nodes in the Italian airspace

Betweenness

Betweenness is a measure of a node's centrality in a network. It is equal to the number of shortest paths from all vertices to all others that pass through that node [3]. Betweenness is a useful measure of both the load and importance of a node. A fix with a high value of betweenness means that there are a lot of airways that pass through it. Table 8 reports the ten nodes of the Italian airspace with the highest value of betweenness. Their geographical location is shown in Figure 23: Map of the betweenness of the nodes in the Italian airspace. The air traffic controller confirmed our results especially for the navpoints in the first three positions of the ranking. The navpoint “ELB” is in fact considered the most critical of the entire Italian airspace as there are multiple airways converging in it.

Rank	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
Navpoint	ELB	FRZ	SOR	ANC	PES	CHI	BOL	TEA	PEMAR	TINTO

Table 8: Rank of the ten most critical nodes in terms of betweenness

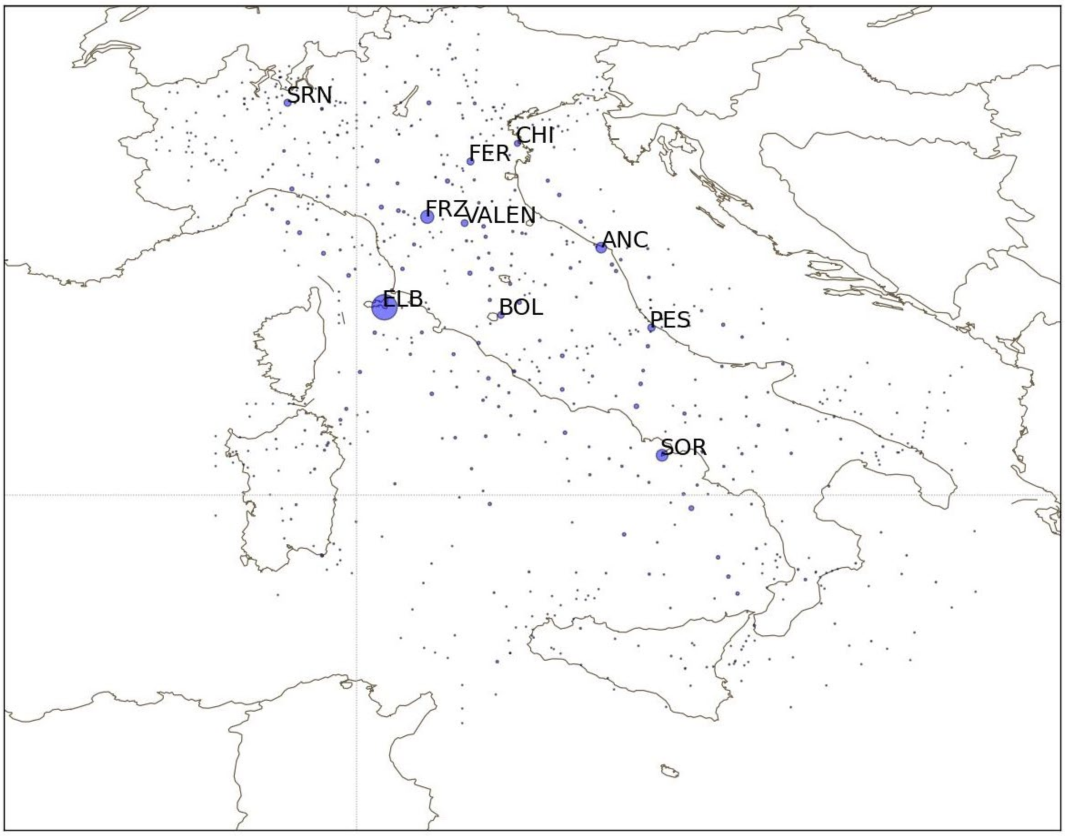


Figure 23: Map of the betweenness of the nodes in the Italian airspace

Origin-Destination Betweenness

The Origin-Destination Betweenness has the same definition of the previously defined betweenness. However for its computation only the nodes that are points of entry and exit from the national airspace are considered [4].

Table 9 reports the ten fixes of the Italian airspace with the highest value of Origin-Destination betweenness. Their geographical location is shown in Figure 24. Also in this case the navpoint "ELB" is at the first position meaning that it is also critical because it also an entry/exit point from the Italian airspace (flights climbing or descending form the en-route portion of their flight are considered as entering/exiting the airspace) thus increasing the need of coordination between sectors of different ANSPs.

Rank	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
Navpoint	ELB	BOA	GEN	TOP	CHI	FRZ	PAR	SRN	BZO	BRD

Table 9: Rank of the ten most critical nodes in terms of Origin-Destination betweenness

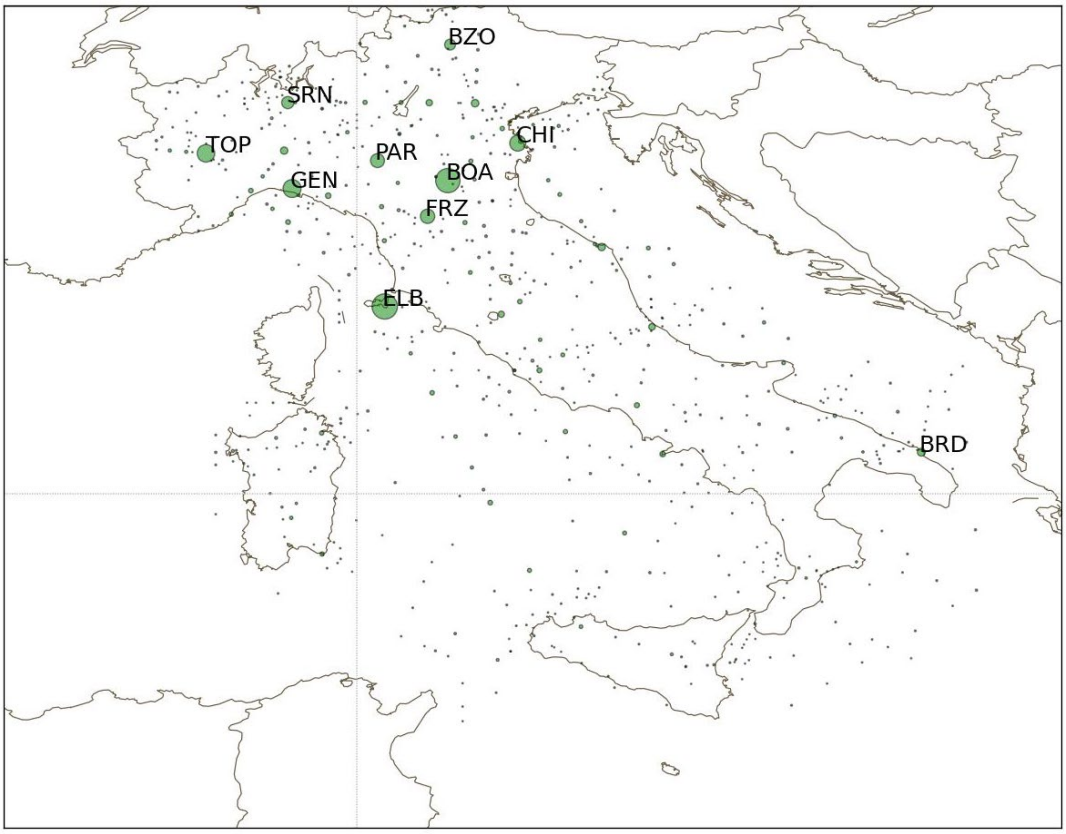


Figure 24: Map of the Origin-Destination betweenness of the nodes in the Italian airspace

Random-Walk Betweenness

The Random-Walk Betweenness is another variant for the measure of nodes centrality in a network. It is based on the traffic flows on the network. Imagine to travel on the network numerous times going from a navpoint to another on the most trafficked airways; a node will have a high value of Random-Walk many more times you go over it [5].

Table 10 reports the ten fixes of the Italian airspace with the highest value of Random-Walk betweenness. Their geographical location is shown in Figure 25.

Rank	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
Navpoint	SOR	ELB	PNZ	VALEN	GEN	CHI	TAQ	VERNA	SRN	GIANO

Table 10: Rank of the ten most critical nodes in terms of Random-Walk Betweenness

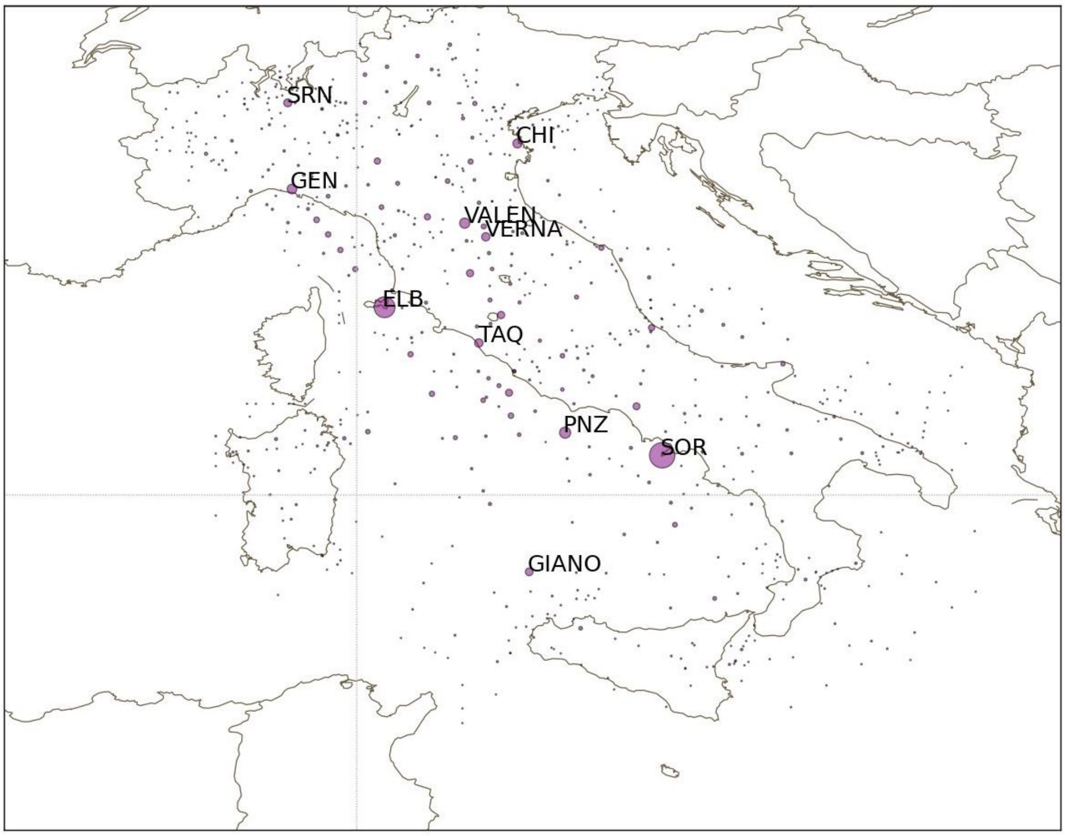


Figure 25: Map of the Random-Walk betweenness of the nodes in the Italian airspace

Table 11 shows an overview of the different rakings for the different metrics analysed. The most critical navpoint of the Italian airspace seems to be "ELB". In fact it is at the fourth place of the strength ranking and at the first or second place for the betweenness measurements. This means that "ELB" is critical for both the traffic load around it and its centrality in the network. This implies that a large number of conflicts may arise around it due to the combination of these factors. This conclusion has been confirmed also by the air traffic controller who identified "ELB" as a critical navpoint where particular effort is required to separate conflicting aircraft. He also recognized that the navpoints that Italian air traffic controllers consider critical, all appear in our ranking thus demonstrating the ability of our analysis to effectively identify the most critical navpoints without requiring operational input.

Rank	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
Strength	BZO	GEN	SRN	ELB	SOR	TOP	BOL	VIC	CHI	FRZ
Betweenness	ELB	FRZ	SOR	ANC	PES	CHI	BOL	TEA	PEMAR	TINTO
OD Betweenness	ELB	BOA	GEN	TOP	CHI	FRZ	PAR	SRN	BZO	BRD
RW Betweenness	SOR	ELB	PNZ	VALEN	GEN	CHI	TAQ	VERNA	SRN	GIANO

Table 11: Overview of navpoints ranking with respect to the different metrics

How

We analyzed the Italian Navpoints network in the timeframe spanning from May to December 2011. The resulting network was composed by 792 nodes (including airports) and 2242 edges. Each edge represent a link between two navpoints that were actually travelled by one or more aircraft in the timeframe analyzed. Then we computed the four selected metrics for the nodes of our network and derive a ranking of the nodes with the highest values.

References

[1] B. Monechi, M. Ducci, M. Cipolla, S. Vitali, S. Micciché, R. N. Mantegna, G. Gurtner, F. Lillo, L. Valori, S. Pozzi (2013). Exploratory analysis of safety data and their interrelation with flight trajectories and network metrics. Paper presented at the ATOS-ISI ATM conference.
[2] M. Newman. The structure and function of complex networks. SIAM review, 45:167–256, 2003
[3] Freeman, Linton (1977). "A set of measures of centrality based upon betweenness". Sociometry 40: 35–41
[4] Kazerani, A.; Winter, S., 2009: Can Betweenness Centrality Explain Traffic Flow?, AGILE, Hannover, Germany
[5] Newman, Mark EJ. "A measure of betweenness centrality based on random walks." Social networks 27.1 (2005): 39-54.

A.4 Exploratory analysis of safety data and their interrelation with flight trajectories and network metrics

Introduction

The objective of this analysis is to look for interrelations between the occurrence of Short-Term Conflicts Alerts and Loss of Separation events, gathered with ASMT, and the so called “complex network metrics” measured over the navpoint network. These metrics are essentially a mathematical way to describe the structure of the navpoint network. They are able to identify in a simple and synthetic way crucial characteristics of the network such as its most congested areas or the nodes that are more important in terms of airways passing through it. However it is still not clear the relation between these properties of the airspace and the probability of having an adverse occurrence in some of its spot. The analysis we carried out aims therefore at highlighting some of the correlations between these two aspects and thus providing a tool to help in identifying the critical spots of the airspace. These information could then be used both in the design of a safer airspace and in the management of the current one.

The events used in the analyses were gathered over the Rome ACC and covers a limited period of time, thus the navpoint network has been built using traffic data of the same part of the Italian Airspace and in the corresponding days.

Network metrics of the Italian airspace

In order to compute network metrics on the Italian navpoints network we analyzed it in the timeframe spanning from May to December 2011. The network resulted to be composed by 792 nodes (including airports) and 2242 edges. Each edge represents a link between two navpoints that were actually travelled by one or more aircraft in the timeframe analyzed.

We computed the following metrics:

- **Strength:** The strength of a node is essentially a measure of the traffic load on its adjacent links. It is the most direct measure of the traffic load on the node, then the nodes with high strength should also be the most busy.
- **Betweenness**: It is a measure of a node's centrality in a network. It is related to the number of shortest paths from all vertices to all others that pass through that node. A node with a high value of betweenness means that there are a lot of airways passing through it.

We measured the above defined metrics for all the navpoints in the Italian airspace and we derived a ranking of the ten most critical ones for each of them (Table 7). The rankings obtained were then validated with an air traffic controller from the Italian ANSP in order to confirm the ability of our analysis to identify correctly the most critical navpoints in the Italian airspace.

Rank	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
Strength	BZO	GEN	SRN	ELB	SOR	TOP	BOL	VIC	CHI	FRZ
Betweenness	ELB	FRZ	SOR	ANC	PES	CHI	BOL	TEA	PEMAR	TINTO

Table 12: Rank of the ten most critical nodes in terms of strength and betweenness

By looking at the ranking and at the corresponding maps (Figure 22 and Figure 27) it is clear that a node can have a high value of strength (i.e. being very congested) but a lower value of betweenness meaning that there are not a lot of airways passing through it. In addition while the value of the strength on a particular node may vary more often with the traffic load, the corresponding value of betweenness will be more stable, being an intrinsic property of the network structure.

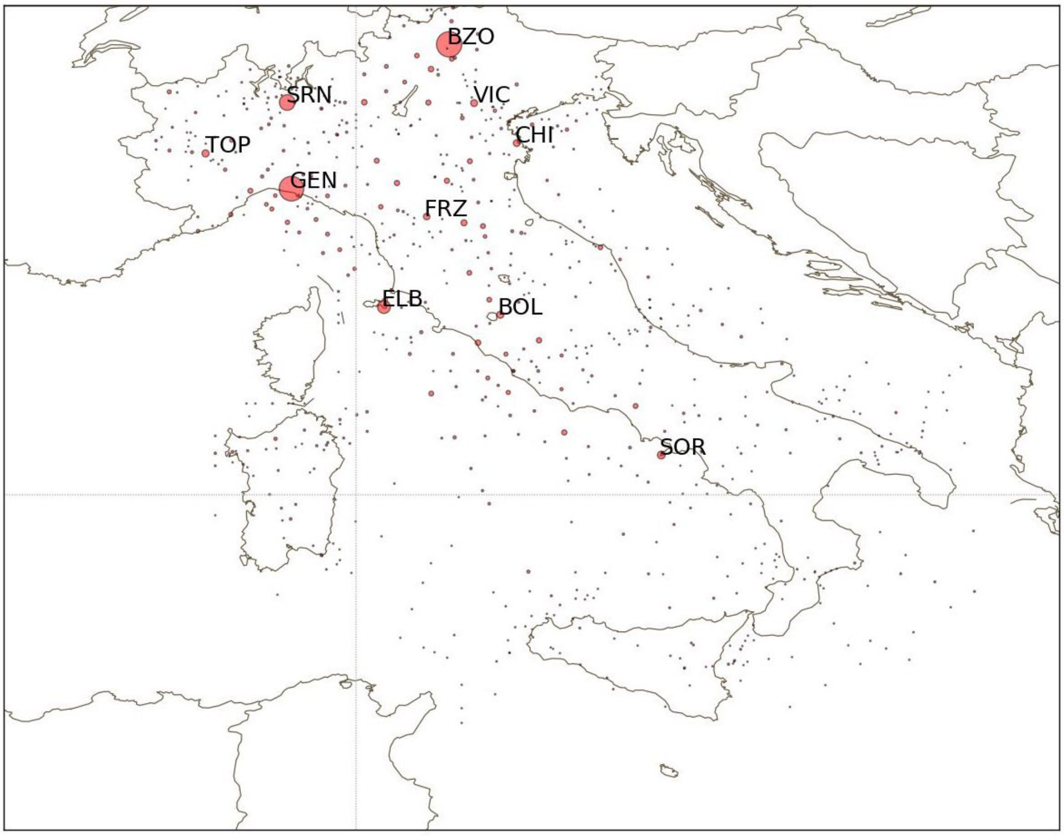


Figure 26: Map of the strength of the nodes in the Italian airspace

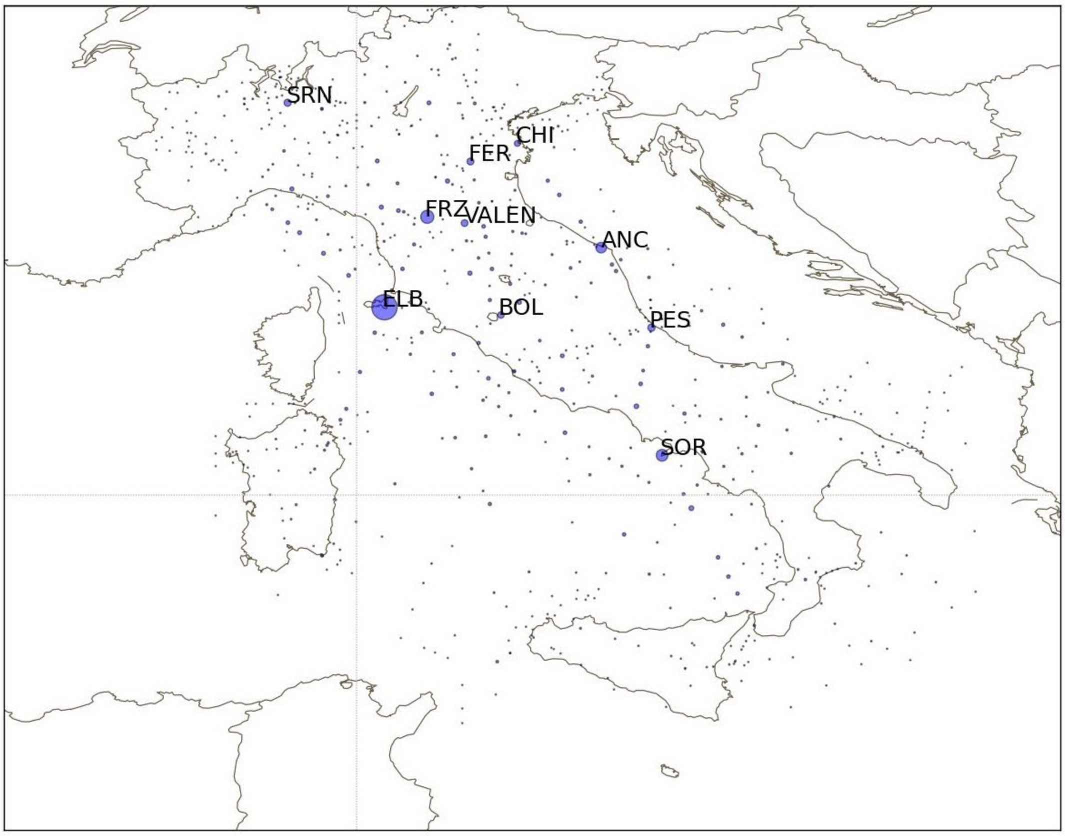


Figure 27: Map of the betweenness of the nodes in the Italian airspace

Data processing

Being an automatic tool, ASMT does not only collect relevant safety events, but it can also record some nuisance events that need to be filtered out “a posteriori”. Thus three types of filtering were applied to the data set:

- ASMT filtering: events were filtered out using ASMT parameters and filters. For instance, they may be activated because of the low track quality (the event shows characteristics connected to a false radar message), or because the event happened in a permanent military area, or under visual control close to an airport.
- logical filtering: it applies if− >then rules to the collected events. The event is filtered out if it matches one of the if−>then condition.
- operational filtering: it involves ENAV safety experts and it is a case by case analysis of the events.

Using the real trajectories in the same days of occurrence of the STCAs, we were also able to apply another filtering to the safety events and to assign to each couple of aircraft involved in each event a unique couple of trajectories in the traffic dataset. By considering one STCA and one of the aircraft involved in the events, we were able to check if there were trajectories compatible with it, i.e. if there were aircraft whose trajectory was not too far from the point of occurrence at the time of occurrence. The trajectories stored in our traffic database have a temporal resolution of about 2 minutes while the typical duration of an STCA is less than 1 minute. This means that in order to see if an aircraft is far or close to the point of occurrence, we had to linearly interpolate its position. Considering the distance between the interpolated points and the point of occurrence of the event, we were able to rule out all the trajectories that were not compatible with it (the criteria used to rule out trajectories has been tested over a small dataset of STCAs for which the matching trajectories were known). If no compatible trajectories were found, the event was filtered out and considered as a false positive. If there were multiple compatible trajectories, we choose among them the one with the smallest distance to the events.

At the end of the procedure we were able to match 750 STCAs with 1307 distinct trajectories. Since we have matched a couple of trajectories to every STCAs (those that have not been filtered out in the process), we were able to know where each aircraft of the couple was heading at the time the event occurred, i.e. we know from and to which nodes the aircraft was going. We used this information to assign each event to a link of the network. Considering one of the aircraft involved in an STCA, we computed all the shortest paths on the network connecting the node where the aircraft was coming from and the node towards which it was flying. Then we measured the distance between the occurrence point of the event and all the segments in the paths and assigned the event to the closest one. We repeated this procedure for all the aircraft involved in an STCA. At the end of this process we obtained for each node a measure of the number of STCA assigned to it that we called n_{STCA}, corresponding to the sum of all the events assigned to the links connected to it.

Results

Classical Network Metrics

Since the events were gathered within the Rome FIR, we restricted the Navigation Point Network to this airspace. Because the network is rather homogenous, this restriction did not change qualitatively the distributions and correlations of metrics presented before, even though the number of nodes and edges was reduced to 525 and 1393 respectively.

We found a positive correlation between the strength and the betweenness of a node and the number of events assigned to it. Since the strength is a measure of the traffic load over a node is clearly linked to the traffic load in a certain point of the airspace. However there is a significant correlation also with the betweenness meaning that the probability of occurrence of STCA should also be linked to the network topology and not just on how the traffic load is deployed over it.

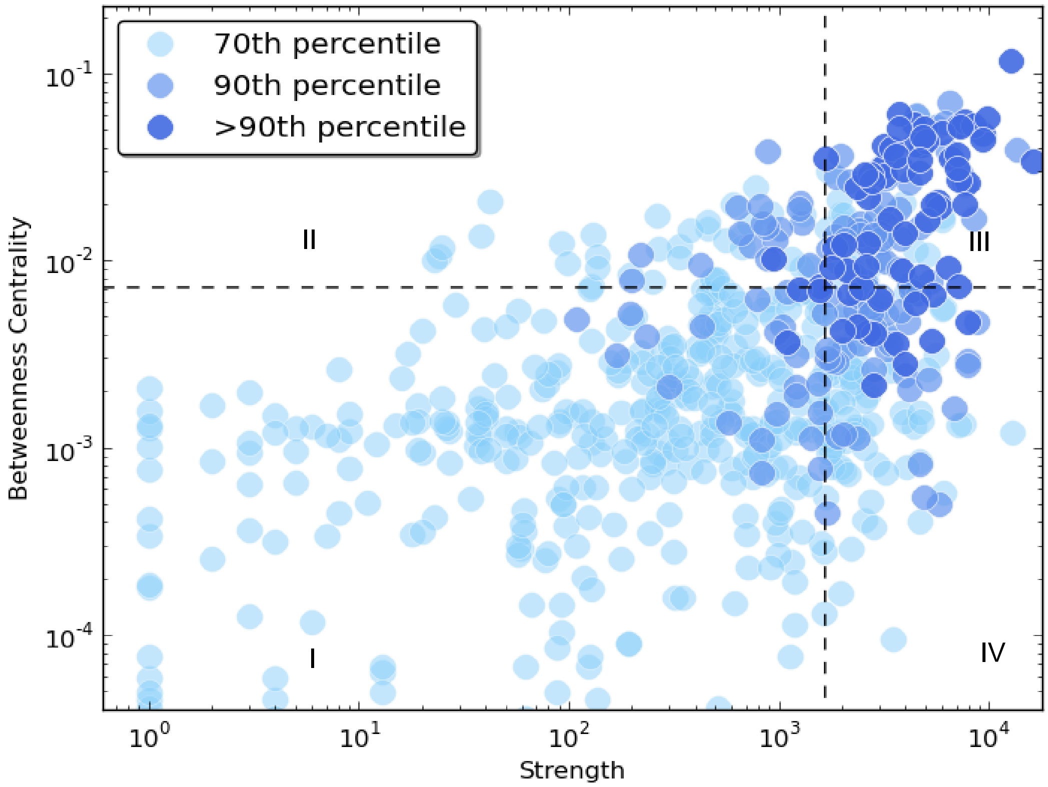
	Strenght	Betweenness
Correlation with n_{STCA}	0.59	0.63

In complex network theory there are analogous definitions of strength and betweenness for the links the network instead of its nodes. The analogous of strength for links is called generically “weight” and in this kind of network represents the number of aircraft that has flown over a certain segment connecting two navigation points. The “betweenness of a link” instead represent roughly the number of airways that are passing over the considered segment. We found positive correlations between these two metrics for the links and the number of STCAs associated to each link (indicated with w_{STCA} in the table below), although these correlations are slightly lower than before.

	Weight of links	Betweenness of links

Correlation with w_{STCA}	0.43	0.35

Figure 28 presents an overview of the results. Darker points (> 90th percentile) represent the navpoints with the higher number of STCA assigned to them. It is possible to notice that most of them are located in the upper-right corner of the diagram meaning that the navpoints with high traffic load (high strength) are also likely to be the intersection of multiple airways (high betweenness) and are likely to being the place where more STCA might happen. However this is not a general rule. There are in fact navpoints in the 90th percentile that are located in the upper-left part of the diagram. These points are not very congested but are probably at the intersection of several airways and the number of assigned STCA is quite high.



70th percentile	90th percentile	> 90th percentile
No events associated	1-7 events associated	More than 7 events associated

Figure 28: Navpoints in the Rome FIR with their values of strength and betweennes related to the number of STCA assigned to them. The vertical and horizontal dashed lines correspond to the average value of strength and betweenness respectively

Using the information presented in Figure 28 it is possible to define "outlier nodes". Ideally nodes with a high number of safety events associated should also have high strength and betweenness and thus lay in the upper right part of the graph (identified with the roman number III in the figure). Similarly nodes with high values of these metrics should have a large number of conflicts occurring near them. So we defined as "outliers" all the nodes that do not correspond to this ideal scheme:

- Nodes in III below the 70th percentile (high metrics, low conflicts)
- Nodes in II over the 70th percentile (critical nodes with high betweenness but low strength)
- Nodes IV over the 70th percentile (critical nodes with high strength but low betweenness)
- Nodes in I over the 70th percentile (critical nodes with low metrics)

Having classified these outliers, we can plot them on a map in order to spot for possible patterns that could indicate the reasons of the disagreement between the network metrics and the safety events. Figure 4 shows the distribution of the first kind of outlier nodes (in blue in the picture) together with the nodes correctly classified, i.e. nodes lying in III with large number of events associated (in red in the picture). From figure 4 emerges that many of these nodes are lying in areas close to the border of the airspace, indicating that they are probably not spotted due to effects of the airspace boundaries. Similarly these nodes can be found following airways of critical nodes close to where they cross the boundary of a sector (Fig. 5). These could indicate the presence of particular procedures taken in order to prevent the occurrence of critical events when aircraft are crossing the boundary from a sector to another one.

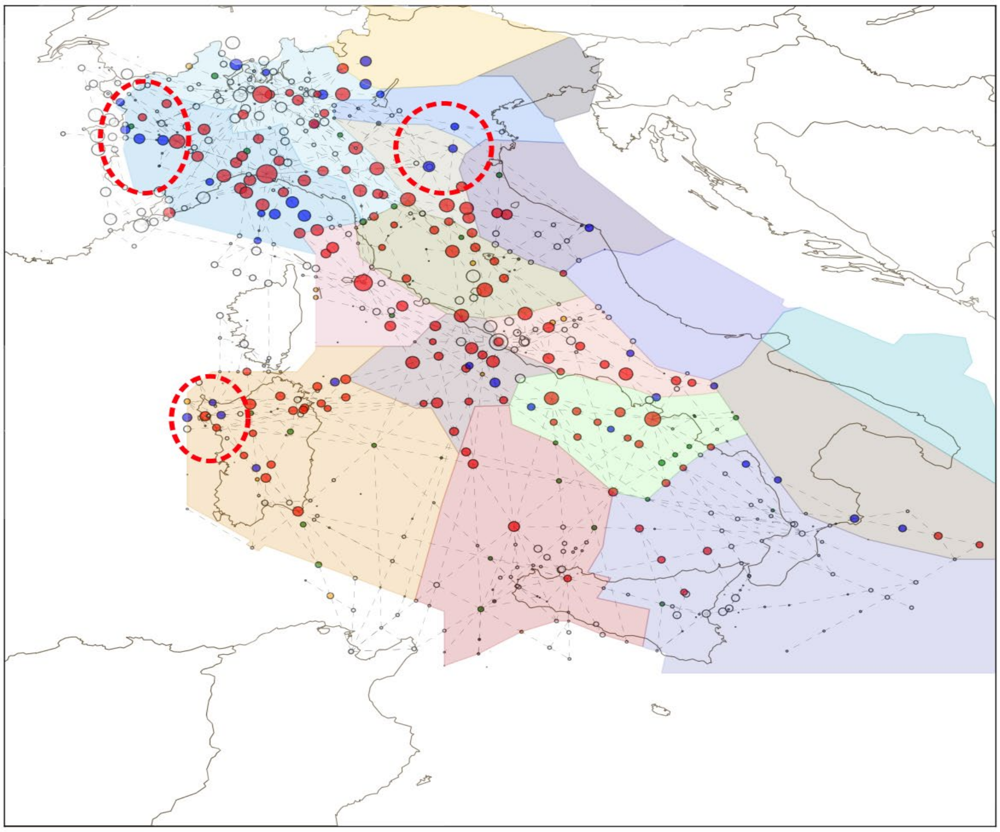


Figure 29: Navpoint Network in the Rome FIR. Red nodes are nodes correctly identified as critical crossing the strength and betweenness centrality metrics, blue nodes are critical nodes with small metrics. The size of the nodes is proportional to their value of strength. The red dashed circles highlight the presence of outliers in some boundary areas.

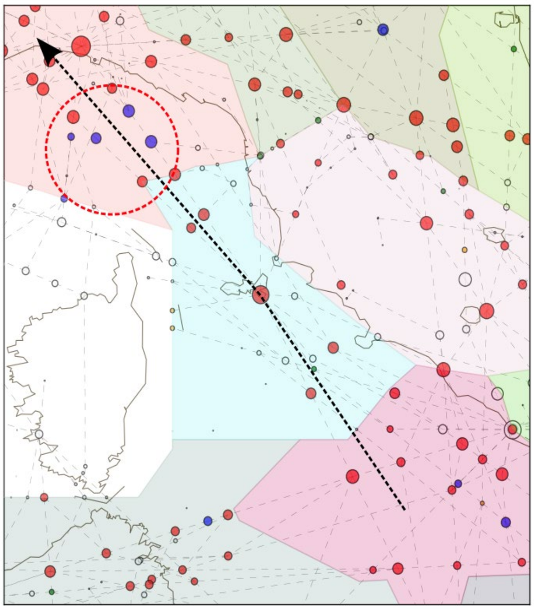


Figure 30: Critical Nodes across airways in the central part of the network. The red circle highlights outliers close to the boundary of a sector.

Dynamical Network Metrics

By comparing the last filed flight plans with the actual trajectories recorder in the radar tracks, we have been able to define new metrics for the nodes in the network that are not related to its topological structure but to the dynamics taking place over it during the air traffic control operations. These metrics have been measured considering a different time frame with respect to the one used for gathering the safety events. This has been done in order to prove that any possible correlation between these dynamical metrics and the safety events are intrinsic properties independent from the time period used to compute them. The new metrics defined are:

- Frac: fraction of flight that should have crossed the node but have not.
- Fork: fraction of flights for which a deviation begins after this point.
- Antifork: contrary of the previous one, points which ends a deviation.
- AfterFork: fraction of flights which had a “fork” at the previous point.
- Alt: absolute difference of altitude at this point between the planned and actual one.
- Delay: amount of en-route delay generated by the point (this metric can be divided in “Positive Delay” and “Negative Delay”)

Surprisingly the study of the correlations with the events highlighted how all the metrics linked with horizontal movements are completely uncorrelated with the occurrence of safety events. On the other hand metrics related to the generated delay are weakly correlated. The metric regarding vertical deviations (Alt) is the most correlated one, suggesting that at least in the considered airspace, vertical movements are often associated with critical points and safety events.

	Alt	Fork	Pos. Delay
Correlation with n_{STCA}	0.43	0.04	0.23

Note that the previous analysis of outliers can be performed again, crossing some of these metrics with strength or betweenness centrality. Figure 6 shows the analogue graph presented in Figure 3, using "Alt" instead of betweenness centrality. The same categorization of outliers is still possible and patterns similar to those presented can be found on map. However in this case there are a few outliers indicating that the most congested nodes are also the areas where vertical movements are more likely to happen and both these characteristics are related to safety events occurrences.

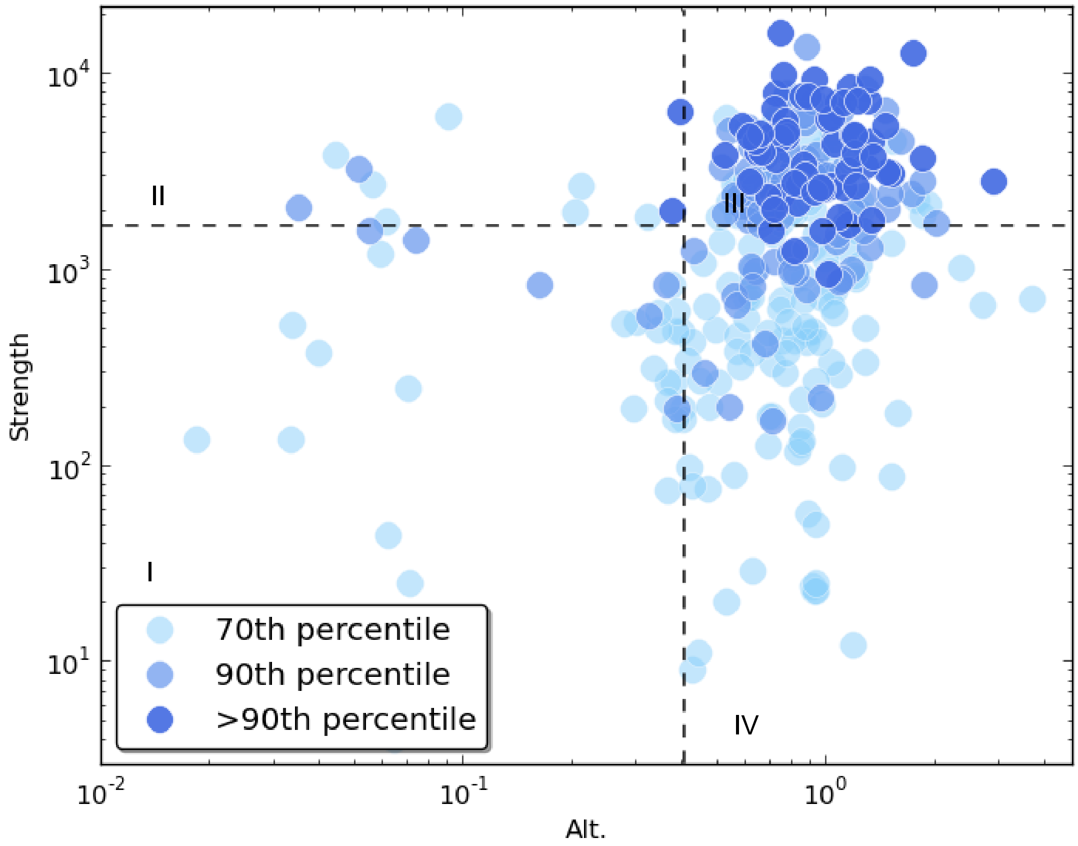


Figure 31: Navpoints in the Rome FIR with their values of strength and Alt related to the number of STCA assigned to them. The vertical and horizontal dashed lines correspond to the average value of Alt and strength respectively

Another Criticality Measure

While the assigned number of STCAs to a link of the network measures the frequencies of occurrence of STCAs over it, we can rescale this quantity with the number of aircraft that has travelled over the link in the corresponding period of the gathering of the events. Similarly as we have done before we can use this measure over the links to define a new criticality measure for the nodes, which in this case is independent from the amount of traffic that has crossed them. This metric, denoted as $P_n(STCA)$, represent the probability that an aircraft that is travelling through the navigation point *n* incurs in an STCA with another aircraft.
This quantity can be related to the network metrics as we did with the frequency of

STCAs, in order to verify such metrics are able to identify critical nodes also with this new definition of criticality.
Surprisingly strength, which is related to the traffic over the network, is still quite uncorrelated with this quantity, while a slight correlation can still be observed with the betweenness that is related just to the network topology. This indicates that a high traffic load in a certain node of the network is correlated to a high frequency of STCAs, but it does not increase the probability of incurring in an STCA for a single aircraft since this probability is independent from it.
On the other hand the slight correlation with the topological metrics indicates that this probability is influenced by the topology of the network.
Concerning dynamical metrics related to the air traffic control activity, the previous results obtained with n_{STCA} are confirmed: metrics related to horizontal movements are uncorrelated, while vertical movements and generated delay are still relevant.

	Strenght	Betweenness	Alt	Fork	Pos. Delay
Weighted Correlation with $P_R(STCA)$	0.10	0.43	0.40	0.03	0.32

Conclusions

We found that correlations between the classical network metrics and the safety occurrences have been observed. The analysis so far pointed out that the metrics linked to the deployment of the traffic load over the network seems to be the most correlated with the number of events occurred on a node, but also Betweenness (for the calculation of which less information about traffic is required) is well correlated with the number of STCAs.

We have also developed non-classical networks metrics related to the variations between the real trajectories of the aircraft and their flight plans due to the air traffic control activity. The analysis pointed out that metrics related to horizontal movements are not correlated with the safety events. On the other hand metric related to vertical movements and delay generation show positive correlations with the occurrence of STCAs.

While the presented metrics are correlated with the number of assigned STCAs, the correlation with the probability of incurring in an STCA (a time and traffic independent metric) are slightly lower. In particular strength seems to be uncorrelated with this metric. This indicates that the probability that an aircraft travelling through a node in the network incurs in a STCA is not depending on the traffic load of that node, but our measures seem to indicate that other topological and dynamical properties might be relevant, such as the way the main routes intersects each other over the navigation points and the amount of vertical movements present in a certain area.

It is important to notice that our analysis presents two major limitations: it has been performed in a very limited part of the European Airspace and the number of recorded events was small. In our opinion, a more global analysis performed in a larger timeframe (in order to increase the number of safety occurrences) could help to clarify which are the best-suited metrics to predict the critical areas of the airspace.

References

[1] ELSA deliverable D3.1
[2] M. Newman. The structure and function of complex networks. SIAM review,45:167–256, 2003

[3] Freeman, Linton (1977). "A set of measures of centrality based upon betweenness". Sociometry 40: 35–41
[4] Kazerani, A.; Winter, S., 2009: Can Betweenness Centrality Explain Traffic Flow?, AGILE, Hannover, Germany
[5] Newman, Mark EJ. "A measure of betweenness centrality based on random walks." Social networks 27.1 (2005): 39-54.

END OF DOCUMENT-