Complex Networks and Transport Systems: Application to Air Transport and Urban Mobility

Doctoral Thesis

Supervisor
Prof. Vittorio Loreto

PhD Candidate
Bernardo Monechi

XXVII Cicle
# Contents

1 Introduction ................................................................. 7
  1.1 Complex Network Analysis in Transportation Systems and ATM . 10
  1.1.1 Complex Networks Modeling in ATM .......................... 18
  1.2 The Dataset and ATM Structure .................................... 21
  1.2.1 Current Airspace Structure and the SESAR Scenario ........ 21
  1.2.2 The Demand Data Repository Dataset .......................... 24
  1.2.3 ASMT Safety Dataset ........................................... 27
  1.3 Network and Delays Analysis .................................... 29
  1.3.1 The Navigation Point Networks .............................. 29
  1.3.2 Delays ......................................................... 35
  1.3.3 Short-Term Conflict Alerts and Network Metrics ............ 38
  1.4 Air Traffic Control Model ....................................... 54
  1.4.1 Definition of the ATC Model .................................. 54
  1.4.2 High traffic transition in Synthetic and Realistic Airspaces 61
  1.4.3 Model Validation ............................................. 77
  1.4.4 Model Refinements ............................................ 89
  1.5 Global Optimization .................................................. 95
  1.5.1 Extremal Optimization Algorithm ............................ 96
  1.5.2 Application to Trajectory Optimization ..................... 98
  1.5.3 The Optimal Solution ......................................... 99
  1.5.4 Efficiency of the Optimal Solution .......................... 103
  1.6 Conclusions and Perspectives ................................. 107

2 Introduction ................................................................. 111

3
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2</td>
<td>Analysis of GPS Tracks</td>
<td>114</td>
</tr>
<tr>
<td>2.2.1</td>
<td>GPS Data</td>
<td>115</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Space-Time Relation</td>
<td>116</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Daily Dynamics</td>
<td>123</td>
</tr>
<tr>
<td>2.3</td>
<td>Grid Model with Shortcuts</td>
<td>128</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Definition of the Model</td>
<td>129</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Reproducing the Space-Time Relations</td>
<td>132</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Reproducing Daily Dynamics Patterns</td>
<td>141</td>
</tr>
<tr>
<td>2.4</td>
<td>Conclusions and Perspectives</td>
<td>146</td>
</tr>
</tbody>
</table>
Overview

This thesis is devoted to the study of transportation systems by means of Complex Systems and Complex Network Theories. Complex Networks are a tool of inestimable value in human transportation studies since in most of the cases the means of transportation used by individuals to move in space are bounded to move on a complex network. The topological properties of transportation networks can influence both the ability of individuals to move as well as their behavior in the environment, thus a characterization of the network is mandatory in order to understand the properties of the considered system. The two transportation systems that have been studied in this work are the Air Transport System and the mobility of cars in a urban environment.

The analysis and modeling of the Air Transport System is the first and most extensive part of this thesis. In particular we will try to characterize and study the networks in which aircraft fly, exploiting these results to build a data-driven model of Air Traffic Control.

The second part of the thesis is a continuation of the studies performed during by Pierpaolo Mastroianni during his Master Thesis. His work concerned the analysis of GPS tracks data in the City of Rome and the inference of statistical laws characterizing the behavior of car drivers. My contribution to his work is the development of a model capable of explaining some of the results presented in the Master Thesis.

Despite the fact that the two systems have very different features, two main analogies link them:

- An optimization dynamics performed by controllers in the case of Air Traffic and by car drivers themselves in the case of Urban Traffic.

- The presence of a network structure over which the dynamics takes place, which is less bounding in the case of Air Traffic with respect to Urban Traffic.
As we will briefly discuss in par. 1.2.1, the air traffic control system in Europe may reach its capacity limits within the next years and a complete redesign of its structure has been proposed by EUROCONTROL\textsuperscript{1} within the SESAR Project [1]. Thus a better understanding of its functioning and criticality is mandatory in order to inform correctly the building of this future scenario. Similarly, a characterization of the human behavior in a urban environment could inform a reorganization of human mobility, providing information to each individual in order to improve the way in which they move in the city [2]. Before introducing the results of my research activity I will present in each part a brief review of works in which complex networks have been applied to problem related with Air Traffic an Human Mobility respectively, focusing also on the more general framework of Complex Networks analysis in the first part.

\textsuperscript{1}www.eurocontrol.org
Chapter 1

Analysis and Modeling of the European Air Traffic Control System

1.1 Introduction

This part of the Thesis, the most extensive one, is devoted to the study of the Air Traffic Management (ATM) system within the framework of Complex Networks and more in general Complex Systems Physics.

Air Transport is nowadays one of the most important means of transportation and one of the main actors of the economic and social development of every country in the world [3]. Despite the competition with other means of transportation, e.g. high speed railways, air transport is still the best suited way of connecting areas of the same continent of in different ones. The recent economic crisis affected the ATM systems, reducing the travel demand in Europe and thus impacting negatively on the constant traffic growth observed in the previous years. Although the emergence on new economies outside the European region could bring new traffic demand and contribute to the reestablishment of the growth. For these reasons a growth is still forecast in the European [4] and also American [5] airspaces within the next years. This increase of the traffic load could contribute to bring the ATM over its capacity limits. One of the main concerns regarding the increase of traffic is surely at the airport level, since airport capacity is one of the major constraints that have to be taken into account at every level of the hierarchical structure of the ATM system. Moreover, a higher density of flights within the airspace could lead to reduced performances and to the violation of safety standards, which will eventually
result in monetary loss both for national service providers and for stakeholders.

To the present day, safety is entirely provided by service providers, whose duty is to guarantee the correct execution of each flight and to avoid adverse occurrences. Safety occurrences are not unlikely to happen even though they are very rare events. EUROCONTROL \(^1\), an international organization composed of Member States from the European Region dealing with almost every aspect of air traffic management, in its annual report in 2012 estimated about 110 safety occurrences per million flight hours in 2011 within the European Airspace [6] with a slight increase of 12% with respect to the previous year.

In order to provide the correct functioning of the system also in a possible future scenario of increased traffic, new solutions and improvements are needed. Thus a better understanding of the capacity limits and criticality of the current system are crucial to inform the design of an airspace capable of sustaining the increase in traffic load. Air Transportation is a paramount example of an increasingly interconnected techno-social system [7] that requires a strong interoperability of the technological infrastructure with a human and social component. The evaluation of the effects of increasing traffic and the introduction of new scenarios is not directly possible due to the lack of high traffic data. Thus the only possibility is the use of theoretical models informed by the data analysis. This “data-driven” approach is now feasible thanks to the recent increase in the available information due to the spreading of ICT technologies within everyday life and industries that are now able to collect huge amounts of data [8,9].

Concepts borrowed from Complex Systems science have been successfully applied in many areas far from traditional physics. Many biological, social and techno-social systems have been successfully studied and modeled within this framework. Some of the most notable and successful examples of its application are protein folding [10], social dynamics [11], jamming of vehicular and pedestrian traffic [12]. One of the most recent and successful application of the data-driven modeling approach is the study of epidemic spreading [13–16]. It is worth noting that studies of the ATM system have been crucial in the development of realistic epidemic model. In this process air transport plays an important role, allowing the infection to spread across distant locations [17]. However, as we will see in the next paragraphs, few studies dealt with the geographical structure of the routes of the aircraft within the airspace and the vast majority focused on the interconnections between airports.

In this work we will try to extend this approach by studying the actual trajectories of

\(^1\)www.eurocontrol.org
the flights within the European airspace trying to infer how the ATM system manage them during normal operations. The data analysis will then lead to the development of an model capable of reproducing the statistical properties found in the data. This part of the thesis is structured as follows:

**Section 1** In section 1 we will present and review some of the major results concerning the application of Complex Systems and Complex Networks framework to the study of ATM related problems. In particular we will distinguish between studies concerning data and complex networks analysis from those concerning the development of models in ATM.

**Section 2** In section 2 we will introduce and describe the set of data provided by EUROCONTROL for the realization of this PhD project. Such dataset contains historical information about the trajectories of all the flights within the European Airspace in a certain period of time in 2011 as well as information about some kinds of critical occurrences recorded in the same period.

**Section 3** In section 3 we will present the main results concerning the analysis of the dataset presented in chapter 2. In particular we have focused our attention in the analysis of the delays of each flight and on the geographical structure of the routes generated by the deployment of the traffic over the airspace. The central tool of such analysis is the so called “Navigation Point Network”, a geographical network [18] built with the navigation aids that the aircraft use during their flight. We will use the data about the critical occurrences in order to understand the relation between the properties of these networks and the criticality of the ATM system.

**Section 4** In section 4 we will define and study a model of Air Traffic Control based on a dynamics taking place over a Navigation Point Network. The results of the previous chapter will be used to define a realistic setup of the model and to validate it. Moreover we will study its behavior in high traffic conditions. The model shows a transition from a state in which all the safety occurrences are correctly managed to a high density phase in which many are not. The transitions present a scaling property with the size of the system, that depends on the topological properties of the airspace.

**Section 5** In section 5 we will present an adaptation of a well-known optimization algorithm, *Extremal Optimization* [19,20], to the problem of trajectory optimization. In the
current scenario conflicts between trajectories are managed during the operations by air traffic control. We used the algorithm to build a new kind of airspace in which possible conflicts are taken into account during the planning phase of the trajectories. We will then use the ATC model in order to test its efficiency with respect to the current one.

Conclusions We will review the results and discuss possible future developments of the work.

1.1.1 Complex Network Analysis in Transportation Systems and ATM

The analysis and representation of complex systems as complex networks has been a growing trend in the last years [21]. This is due mainly to the large variety of tools that have been developed within this framework that are often able to highlight the topological features of such systems and to provide useful information about their functioning, their structure and their dynamics [22, 23]. Moreover, complex networks have been crucial in order to understand many emergent phenomena in systems with a large number of interacting actors. Some remarkable examples are the diffusion of rumors in social media [24], the diffusion of diseases [15, 25], synchronization of neural networks [26], breakdown of power grids [27, 28]. Moreover such formalism has been successfully applied to the study of many transportation systems, including streets [29], railways [30], subways [31] and air transport [32]. Note that the transportation networks used to describe these systems are usually embedded in a metric space [33] and such embedding have many effects on the resulting topology. In the following paragraph we will review some of the most relevant studies concerning the complex network analysis focusing on this kind of systems. A characterization of the topological features of a network is the first step that is usually performed when analyzing a system, since topology influences or is influenced by the dynamical processes taking place in the network. Nowadays there are some standard metrics that are usually considered when a network has to be characterized. Considering the Adjacency Matrix $A_{ij}$ associated with the network, i.e. a matrix so that $A_{ij} = 1$ if the node $i$ is connected with the node $j$ and $A_{ij} = 0$ otherwise, the degree is the number of links connected to a node,

$$k_i = \sum_j A_{ij}.$$  (1.1)
1.1. INTRODUCTION

The distribution of the degree $P(k)$ is a quantity that can give interesting information about the system. For example when considering the Internet or Airline Networks, such distribution follows a power-law indicating the presence of a relevant number of hubs, i.e. highly connected nodes that are thus particularly important for the network [34]. These two networks are spatial based since each airport and each node of the Internet has a specific geographical position, although spatial constraints are not particularly strong for these kind of networks. In that case a cut-off usually appears in the distribution of $P(k)$ or alternatively the distribution may become highly peaked. In [35] has been shown that embedding two particular preferential attachment network models in space could lead to power-law degree distributions with a cut-off and could reproduce some of the correlations observed in real spatial networks. The fact that $P(k)$ can be peaked in spatial networks makes this distribution less interesting for their characterization. An example of another metrics that can be considered is the clustering coefficient, defined as

$$C(i) = \frac{2E_i}{k_i(k_i - 1)},$$  

(1.2)

where $k_i$ is the degree of the node $i$ and $E_i$ is the number of links connecting the neighbors of $i$. This quantity is useful since it can give information about the spatial structure of the network and depends with the number of triangles present in the network. Another version of this coefficient allows to study the clustering of classes of degree and it is defined as

$$C(k) = \frac{1}{N(k)} \sum_{i: k_i = k} C(i),$$  

(1.3)

but is not very useful with non heterogeneous networks in which many classes of degree are explorabale.

Another quantity that gives information about classes of degree is the average nearest-neighbors degree. This quantity is related to the correlations between the degree of connected vertices and can be defined as

$$k_{nn} = \sum_{k'} k' P(k'|k)$$  

(1.4)

or alternatively for each node as

$$k_{nn} = \frac{1}{k_i} \sum_{j \in \Gamma(i)} k_j,$$  

(1.5)
where $\Gamma(i)$ is the set of the neighbors of $i$. This quantity can identify the presence of assortative or disassortative mixing of the degree, i.e. if nodes with large degree tend to connect to other high degree nodes or with low degree nodes, depending on its behavior as $k$ grows.

These metrics concern the topological structure of the network, but they disregard the information about the weights of the links. Weights are a valuable instrument to characterize a network since they describe the intensity of a connection (e.g. in transportation network the weight is usually the flow of traffic over the link). The most common metrics measured in weighted networks is the strength,

$$s_i = \sum_{ij} w_{ij},$$

(1.6)

where $w_{ij}$ is the weight on the link $(i, j)$. This metric is a generalization of the degree of a node and in transportation networks is usually a measure of the traffic handled by a node. As well as the degree, also the strength is usually power-law distributed in many real world networks. In [36] many generalizations of the standard topological metrics for weighted networks were introduced, in order to take into account the strength of the connection in topological measures. These metrics have been studied and applied to two different networks: the World-wide Airport Network [37] (the most common type of network studied in ATM complex networks) and the scientist collaboration network [38].

The weighted clustering coefficient is the generalization of equation 1.2 that measure the local cohesiveness taking into account the intensity of the connections of the triplets,

$$C_w(i) = \frac{1}{s_i(k_i) - 1} \sum_{jh} \frac{w_{ij} + w_{ih}}{2} A_{ij} A_{ih} A_{jh}.$$  

(1.7)

Note that it is also possible to define a weighted clustering coefficient for a degree class $C_w(k)$ similarly to equation 1.3. Another generalization is the weighted average nearest-neighbors degree defined as

$$k_{nn}^w = \frac{1}{s_i} \sum_j A_{ij} w_{ij} k_j.$$  

(1.8)

These metrics have been studied for the two previously mentioned networks, pointing out the assortative nature of the Scientific Collaboration Network and the rich club phenomenon [39] of the World Airport Network. This phenomenon is particularly interesting for the Airport Network since it indicates that highly connected airports are preferably connected with other highly connected airports.
A central issue in networks characterization is the identification of the most important nodes according to some given criterion. Usually this is done by means of Centrality Metrics used to rank the nodes of the network. The degree of the nodes is one of the most natural measure of centrality that can be considered but it could lead to misleading classifications since low degree nodes could be important since they may be “bridges” connecting different part of the network. Probably the most famous centrality metrics is Betweenness Centrality \[40,41\]. This metric is based on the shortest-paths on the network and its defined as:

$$b(n) = \sum_{i,j;i\neq j\neq n} \frac{\sigma(i,j|n)}{\sigma(i,j)},$$

where $\sigma(i,j)$ is the total number of shortest-paths on the network connecting the nodes $i$ and $j$, and $\sigma(i,j|n)$ is the total number of shortest-paths on the network connecting the nodes $i$ and $j$ and passing through $n$. Initially introduced for social networks, it has been found to be important also in transportation networks. Despite the fact that its definition relies only on topological quantities (weights are not necessarily considered when computing shortest-paths), it is often used as an estimate of the load of the nodes in many kind of networks assuming that the number of shortest-paths is a zeroth order approximation of the frequency of usage of a node. Many other definitions and modifications of the betweenness centrality have been proposed in literature \[42\], as well as methods to make its computation faster \[43\]. In particular in the field of transportation networks, these modifications aimed to improve the predictive power of traffic \[44,45\] or at identifying critical nodes that are particularly relevant for the correct functioning of the network \[46\]. In par. 1.3.3 we will use some of these metrics for our analysis of the topology of the airspaces.

The application of these concepts to the study of ATM systems has been focused on the Airport Networks. The availability of data provided by national and international organizations and the interesting features of these networks are certainly the main reasons that have driven all the interest toward this representation of the system.

The studies of the airport network focused on restricted areas of the world \[47–52\] as well as on the World-Wide Airport Network \[37,53\]. In most of the cases some common qualitative features emerged, independently of the location of the considered network. Probably the most important of these features is the power-law distribution of the degrees of the nodes which is usually a Pareto, a Double-Pareto or a power-law with an exponential cut-off. The exponents of the power-law distributions range in every case from $-2$ to $-3$. 
This indicates the presence of a large number of hubs, i.e. highly connected airports that exhibit large degrees, but that however suffer by the effects of geographical constraints leading to the exponential cut-off of the distribution or the change in the exponent of the power-law. Particularly interesting is the case of the Italian Airport Network in [50], since the authors have suggested that the network is a “fractal scale-free network” by looking at the exponents of its Double-Pareto distributions, resulting in an increased robustness against targeted directed attacks to hubs with respect to simple scale-free networks. In [49] the authors studied the efficiency of the sub-clusters (defined by an hub and its neighbors) of the Chinese Airport Network, besides its trivial topological features. They found that the efficiency of these sub-clusters increased with the density of links within it.

Another feature that national airport network exhibits is disassortativity. This is a signature of the hub’n’spoke structure of the network, i.e. highly connected airports are usually connected with less connected ones. Surprisingly, this effect is not present when considering the World-Wide Network, since in [36] it has been found to be assortative. Moreover the authors studied the correlations between degree and clustering coefficient, highlighting the presence of a rich club phenomenon so that high degree airports tend to form cliques with equal or higher degree airports.

Besides static measures the analyses have also focused on the dynamical evolution of the airport network in several airspaces. In [54] the evolution of the US airport from 1976 to 2005 network was performed in order to analyze its scaling behavior in time. They found that the rate of growth of single airports deviates from the linear dependence with traffic shares in the network due to the presence of capacity constraints. Although aggregating the airports that serve the same region considerably reduced this effect, making the network scalable. This suggest that, as capacity constraints are reached, the growth and emergence of secondary airports will be triggered in order to maintain the scaling property typical of preferential attachment networks.

Many studies also focused on the shift between point-to-point to hub’n’spoke due to deregulations of the air traffic in several countries. As we said before hub’n’spoke is a structure in which hubs are connected to low degree nodes (the spokes), so that the network is a collection of star-like structure. Passengers usually fly from a spoke and a hub and have to take at least two flights to reach their destination. This structure presents several benefits for the airlines (e.g., higher occupation rate of flights, just one connection is needed to reach new airports etc.) and it is the structure used usually by major airlines. In point-to-point structure on the other hand a single aircraft connects two different airports if the
demand is high enough to justify the presence of the route. This means that an higher number of aircraft is needed in order to satisfy the demand. This structure is still used by low cost companies, but it was common to all the companies before the deregulations occurred in the '70s in the USA and in the '80s in Europe. The shift between these types of networks was studied in [55,56] for the European Airspace, investigating the evolution of the nodes connectivity and the number of hubs. The analysis also identified different growth patterns, since it pointed out as intra-European traffic concentrated in medium size airports that became hubs specialized in internal traffic. On the other hand traffic coming from outside the continent was directed towards big airports, creating a new type of hubs for inter-continental travels. Recently these studies have been extended also to the Chinese [57] and Brazilian Airspaces [58], confirming that deregulations lead to a hub’n’spoke structure.

As we said before, centrality studies in transportation networks are frequent when considering urban systems. Many of these studies focus on highlighting the importance of certain areas or on the capability of centrality measures to predict traffic flows. Studies of centrality have been very important also in the study of the ATM system. In [59] the authors studied the degree, closeness and betweenness centrality of the Chinese Airport Network. They found that all these centrality metrics were correlated with other socio-economical metrics (air passenger volume, population and gross regional domestic product). In particular they found that traffic grows with degree and closeness as a power-law, while showed a linear growth with betweenness centrality. In [37] an extensive study of the World-Wide Airport Network is presented for the first time. Besides its well-known properties, the network have some highly central but scarcely connected nodes. They explained the occurrence of these nodes as the result of local community structure of the network resulting from geo-political constraints. Finally they proposed a method for classifying the nodes based on their connections inside and outside their community. “Peripheral” nodes, i.e. nodes connected mostly with other nodes in the community, were the most common ones, while the nodes connecting different communities were usually hubs within their community (but not necessarily globally). Centrality measures are also connected to the resilience of a network [60,61], i.e. the capability of the network to function also under disruptions. The typical framework for resilience studies is percolation theory [62]. A fraction of nodes is removed from the network and its resilience is studied by looking at its structural responses as function of this fraction. Centrality measures come into play when the removal (or “attack strategy”) has to be defined. In fact the
attack can be random, directed to high degree nodes or to the most central ones. Scale-free networks are particularly resilient to random attack, while attacking central nodes seriously compromises its functioning. These approach has been used also in ATM for the Australian [63] and the US [64] airport networks. In [65] however has been pointed out how the percolation threshold for the airport network, that identifies the fraction of nodes to be removed in order to highly affect its structural properties, is well beyond possible real world perturbations. Thus the authors proposed a new method for resilience studies in transport network based on shortest-path trees.

The recent eruption of the Eyjafjallajokull volcano has been a remarkable opportunity to study the resilience of the ATM to major disruptions. In [66] a model has been developed in order to study the resilience of the European Airport Network when disruptions similar to those of the volcanic ashes occur. They found that, despite scale-free network like the airport network are resilient to random disruptions, when such disruptions are spatially coherent like the volcanic ashes, the tolerance of the system is considerably smaller. In [67] the temporal evolution of the European ATM system during the volcanic ashes crisis has been studied, considering three different level of aggregation of the system corresponding to three different types of network: the airports network, the bipartite flight-airway network and the bipartite flight-navpoint network.

The studies presented at this point mainly concerned the same approach when dealing with complex networks applied in ATM, that is the study of the airport network and its features. Despite the interesting results and the importance of such network in many real world applications (e.g. epidemic spreading [17]), this is not the only possible approach. For example the same network can be seen as a multiplex network [68], each layer of the multiplex being made by a different air company. In [68] it has been shown how the global properties of the network emerges as more layers are aggregated together. Properties like the rich club effect and small worldness are the result of the interplay between low-cost and major company layers.

In [69] many other layers of the ATM system were studied together with their community structure. Besides the European airport network, the network of sectors and the navigation point network of the whole European Airspace was studied. In particular the latter will be central in the following thesis as it will be a fundamental part of the analysis and modeling parts (see 1.3.1 and 1.4 for more details). Using two different algorithm in order to identify communities, the authors proposed a new method to inform the design of the airspace, comparing the resulting communities with pre-existent structures like sectors,
national airspaces and functional airblocks.

A completely different usage of complex networks has been proposed in [70]. In this work networks of safety adverse occurrences recorded in the Italian Airspace has been build. These occurrences are “Short-Term Conflict Alerts” (STCA), alarms triggered when a possible conflict between two aircraft is automatically detected in order to alert air traffic controllers during their management activity. These networks of STCA shows statistical regularities and seem to propagate through the airspace.

Another attempt to use complex networks in air traffic safety studies is [71]. By analyzing a dataset of aircraft trajectories with a small temporal resolution the authors studied the occurrence of conflicts and avoided conflict situations. Using a network representation of the traffic pattern before the occurrence or the avoidance of the conflict and used complex network metrics in order to find features that can discriminate between the two cases. Finally in [72] complex networks are used in order to measure the complexity of air traffic pattern, i.e. to quantify the level of effort needed by air traffic control to manage it.


1.1.2 Complex Networks Modeling in ATM

Complex Systems Physics has devoted much attention to the study of transportation systems. These systems are usually composed by self-driven particles, such as pedestrians and cars on a freeway [73, 74] or Internet traffic [75]. The investigations of these systems usually concerned the emergence of collective physical phenomena like phase transition [76, 77] and criticality [78].

Probably one of the first and more significant example of complex systems modeling applied to vehicular traffic is the Nagel-Shreckenberg model [79, 80], a cellular automata model describing the dynamics of car on a single lane. In this model cars move on a lane divided in cells, jumping from a cell to another one. Simple local rules are defined in order to describe the behavior of each driver and its interaction with the others. The model shows how congestion can emerge spontaneously without external interferences as the traffic increases. In fact, as the density of cars grows, the system undergoes to a phase transition from a free-flow phase to a congested phase where the average speed decreases until all the cars are stuck in their position. Many generalizations of these model as been proposed in order to explain how other traffic phenomena as traveling jam waves emerges at a global level [81].

Pedestrian dynamics has also been studied within this framework. The most successful mode describing pedestrian behavior is the social force model [82] where the interaction between the pedestrians is described by non-newtonian forces. Despite its simplicity the model is able to explain many emergent phenomena like the formation of ordered lanes of pedestrians or panic effects, i.e. jamming of pedestrians driven in opposite direction subject to a certain level of noise [83].

Another type of traffic dynamics that has been studied is the dynamics of Internet Traffic. For this kind of system a similar transition from a free state to a congested phase has been observed [75] and also analogies with free-way traffic has been proposed [84]. The main difference between this kind of dynamics and those of car and pedestrians is the role that the topology of the network over with the dynamics takes place [85]. In fact the traffic dynamics of the Internet is strongly coupled to the underlying topology and this effect is common to many other physical systems (e.g. the epidemic threshold in epidemic spreading is strongly affected by the topology [15]).

Similar studies concerning air traffic are somehow lacking due to the unavailability of data and just in very recent times the problem has started being tackled by the complex systems community. Similarly to the Internet, also the ATM system is highly influenced by
1.1. INTRODUCTION

its topology. Many studies do not deal with the actual traffic dynamics over the network, but how from the traffic demand and the geographical constraints the world-wide network topology emerges. In [35] the authors developed a growing network model with an attachment rule so that the nodes with higher strength are more likely to be linked by the new nodes. Although the attachment probability also depends on the distance between the nodes in an embedding space, so that too far connections are costly. Finally the model has a weight dynamics expressing the local perturbation produced by adding a new link in the attachment process. The model is able to reproduce many correlations and features observed in the North-American Air Transport network like non-linear correlations between traffic and topological features. An interesting model has been proposed in order to reproduce the microscopic dynamics of the links in the US Airport Network in [86]. This network exhibits an intense dynamics at a microscopic level, while macroscopic distributions are stationary. Traffic changes over the links of the network are modeled as random changes in their weight, while rules depending on these weights are present in order to suppress or create other links.

The previous examples are clearly more general than the simple application to air traffic systems, since they can be adapted to other similar problems. An example more oriented to ATM studies is [87] in which scheduled network are introduced. A scheduled network is a standard airport network with secondary nodes added on the links. The number of secondary nodes on a link is proportional to the travel time of the route associated to it. These networks allow to take properly into account the indirect connectivity between airport and many standard network metrics (mean path length, giant component, clustering coefficient, and tolerance to errors and attacks) can be adapted to their new formalism. This formalism has been applied to the study of the European Airport network finding that the period of least efficiency is during nighttime when travelers have to wait for long times for some connections to be active.

The first and probably the only one example of study concerning jamming transition in air traffic is [88]. The model describes a set of interconnected airports (initially they are nodes in a random network but European Airport Network has also been considered). Aircraft diffuse from an airport to the other and the weight of the link connecting these two airports is the travel time. Each airport has an assigned value of capacity that is perturbed by random noise, in order to simulate airport operational problems and adverse weather conditions. A set of queuing rules describe the traffic management activities at the airports, trying to regulate the incoming and outcoming aircraft flows. These set of
rules allow the aircraft to diffuse freely if the capacity limits are not exceeded, while if
the input flow becomes larger that such limits some aircraft gets stuck in a queue and can
be released just when capacity limits are satisfied again. Similarly to what happens in
the Nagel-Shreckenberg model, the system undergoes a jamming transition as the aircraft
density increases above a certain value. In this case the observable associated with this
transition is the percentage of aircraft that are not stuck in a query. This quantity is
equal to 1 for small density values, while decreases rapidly to 0 in the congested phase.
A similar approach has been proposed in [89] to explain the role of delay propagation
in airport congestion. Following a data-driven approach, the authors propose a model
that describes aircraft flying on an airport network according to their “rotation”, i.e. the
sequence of airports that each aircraft connects during the operations. Moreover other
two independent processes take place: crew connectivity, i.e. the crew has a certain prob-
ability to connect with another flight within a time range of 3 hours from the scheduled
arrival of their flight, and airport capacity, i.e. the maximum number of aircraft per hour
that an airport can manage without producing further delays. The model showed a good
agreement with experimental data, highlighting how macroscopic congestions can arise
also during normal operations without the influence of external events. This model has
also been applied in [90] to the study of the congestion produced by the 2010 “Super-
storm” that took place in the USA.
As the complex network analyses, also the modeling in ATM mainly concerns with the
airport layer of the systems. In all the previous studies aircraft were diffusing on routes
represented by links of the network, while the underlying structure of the airspace was
completely disregarded. To the best of our knowledge, the only work so far concerning
this aspect in the field of complex systems is [91]. The model presented has two layers
called the strategic and tactical layers. The strategic layer models the interaction between
Airline Operators during the planning phase of the trajectories of the aircraft. Airline Op-
erator are competing to get the best slot and trajectory for their flight, limited by airspace
capacity constraints. The other agent of the model is the network manager whose duty is
to guarantee safety forbidding the violation of those constraints. The simulations are per-
formed over a synthetic network of sectors with an assigned capacity, i.e. the maximum
number of aircraft per hour that are allowed to travel inside it (see chapter 1.2.1 for a
definition of sectors). Airline Operators produce a series of trajectories with an assigned
departure time and rank them according to a cost function that modulates between the
length of the path and the difference between the proposed departure time and the one
that the Airline desires. The Network Manager picks Airline Operators in a random order and for each one tries to fill the airspace with the proposed trajectories (following the ranking proposed by Airline Operators). If a capacity is exceeded by a trajectory, the Network Manager rejects it and tries the next one in the ranking order. Introducing a metrics of satisfaction of each airline operator, measuring how far the accepted trajectory is far from the best possible one, the overall satisfaction has been studied using waves of flights, a typical organization of departing traffic in the European airspace [92]. The tactical layer of the model concerns the management of the air traffic during when aircraft are effectively flying. The agents of the model are Air Traffic Controllers (see chapter 1.2.1), responsible of providing safety standards between the aircraft flying inside a sector. At each time step of simulation, the controller checks for possible conflicts between trajectories (meaning that two aircraft are too close from one another) within a certain lookahead time. If conflicts are found the conflicted aircraft are redirected according to some rules. External disturbances, like adverse weather conditions are also included in the model as shocked areas that the aircraft cannot cross. The simulations are performed within a single sector of the Italian Airspace and the effects of the shocked areas and the interplay with rerouting strategies is studied. The tactical layer of [91] is the model which is closely related to the topic of this thesis, despite the fact that its application is up to now limited to a very small portion of the airspace.

1.2 The Dataset and ATM Structure

1.2.1 Current Airspace Structure and the SESAR Scenario

The current European Airspace Structure is organized in a hierarchical way. At the top of the hierarchy are the National Airspaces, corresponding more or less to the border of the European Nations, in which every National Service Provider can apply its own policy. It is up to the service provider to decide the costs of crossing the airspace (which usually depends on the time spent inside it) and to provide an efficient travel to the aircraft that are served. Each National Airspace in divided volumes managed by different Area Control Centers (ACC) which are centers responsible for the safety and the effective tactical management of all the traffic in their volume. At the bottom of the hierarchy of the airspace structure there are the sectors. These sectors are volumes in which every ACC is divided in order to facilitate the traffic management. Each sector is in fact under the
responsibility of a pair of Air Traffic Controllers, whose duty is to monitor and manage the traffic that is crossing the sector. Figure 1.1 panel (a) shows the sector structure of the Italian Airspace. Note that the sector structure could be more complex than the presented one, since in this case it looks like a bi-dimensional tessellation of the ACC while in general it could be the result of many joint 3-dimensional structures. Moreover the structure of the sector is dynamical, since a single sector could be split in two or more other sector due to management reasons before or during the operations.

In the current scenario the aircraft are not free to move in a fully 3-dimensional environment. From days to hours before the actual take-off of the flight, the Stakeholders negotiate with Service Providers a planned route to be followed by aircraft. This planned route, called flight plan, is basically a series of fixed geographical references called navigation points (navpoints) that the aircraft is supposed to cross at a certain time and at a certain altitude. The navpoints are usually placed along fixed routes that each flight must follow and the structure of these fixed routes are called airways (figure 1.1 panel (b) shows the navigation point in the Italian Airspace). However the flight plans are not only bounded to follow the airways, but must respect another fundamental constraint provided by the “capacity” of the crossed sectors. In order to provide the correct management of the traffic, the workload for the air traffic controllers must not be to high so that human errors are limited. The workload of a controller is measured by the number of aircraft that cross his sector every hour, so a threshold is set by the service providers for each sectors. This threshold indicates the maximum number of aircraft per hour that a sector is supposed to handle, based on the structure of the sector and the kind of traffic traveling inside. As stated before, the structure of the sectors is not static. If a certain sector exceeds its capacity limits, it is usually possible to split it in some sub-sectors so that in the end the same volume is managed by more controllers. This increases the total capacity of the sector, since if a single sector with a capacity of 40 aircraft per hour is split in two sub-sectors with the same capacity, the total capacity of the sector is doubled. Thus flight plans are also supposed to meet the capacity limits of the sectors that they are crossing in order to not exceed the maximum workload for the controllers. However the flight plans do not take into account two common occurrences: adverse weather conditions and possible conflicts with other aircraft trajectories. A conflict or loss of separation occurs when there is an infringement of the separation between one or more aircraft. This separation minima may vary from an airspace to another or between different parts of an airspace, but are usually defined as 5 Nautical Miles (1 NM= 1.852 km) in the horizontal direction.
1.2. THE DATASET AND ATM STRUCTURE

Figure 1.1: (a) Italian sector structure projected on the horizontal plane. (b) Navigation Points in the Italian Airspace (white points). Big red points are the airports within the airspace.

and 1000 feet (1 feet ≈ 0.3 m) in the vertical direction. If a pair of aircraft are closer than these two thresholds in both the directions a conflict occurs. A conflict is a dangerous situation that must always be avoided since it could result in a collision between the aircraft
involved. It is the duty of the ATC to provide separation, spotting possible conflict and redirecting the aircraft making them deviate, horizontally or vertically, from their flight plan. The existence of the airways is an heritage from the past that was useful to ATC in order to manage and reroute the traffic easily. Thanks to modern instruments, like radar tracks, nowadays controllers are capable of rerouting an aircraft completely disregarding the underlying airways structure if it is convenient in terms of fuel consumption (i.e. if it shortens the trajectory). Note that the current ATC system is very efficient in preventing this kind of occurrences and, according to the Eurocontrol annual report of 2012 [6], no registered occurrence has been due to an error of the air traffic control itself in Europe during 2011.

Within 2015 an increase of traffic load is expected within the European Airspace, driven by the emergence of new economies outside the European Area [4]. Such increase of traffic could lead the current ATM system to exceed its capacity limits so that safety and efficiency could not be guaranteed anymore to all the flights. For this reason EUROC-ONTROL launched a programme called SEASAR that aims at improving the overall productivity of the European ATM system. One of the key concept of the SESAR program is the business-trajectory scenario. In this possible future scenario each aircraft will fly in a less-structured airspace, i.e. the airways structure will be disregarded during the planning and the tactical phases [1]. Consequently, the aircraft will be able to fly along a fully 4-dimensional trajectory, exploiting all the available airspace. Such trajectory will be built according to the need of costs and efficiency of the owner of the aircraft and moreover it will be safer and easier to manage for the ATC. Thus this shift from the current system to the SESAR scenario should enhance the overall performances of the ATM system and also the safety and resilience standards by a global rearrangement of the airspace and routes structure.

1.2.2 The Demand Data Repository Dataset

In order to gain basic information about the deployment of the air traffic within the European Airspace, we analyzed historical data coming from the Demand Data Repository (DDR) [93] developed by EUROCONTROL. This data has been organized into a database within the ELSA Project [94], also financed by EUROCONTROL, that can be queried in order to extract information about a particular time frame or a particular area of the European Airspace.

In particular we analyzed data concerning all the flight in Europe in a time frame going
1.2. THE DATASET AND ATM STRUCTURE

from March to August 2011. The dataset covers all the kinds of flight, including those from civil and commercial aviation to military flights. The information stored by DDR includes many aspects of each single flight, such as the arrival and destination airport, the departure time, the requested flight level, and also technical information like the type of aircraft and the number of people in the crew. Since it is possible to identify the kind of each flight this information has been used to filter out all the military flights, since they do not follow the rules of civil aviation. Moreover we filtered out all the aircraft with the same departure and arrival airport and those with a required flight level, i.e. the altitude at which the aircraft is supposed to fly, below a certain threshold. Since the minimum en-route altitude (the minimum “safe” altitude that each flight must have to be considered in its en-route) is usually 2000 ft, all the flight with a required flight level below this level have been excluded from the dataset. At the end of this simple filtering process, we end with a total number of 345490 flights in 14 days in the whole European airspace. The number of aircraft per day is quite constant and close to the average value of 24678 flights per day.

Beside these basic information, our dataset stores also the last filed flight plan and the radar updated trajectory for each flight. Last filed flight plans are submitted by the air companies to the Central Flow Management Unit from six months to an hour before the departure of the flight and are the result of the planning process and agreement between these two actors. Instead, the radar updated trajectories are the trajectories that aircraft did actually flow after the take off. Both the last filed flight plan and the radar updated trajectory consist of a sequence of geographical coordinates with the time and the height (i.e. the flight level) at which the aircraft is supposed to cross or did really cross them. Once again, we stress that the last filed flight plans usually do not coincide with the actual trajectory flown, since the ATC is likely to apply deviations and flight level changes for safety and management reasons. The geographical points that compose both the flight plans and the radar updated trajectories could be either fixed navigation points (usually identified by a 5 letter name) or a temporary point that are not part of the airways structure but were used during the flight. Temporary points are uncommon in the flight plans trajectories but their number is large in the real trajectories, reflecting the fact that aircraft do not follow the airways in their flight. Using the sequence of navpoints and temporary points it is possible to reconstruct for each flight the 4-dimensional trajectory of its flight plan and the 4-dimensional trajectory that it has effectively traveled.

The data delivered by EUROCONTROL are characterized by a variable temporal sample
rate of aircraft trajectories, with an average value of about two minutes. Figure 1.2 shows the distribution of the sample rates for the all the flight plans and radar updated trajectories in the whole European Airspace on the 9\textsuperscript{th} of June 2011. Despite the presence of a fat tail, both distributions are highly peaked around a value of about 120 seconds with an average value of about 165 seconds. Since, as a matter of definition, a conflict occurs when at least two aircraft are closer than 5 NM in the horizontal direction and 1000 feet in the vertical direction, the typical time duration of a conflict is about 1 minute considering that the average speed of an aircraft with respect to the ground is about 800 km/h. Thus, the temporal resolution of the data set is not sufficient to spot the possible losses of separation between couples of aircraft, which can be expected to be many in the planned data and very few in the radar updated trajectories. Moreover, we do not have information about the meteorological conditions and the possible turbulences of the days of the dataset. Thus, while we are able to see the deviations produced by the ATC over the single trajectories, it is not possible to extract from the data the causes of such deviations. In 1.2.1 we mentioned that the structure of the airspace is dynamical

![Figure 1.2: Distribution of the temporal distance between two consecutive sampled points of a trajectory for the flight plans and the radar updated trajectories on the 9\textsuperscript{th} of June 2011 in the whole European Airspace.](image)
1.2. THE DATASET AND ATM STRUCTURE

and hierarchically organize. The dataset also contains files regarding the structure of the airspace in the corresponding days of the DDR data. In these file can be found the structure of the airways and of the sectors inside national airspaces, i.e. their boundaries in the horizontal and vertical directions. Moreover, the opening scheme of sectors is present, so in principle it is possible to reconstruct whether a sector has been split or closed due to traffic management reasons. In the following, we will completely disregard this possibility using this data in order to build a static sector structure that could be used in simulations. Figure 1.1 shows the reconstructed sector structure for the Italian Airspace. Note that since the sector splitting is performed in the vertical direction, we decided to project each sector in the horizontal direction, building a bi-dimensional static sector structure.

1.2.3 ASMT Safety Dataset

In the past 15 year, EUROCONTROL has developed the Automatic Safety Monitoring Tool (ASMT) [95,96], one of the most advanced tools for Automatic Safety Data Gathering (ASDG) of the ATM system. ASMT can be connected to the operational ATM system in an on-line or off-line mode to elaborate, in quasi real-time, data on radar tracks, flight plans and system alerts. It automatically detects operational and technical occurrences according to user-defined parameters. ASMT detects events through the computation of the current air traffic situation, continuously updated from the tracks and flight plan inputs. In the ELSA project, ASMT was used to provide data on two types of safety events, i.e. losses of separation (named Proximity in ASMT) and Short-Term Conflict Alerts (STCA). STCAs are alarm triggered when a possible loss of separation is automatically predicted within a look ahead time of 2 min, aiding the ATC to spot and solve possible conflicts.

For each detected occurrence, ASMT stores the relevant data (shortly before, during and shortly after the event), which can be later queried to extract the data or to review the occurrence in a dedicated replay window. The unique opportunity offered by ASMT is to have an objective and comprehensive safety data collection, which would not be feasible just by relying on manual voluntary report (which is currently the largest source of safety data in the European states).

In collaboration with the ELSA Project, we had access to a dataset of STCAs collected from March to August 2011 in the Rome Area Control Center. 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:

1. ASMT filtering: events are filtered out using ASMT parameters and filters. For instance, these 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.

2. 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 (e.g. an event is excluded if the two aircraft involved have the same ID, meaning that it is a nuisance event due to a double track created by the radar).

3. Matching with the traffic data: data coming from ASMT is anonymized, meaning that it is not possible to identify the aircraft involved in the occurrence. Thus we used the traffic data to match each event with a couple of trajectories from the database and filtered out all the unmatched events.

We will not go much further in the details of the filtering (1) and (2). At the end of these two procedures we had 983 STCAs events left in the dataset. Since filtering (3) is useful in order to find correlations between network metrics and the occurrence of these events, we will discuss the details in par. 1.3.3.
1.3 Network and Delays Analysis

In this section we will present the analysis performed with the traffic and safety dataset. In particular the traffic dataset has been analyzed in order to provide input and validation results for an agent based model of ATC presented and discussed in chapter 1.4.

1.3.1 The Navigation Point Networks

There are many possible network representations of an airspace due to the multiple layers that compose its structure. Though the most natural way to study the properties and the deployment of the traffic through the airspace is by using the Navigation Point Networks (also Navpoint Network). The topology of such network has been already studied in the case of the Chinese Airspace [97] and recently in the European Airspace [69].

This network is built by using the fixed geographical references as nodes and by linking them if at least an aircraft flew from one to another in the considered period of time.

The result of this simple process is a geographical network, whose links can be weighted either with the amount of traffic, i.e. the number of aircraft that has traveled over the link in the considered time frame, or the geographical distance between the nodes that they connect. Since we have two sorts of trajectories, it is possible to build two different networks: a planned navigation point network, built with the last filed flight plans, and a real navigation point network built with the actual radar updated trajectories. The information of carried by these two networks are somehow complementary. Last filed flight plan are usually built following the structure of the existing airways, so the planned navpoint network will reflect the current airways structure and topology. On the other hand, ATC is not bounded to follow the airways in their activity and the deviations produced by it will modify the structure of the routes. In this sense the topological variations between the planned and the real navpoint networks can be studied in order to understand the action of the controllers in an aggregated way an not just considering the single trajectories.

For every navigation point network we build, we will focus our attention over the following nodes and links metrics:

- The degree $k$ of a node, i.e. the number of nodes connected to the considered one.

- The strength $s$ of a node, i.e. the sum of the weights of the links connected to a node. Note that for this metric we will weight the links using the traffic. In this
sense the strength of a node will always represent its traffic load.

- The betweenness centrality $b$ of a node, a well known centrality measure. For a node $n$ it is defined in (1.9). For the calculation of the shortest-paths we will weight the links with the geographical distance, so that a shortest-path on the network will be a shortest-path in the embedding space of the network.

- Amount of traffic over a link $w$, i.e. the traffic load over a link,

- The geographical length of a link $d$.

The planned and real networks have many similar qualitative features. Considering the networks built using the whole European Airspace and a time frame of a week, these two network have similar topological features like an exponential-like degree distribution (Fig. 1.3), probably resulting from the spatial embedding of the network [33], an exponential-like distribution of the strength of the nodes (figure 1.3) and of betweenness centrality (Fig. 1.3). From these distributions emerges that the real navigation point network is more homogeneous than the planned one, i.e. the action of the controllers reduces the number of nodes with high degree, strength and betweenness, creating more nodes with low values of these metrics. Another similar feature is the shape of the distribution of the weights of the links (Fig. 1.5) that is a power-law with an exponent close to $-1.3$ and an exponential cut-off. Despite this variations the values of strength for the corresponding nodes in the planned and real network are well correlated, indicating that the action of the ATC does not produce large changes in how the traffic is deployed over the network (Fig. 1.4).

The two networks have about the same number of nodes, 13528 nodes in the planned and 14951 nodes in the real network considering the airports as nodes. However the real network has a number of links that is more than doubled with respect to the planned case, 44052 links in the planned and 116879 in the real case. This is a consequence of the action of the controllers that redirecting the aircraft create new connections in the network. Moreover as can be seen in figure 1.5 the new links are usually longer than those present in the planned case, since longer connection are better in order to shorten the trajectories. The fact that we have many links longer than the typical dimension of a sector ($100 NM$) indicates that some times aircraft are directed from their current sector to another one without passing through any navigation point. This is a common action performed by the controllers called “direct” that is used to lower the traffic load inside the sector sending some aircraft away and thus easing their monitoring activity.
1.3. NETWORK AND DELAYS ANALYSIS

Figure 1.3: Cumulative distributions of the degree (a), the strength (b) and the betweenness centrality (c) of the nodes of the planned (red) and real (black) European Navopoint Networks.

Another interesting feature of the real navigation point network is the negative correlation between the length of the links and their traffic load. Figure 1.5 shows a histogram of the scatter-plot between these two quantities. As the distance increases the binned traffic load starts to decrease, meaning that longer links are harder to be traveled because they are more likely to intersect many trajectories and thus generating conflicts. The qualitative features observed in the whole European Navpoint Networks can also be found considering the networks restricted to national airspaces. Figure 1.6 presents all the previously seen distributions for some national planned route networks of different sizes. Note that, when considering the planned navpoint network of a nation, since the distribution of the degree
is exponential-like and the distribution of the lengths of the links is peaked around a value of 20 NM, the network can be considered a grid in a first approximation. The total surface of a grid with length of the links equal to $a$ grows as $A \sim a^2N$, where $N$ is the number of nodes of the grid and the coefficient of the linear relation depends on the geometry of the grid. If we consider the area of a certain airspace, we find that its area grows with a similar law, i.e. $A \sim \langle d \rangle^2N$, where $\langle d \rangle$ is the average length of the links of the planned navigation point network inside the airspace and $N$ is the number of its navigation points (Fig. 1.7). This is a very important result for the simulations, since we can consider every single planned navigation point network as a different realization of the same network whose size corresponds to the number of nodes. In this sense we will be able to simulate an increasing size of the system just by considering national airspaces with an increasing number of navigation point.

Figure 1.4: Correlations between the strength of the nodes in the planned European Navpoint Network and the corresponding values of strength in the real European Navpoint Network.
1.3. NETWORK AND DELAYS ANALYSIS

Figure 1.5: Distributions of the weights (a) and the lengths (b) of the links of the planned (red) and real (black) European Navpoint Networks. (c) Binning of the scatter-plot between the length and the weights of the links for the real European Navigation Point Network.
Figure 1.6: Cumulative distributions of the degree (a), the strength (b) and the between-ness centrality (c) of the nodes of the planned navpoint networks for some European National Airspaces.

Figure 1.7: Total area of national airspaces as a function of $\langle d \rangle^2 N$. 
1.3. NETWORK AND DELAYS ANALYSIS

1.3.2 Delays

Flight delays are one of the primary concerns of the various actors involved in ATM. On one side, delays can damage companies by increasing their costs and can make them miss arrival or departure slots at the airports, on the other, delays can create congestions and unbalances of the traffic load, which controllers have to cope with, and increase the emissions of CO$_2$ due to attempts to recover from them. With the data set at our disposal, we can measure the delay $\delta t_{\text{tot}}$ of each flight as the difference between the arrival time in the flight plan and in the radar updated trajectory, i.e. the real experienced delay. We can also measure the departure delay $\delta t_{\text{dep}}$ as the difference between the planned departure times and the actual one. We wish to add here that the last filed flight plans may be rescheduled shortly before the departure in response to ground generated delay. Unfortunately, our data set does not allow to detect this possibility so that some departure delays occurred in reality might not be deduced from the data set. Thus we will disregard this possibility and consider the departure delay we can measure as the only possible. Note that it is possible to define a departure delay also considering the trajectories inside a single nation. In this case if a flight does not depart from an airport inside the airspace and thus comes from outside the airspace, the departure delay is measured considering the difference between the planned and real airspace entering times.

We define the difference $\delta t_{\text{enr}} = \delta t_{\text{tot}} - \delta t_{\text{dep}}$ as the en-route delay, i.e. the delay generated between the take off and the arrival phase of the aircraft. The histogram of $\delta t_{\text{enr}}$ displayed in figure 1.8 panel (a) shows the action of the controllers on the flights. In case of no safety infringements or other external events, this histogram would be highly peaked around the null value. The distribution of en-route and departure delays are quite robust in time and space, meaning that in different nations and different day their qualitative features are almost the same. Figure 1.8 panels (b) and (c) shows the distributions of these delays for a week of flights in the Italian Airspace. From one day to another small differences can be observed just in the departure delays distribution, but the en-route one is almost stable. Considering instead different nations (1.8 panels (d) and (e)), it is evident that while the distribution of $\delta t_{\text{dep}}$ is again stable, the amplitude of the distribution of $\delta t_{\text{enr}}$ depends on the size of the nation. These two quantities are uncorrelated in every day of the dataset (the correlation coefficient is always smaller than 0.05), indicating that the delay acquired before the take-off is not managed by the Air Traffic Control (figure 1.9). This information may be surprising, but usually controllers completely disregard the delay of a flight when they have to decide to shorten its trajectory or not. All the redirections are applied just
Figure 1.8: (a) Comparison between the total, en-route and departure delays in the Italian Airspace on the 9th of June 2011. (b) and (c) Comparison between the en-route and departure delays in different days of the dataset in the Italian Airspace. (d) and (e) En-route and Departure delays distribution for several European Airspaces on the 9th of June 2011.
for safety reasons or to reduce their workload. Again, this feature is constant in almost every European National Airspaces (table 1.1 shows the correlations between all the kinds of delay for some airspaces).

<table>
<thead>
<tr>
<th></th>
<th>Italy</th>
<th>Spain</th>
<th>Germany</th>
<th>Greece</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho(\delta t_{dep}, \delta t_{tot})$</td>
<td>0.96</td>
<td>0.90</td>
<td>0.96</td>
<td>0.91</td>
<td>0.97</td>
</tr>
<tr>
<td>$\rho(\delta t_{enr}, \delta t_{tot})$</td>
<td>0.37</td>
<td>0.50</td>
<td>0.37</td>
<td>0.31</td>
<td>0.32</td>
</tr>
<tr>
<td>$\rho(\delta t_{dep}, \delta t_{enr})$</td>
<td>0.05</td>
<td>0.09</td>
<td>0.11</td>
<td>0.03</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Table 1.1: Correlation coefficients between $\delta t_{dep}$, $\delta t_{enr}$ and $\delta t_{tot}$ for several European Airspaces.
Figure 1.9: Correlations between $\delta t_{\text{dep}}$ and $\delta t_{\text{tot}}$ (a), $\delta t_{\text{enr}}$ and $\delta t_{\text{tot}}$ (b), $\delta t_{\text{dep}}$ and $\delta t_{\text{enr}}$ (c). These delays are those of all the flight that have crossed the Italian Airspace on the 9th of June 2011.

1.3.3 Short-Term Conflict Alerts and Network Metrics

In 1.2.3 we introduce a safety dataset, containing Short-Term Conflict Alert events. Information about the occurrence time of the event, the position and altitude of the two aircraft involved and the duration of the event was stored. However the dataset has been anonymized so that it is not possible to match automatically each event to a couple of trajectories in the database. Thus we developed a simple algorithm in order to do the matching and we tested it to a very small dataset of just 4 STCAs in the Romanian Airspace [98]. Note that all the STCAs in our dataset, including those in the Romanian Airspace, have occurred at high altitudes so they can be considered as en-route occurrences, i.e. the aircraft involved were not climbing or descending towards an airport at
the occurrence time (Fig. 1.10 panel (c))

**STCA Matching and Filtering**

Let’s call \( A \) and \( B \) the two aircraft involved in an STCA event recorded by ASMT. Their respective position will be \((\text{lat}_A, \text{lon}_A)\) and \((\text{lat}_B, \text{lon}_B)\) and their respective flight levels (i.e. the altitude at which they are flying) \( FL_A \) and \( FL_B \). The occurrence time of the event is indicated by \( t_{\text{occ}} \).

We start considering the set of all the radar updated trajectories within the considered airspace in the day of occurrence of the event and one of the two aircraft involved, say \( A \). We exclude from the set all the trajectories not containing at least a segment with \( t_1 \leq t_{\text{occ}} \leq t_2 \) and \( FL_1 + 10 \text{ ft} \leq FL_A \leq FL_2 - 10 \text{ ft} \), where \( t_1 \) and \( FL_1 \) are the time and the flight level at the beginning of the segment and \( t_2 \) and \( FL_2 \) are the time and the flight level at the end of the segment (provided that \( FL_1 \leq FL_2 \)). The \( \pm 10 \text{ ft} \) in the condition for the flight levels is included in order to not filter out cruise segments with small deviations from the flight level of the event (e.g. an aircraft cruising at \( FL = 3500 \text{ ft} \) will not be excluded from an event with \( FL_A = 3490 \)). Being \( \text{set}_A \) the set of the non-excluded trajectories for the aircraft \( A \) and \( \text{set}_B \) the corresponding set of non-excluded trajectories for the aircraft \( B \), for each trajectory in these two sets are recorded the segments fulfilling the constraints above and the distance from the coordinates of the corresponding event. These distances are computed interpolating the coordinates of the trajectory at \( t_{\text{occ}} \) using the coordinates at the border of the recorded segment. The interpolation is performed assuming that the aircraft will fly at the same constant speed from through all the segment. Similarly the altitude at \( t_{\text{occ}} \) is linearly interpolated assuming a constant speed of climb or of descent from \( FL_1 \) to \( FL_2 \). Note that the assumption of constant speed travel might be realistic in the horizontal directions, since big speed changes are physically impossible during the en-route part of a flight. On the other hand vertical variations of 100 feet or more can occur rapidly or slowly depending on the needs of the ATC, so in this case the hypothesis is not completely valid.

We then compute a “distance measure” for all the trajectories in \( \text{set}_A \) and \( \text{set}_B \) from the corresponding recorded aircraft:

\[
\hat{d} = \alpha d_{\text{geod}}((\text{lat}, \text{lon}), (\text{lat}_A, \text{lon}_A)) + (1 - \alpha) |FL - FL_A|,
\]

where \( d_{\text{geod}}((\text{lat}, \text{lon}), (\text{lat}_A, \text{lon}_A)) \) is the geodesic distance between the interpolated position of the trajectory at \( t_{\text{occ}} \) and the recorded occurrence position of the aircraft involved
in the STCA, $FL$ is the interpolated altitude of the considered trajectory at $t_{occ}$ and
$\alpha \in [0, 1]$ is a parameter that modulates the importance of the vertical distance. From all
the distances of the trajectories in set A and set B , the smallest is taken and the callsign
of its trajectory is assigned to the event. From the set of trajectories of the unmatched
aircraft is taken again the one with smallest distance and it is associated.
The algorithm has been tested with the Romanian dataset of just 8 STCAs, being able
to match every one of them correctly if $\alpha = 0.5$. Moreover all the trajectories have been
matched within a maximum value of distance $\hat{d}$ of 15 NM. We will use this value as a
threshold, so that every matching with a value of $\hat{d}$ larger than the threshold is not pos-
sible. Since we do not have a control sample to use in order to test the efficiency of the
matching, we used $\alpha = 0.5$ to match all the STCAs in the ACC of Rome. The introduc-
tion of a threshold for $\hat{d}$ left many events with just one or no matching trajectories at all.
These events have also been filtered out as nuisance events. At the end of the process we
excluded 672 STCAs that were not filtered before and ended up with 750 STCAs matched
with 1307 distinct trajectories.
A clear check of the goodness of the method cannot be done due to lack of other dataset
with already identified aircraft. However plotting the events excluded with this method
on a map reveals that the vast majority of them occurred over the sea (Fig. 1.10 panel (b)). Some STCAs may be triggered by the creation of “ghost aircraft” by the reflection
of the radar signal over the surface of the sea, appearing to be close to the real ones.
Despite the fact that it is not a clear evidence of the correctness of the algorithm, the
large fraction of unmatched events placed over the sea suggests that many “double tracks”
(i.e. tracks doubled by the reflection on the surface of the sea) has been filtered out.
1.3. NETWORK AND DELAYS ANALYSIS

Figure 1.10: (a) and (b) plot on map of the matched and unmatched STCAs. The colors correspond to the occurrence altitude of the event. (c) Distribution of the occurrence altitudes of the matched and unmatched STCA events.

Criticality Measures of the Navpoint Network

At the end of the matching process we have for each event an associated couple of trajectories and the corresponding link that each aircraft was traveling when the event occurred. Thus we can use this information in order to study possible correlations between the features of the airways structure and the emergence of safety events. We built the planned navigation point network in the Rome ACC using all the flight plans.
in the whole period of STCA data gathering. This network represents the features of the airways structure within the ACC and the way in which the traffic is typically deployed over it. To assign the STCA events to the network we must take into account that an aircraft does not necessarily follow the airways during its travel. Considering each aircraft in each couple assigned to an STCA:

- If the link traveled at the occurrence time of the STCA is part of the planned navpoint network, the event is assigned to the corresponding link.

- If the segment is not part of the planned navpoint network, we compute the shortest-path over the network connecting the nodes at its extremities. We then compute the distance between the occurrence point of the event and each link in the shortest-path and assign the event to the closest link.

At the end of this simple procedure we end up with a new metric associated to each link of the network, the number of assigned STCA, $w_{STCA}$. Note that since we assign both aircraft of the couple involved in an event, each STCA is assigned twice to the network. The number of STCA assigned to a link is a measure of the criticality of the link itself, since it measures the frequency of occurrence of STCAs. Measuring the strength of a node with this weight gives a similar measure for the node, i.e. the frequency of occurrence of an STCA for a node in the network, $n_{STCA}$. The distributions of these two quantities are presented in figure 1.12, like many other features of the navigation point network also in this case we have exponential-like distributions.

These two measures of criticality are “extensive” quantities, meaning that they depend linearly on the number of aircraft that have flown through the considered node or link. Considering the latter, it is straightforward to define an “intensive” measure of criticality simply by rescaling with the weight $w$ of the link

$$P_{STCA}(e) = \frac{w_{STCA}(e)}{w(e)},$$  \hspace{1cm} (1.11)

where $P_{STCA}(e)$ is an estimate of the probability that an aircraft flying through the link $e$ incurs in an STCA with another flight. Using this metric we can define a similar measure $P_{STCA}(n)$ for the nodes of the navpoint network. Considering a node $n$ there are several ways in which an aircraft can pass through it (Figure 1.11). The total probability of incurring in an STCA traveling through $n$ must be the result of the probabilities of
1.3. NETWORK AND DELAYS ANALYSIS

incurring in an STCA traveling through a path $\gamma$ that crosses the node

$$P_{STCA}(n) = \sum_{\gamma} P(STCA|\gamma)P(\gamma).$$

(1.12)

A path $\gamma$ through the node must cross one of its in-going links and one of its out-going links. For sake of simplicity we will assume that the probability of traveling through a link depends only on the weight of the link itself and its uncorrelated with the previously traveled link (i.e. we are assuming that an aircraft travel over the network with a 1st order Markovian process). Thus the probability of traveling following the path $\gamma = (e_i, e_j)$ in figure 1.11 is given by

$$P(\gamma) = \frac{w(e_i) w(e_j)}{N_{tot}}$$

(1.13)

where $N_{tot}$ is the total number of aircraft that has flown through $n$ and $w(e_i)$ and $w(e_j)$ are the weights of the links $e_i$ and $e_j$. Similarly the conditional probability $P(STCA|\gamma)$ will be given by 1 minus the probability of not incurring in an STCA when flying through $\gamma$

$$P(STCA|\gamma) = 1 - (1 - P_{STCA}(e_i))(1 - P_{STCA}(e_j)).$$

(1.14)

Combining equation (1.13) and (1.14) into (1.12), we obtain the expression of the probability of incurring in an STCA when traveling through $n$

$$P_{STCA}(n) = \sum_{e_i, e_j} [P_{STCA}(e_i) + P_{STCA}(e_j) - P_{STCA}(e_i)P_{STCA}(e_j)] \frac{w(e_i) w(e_j)}{N_{tot}}$$

$$N_{tot}$$

(1.15)

The distributions of $P_{STCA}(n)$ and $P_{STCA}(e)$ are presented in figure 1.13. Note that while the extensive quantities have exponential-like distributions, in this case both the distributions of these metrics have power-law tails. The statistical test described in [99], based on the Kolmogorov-Smirnov statistics has confirmed this hypothesis, so $P_{STCA}(e)$ has a power-law tail with exponent $\simeq -1.7$ for values larger than 0.0014 and $P_{STCA}(n)$ has a power-law tail with an exponent $\simeq -2.74$ for values larger than 0.0045. The fact that these distributions are heterogeneous indicates that this quantity are the best suited in order to classify the nodes according to their criticality.

The main issue with this kind of measures is that they are based on the estimator of a frequency $p = \frac{n}{N}$ so their standard error decreases with $\sqrt{N}$. Considering $P_{STCA}(e)$, this means that the error decreases with the number of aircraft that have traveled through the link $e$, which in some cases could be very small. This could lead to very large errors for $P_{STCA}(e)$, propagating to the errors of $P_{STCA}(n)$, making these two estimate to be affected
by very large measurement errors. Figure 1.14 shows the distribution of the relative errors of $P_{STCA}(n)$ and $P_{STCA}(e)$.

Figure 1.11: Example of a path $\gamma$ crossing the node $n$ for the calculation of $P_{STCA}(n)$. Each path is a combination of an in-going link $e_i$ and an out-going link $e_j$.

Figure 1.12: Cumulative distributions of $w_{STCA}$ (a) and $n_{STCA}$ (b) for the Rome ACC Navpoint Network.
1.3. NETWORK AND DELAYS ANALYSIS

Figure 1.13: Cumulative distributions of $P_{STCA}(e)$ (a) and $P_{STCA}(n)$ (b) for the Rome ACC Navpoint Network. Dashed lines are guides to the eye for the power-law tails of the distributions. Note that since these are cumulative distributions the exponent of the distribution is the showed one minus 1.

Figure 1.14: Distributions of the relative errors for the measures of $P_{STCA}(e)$ (a) and $P_{STCA}(n)$ (b) for the Rome ACC Navpoint Network.

Criticality Measures and Network Metrics

The criticality measures of the links and the nodes of the Planned Navpoint Network might be relevant in identifying the critical areas of the airspace in order to take mea-
sures aimed at improving the overall safety of the ATM system. The choice of assigning the STCA to the planned navigation point network is not arbitrary since such network is more related to the airspace design than the real one. Although having information about occurrences such as STCAs or losses of separation might be difficult due to the fact that this kind of data are not always available, so we could ask if other network metrics are predictive in order to spot critical nodes or links. There are many examples in literature of metrics aimed at identifying the most important or critical nodes in complex networks. In [44] for example, the authors introduced a modified betweenness centrality capable with the aim to identify the most trafficked nodes (and thus the most likely congested ones) in a urban network.

Thus, we tried a similar approach in order to find correlations between some network metrics and the occurrence of STCAs. A strong correlation between STCA occurrences and some topological metrics of the airspace network could be an important indicator of airspace design criticality. Moreover it could also drive a proper redesign of the airspace topology to reduce the risk of STCA events. In our analysis we identified two classes of metrics both for the links and the nodes: traffic related metrics, i.e. computed using weighting the links with the traffic, and topology related metrics, i.e. computed weighting the links with the geodesic distance between the nodes that they connect and no information about the traffic deployment is needed. For the links the only traffic related metric used is the weight, while the topology related one is the link betweenness centrality defined as 1.9 using shortest-paths over the network. For the nodes the variety of possible metrics is larger and thus we took, for the traffic related ones:

- The strength of a node $s$, indicating the local traffic load over the node.
- The Random-Walk Betweenness Centrality defined in [100]. The idea was to generalize the usual betweenness centrality taking into account the fact that other paths beside the shortest ones could be relevant in order to rank the nodes of a network. This new betweenness centrality is defined as

$$b^{RW}(n) = \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} r_{ij}(n), \quad (1.16)$$

where $r_{ij}(n)$ is the probability that a biased random walk originated in $i$ will pass through $n$ before reaching its sink $j$. Note that the random-walk will jump from a node to another one with a probability proportional to the weight (traffic in our
1.3. NETWORK AND DELAYS ANALYSIS

case) connecting them. Thus for the navpoint network this metric measures the importance of a node in relation to the traffic flows over the network.

The topology related metrics are:

- The degree of a node, indicating its number of connections.
- The betweenness centrality, as defined in equation (1.9). Roughly this metric indicates how many airways are crossing over a node (assuming that the airways are always following the shortest-paths of the network).
- The Origin-Destination Betweenness Centrality which is a betweenness centrality computed taking into account the fact that not every node is the origin or the destination of an actual path taking place on the network. If the information about the actual starting and ending points of paths on the network is available, it is possible to build a Origin-Destination matrix $OD_{ij}$ whose elements are equal to 1 if there are real paths going from the node $i$ to the node $j$ and equal to 0 otherwise. Using this matrix, the centrality is defined as

$$b^{OD}(n) = \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} OD_{ij} \frac{\sigma(i,j|n)}{\sigma(i,j)},$$  

(1.17)

which is similar to (1.9) with the term $OD_{ij}$ that exclude the contribution of every couple of nodes $i$ and $j$ that are not origin and destination of real path over the network.

All these metrics are related to the static properties of the navigation point network, but it is also possible to define metrics that are related to the dynamics taking place over it during the air traffic control operations. Within the ELSA project some metrics of this kind have been developed comparing each flight plan with its corresponding radar updated trajectory [101]. Using Dynamic Time Warping [102] it has been possible align the planned and the real trajectory and to identify the nodes where deviations started and ended in order and to assign each deviation to a node of the network. Thus several dynamical metrics have been defined:

- 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. Basically it is a measure of the intensity of the vertical deviations occurring on this node.

• **Delay**: amount of en-route delay generated by the point (this metric can be divided in *Positive Delay* and *Negative Delay*).

These metrics are useful in order to gain information about the typical dynamics correlated with the occurrence of safety events, e.g. if they happen mostly in nodes with a large number of horizontal movements or vertical movements.

**Link Metrics Correlations**  Considering the criticality measures, it is evident that node metrics are an aggregation of the metrics defined for links. This is particularly important in our case since the paucity of safety data due to their low occurrence ratio may prevent some patterns from being clear. Thus considering link metrics, we expect to see small correlations that are probably due to this kind of effect (that should be larger when considering $P_{STCA}(e)$ because of its large estimate errors).

When considering the correlations with $w_{STCA}$, weak correlations are present with both the betweenness centrality and the weights (Figure 1.15). Although the fact that these correlations are weak suggests that these metrics are not well-suited to identify links with a high frequency of occurrence of STCAs.

The situation is worse when considering $P_{STCA}(e)$, since in this case the Pearson Correlation Coefficient is close to 0 for both the correlations with weights and betweenness. Moreover the $p$-value of the significance test of the coefficient is always larger than 0.05 indicating that the coefficients observed could be the result of random fluctuations. Since this results may depend on the large errors associated with the measures of $P_{STCA}(e)$, we decide to the *Weighted Pearson Correlation Coefficient*. Since the errors associated to $P_{STCA}(e)$ decrease with the number of aircraft traveled over a link, we consider this quantity as the weight for the calculation of the coefficient. In this way links with high traffic and thus with small errors are more relevant in the calculation of the coefficient, so that effect of the errors should be reduced. Although also with this way of computing the correlation coefficient the result is the same for the correlation with weights. For betweenness centrality instead the coefficient is still close to 0, but the $p$-value was smaller than
0.1 indicating that this uncorrelation is not due to fluctuations and thus betweenness is not predictive when the probability of incurring in an STCA is considered.

![Figure 1.15: Correlations between $w_{STCA}$ and the weights of the links (a) and between $w_{STCA}$ and the betweenness centrality of the links.](image)

**Node Metrics Correlations** The correlations for the extensive criticality measure $n_{STCA}$ with the node metrics are usually stronger with respect to its corresponding $w_{STCA}$. Table 1.2 shows all the correlation coefficients between this metric and the topological and traffic related ones. All these coefficients have a small associated $p$-value, so they are all significant. Despite the fact that every metric is correlated with $n_{STCA}$, the highest correlations are observed for the random-walk betweenness indicating that the information about traffic flows is the most relevant when trying to predict the frequency of occurrence of STCAs. Also the other two topological betweennesses are quite well correlated, meaning that the topology of the navpoint network itself and the way in which the airways intersect each other might play an important role in the occurrence of safety events. Moreover this means that the betweenness centrality is a good indicator of the critical nodes in the network and that the deployment of the traffic is not necessary to have this kind of information. Weaker correlations are observed also for the strength and the degree. The situation is quite different when considering the intensive criticality measure $P_{STCA}(n)$ (Table 1.2). There is no correlation with the strength, indicating that the probability of incurring in an STCA is independent from the local traffic load. Weak correlations with the centrality measures and the degree are still present but the correlation coefficients are
<table>
<thead>
<tr>
<th></th>
<th>Strength</th>
<th>Degree</th>
<th>Betweenness</th>
<th>OD Betweenness</th>
<th>RW Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_{STCA} )</td>
<td>0.60</td>
<td>0.50</td>
<td>0.67</td>
<td>0.61</td>
<td>0.71</td>
</tr>
<tr>
<td>( P_{STCA}(n) )</td>
<td>0.08</td>
<td>0.15</td>
<td>0.24</td>
<td>0.18</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Table 1.2: Correlation coefficients between the node criticality measures and the static network metrics.

<table>
<thead>
<tr>
<th></th>
<th>Strength</th>
<th>Degree</th>
<th>Betweenness</th>
<th>OD Betweenness</th>
<th>RW Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_{STCA} )</td>
<td>0.52</td>
<td>0.52</td>
<td>0.73</td>
<td>0.64</td>
<td>0.70</td>
</tr>
<tr>
<td>( P_{STCA}(n) )</td>
<td>0.10</td>
<td>0.22</td>
<td>0.40</td>
<td>0.28</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Table 1.3: Weighted correlation coefficients between the node criticality measures and the static network metrics.

always \(< 0.3\). Similarly to what we have done for \( P_{STCA}(e) \), it is possible to argue that these coefficients suffer from the effects due to the large errors of the estimate of \( P_{STCA}(n) \). Thus we recomputed the correlations with the Weighted Pearson Correlation Coefficient, using the strength of the nodes as weight since the errors surely decrease with the amount of traffic that passed through a node. Table 1.3 shows the weighted correlation between all the metrics and the two criticality measures \( n_{STCA} \) and \( P_{STCA}(n) \). In both cases we see an increase of the correlation coefficients with respect to the unweighted case, except that for strength. In fact strength seems to be still quite uncorrelated with \( P_{STCA}(n) \) and its correlation coefficient with \( n_{STCA} \) is lowered. Considering the centrality measures, betweenness centrality seems to be the most strongly correlated with both \( n_{STCA} \) and \( P_{STCA}(n) \) and thus it is the best criticality indicator for the navigation point network.

As for unweighted correlations with the dynamical metrics, \( Alt \) is the most correlated one with both \( n_{STCA} \) and \( P_{STCA}(n) \) (Table 1.4) despite the fact that these correlations are still weak (i.e. the coefficients are \(< 0.5\)). Surprisingly, all the metrics related to horizontal movements (\( Fork \), \( Antifork \) and \( Afterfork \)) are uncorrelated with \( P_{STCA}(n) \), while the \( p \)-values of the correlations with \( n_{STCA} \) are larger than 0.1 indicating that the observed coefficients are the result of random fluctuations. Small correlations are observed for \( Positive\ Delay \) and \( Negative\ Delay \). Weighted correlations (Table 1.5) confirm these findings increasing all the correlations coefficients but the ones of the horizontal movements metrics. At this point we have a certain number of features for each node, so the question might be how to group the nodes with similar features together and how
1.3. NETWORK AND DELAYS ANALYSIS

<table>
<thead>
<tr>
<th></th>
<th>Frak</th>
<th>Fork</th>
<th>Antifork</th>
<th>Afterfork</th>
<th>Alt</th>
<th>Pos. Delay</th>
<th>Neg. Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{STCA}$</td>
<td>0.07</td>
<td>0.03 (rej.)</td>
<td>0.05 (rej.)</td>
<td>0.05 (rej.)</td>
<td>0.40</td>
<td>0.23</td>
<td>0.14</td>
</tr>
<tr>
<td>$P_{STCA}(n)$</td>
<td>0.06</td>
<td>0.08</td>
<td>0.09</td>
<td>0.10</td>
<td>0.27</td>
<td>0.12</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Table 1.4: Correlation coefficients between the node criticality measures and the dynamical network metrics.

<table>
<thead>
<tr>
<th></th>
<th>Frak</th>
<th>Fork</th>
<th>Antifork</th>
<th>Afterfork</th>
<th>Alt</th>
<th>Pos. Delay</th>
<th>Neg. Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{STCA}$</td>
<td>0.02 (rej.)</td>
<td>0 (rej.)</td>
<td>0.04 (rej.)</td>
<td>0.01 (rej.)</td>
<td>0.42</td>
<td>0.34</td>
<td>0.12</td>
</tr>
<tr>
<td>$P_{STCA}(n)$</td>
<td>0.04 (rej.)</td>
<td>0.04 (rej.)</td>
<td>0.05 (rej.)</td>
<td>0.06 (rej.)</td>
<td>0.43</td>
<td>0.34</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 1.5: Weighted correlation coefficients between the node criticality measures and the dynamical network metrics.

critical these groups are. Thus we performed a Principal Component Analysis (PCA) with all the metrics introduced in this section excluding the criticality ones. This technique allows us to reduce the dimensionality of our dataset so that the nodes in the network could be described in just 2 or 3 dimensions. In our case we checked that the explained variance of the first 2 principal components combined together is the 60%, so a bi-dimensional approximation is reasonable enough to describe our sample. Since PCA allows us to group together many features in just two principal components, it is possible to know how much every single feature weights on every component considering the scores on each component. For example if a feature have a score close to 1 in a component, it means that in the direction of the component the features grows in value. Vice-versa a negative value indicates that the feature grows in the opposite direction. Figure 1.17 panel (a) shows the scores of all the feature in both the considered principal components. From the scores it is evident that nodes with large values on the principal components have large static metrics, while the dynamical metrics grows slower in this direction. Considering the other component, all the dynamical metrics grow along its negative direction, especially Fork and Afterfork linked to the beginning of horizontal movements. Static metrics instead grows slowly in the other direction. Figure 1.17 panel (a) shows the plot of the nodes in the plane identified by the two principal components. In order to group together the nodes, we applied a clustering algorithm in the plane of the principal components, so that two elements in a single cluster could not be more far than 0.1 (a reasonable value in order to generate large clusters since 0.1 is the order of magnitude of both principal
The clustering procedure left two isolated nodes that where added to closest cluster manually. Thus we identified 5 different clusters indicated with different colors in figure 1.17 panel (a). Of these 5 clusters, the largest one contains all the nodes with small metrics both dynamical and static ones. These nodes can be identified as the “regular nodes” of the network. Two other clusters are displaced along the first principal component, so they are in the direction where static metrics grows faster. The other two are instead along the other principal component in the negative direction, so they have large dynamical metrics and small static ones.

Once we have defined these clusters we can compute the average value of $n_{STCA}$ and $P_{STCA}(n)$ for each one of them. In both case it is evident that the clusters along the first principal components are the most critical ones and that the average value grows with this component (Figure 1.17 panels (b) and (c)). The clusters along the second component have in both cases a small value of criticality metrics, indicating that this component is not relevant in order to predict critical nodes of the network. This result is not surprising since node with low network metrics and many horizontal movements are nodes in which the controllers apply redirection to speed up and not to solve critical situations. On the other hand nodes with large network metrics have also small horizontal movements and large critical criticality measures, indicating that in critical areas the Air Traffic Control rather apply less redirection on the horizontal plane and more vertical deviations, since vertical separation is easier to achieve and manage.
Figure 1.17: Plot of the nodes of the Rome ACC Navpoint Network on the plane of the two principal components. In (a) different colors identify different clusters obtained with a hierarchical clustering algorithm. In (b) different colors identify the average value of $n_{STCA}$ between all the nodes in the cluster, while in (c) they identify the average value of $P_{STCA}(n)$. 
1.4 Air Traffic Control Model

In the previous section we analyzed the traffic and safety data in order to gain information about ATC. We will now use this information in order to build and validate a model of Air Traffic Control based on a dynamics taking place over a Navigation Point Network. The analyses of the features and criticality of such network are crucial in order to define a simple but realistic model. Moreover the comparison between the planned and real navpoint network will be useful in order to validate the model and verify if it is able to reproduce correctly the modification that the ATC performs over the flight plans. The chapter is structured as follows: in section 1.4.1 we will introduce and define the ATC Model, together with the strategies of conflict resolution and the protocols in which they are combined; in we will 1.4.2 present the results of simulation with increasing traffic load; in 1.4.3 we will present the validation performed simulating one day schedules of flights in some European Airspaces; finally in 1.4.4 we will present other refinements that can be added to the model in order to increase its realism, but that we have not taken into account in the validation part.

1.4.1 Definition of the ATC Model

From the analysis of the delays and the navigation point networks two facts are evident: controllers are not aware of the delay of an aircraft and when an aircraft is re-routed the aim is to shorten its trajectory as much as possible. Moreover the fact that new links are created in the navpoint network, indicates that the ATC disregard the structure of the network and thus can create new paths in order to reroute the aircraft efficiently. Thus we model the action of the controllers as a local search over a (planned) navigation point network disregarding its structure in the search process, meaning that new connections can be created in order to send an aircraft from a node to another one despite the fact that these two nodes are not linked or that a path does not exist between them. In our model aircraft fly according to a planned trajectory with a departure time assigned randomly from a uniform distribution in the time frame $[0 \ h, 2 \ h]$. The planned trajectories will be a series of navigation points $\{n_1, \ldots, n_l\}$ connecting the origin $n_1$ and the destination $n_l$ of the flight (which can be either airports or boundary points of a national airspace). We disregard the take-off and landing part of the flight, so we assume that all the aircraft fly at the same average speed of 800 $km/h$ and constant flight level. These are good approximations for the en-route phase of the flight since variations in speed are rare and
not efficient in terms of fuel consumption and horizontal deviations are considered more frequent than the vertical ones. Note that both these two hypotheses can be relaxed in order to improve the realism of the model and in 1.4.1 and 1.4.4 the extensions of the model in order to take into account altitude and speed changes will be discussed. Moreover flight level changes will be crucial in order to build a realistic validation simulation (see the end of par. 1.3.3).

At every time step of simulation (corresponding to 1 second of real time) for every aircraft that has arrived to one of the navigation point in its flight plan, controllers check the next one for conflicts. If a possible conflict is detected, the controllers apply avoidance strategies trying to redirect the aircraft towards a point of the airspace where no conflict are generated. When all the strategies fail, the aircraft is not redirected and the conflict occurs. Note that in real operation usually this will not occur, since other safety nets are enabled in order to prevent the conflict from happening if the ATC fails.

**Conflict Conditions**

The current safety standards for separation between aircraft say that two aircraft must not come closer one another less than 5 NM on the horizontal direction and \( \approx 300 \) m on the vertical one. If these two conditions are violated, the aircraft are “under-separated” and a conflict has occurred. Since we are not considering the vertical component of the trajectories at this stage, horizontal separation is the one that has to be assured. Moreover the assumption of constant equal speed of all the aircraft allows us to translate the spatial separation into a temporal one. Considering a couple of aircraft flying \( M \) and \( N \) on two different segments, say \((n_1, n_2)\) and \((m_1, m_2)\) respectively, we indicate with \( t_n \) and \( t_m \) the times at which they will cross the \( m_2 \) and \( n_2 \). The conflict between these two aircraft is computed geometrically, requesting that at every point on the segment over which they are flying their relative distance is always larger than 5 NM (assuming that every aircraft has the same speed this spatial threshold can be translated into a temporal one, \( \overline{\delta t} = 40 \) seconds). There are 4 possible situations, corresponding to different patterns of the segments over which the aircraft are flying:

1. \( n_2 = m_2 \), i.e. the aircraft are flying towards the same node (Fig. 1.18 panel (a)).

Assuming that \( t_m > t_n \) the condition for the occurrence of a conflict is

\[
|t_m - t_n| < \max \left( \frac{\sqrt{2} \overline{\delta t}}{\sqrt{1 - \cos \alpha}}, \overline{\delta t} \right),
\]

(1.18)
where $\alpha$ is the angle between $(m_1, m_2)$ and the segment over which the aircraft $N$ is supposed to fly after having crossed $n_2$.

2. $n_1 = m_2$ (as well as $n_2 = m_1$), i.e. an aircraft is flying towards the last navpoint that the other aircraft has crossed (Fig. 1.18 panel (b)). In this case the condition for the occurrence of a conflict is

$$|t_m + \bar{t}_n| < \max \left( \frac{\delta t}{\sqrt{1 - \cos \beta}}, \delta t \right),$$

(1.19)

where $\bar{t}_n$ is the time at which the aircraft $N$ has crossed $n_1$ and $\beta$ is the angle between $(m_1, m_2)$ and $(n_1, n_2)$.

3. The segments $(m_1, m_2)$ and $(n_1, n_2)$ intersects each other (Fig. 1.18 panel (c)). In this case the previous conditions (1.18) and (1.19) still hold considering the intersection as a node and interpolating the corresponding crossing times.

4. The segments $(m_1, m_2)$ and $(n_1, n_2)$ do not intersect each other (Fig. 1.18 panel (d)). Also in this case (1.18) and (1.19), but now the intersection of the extensions of the segments has to be considered as a node and the crossing times has to be interpolated as the two aircraft fly on a straight line. Note that even though the conditions of conflict occurrence are satisfied it has to be checked that the conflict occur when both of the aircraft are on the segments and not when one of them is on the extensions.

All these conditions are computed simply by considering the relative distance between the aircraft at a generic time $t$ and applying the Cosine Theorem. For example when considering figure 1.18 panel (a), the relative distance between the two aircraft at a time $t < t_m, t_n$ will be

$$d^2(t) = (t_m - t)^2 + (t_n - t)^2 - 2(t_m - t)(t_n - t)\cos \alpha,$$

(1.20)

where $d(t)$ is expressed in seconds. Since the conflict occurs if there is at least an instant $t$ where $d^2(t) < \delta t^2$, it is sufficient that the minimum of the distance is below this threshold. Thus if we derive (1.20) in order to find the time $\hat{t}$ that minimize this expression and we require that $d^2(\hat{t}) < \delta t^2$, we obtain equation (1.18). The condition (1.19) can be computed in a similar way.

The conditions (1.18) and (1.19) are necessary for the occurrence of a conflict, but other
conditions are to be considered if (1.18) and (1.19) are satisfied in the case at point 4. In fact since we are considering the extensions of the segments, conditions (1.18) and (1.19) are computed as if the aircraft will fly also on such extensions. Thus it is necessary to compute the time range at which the aircraft are under separated. This time range can be computed easily solving the inequality

\[ d^2(t) < \delta^2 \]  \hspace{1cm} (1.21)

(note that (1.18) is basically the condition \( \Delta > 0 \) of this inequality). Once the time range \([t_1, t_2]\) has been found the conflict occur if the intersection between it and the time ranges at which each aircraft flies over the considered segment is not empty.

**Basic Rerouting Strategies**

Whenever a conflict has been spotted the controllers must look for possible redirections to apply to the aircraft and thus sending it from \( n_1 \) to another \( n_r \neq n_2 \). This process has been modeled as a local search between the nodes of the network so that the controller look for all the possible nodes toward which redirect the aircraft without generating other conflicts within a certain subset of nodes of the network and then choses the one that minimize a certain cost function. Since controllers can redirect aircraft both inside their current sector and towards a nearby one, we defined two main strategies of conflict resolution:

- **IN strategy:** the aircraft is redirected towards a node inside its current sector. This event represents in fact, the redirection that controllers can usually perform inside the sector under their responsibility (figure 1.19 panel (a)).

- **OUT strategy:** the aircraft is redirected towards a node in a sector adjacent to its current one. This represents the so-called directs i.e. the pilot of an aircraft can ask the controllers to be redirected directly to another sector, if this does not cause problem in the traffic management. Although directs are not used to solve conflicts as in this case, but just to reduce the traffic load inside a sector. This maneuver usually requires the coordination of the controllers working in the two sectors involved (figure 1.19 panel (b)).

Once a certain set \( \{n_r\}_{r=1, \ldots, k} \) of navigation points has been found, the controller choses the one that minimize a cost function representing the efficiency of the flight. In the simple bi-dimensional case this cost function is just the temporal distance to the arrival:

\[ C_0(n_1, n_r, s) = d_{n_1, n_r} + d_{n_r, s}, \]  \hspace{1cm} (1.22)
Figure 1.18: Geometries used to spot the conflict of a pair of aircraft.

where $d_{n_1,n_r}$ is the distance between $n_1$ and $n_r$ and $d^{sp}_{n_r,s}$ is the length of the shortest-path on the navigation point network between $n_r$ and the arrival node of the aircraft $s$. Since the controllers are not forced to follow the predefined airways to redirect aircraft,
redirects between nodes that are not connected in the original navigation point network are possible.

**Vectoring and Flight Level**

So far we are assuming that all the aircraft fly from a node to another one always following a straight line. This is a good approximation since, looking at the new links created in the real navigation point network more than the 80% of them has been traveled in this way. Controllers can actually make aircraft move in a more complicated way, “vectorizing” them towards geographical references not present in the network. In order to consider also this possibility, we introduce the “vectoring” strategy:

- **Vectoring**: some nodes are added to the network inside the current sector of the aircraft such as they are not closer to each other and the other nodes less than 7 NM. Every temporary node is linked to each other and to every node in the sector. If the conflict is occurring closer than 5 NM from the border of the sectors, nodes are generated also in the next sector of the flight plan of the aircraft. The local search is performed between all the temporary nodes and the other nodes in the sector. At the end of the process all the temporary nodes that are not part of the new trajectory are removed from the network (figure 1.19 panel (c)).

During the en-route phase the flight levels available for the aircraft are discretized in order to simplify the work for the controllers. The flying level at which an aircraft can fly lies in a range going from 2400 ft (where 100 ft ≈ 300 m) to 4000 ft at discrete steps of 100 ft. In this way two aircraft that are not at the same flight level and that are not descending or ascending are automatically (vertically) separated. Then we assigned to every aircraft in the system a required flight level, i.e. the flight level that the owner of the aircraft required for his flight. We assume that each aircraft can ascend or descend with the same speed of 100 ft/min. Now every time an horizontal loss of separation is detected also the vertical separation is checked. This means that, considering the time frame \([t_1, t_2]\) at which the horizontal under-separation occurs, we interpolate the altitude of each flight at every time step in that range and check if there is at least a time \(t \in [t_1, t_2]\) at which their altitude distance is < 100 ft. The IN and OUT strategies are slightly changed. Whenever controllers are searching for possible nodes to perform a re-routing, they check in each node all the flight level that are reachable considering the vertical speed of the aircraft and assign them the one
that is closest to the required flight level of the aircraft. A strategy fails if there are no available and reachable flight levels in every checked node. Once they found some available node, they chose the one that minimizes a new cost function depending on a parameter $\alpha \in [0, +\infty)$,

$$C_\alpha(n_1, n_r, s) = C_0(n_1, n_r, s) + \alpha |FL_{n_r} - FL_{req}|,$$

where $C_0$ is defined as in (1.22) and $FL_{req}$ is the requested flight level of the considered aircraft and $FL_{n_r}$ is the closest available flight level on node $n_r$ to it. The parameter $\alpha$ has been introduced in order to model the fact that some in some airspaces controllers may be more reluctant to large flight level changes and prefer horizontal deviations (in that case $\alpha$ is high) and in some other they may care less ($\alpha$ is close to 0). Since the distances used in the cost function $C_0$ are converted to the time needed to travel them, an $\alpha = 1 \text{ sec. } \frac{\text{ft}}{\text{ft}}$ with a variation in flight level of 10 ft is equivalent to a penalty of 10 sec. to the total cost function. In other words, since the new term added converts a flight level change in travel time, it can be considered as how much the increase of fuel consumption due to the vertical deviation is taken into account by the controllers.

**Conflict Resolution Protocols**

All the defined strategies are combined together into different protocols in order to test their effectiveness in case of high traffic. We call the protocols as

- **IN-OUT protocol**: the IN strategy is used first and then the OUT strategy if the first one fails. This strategy is the basic pattern of conflict resolution since we expect to have usually reroutings within the sector. The OUT strategy in this case is used when an aircraft is close to the border of the sector and the options of rerouting within the sector as few.

- **OUT-IN protocol**: the opposite of the previous one. This protocol is useful to study the effect of long deviations over the trajectories since reroutings toward another sector are usually long.

- **Vectoring-OUT**: similar to the IN-OUT protocol, but vectoring is used instead of the IN strategy.

- **IN-OUT(FL)**: the IN-OUT protocol with vertical deviations taken into account.
1.4. AIR TRAFFIC CONTROL MODEL

We will see than in the validation simulation we will use only the IN-OUT(FL) protocol, which in fact, is the most used by controllers since it requires the least effort. Although the vectoring strategy allows more freedom since controllers redirect aircraft towards any geographical point of the sector, it is the least used because it forces controllers to follow the aircraft singularly in a more detailed way. In case all strategies of conflict resolution at disposal fail to solve a conflict, the conflict simply occurs without affecting the trajectory of the aircraft.

1.4.2 High traffic transition in Synthetic and Realistic Airspaces

The first test of the model that can be perform is to study its behavior with increasing traffic densities. As already mentioned, we defined three protocols (IN-OUT, OUT-IN, vectoring-OUT, IN-OUT(FL)) composed by a main strategy and a backup strategy. Whenever an aircraft has to be rerouted to solve a conflict the main strategy is used and if it fails, i.e., if no available nodes are found, the backup is used. This assures that the main strategy is mostly used, thus allowing to study its effects individually. Since we aim at increasing the traffic far beyond the actual limits, no capacity limits are assigned to the sectors and no direct assignment is used in these simulations. We first perform the simulations of traffic growth on a synthetic airspace with periodic boundary conditions in order to exclude border effects and then we perform similar simulation in realistic airspaces. Since the introduction of synthetic airspace is functional only in order to understand the effect of the boundaries, we will limit the study only to the IN-OUT and OUT-IN protocols.

Synthetic Airspaces

In order to study the model removing possible finite-size effects on realistic networks, we developed a synthetic airspace with periodic boundary conditions. In 1.3.1 we have seen that the planned navpoint network can be approximated by a grid, due to its exponential-degree distribution, its highly peaked lengths of links distribution and the fact that its surface grows linearly with $\langle d \rangle^2 N$ like bi-dimensional lattices, where $\langle d \rangle$ is the average length of the links and $N$ is the number of nodes. Thus we built the synthetic airspace as a grid on a sphere of fixed radius $R$. The navigation points are built using a Fibonacci Grid [103], based on the Fibonacci recursion. The grid has $N = 2n + 1$ nodes, where
$n \in \mathbb{N}$, with latitudes and longitudes defined by the relation

\[
\begin{align*}
lat_i &= \arcsin \left( \frac{2i}{N} \right), \\
lon_i &= \frac{2\pi i}{\Phi},
\end{align*}
\] 

(1.24)
where $i = -n \ldots n$ are the indexes of the nodes and $\Phi$ is defined by $\Phi = 1 + \Phi^{-1}$. The navigation point network is built triangulating the grid, so that in the end we obtain a lattice of degree 6 (Fig. 1.20). We tried to keep the proportion between the surface of the airspace, its navigation points, its number of airports and its number of sectors making it as similar as possible to real airspaces. Table 1.6 reports the number of nodes selected and their corresponding number of airports, number of sectors and surface of the sphere together with the same values for some real planned navigation point networks. With the chosen values, the distribution of the length of the links in the grid is peaked around 28 NM similar to that observed in figure 1.5. In order to assign the sectors in the more homogeneous possible way we divide the latitudinal direction in $k$ equal parts, where $k$ is the larger even divider of the chosen number of sectors $N_{\text{sectors}}$, and the latitudinal direction in $m = \frac{N_{\text{sectors}}}{k}$ equal parts. We checked that with the chosen values of $N$, this division of the airspace guarantees that each sector contains about 37 nodes. At the

<table>
<thead>
<tr>
<th>Name</th>
<th>$N$</th>
<th>$N_{\text{airports}}$</th>
<th>$N_{\text{sectors}}$</th>
<th>Surface (NM²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estonia</td>
<td>53</td>
<td>3</td>
<td>3</td>
<td>242912</td>
</tr>
<tr>
<td>Greece</td>
<td>371</td>
<td>45</td>
<td>9</td>
<td>852729</td>
</tr>
<tr>
<td>Italy</td>
<td>725</td>
<td>43</td>
<td>18</td>
<td>1299055</td>
</tr>
<tr>
<td>France</td>
<td>1199</td>
<td>63</td>
<td>44</td>
<td>2106885</td>
</tr>
<tr>
<td>Synthetic airsp.</td>
<td>185</td>
<td>9</td>
<td>5</td>
<td>425000</td>
</tr>
<tr>
<td>Synthetic airsp.</td>
<td>371</td>
<td>18</td>
<td>9</td>
<td>850000</td>
</tr>
<tr>
<td>Synthetic airsp.</td>
<td>789</td>
<td>37</td>
<td>20</td>
<td>1784095</td>
</tr>
</tbody>
</table>

Table 1.6: Number of nodes, airports, sectors and surface area of some national airspaces and some synthetic airspaces.

beginning of every simulation we chose a certain number $N_{\text{aircraft}}$ of flights, that will be the control parameter of the model. For each realization of a chosen value of $N_{\text{aircraft}}$, we randomly chose $N_{\text{airports}}$ nodes and use them as airports from which the aircraft depart and to which arrive. We assume that, from an airport to another one, the planned trajectory will always be the shortest-path connecting them. In order to reproduce the fact that some routes may be more exploited than other we label each airport with a different values $k \in \{1, \ldots, N_{\text{airports}}\}$ and use them as “departure fitness”. Similarly we assign an “arrival fitness” $q$ randomly reshuffling the departure labels. Considering a flight, the departure airport will be randomly chosen with a probability $p_{\text{dep}} \propto k^{-\tau}$ and the arrival
with a probability $p_{\text{arr}} \propto q^{-\gamma}$, where $\tau$ and $\gamma$ are always arbitrarily chosen equal to 1 in the following. This is done in order to reproduce the fact that some routes could be preferred with respect to others, choosing $\tau = \gamma = 0$ would have meant that the routes are uniformly chosen.

For each size of the system, i.e. the dimension of the synthetic airspace, we found a transition from a phase in which all the conflicts are solve to a phase in which many are not as the control parameter increases. Since with the same value of $N$ but different sizes of the airspace the density of aircraft flying in the system may vary, $N_{\text{aircraft}}$ it is not a good parameter itself to compare the results of the simulation as the size of the system grows. Although the average number of flying aircraft $\bar{N}_f(t)$ per sector grows linearly with $N_{\text{aircraft}}$ in every case and thus it can be used as the free parameter of the model (figure 1.21 panel (a)). This transition is triggered by high local densities of aircraft that make the optimization algorithm fail to solve conflicts due to lack of rerouting options. This can be seen looking at the average number of unresolved conflicts in figure 1.23. For small values of the free parameter we have $\bar{N}_{\text{conflicts}} = 0$, while above a certain threshold depending on the size of the system $\bar{N}_{\text{conflicts}} = 0$ grows with $\bar{N}_f(t)$ in power-law fashion with exponent $\sim 3.6$. Note that this behavior is the same for both the IN-OUT and OUT-
IN protocols and also the transition curve is equal in the two cases. This happens despite the fact that the OUT-IN protocol seems to be more convenient in terms of trajectory shortening as can be seen in figure 1.21 panel (b) that shows the average delays as a function of $N_{\text{aircraft}}$. It is evident that, as the traffic increases and the main strategy is more used, the OUT-IN protocol generates negative delays while the IN-OUT protocol generates positive delays.

Since we are considering a phase transition generated by the lack of possibilities and not by human error, it has to be expected that at a certain point such possibilities are so few that most of the time the strategies are not used. In that case we expect the system to behave as there are no controllers and each aircraft just flies according to its flight plan. This is evident looking at figure 1.22 in which the comparison between the curves of $n_{\text{conflict}}$ for the IN-OUT protocol and for the case in which no strategy is applied. For very high value of $n_{\text{aircraft}}$ the curve for the IN-OUT protocol starts to bend and eventually merges with the other one for higher values.

For both the protocols the transition presents a scaling property with the system size. Assuming that the size is represented by the number of navigation points $N$ within the airspace, all the shown curves collapse into one if $\overline{N_f(t)}$ is rescaled with $N^a$ with $a \simeq 0.4$ (figure 1.21). This transition represent the validity limit of the model and the boundary in which it performs “safely”. Every simulation of realistic situations must take into account the possibility of conflict generation, so that it must be assured that the density of aircraft must always lie below the transition point.
Figure 1.21: (a) $N_f(t)$ as a function of $N_{\text{aircraft}}$ for the IN-OUT and the OUT-IN protocols. (b) Average delay as a function of $N_{\text{aircraft}}$ for the IN-OUT and OUT-IN protocols. These pictures correspond to simulations performed over a synthetic airspace with $N = 185$.

Figure 1.22: $\overline{N_{\text{conflicts}}}$ as a function of $\overline{N_f(t)}$ for the IN-OUT protocol and the case in which no redirection is applied in the synthetic airspace with $N = 185$
1.4. AIR TRAFFIC CONTROL MODEL

Figure 1.23: $\overline{N_{\text{conflicts}}}$ as a function of $\overline{N_f(t)}$ for the IN-OUT (a) and OUT-IN (b) protocols in the synthetic airspaces with $N = 185, 371$ and 789.

Figure 1.24: $\overline{N_{\text{conflicts}}}$ rescaled with $N^{0.4}$ as a function of $\overline{N_f(t)}$ for the IN-OUT (a) and OUT-IN (b) protocols in the synthetic airspaces with $N = 185, 371$ and 789.
CHAPTER 1.

Realistic Airspaces

Similar simulation in a more realistic airspace can be built with our dataset in a quite simple way. Since the planned navpoint networks resemble the real structure of the airways within a national airspace, its choice as airways network over which the aircraft fly is the most natural one. Having chosen a national airspace it is usually simple to build a sector structure to perform the simulation. In the case of the synthetic airspace this has been done dividing the airspace in a number of equal areas growing with the size of the system. As explained in 1.2.1, the structure of the sectors is dynamical meaning that a sector can be split in two or more other sectors in order to ease the management of the traffic. We decided to disregard this possibility also in realistic airspaces and use our dataset to build a simplified static structure of each considered airspace. Basically this structure has been built projecting all the sectors into the bi-dimensional plane in order to obtain a tessellation of the airspace. Figure 1.25 shows the sectors structure of the Estonian, Greek and Italian airspaces used in our simulations. Thus we choose a certain number of aircraft, say $N_{\text{aircraft}}$, and we randomly assign to each one of them a flight plan from the dataset. Since with this procedure the most frequent flight plans are the most likely to be chosen, we are sure that the most trafficked routes in the real system will also be the most trafficked in the simulation. Then we assign to each flight a departure time from a uniform distribution in $[0, 2h]$.

When dealing with realistic airspaces, the effects of the boundaries of the airspace must be taken into account. Suppose that an aircraft has a path that leads a certain boundary point of the airspace and its getting close to it. When the aircraft is close enough the possibilities of redirection become very small and it is not possible to redirect it to a node external to the airspace because everything outside it has been cut out. In this case we have an effect of “conflict generation” just before the transition, caused by the presence of the boundaries. We decided to cure this problem in realistic networks “extending” the airspace including not just the sectors inside it, but also those that lies just outside. Figure 1.26 panel (a) shows the extended sector structure for the Greek Airspace. We assume that no conflicts can occur in these sectors and that an aircraft that is inside the airspace can be redirected outside only if the next sector in its flight plan is outside.

Figure 1.26 panel (b) shows the comparison between the transition in the Greek Airspace and in the “extended” Greek Airspace for the IN-OUT protocol. The effect of the boundary has been considerably reduced, even though it has not been completely eliminated maybe because some refinements on the protocol for the redirection to the outside sectors.
Figure 1.25: Sector structure of the Estonian (a), Greek (b) and Italian (c) airspaces. The sectors have been projected on the bi-dimensional plane and the intersection between the resulting polygons have been excluded.
need to be made.

Using the IN-OUT and the OUT-IN protocols as in the previous paragraph for three different airspaces, the Estonian, the Greek and the Italian we obtain a transition similar to the one found in synthetic airspaces (figure 1.27 panel (a) and (c)). In this case we still have a power-law growth of \( N_{\text{conflicts}} \) with the traffic density, but the exponent of the power law is close to 4.5, higher than in the previous case. Note that however this exponent is still the same for both the IN-OUT and OUT-IN protocol, indicating that its value depends on the topology of the airspace. The transition has still a scaling property analogous to the previous one. Considering \( N \) (the number of navpoints) as the size of the system, the curves of \( N_{\text{conflicts}} \) scale with \( N^a \) with \( a = 0.43 \) (figure 1.27 panel (b) and (d)). Once we have confirmed that the simulations with the realistic airspaces have the same qualitative features of the ones with the synthetic airspaces using the IN-OUT and the OUT-IN protocols we can proceed to test the other defined protocols. In particular the IN-OUT(FL) protocol is the most realistic one since from 1.3.3 it is evident that vertical deviations is the favorite strategy to solve conflicts in critical and highly trafficked areas (at least in the Italian Airspace where the analysis presented in 1.3.3 has been performed). Since with the IN-OUT(FL) protocol it is necessary to assign a requested flight level to each trajectory. In order to maximize the interactions between
the trajectories we assigned to each one of them the same requested flight level of 3600 ft. This is a strong assumption but it is not far from the reality since the distribution of the requested flight levels is highly peaked around that value (Fig. 1.28). Figure 1.30 panel (a) shows the comparison between the IN-OUT and IN-OUT(FL) strategies in the
Italian Airspace. We have chosen $\alpha = 0 \frac{sec}{ft}$, so a variation in flight level is not costly in terms of consumption and is always preferred to a horizontal movement. As it has to be expected the transition point has been shifted towards higher values of $N_f(t)$, from $\simeq 11.5$ aircraft to $\simeq 30$ aircraft. Moreover the exponent of the power-law growth of the curve is smaller. While for the IN-OUT protocol is $\simeq 4.5$, in this case is $\simeq 3.7$. Since this protocol is the most realistic one, we can compare the transition point (defined as the first point at which we see at least a conflict in a realization) with the value of $N_f(t)$ measured with the dataset. Considering a time frame of two weeks, from the $9^{th}$ of June to the $21^{st}$ of June 2011, we compute $N_f(t)$ for each day. The average has been done considering all the flight from 6 am to 20 am, thus excluding the night time when the traffic load is very low. If we consider all the flight levels together we have a value of $N_f(t)$ definitely larger than the thresholds of the IN-OUT of OUT-IN protocols, with a maximum value of $\simeq 180$ aircraft on the $19^{th}$ of June (Figure 1.29). Assuming that also in the real system a trajectory interacts mainly with other trajectories with the same flight level, we can compute $N_f(t, FL)$ for each requested flight level, i.e. the average number of flying aircraft with requested flight level $FL$. Figure 1.29 shows $N_f(t, FL)$ in the considered range of days for some values of requested flight level. In this case it is evident that each curve of $N_f(t, FL)$ is usually below the threshold value of IN-OUT(FL) and only occasionally they get closer (the maximum value of $N_f(t, FL)$ is 32 aircraft on the $12^{th}$ of June for $FL = 3800$ ft). Note that the IN-OUT(FL) protocol is close to the real one, but directs must also be taken into account. The purpose of directs is to speed up the traffic, reducing possible local areas of high traffic density and contributing to move the transition point to higher values of $N_f(t)$. The red dotted line in figure 1.29 is the threshold value for the IN-OUT(FL) protocol with directs. This threshold is higher than the one of the simple IN-OUT(FL) protocol so that the traffic load is always below it. For more information about the direct assignment procedure see 1.4.3.

Similarly to the other protocols, the transition curve for the IN-OUT(FL) protocol scales with the size of the system $N$ (figure 1.30 panels (b) and (c)). Moreover the scaling law is the same with the previous protocols, i.e. $N^a$ with $a = 0.43$. Similar results can be found also for the vectoring-OUT strategy. Despite the fact that the transition point does not seem to be shifted, the curve of $N_{\text{conflicts}}$ grows slower, since the exponent drops to $\simeq 2.7$ (figure 1.31 panel(a)). So the decrease of the exponent of the power-law of the number of conflicts above the transition is a natural effect due to the increased possibility of resolving the conflicts by enhancing the strategy repertoire both. In particular, the
vectoring strategy allows to redirect flights towards any point of the map and enlarges considerably the search space of conflict resolution while the flight level allows the ATC to exploit the vertical direction.

Finally the scaling law of the transition curve is the same also with this protocol (figure 1.31 panel (b) and (c)). This suggest that, while the exponent of the power-law curve of $N_{\text{conflicts}}$ depends on the protocol used, the scaling of the curve and then of the transition point depends just on the topology of the airspace.
Figure 1.29: \( \overline{N_f(t)} \) for various days of the dataset in the Italian Airspace. Black curve is the value of traffic considering all the requested flight levels together, while the others are the values at different flight levels. The horizontal dotted line is the estimated threshold value that separates the conflict-free and the conflicted phases with the IN-OUT(FL) protocol.
1.4. AIR TRAFFIC CONTROL MODEL

Figure 1.30: (a) $N_{\text{conflicts}}$ as a function of $N_f(t)$ for the IN-OUT(FL) and IN-OUT protocols in the Italian Airspace. (b) $N_{\text{conflicts}}$ as a function of $N_f(t)$ for the IN-OUT(FL) in the Estonian, Greek and Italian Airspaces. (c) Scaling of the transition curve for the IN-OUT(FL) protocol. The abscissa has been rescaled by $N^{0.43}$, where $N$ is the number of navpoints in the considered airspace.
Figure 1.31: (a) $\overline{N_{\text{conflicts}}}$ as a function of $N_{\text{aircraft}}$ for the vectoring-OUT and IN-OUT protocols in the Italian Airspace. (b) $\overline{N_{\text{conflicts}}}$ as a function of $N_f(t)$ for the vectoring-OUT in the Estonian, Greek and Italian Airspaces. (c) Scaling of the transition curve for the IN-OUT(FL) protocol. The abscissa has been rescaled by $N^{0.43}$, where $N$ is the number of navpoints in the considered airspace.
1.4.3 Model Validation

The model so far exhibits an interesting behavior when the traffic is higher than the current one, i.e. undergoes to a transition where the optimization algorithm performed by the ATC is not able to solve conflicts. As we stated before, this failure is due to the lack of possibilities to solve a conflict and not to human error.

In 1.2.1 we have seen that traffic load is about to increase in the forthcoming years, but at the present day we have no information about a possible transition to the highly conflicted state of the system (figure 1.29 shows that we are actually below the transition also not considering the possibility of assigning *directs*). Thus the only way to validate the model and to assess its level of realism is to study if it is able to reproduce the modification made by the ATC in normal operations.

By using a data-driven approach we set up a simulation in the Italian Airspace in which flight schedules are real schedules extracted from the data set. The days of the simulation goes from the 8\textsuperscript{th} to the 14\textsuperscript{th} of June 2011, so we are simulating both days in the weekend and working days. Each flight plan in the schedule is simulated with its real requested flight level and its real departure delay. In the following paragraph we will define a new protocol for the validation simulation, based on the analyses performed in 1.3.1 on the whole European Airspace. Nevertheless, it is possible that in some nations such protocol is less valid since the ATC strategies may vary from a nation to another. Thus we will perform the validation simulation also in other national airspaces, in order to verify the limits of validity of the model.

Validation Protocol

The protocol used since now are quite realistic, but some ingredients are missing in order to get close to the real ATC activity. In 1.3.1 we have seen that, from the planned to the real navpoint network, the number of long links increases while the traffic load on such long links decreases with their length (figure 1.5). This seems to be an effect of the procedure of *direct assignment*. A direct assignment is a maneuver a controller may perform to speed up the traffic and reduce his workload. It is performed by sending an aircraft to a nearby sector, i.e. it relies on an OUT procedure, and requires the coordination of the controllers in the two sectors involved. Although very similar, it differs from an OUT strategy since the latter is used only to solve conflicts. In our model we simulate the direct assignments by considering a sector dependent probability $p_{\text{direct}}$ deduced from our data set, so that
each time an aircraft crosses a navigation point in its flight plan, it can be directed towards a nearby sector with such probability. This kind of redirection occurs only if it shortens the trajectory of the aircraft allowing it to arrive at destination earlier. Further, in the real-world ATC, each sector has an assigned capacity, i.e. an assigned maximum number of flying aircraft per hour that can fly through it. To prevent the possibility of getting very high traffic loads inside a sector, these capacities must not be exceeded during both the planning phase and the management activity of the controllers. In the simulation, sector capacities are taken always into account so that no redirection that would exceed the capacity limits is accepted.

In summary, the simulation of the Italian airspace runs as follows:

- When an aircraft leaves a node and heads towards the next one the controllers assigns a direct with a sector dependent probability $p_{\text{direct}}$, but only in case that it shortens the flight and does not cause conflicts.

- If the direct is not assigned the controller checks whether in the next node a conflict would arise; if not, the aircraft proceeds to the planned node.

- Otherwise the controller chooses first the IN strategy and if it does not succeed, the OUT strategy. Flight level changes are possible (i.e. we are using the IN-OUT(FL) protocol.)

- If all these strategies fail, the aircraft goes on straight to the next node, a conflict arises and is recorded.

The probabilities of assigning a directs are measured as the fraction of aircraft that has gone to a node inside their current sector, towards a node in another sector without passing through any other navigation point and such that the the distance between them is higher than $80 \text{NM}$ at least one time in its current sector (the typical dimension of a sector). We checked that for each considered sector the fraction of directed aircraft was about constant in every hour of the day for each day where the traffic is constant (from 6 am to 8 pm) in our dataset (Figure 1.32 shows some examples for some highly trafficked Italian Sectors), so we assumed that all the probabilities of direct assignment are constant during the simulation. Since we do not have any certain information about the real assigned capacities of each sector, we decided to estimate them from our dataset. Considering a sector we measured the number of flying aircraft per hour in each hour of our dataset and use the maximum between these values as its capacity. We can now
1.4. AIR TRAFFIC CONTROL MODEL

Figure 1.32: Estimates of $p_{\text{direct}}$ for each hour of the day in two sectors within the Italian Airspace.

assume that, as long as capacities are not exceeded, human errors are not possible. Table 1.7 shows the value of capacity and $p_{\text{direct}}$ for every considered sector within the Italian Airspace.

<table>
<thead>
<tr>
<th>Name</th>
<th>$p_{\text{direct}}$</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIPPC3</td>
<td>0.07</td>
<td>75</td>
</tr>
<tr>
<td>LIPPD5</td>
<td>0.16</td>
<td>66</td>
</tr>
<tr>
<td>LIRRTS2</td>
<td>0.24</td>
<td>70</td>
</tr>
<tr>
<td>LIPPB5</td>
<td>0.05</td>
<td>144</td>
</tr>
<tr>
<td>LIRRMIE</td>
<td>0.13</td>
<td>188</td>
</tr>
<tr>
<td>LIRRNW2B</td>
<td>0.05</td>
<td>69</td>
</tr>
<tr>
<td>LIRREW2A</td>
<td>0.09</td>
<td>72</td>
</tr>
<tr>
<td>LIBBES3</td>
<td>0.09</td>
<td>51</td>
</tr>
<tr>
<td>LIRRES2</td>
<td>0.18</td>
<td>54</td>
</tr>
<tr>
<td>LIRRNE3</td>
<td>0.09</td>
<td>108</td>
</tr>
<tr>
<td>LIRRSU2</td>
<td>0.16</td>
<td>48</td>
</tr>
<tr>
<td>LIBBND4</td>
<td>0.18</td>
<td>56</td>
</tr>
<tr>
<td>LIBBMD1</td>
<td>0.07</td>
<td>53</td>
</tr>
<tr>
<td>LIRRUS1</td>
<td>0.07</td>
<td>81</td>
</tr>
<tr>
<td>LIPPE3</td>
<td>0.15</td>
<td>59</td>
</tr>
<tr>
<td>LIRROV1</td>
<td>0.16</td>
<td>125</td>
</tr>
<tr>
<td>LIPPA6</td>
<td>0.05</td>
<td>115</td>
</tr>
<tr>
<td>LIRRMIW</td>
<td>0.10</td>
<td>167</td>
</tr>
</tbody>
</table>

Table 1.7: Probability of direct assignment and capacity of each sector within the Italian Airspace used in our simulation.

Validation in the Italian Airspace

As we said before we simulate all the full day schedules of all the days of a week within the Italian Airspace. Thus for each considered day we assign to each aircraft its fight plan
in the dataset, using as departure time the real departure time and not the planned one to reproduce the traffic pattern as it occurred in the considered day. Note that we do not have any information about the rotation of the aircraft. In fact an aircraft does not perform just one single flight per day and it is likely that it will connect more than a couple of airports, i.e. it will go from the airport A to the airport B and then from B to another airport C after passengers and fuel load. So different flights might not be independent from one another and the delay of a flight could be spread during the rotation. Although without this kind of information it is not possible to understand when this phenomenon has occurred during the operation and thus we will consider every flight as independent from the other and no delay diffusion is possible at the airports. Finally we assign to each flight it required flight level, but we assume $\alpha = 0$ in the cost function 1.23. Thus vertical deviations are not costly and flight levels are used to prevent interactions from aircraft at too far vertical layers (the bi-dimensional approximation would be too strong since aircraft at 4000 ft would hardly interact with aircraft at 2000 ft).

The first check performed is that no conflict are generated in any of the simulations done. This is a fundamental condition to assure that the model is reproducing correctly the real situation, where the very few conflicts recorded do not stem from the ATC management. Then we proceeded the validation in two ways. In 1.3.1 we have seen that the ATC produces modifications in the topological structure of the navpoint network of the airspace. Thus we can perform a coarse grained validation checking if our model is able to reproduce these macroscopic changes. Moreover a microscopic validation is possible comparing the statistics of the variations on the single trajectories.

The coarse grained validation is performed building a “real” navpoint network using the output trajectories of the simulation and comparing it with the one built using the radar updated trajectories of the corresponding day of the simulation. As explained in 1.3.1, the action of the controllers generates new links in the network, making the real networks more dense than the planned one. Table presents the number of links in the planned and real navigation point networks, built in each day going from the $8^{th}$ to the $14^{th}$ of June 2011, together with the corresponding number of links in the real navigation point network built using the output trajectories of our simulations. We found that the number of links are always very similar (table 1.8), even though those belonging to the networks built using the simulations are usually slightly larger. Moreover the distribution of the length of the links of the generated network is in good agreement with that of the real network presented, meaning that our model reproduce correctly the length of the new links created.
by the controllers. Note that the agreement between these distributions depends on the presence of *direct assignment*: without this procedure the agreement is worse, confirming our hypothesis that long links are generated by this process (figure 1.33 panels (a) and (b)). In figure 1.5 we presented the anti-correlations between the length and the weight of the links in the real (European) network. This anti-correlation are a common phenomenon in most airspaces in Europe and the Italian one is not an exception. Again our model is able to reproduce what found with the data (figure 1.33 panel (c)). Then we can compare the node metrics presented in 1.3.1, i.e. the variations in strength ($\delta s$), degree ($\delta k$) and betweenness centrality ($\delta b$). These variations must be consistent with those measured using the real navpoint network build using the dataset. Figure 1.34 shows the comparison between the distributions of the variations of these metrics measured with the dataset and the simulations. Only one day of the whole week is presented as example, since different days produces the same results. Due to the action of the controllers we usually have an increase of the degree of the node, up to an increment of about 20 new connections in some cases. The increment of the number of links instead generates new shortest-paths on the network and this usually leads to a decrease of the values of centrality of the nodes. Finally the distribution of the variations in strength is highly peaked around 0 with a slightly longer tail towards negative values, meaning that the action of the controllers naturally tries to decrease the traffic load over the nodes without creating new highly trafficked ones. We found a good agreement between the distributions obtained by the simulation and those of the data set, meaning that the model also reproduces this kind of topological variations between real and planned networks. Distributions by themselves are not sufficient to assess if the model is reproducing the variations correctly. For example the correct variations in strength (i.e. in traffic load) could be assigned to the wrong nodes by the model, producing a wrong rearrangement of traffic load over the network. Table 1.8 presents the correlation coefficients for the variations in betweenness centrality, degree and strength of the nodes for each day of simulation. All the coefficients are quite constant through the days, the correlation coefficient for the variations in betweenness centrality of the nodes is usually the highest (about 0.8), while the other are always larger than 0.5. This clearly indicates that not only the distributions of the variations are consistent, but that they are also placed correctly over the nodes of the network. Concerning the *microscopic validation*, we defined three variation metrics between the flight plans and the radar updated trajectories: the variation of the number of crossed navpoints ($\delta n$), the variation in length of the trajectory ($\delta l$) and the en-route delay ($\delta t_{enr}$). The distributions
Figure 1.33: (a) Distribution of the lengths of the links for the Italian Navigation Point Networks built with planned, radar updated trajectories and with trajectories generated by our simulation. (b) Same distributions, but the direct assignment has been suppressed in the simulation. (c) Binning of the scatter-plot between the length of the links \(d\) and the number of aircraft that traveled over a link \(w\) for the Italian Navigation Point Networks built with radar updated trajectories and with trajectories generated by our simulation. These figures correspond to the 9\textsuperscript{th} of June 2011.

of the first two quantities are well reproduced by the model (figure 1.35) indicating that the redirections applied are likely to modify the trajectories in a similar way of the action of the controllers. On the other hand the agreement when considering the en-route delay
Figure 1.34: Distribution of the variations in strength (a), degree (b) and betweenness centrality (c) between the Italian Navigation Point Network built with radar updated trajectories and the trajectories generated by our simulation. These figures correspond to the 9th of June 2011.

distribution is not so good as in the previous cases (figure 1.36 panel (a)). In fact, since redirections are the result of an optimization process, they are likely to produce negative delays instead of positive ones. To reproduce correctly the behavior of the en-route delay distribution, a random external effect affecting the action of the controllers has to be introduced. The ATM system is far from being isolated, since adverse occurrences like bad weather conditions affect the system on a daily basis. Therefore, we introduce in
Table 1.8: Comparison between the real navigation point networks built with the data and with the trajectories generated by simulations, for each day of simulation.

<table>
<thead>
<tr>
<th>Days</th>
<th>$n_{\text{links}}$ (data)</th>
<th>$n_{\text{links}}$ (simulation)</th>
<th>$\delta d$ corr. coeff.</th>
<th>$\delta s$ corr. coeff.</th>
<th>$\delta b$ corr. coeff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>08/06/2011</td>
<td>3318</td>
<td>3151</td>
<td>0.64</td>
<td>0.67</td>
<td>0.86</td>
</tr>
<tr>
<td>09/06/2011</td>
<td>3056</td>
<td>3160</td>
<td>0.60</td>
<td>0.81</td>
<td>0.87</td>
</tr>
<tr>
<td>10/06/2011</td>
<td>3522</td>
<td>3598</td>
<td>0.66</td>
<td>0.69</td>
<td>0.87</td>
</tr>
<tr>
<td>11/06/2011</td>
<td>3323</td>
<td>3419</td>
<td>0.66</td>
<td>0.71</td>
<td>0.86</td>
</tr>
<tr>
<td>12/06/2011</td>
<td>3562</td>
<td>3251</td>
<td>0.67</td>
<td>0.64</td>
<td>0.88</td>
</tr>
<tr>
<td>13/06/2011</td>
<td>3178</td>
<td>3350</td>
<td>0.64</td>
<td>0.67</td>
<td>0.83</td>
</tr>
<tr>
<td>14/06/2011</td>
<td>3163</td>
<td>3230</td>
<td>0.65</td>
<td>0.69</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Figure 1.35: Distributions of the variations in length (a) and variations in number of crossed navigation points (b) between the planned trajectories and the radar updated ones and between the planned trajectories and the trajectories generated in our simulations. These figures correspond to the 9th of June 2011.

the model some external disturbances as delay generating perturbed areas. To simulate external disturbances we introduced the first un-fixed parameter of the model, the number of external disturbances $n_{\text{ext}}$. Each disturbance is a round area with a randomly assigned radius $r_{\text{dist}}$ in a range between 1 NM and 15 NM. Every time an aircraft is flying on a link that crosses one of the perturbed areas it gains a random penalty delay from a uniform distribution in a range between 1 min. and $0.14\ \text{NM}^{-1}\ \text{min}$. $r_{\text{dist}}$ ($0.14\ \text{NM}^{-1}\ \text{min}$ is the
inverse of the fixed velocity of the aircraft). Note that if a link crosses more than one disturbance, the delays are cumulated together. This in order to model the fact that it is not convenient to fly over highly perturbed areas. Every disturbance affects all the flight levels inside its area and is static. To reproduce the fact that weather changes during the day, every 60 min of simulation the displacements and dimensions of the disturbances are randomly reassigned. The values of $n_{ext}$ used are 0, 200 and 2000 for each day of simulation. From our data we have no clues on the typical number of disturbances during the day therefore we analyzed the response of the model to various values of $n_{ext}$. We find that all the measures of topological variation in the navigation point network and the variations of length and crossed navigation points of the trajectories depend slightly on $n_{ext}$, while the shape of the distribution of the delays $\delta t_{enr}$ is very sensitive to it. In Fig. 1.36 it is shown that with a relatively small number of disturbances ($\approx 200$) the agreement improves, while with a higher number ($\approx 2000$) the agreement is not recovered.
Figure 1.36: Comparison between the en-route delay distributions measured with the dataset and from the simulations without external disturbances (a), with \( n_{\text{ext}} = 200 \) external disturbances (b) and with \( n_{\text{ext}} = 2000 \) external disturbances (c). These figures correspond to the 9\(^{th}\) of June 2011.

Validation in Other Airspaces

Similar simulations can be build in other European Airspaces in order to test whether the model works or not, i.e. if the defined protocol is general enough to be used in every airspace. This is not a trivial question since, for example, the classification of the nodes of the navpoint network presented in 1.3.3 to explain the occurrence of STCA's that partially
1.4. AIR TRAFFIC CONTROL MODEL

Figure 1.37: (a) Distribution of the lengths of the links for the German Navigation Point Networks built with planned, radar updated trajectories and with trajectories generated by our simulation. (b) Binning of the scatter-plot between the length of the links \((d)\) and the number of aircraft that traveled over a link \((w)\) for the German Navigation Point Networks built with radar updated trajectories and with trajectories generated by our simulation. These figures correspond to the 9th of June 2011.

informed the protocol, might not be valid in a different airspace. Moreover the inclusion of the Vectoring strategy could be crucial in some airspaces. Thus we performed the validation simulations also in the Estonian, Greek, German and Spanish airspaces using the flight plans of the 9th June 2011. The results presented in the previous paragraph are confirmed for most of the other airspace used. The distributions of the length of the links are well reproduced in all of the cases and also the anti-correlation between the length of the links and the traffic over them (Fig. 1.37). Concerning node metrics, again we have good agreement in the distributions of the variations of these quantities (Fig. 1.38). Although the correlations of these variations are not large or are absent for some countries as can be seen in table 1.9. This indicates that the protocol is not suited to reproduce the macroscopic changes performed by the ATC in these case, but some refinements or modifications have to be taken into account.

Microscopic metrics are instead always well reproduce, considering that external disturbances must be added in every airspace in order to improve the agreement with the experimental distribution of \(\delta t_{enr}\).
Figure 1.38: Distribution of the variations in strength (a), degree (b) and betweenness centrality (c) between the German Navigation Point Network built with radar updated trajectories and the trajectories generated by our simulation. These figures correspond to the 9\textsuperscript{th} of June 2011.

<table>
<thead>
<tr>
<th>Days</th>
<th>$n_{\text{links}}$ (data)</th>
<th>$n_{\text{links}}$ (simulation)</th>
<th>$\delta d$ corr. coeff.</th>
<th>$\delta s$ corr. coeff.</th>
<th>$\delta b$ corr. coeff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greece</td>
<td>1022</td>
<td>1181</td>
<td>0.55</td>
<td>0.83</td>
<td>0.87</td>
</tr>
<tr>
<td>Austria</td>
<td>430</td>
<td>495</td>
<td>0.40</td>
<td>0.53</td>
<td>0.82</td>
</tr>
<tr>
<td>Spain</td>
<td>2080</td>
<td>1230</td>
<td>0.326</td>
<td>0.14</td>
<td>0.78</td>
</tr>
<tr>
<td>Germany</td>
<td>4702</td>
<td>6759</td>
<td>0.53</td>
<td>0.61</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Table 1.9: Comparison between the real navigation point networks built with the data and with the trajectories generated by simulations for some European Airspaces. These value correspond to the 9\textsuperscript{th} of June 2011.
1.4.4 Model Refinements

The protocols and conflict resolution strategies that have been considered until now are quite realistic and the models is able to reproduce many of the experimental findings despite its simplicity. Although many refinements are possible in order to improve it. In this section we will discuss to possible modification of the IN and OUT rerouting strategies, speed changes and lookahead time. Note that although we will consider just the IN-OUT protocol, but the extension to IN-OUT(FL) and Vectoring-OUT protocols is straightforward.

**Speed Changes**

In the basic model each aircraft is flying at the same speed \( v = 800 \text{ km/h} \), which is about the average en-route speed for an aircraft. This fact did not influence too much the agreement with experimental data, so constant average speed for every aircraft is a good approximation that can be made in simulations. However, our model allows the introduction of different speed and speed changes maneuvers in conflict resolution. We can assign to each aircraft a different speed \( v_m \), both measuring it from the data or taking it from a defined distribution. Indicating with \( r_m \) the ratio between \( v_m \) and \( v \) so that

\[
v_m = r_m v,
\]

we will assume that \( r_m \) are randomly assigned from a uniform distribution. In order to keep the model discrete, we will assume that \( r_i \) will be randomly taken from the set \( V = \{0.95, 0.94, \ldots, 1.05\} \). With different speeds the conditions for the occurrence of a conflict must be rewritten. Equation (1.18) for the distance of two aircraft on different segments becomes

\[
d^2(t) = r_m^2(t_m - t)^2 + r_n^2(t_n - t)^2 - 2r_m r_n(t_m - t)(t_n - t) \cos \alpha,
\]

where every term has been rescaled by \( \frac{1}{v} \) so \( d(t) \) is expressed in seconds. It is then easy to calculate the analogous of the equations for the occurrence of a conflict. For example equation (1.18) becomes

\[
|r_n t_m - r_m t_n| < \max \left( \frac{a_{mn} \delta t}{\sqrt{1 - \cos^2 \alpha}}, \delta t \right),
\]

where \( a_{mn}^2 = r_m^2 + r_n^2 - 2r_m r_n \cos \alpha \). Then it is possible to calculate the time range in which the aircraft are under separated as before, solving \( d^2(\hat{t}) < \delta t^2 \).

Thus we can proceed modifying the IN and OUT strategies of conflict resolution as follows:
Whenever a navpoint has to be checked for rerouting, the conflict conditions are check for every $v_m \in \mathcal{V}$. A list of possible traveling speed on the next link is recorded. In case this list is empty the node is not available for a rerouting.

Between all the possible speeds assigned to a navpoint, one is chosen with a certain criteria. In our case we chose the speed which is closer to the one initially assigned to the aircraft, considering it as some sort of required speed of the aircraft. We indicate with $v^{req}$ the required speed and $r^{req}$ the ratio with $v = 800 \text{ km/h}$.

At the end of the search procedure, we will have a list of nodes with an assigned traveling speed. The one that minimizes the cost function

$$C_{\text{speed}}(n_1, n_r, s) = C_0(n_1, n_r, s) + \alpha |r_{n_r} - r^{req}|$$  \hspace{1cm} (1.28)

is chosen and the corresponding speed is assigned to the aircraft ($C_0(n_1, n_r, s)$ is the cost function defined in (1.22)). In the following we will assume $\alpha = 6000 \text{ sec}$, so a variation in relative speed of 0.01 is equal to a penalty of 1 minute.

Figure 1.39 panel (a) shows the curve of the average number of conflicts per realization as a function of the number of aircraft injected in the system for the basic model and the model with speed changes in the Estonian Airspace. It is evident that adding this ingredient does not affect much the behavior of the model since the curve of the transition has been slightly shifted towards higher values of traffic load. The power-law shape and the exponent of the curved have not varied from the basic case. On the other hand adding this ingredient allows us to study how the velocities of the aircraft are rearranged in order to guarantee safety. In figure 1.39 panel (b) and (c) the average of the ratio $r_n$ and its standard deviation are presented as functions of the traffic load. These quantities has been measured by computing the average of $r_n$ for each trajectory (averaging over each segment of the trajectory) and then calculating the average and standard deviation of these values. Despite the fact that the parameter $\alpha$ in equation (1.28) should prevent each aircraft from flying at a speed too far from its required one, as the traffic load increases the average velocity grows. On the other hand the standard deviation of the velocities is slowly decreasing, meaning that the distribution of the velocities is becoming sharper around the highest values. In other words in the conflicted phase, high speed are favorite with respect to slow one since they are most likely to solve conflicts and reduce the traffic within the system.
1.4. AIR TRAFFIC CONTROL MODEL

Figure 1.39: (a) $\overline{N_{\text{conflicts}}}$ as a function of $N_{\text{aircraft}}$ for the IN-OUT protocol with and without speed changes. (b) Average relative velocity $r_m$ as a function of $N_{\text{aircraft}}$ for the IN-OUT protocol with speed changes. (c) Standard deviation of $r_m$ as a function of $N_{\text{aircraft}}$ for the IN-OUT protocol with speed changes. All these pictures correspond to simulations performed in the Estonian Airspace.

**Lookahead Time**

In real operations controllers have a typical lookahead time within which they are able to predict if a conflict will occur between a couple of aircraft following a certain path. In the basic model a natural threshold was present due to the sector structure. Since the typical
dimension of a sector is 100 NM with a constant speed of 800 km/h (i.e. 430 NM/h) the lookahead time was about 14 minutes for redirections with the IN strategy. However in the basic model such lookahead time is not the same for every checked node, but it depends on how far the node is from the aircraft at the time of the redirection. Thus if this distance is short the lookahead time can be considerably reduced and the aircraft could be forced to fly through a path that within few minutes will cross a highly trafficked area, leading to more reroutings. Instead it is more realistic to assume that the lookahead time is always the same for each path.

To solve this problem we introduced a new parameter, $t_{thr}$, indicating the lookahead of the controllers (we will assume that this time is the same for every controller in every sector). Each strategy of conflict resolution is then modified in this way:

- Now not just the rerouting node is considered for the rerouting but the entire path $\{n_1, n_2, \ldots, n_l\}$ with its corresponding crossing times $\{n_1, t_r, \ldots, t_l\}$. ($n_1$ here indicates the current node of the aircraft at the time of rerouting $t_1$).

- A first check is performed on the segment $(n_1, n_2)$. As in the basic model, if a conflict is spotted over this segment the path is discarded.

- Conflicts are then checked over successive segments $(n_i, n_{i+1})$ until the condition $t_i \geq t_{thr}$ is satisfied. For each conflicted segment a penalty equal is assigned to the cost function (1.22) used to chose the new path. This penalty will be of 1 minute for every conflicted segment in the following.

- The path with the smallest cost is chosen.

In this way all the paths that generate more conflict are penalized even though they are the shortest ones. Notice that if $t_{thr}$ is small, its effect is just to make the action of the controllers more uniform in time. Otherwise if $t_{thr}$ is big enough (bigger than the natural upper threshold of the system), it simulates the effect of controllers that are able to spot conflicts further in time. Figure 1.40 panel (a) shows $N_{conflicts}$ as a function of $N_{aircraft}$ for different values of $t_{thr}$ in the Estonian Airspace. The protocol used for conflict resolution is the same (IN-OUT) used in the previous section concerning the validation activity. The shape of the curve of the transition is not influenced for small values of this parameter ($t_{thr} \leq 10$) close to the transition point, although the transition point the system increases with it. For large values of $t_{thr}$ the power-law growth is the same (i.e. with the same exponent), while it slows down as the traffic load increases. This is
1.4. AIR TRAFFIC CONTROL MODEL

Figure 1.40: (a) $\overline{N_{\text{conflicts}}}$ as a function of $N_{\text{aircraft}}$ for the IN-OUT protocol and various values of lookahead time $t_{\text{thr}}$. (b) $\overline{N_f(t)}$ as a function of $N_{\text{aircraft}}$ the IN-OUT protocol and various values of lookahead time $t_{\text{thr}}$. All these pictures correspond to simulation performed within the Estonian Airspace.

reflected by the fact that $\overline{N_f(t)}$ does not depend linearly by $N_{\text{aircraft}}$ anymore (Fig. 1.40 panel (b)), since the introduction of a large lookahead time makes the more efficient in terms of speeding up the traffic. Thus with the same value of $N_{\text{aircraft}}$, $\overline{N_f(t)}$ decreases as $t_{\text{thr}}$ grows. If $t_{\text{thr}} > 30$ minutes the transition point does not increase, since this value correspond to the typical length of a trajectory inside this airspace ($\approx 200$ NM).

Combining the Refinements

The two previous ingredients can also be combined together, mixing the possibility to assign a new speed to an aircraft to the existence of a temporal threshold below which the controllers can spot conflicts. In this new case for each possible path all the velocities are checked and the available ones are recorded. Moreover for each recorded velocity, the assigned penalty due to the conflicts generated by the segments within is recorded. Thus each navpoint available for a rerouting will have a set of velocities and the corresponding penalty generated by that velocity. It is possible to define a new criterion for velocity assignment, not linked to a required speed of each aircraft but based on a criterion to avoid safety issues and congestion: from all the available speeds the controllers can for example chose the one that generates the smallest penalty. Figure 1.41 panel (a) shows
the curve of the transition for the model without speed changes and $t_{thr} = 15\,\text{min}$ and the model with speed changes and the same value of $t_{thr}$. In this case it is evident that the presence of a large time horizon and the possibility to adjust speed combined together definitely increase the capacity of the system and the curve of the transition is shifted more the in the previous case. Moreover the effect of speeding up the traffic making as the traffic load increases is even stronger than in the case without lookahead time. In figure 1.41 panels (b) and (c) the average of the ratio $r_n$ and its standard deviation are presented as functions of the traffic load for the model with and without lookahead time. While the behavior of the standard deviations is unvaried, the average value of $r_n$ grows faster with the lookahead time. Note however that this behavior can be modulated by the value of the parameter $\alpha$ in equation (1.28).
1.5 Global Optimization

In the future SESAR scenario the structure of the airspace will be particularly different from the current one, not just because of different regulations, but also because a more flexible planning of the trajectories of the flights will surely result in a new airways topol-
ogy. With this perspective it is reasonable to ask if the new structure will effectively be more efficient than the current one in terms of performances for the flights and in manageability for the controllers. Since our model is able to reproduce the action of the ATC in a normal situation within several European Airspaces, one of its possible applications is the study of new configurations and its response to the typical disturbances that can occur during the operations.

Thus we decide to build a new solution for the arrangement of the planned trajectories in which besides the usual capacity constraints of the sectors also the conflicts between the trajectories are taken into account. In other words, if the controllers act as local optimizer, this solution is a globally optimized one computed with a stochastic optimization algorithm well-known in complex systems physics. Despite the fact that such kind of planning could be too unrealistic, both the method applied for its construction and its testing are interesting from a theoretical and practical point of view. This section is structured as follows: in 1.5.1 we will present and review the main aspects of the Extremal Optimization Algorithm used for the construction of the planned trajectories; in 1.5.2 we will discuss its adaptation to the problem of trajectory planning; in 1.5.2 and 1.5.4 we will present the resulting airways structure within a national airspace and we will test its efficiency using the ATC model presented in 1.5.3.

1.5.1 Extremal Optimization Algorithm

In complex system physics many optimization algorithm have been developed in order to solve heuristically a wide variety of problems. Despite the fact that many of these algorithms have been developed to solve specific problems, they can be applied to situations far from the framework of physics. Some notable examples are the “simulated annealing” techniques [104] and the “genetic algorithms” [105]. In relatively recent times a new algorithm has been proposed, the “Extremal Optimization” (EO) algorithm, based on the avalanche phenomenon present in systems with “Self-Organized Criticality” (SOC) [106]. One of the most notable example of systems with SOC is the Bak-Sneppen model [107], which modeled the evolution of interrelated animal species. In this model some species of animals are positioned on a grid, each one having a certain value of fitness in a certain range. At each time step the species with the smallest fitness is updated and its fitness is randomly reassigned (representing the death of such species and its substitution with a new one). This change in fitness has an impact also on the neighbors, so that their fitness is also randomly reassigned. The system rapidly reaches a state called SOC, in
which all the fitnesses are above a certain value and avalanches take place. These are chain reactions leading to large fluctuations that make every possible configuration of the system virtually accessible. Extremal optimization is based on the same principles of the Bak-Sneppen model. The optimization is performed by updating at each step the element of the system with the lowest “fitness” (or cost in this case). The generation of fluctuations through avalanches in this optimization process, make the system visit much of its accessible configuration preventing it from being trapped in a local minimum (or maximum) of its cost function. Considering a certain set of variables $x_i$ each one with an assigned fitness $\gamma_i$ so that each fitness contributes linearly to the cost function $C(\gamma)$ defined as

$$C(\gamma). \quad (1.29)$$

Indicating with $S$ a generic configuration of the variables $x_i$, $\Omega(S)$ will be the set of neighbor configurations of $S$, i.e. a set of configurations that are close and accessible from the configuration $S$. Note that the definition of $\Omega(S)$ is completely arbitrary. The algorithm proceeds as follows:

- Choose a starting configuration and set $S_{\text{best}} := S$

- For the configuration $S$:
  - Evaluate the fitness $\gamma_i$ of each variable
  - Find the variable with the highest fitness (indicated by the letter $j$)
  - Choose a new configuration $S' \in \Omega(S)$ so that the variable $j$ changes its value
  - Accept $S := S'$ unconditionally
  - If $C(S_{\text{best}}) \geq C(S)$, set $S_{\text{best}} = S$.

- Repeat the previous point as long as desired.

- At the end the sub-optimal configuration will be $S_{\text{best}}$ and the sub-optimal cost will be $C(S_{\text{best}})$.

It is clear that this algorithm is well suited to solve all those optimization problems in which the cost function can be factorized in a sum of individual contributions, provided that its is possible to defined $\Omega(S)$ reasonably. Extremal Optimization has been proposed in [19,20] and tested on some combinatorial problems, showing better performances than SA and Genetic Algorithms. It has also been applied to problems such as community
detection [108] and spin glasses [109]. Analytical studies have been performed to its application to a particular type of spin glass minimization problem in [110].

1.5.2 Application to Trajectory Optimization

Our aim is to adapt the EO algorithm to the problem of trajectory optimization, so that we are able to produce optimal flight plans according to some criterion. A realistic approach to this problem would be too complex and beyond the our purposes so we will assume some simplifications. First we will disregard the departure and arrival part of the flight and will focus on the en-route phase, meaning that the within a certain radius to the airports (≈ 20 NM) all the conflicts between the aircraft are disregarded since they will be considered in the departure or approach phase. We will also assume that each aircraft fly at its required flight level and that height changes are not possible in the optimization process. In the SESAR scenario each trajectory will exploit the airspace freely without geographical constraints. Here for simplicity we will assume that the optimization is still performed using some fixed geographical references, even though no airways structure is present between them and every path connecting them is virtually available. These geographical references can be any set of points in the considered airspace, but in the following we will use the existing navigation points. Thus each variable $x_i$, corresponding to the trajectory of the $i^{th}$ aircraft is

$$x_i = \{(n_{start}, t_{start}), (n_1, t_1), \ldots, (n_{stop}, t_{stop})\},$$

(1.30)

where $n_k$ is the $k^{th}$ crossed navigation point and $t_k$ is the corresponding crossing time. Note that $t_{start}$ is fixed, while $t_{stop}$ depends on the path. As in the ATC model, we will assume that each aircraft flies at the same constant speed of 800 km/h. All the previous simplifications, i.e. constant speed, constant flight level and fixed departure time could be relaxed leading to better optimal solutions, but the behavior of the algorithm will not be affected. Finally we assume that capacity constraints must be satisfied, i.e. configurations $S$ in which the number of aircraft per hour inside a certain sector is above the threshold must be rejected during the optimization. Given a certain configuration of trajectories $S = \{x_1, \ldots, x_N\}$, its cost function will be

$$C(S) = \sum_{i=1}^{N} l(x_i) + \frac{\epsilon}{2} \sum_{i=1}^{N} \sum_{j, j \neq i}^{N} m(x_i, x_j),$$

(1.31)
where $l(x_i)$ is the length of the path $x_i$ divided by the length of the straight line connecting it departure and arrival node and $m(x_i, x_j)$ is the number of conflicts between the aircraft $i$ and the aircraft $j$. The parameter $\epsilon$ is a positive real number that quantifies the weight of conflicts in the optimization process. If $\epsilon = 0$ then the optimal solution is when all the trajectories are a straight line from departure to arrival and $l(x_i) = 1$ for every $x_i$. On the other hand if $\epsilon \gg 1$, we will get a conflict-free solution. In order to apply the EO algorithm to this problem, the cost function must be in the form (1.29). This can easily achieved by defining the fitness of each trajectory $x_i$ as

$$\gamma_i = l(x_i) + \frac{\epsilon}{2} \sum_{j=1,j\neq i}^{N} m(x_i, x_j).$$  \hspace{1cm} (1.32)

Finally we must define $\Omega(S)$ and the process to go from $S$ to $S' \in \Omega(S)$. At the beginning of the optimization process every trajectory is composed by one long segment going from $n_{\text{start}}$ to $n_{\text{stop}}$, i.e. in the optimal configuration if $\epsilon = 0$. At each optimization step, the path of the less fit aircraft is built using a biased random walk process starting from $n_{\text{start}}$ and ending in $n_{\text{stop}}$. Note that the interaction term in (1.32) guarantees that changing $x_i$ will modify all the fitnesses of the others trajectories. The new random path is built starting from $(n_0, t_0) = (n_{\text{start}}, t_{\text{start}})$. At each step $(n_k, t_k)$, the next node $n_{k+1}$ is randomly chose from the set of all the other nodes with a probability

$$p(n_{k+1}) \propto \left( \frac{d_{n_k,n_{k+1}} + d_{n_{k+1},n_{\text{stop}}}}{d_{n_k,n_{\text{stop}}}} \right)^\alpha,$$  \hspace{1cm} (1.33)

where $\alpha < 0$ and $d_{n,m}$ is the geographical distance between the navigation points $n$ and $m$. The process ends when $n_{k+1} = n_{\text{stop}}$ and it is restarted whenever a step violates the capacity constraints of the sectors. Note that (1.33) disadvantages steps that are far from being on the straight line connecting the previous step with the destination. In the following we will fix the parameter $\alpha = -4$ arbitrarily. Other choices may fail to converge (e.g. if $\alpha$ is too close to 0) or may lead to solutions too close to the straight line (e.g. if $\alpha$ is too large).

### 1.5.3 The Optimal Solution

We apply the algorithm to the Greek National Airspace in order to optimize all the trajectories of the flight that took place on the 9th of June 2011. The choice of the airspace and of the day of operations is arbitrary, but different choices will lead to similar
results. We assign to each flight its real departure time, i.e. considering the departure delay acquired that day and we do not put any constraint on the arrival time. Despite its simplicity, the algorithm converges within a small number of steps to a sub-optimal solution. As it is shown in figure 1.42, initially the algorithm generates deviations for many trajectories in order to solve some or all the conflicts and thus increasing the values of $l(x_i)$. After some steps, the sub-optimal number of conflicts is reached and the algorithm proceeds optimizing the length of the trajectories until the final optimal configuration is reached. As we vary the parameter $\epsilon$, we modulate from a trivial solution in which all the aircraft fly on a straight line to a conflict-free solution. Figure 1.43 shows the sub-optimal values of the average $\langle l(x_i) \rangle$ over all the trajectories and the total number of unresolved conflicts $N_{\text{conflicts}} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} m(x_i, x_j)$ as $\epsilon$ varies. Starting from the trivial case $\epsilon = 0$, $\langle l(x_i) \rangle$ starts to grow as the number of conflicts decreases, since the algorithm starts to create deviations in order to solve some of the conflicts. When $\epsilon \in [1, 2]$ the solution seems to be stationary, while for $\epsilon \geq 2$ we always approach a conflict-free solution, indicating that for these value of the parameter each conflict weights more than any redirection in terms of cost in the cost function (1.31). Note that the $\langle l(x_i) \rangle$ is close to 1 for every value of $\epsilon$, so every solution is a set of trajectories that are some small deviations from the straight case.

Figure 1.42: $\langle l(x_i) \rangle$ (a) and $N_{\text{conflicts}}$ (b) as functions of the optimization step for $\epsilon = 2$. The optimization has been performed using all the flight in the Greek Airspace on the 9th of June 2011.
1.5. GLOBAL OPTIMIZATION

Figure 1.43: $\langle l(x_i) \rangle$ (a) and $N_{\text{conflicts}}$ of the sub-optimal solution as functions of $\epsilon$.

In every considered case the resulting structure of the airspace is very different from the current one. Since building a simple navigation point network has no meaning in this new optimal case (navigation points are just arbitrary points in space used for the optimization, they could have been substituted by any other set of geographical references), we proceeded considering all the intersection points between the trajectories and applying a clustering algorithm, so that all the close intersections are clustered together into one single node. Then two nodes are linked if at least two intersections in each cluster belong to the same trajectory, i.e. if an aircraft has flown from a node to the other. We applied the same procedure also to the real historical trajectories in the same day of the dataset in order to obtain two comparable networks. Figure 1.44 panel (a) and (b) shows these two networks with the nodes weighted with the total traffic. It is evident that while the real trajectories are still clustered around the airways, since the ATC only provides minor changes to the flight plans, the optimal solutions exploit all the airspace in a more efficient and homogeneous way since the trajectories are scattered through all the available space. This is reflected also in the distribution of the traffic over the nodes (Figure 1.45 panel (c)), since the number of nodes with high strength is considerably lower in the global sub-optimal solution with respect to the real case. Figure 1.45 shows the distributions of the degree of the nodes (panel (b)) and the length of the links (panel (c)) for the two networks. Despite the fact the distributions are always peaked around a value, in the suboptimal case the distributions of the length of the links show a less fat tail towards
Figure 1.44: Real (a) and Optimal (b) navigation point networks in the Greek Airspace built using all the flights on the 9th of June 2011. The optimal network correspond to the conflict-free case with $\epsilon = 2$.

large values and the distribution of the degree is peaked around a value of about 6. Thus it is clear than the resulting network is some sort of triangulation of points distributed in the airspace with an almost homogeneous traffic load deployed over them.
Figure 1.45: (a) Inverse cumulative distribution of the strength of the nodes for the real and optimal Greek Navigation Point Networks. (b) Distribution of the degree of the nodes for the real and optimal Greek Navigation Point Networks. (c) Distribution of the length of the links for the real and optimal Greek Navigation Point Networks. The optimal network correspond to the conflict-free case with $\epsilon = 2$.

### 1.5.4 Efficiency of the Optimal Solution

The topological properties by themselves are not sufficient in order to understand if a different kind of planning is well-performing. In fact despite the fact that a global sub-optimal solution might be conflict-free, such solution is built in a situation in which many external factors are not taken into account. For example it is not possible to
forecast departure delays due to for example to reactionary delays or the occurrence of every adverse weather condition. In particular in our case none of these conditions have been considered in the optimization process and the solution found is valid in a idealized situation in which everything goes just as planned. Thus it is important to test how the planned routes are working when such external occurrences are taking place and the ATC model we developed is suited to give useful insights about this issue.

In particular we built a new simulation setup similar to that presented in 1.4.3 for the validation of the ATC model. In particular:

- Every aircraft flies according to a flight plan obtained with the EO algorithm for various values of $\epsilon$.

- The structure of the sectors is unvaried with respect to the one used in 1.4.3, but the airways structure is not considered. After every redirection and aircraft is sent directly to its destination following a straight line.

- Since the trajectories are already as straight as possible, directs in order to speed up the traffic are not considered.

- As in the previous case, controllers solve conflicts using the IN-OUT(FL) protocol.

- Capacity constraints are considered as in 1.4.3. The value of capacity are the same presented in table 1.7.

We tested the solution considering two kinds of external stochastic disturbances, i.e. the delays generating perturbed areas presented in 1.4.3 and random departure delays. In the second case we apply to each flight a departure delay from a uniform distribution in a range $[-\tau, \tau]$, where $\tau$ is a free parameter. For both cases we studied the response of the system in terms of generated delays and number of actions performed by the controllers as functions of $n_{ext}$ (the number of perturbed areas) and $\tau$ for different sub-optimal solutions obtained with different values of $\epsilon$. Moreover we performed the same studies, using the validation setup in 1.4.3 in order to compare such results to a case close to the actual situation. The most natural way in order to compare these different situations is, according to our model, the number of actions that have to be performed to manage the traffic. Despite the fact that some actions may require more effort for the ATC with respect to others, we consider them to be equivalent and use their total number as an indicator of the effort of the controllers.
1.5. GLOBAL OPTIMIZATION

Considering the departure delays it is evident that the current system is well-suited to manage this kind of occurrence, at least during the en-route part of the flight. In fact, despite the fact that every delay can cause congestion at the departure or arrival airport, the ATC does not care too much about the delay of an aircraft crossing a sector as long that capacity limits are not overcome. Figure 1.46 panel (a) shows the number of actions performed by the ATC (considering also the directs) as a function of the parameter $\tau$ of the delay distribution. Each value of the curve is averaged over many realizations of the same initial conditions. It is evident that such number is constant and it is not affected by the intensity of departure delays. On the other hand, the situation is quite different considering the sub-optimal planning. The number of actions of the trivial solution ($\epsilon = 0$) is still independent from $\tau$. As $\epsilon$ grows instead the curve presents a clear trend, growing from small values to a constant value for large $\tau$. This indicates that with small departure delays the systems does not need many actions to be managed correctly (no action is needed at all in the case of the conflict-free planning $\epsilon = 2$), while the number of action is always about the same as the departure delays get more intense. Note that the constant value for large $\tau$ depends on $\epsilon$ and it is lower as the planning approaches the conflict-free initial situation. Moreover, for any value of $\tau$ or $\epsilon$, the number of actions in the sub-optimal case is always lower than the constant value for the system in the current situation. Figure 1.47 panel (a) shows the average en-route delay generated by the ATC as a function of $\tau$ for the current situation and for the sub-optimal planning. It is true also in this case that this metric is independent from the external delay in the current situation. The average delay has a constant negative value of $\approx -80$ seconds. Again this is not true anymore for the sub-optimal solutions. While with $\epsilon = 0$ very small delays are generated, when $\epsilon > 0$ the ATC starts to generate negative delays again even though always smaller than the constant value of the current scenario.

Considering the other kind of disturbances, the system in the current situation is not able to function in the same way independently from their number (Fig. panel (b)). In this case the number of actions grows as the number of disturbances grows and it does not seem to approach any constant value. The behavior is the same also considering the sub-optimal solutions. For every value of $\epsilon$, the number of actions grows with the number of disturbances and it does not approach a constant value, although, the number of actions required to manage the sub-optimal cases is always smaller than the number of those needed in the current situation. Moreover, in this case their number gets smaller as $\epsilon$ grows. Figure 1.47 panel (b) shows the average en-route delay as a function of $n_{\text{ext}}$. With
the current planning the generated delay is not independent from \( n_{ext} \), but grows from \(-80\) seconds to values close to 0. This means that with a large number of disturbances the ATC is not able to speed up the traffic. The behavior of this quantity for the suboptimal planning with \( \epsilon = 0 \) is the same seen with the departure delays, since the generated delay is always very small and close to 0. If \( \epsilon > 0 \) instead the generated delays decrease as \( n_{ext} \) grows. Thus in this case the ATC is still able to reduce the travel time when trying to avoid perturbed areas.
1.6 Conclusions and Perspectives

In the current work we presented a data-driven model of Air Traffic Control capable of reproducing the modifications performed by controllers on the flight plans of the aircraft. The model has been developed analyzing historical traffic data in the European airspace in 2011. We used the data in order to infer the strategies adopted by controllers in order to avoid possible conflicts between aircraft as well as to validate the model. We then applied the model to the study of high traffic simulations and to the impact that specific conflict resolution strategies has to the trajectories in high traffic conditions. Finally we developed an optimization algorithm in order to obtain globally optimal flight plans and tested their efficiency with the ATC model. Our work contributes to enrich the scarce literature concerning air traffic in Complex Systems and the even scarcer literature concerning air traffic at the geographical level.

In chapter 1.2 we presented the datasets that we used and introduced the problem of air traffic nowadays. The most important type of data that we used is the historical data coming from the Demand Data Repository, provided us by Eurocontrol. This dataset
contains all the flights in the European Airspace in some months of 2011, both real trajectories and flight plans. Our dataset does not allow to spot conflicts between aircraft nor to have information about the meteorological conditions of each day in the dataset. Although it provides us information about the structure of the airspace. The other dataset is composed of safety events gathered in the same period of the traffic data. These events are Short-Term Conflict Alerts (STCA), alarms automatically triggered when a conflict is spotted in order to alert controllers. This data is very useful since provides information about the critical spots of the airspace.

In chapter 1.3 we presented the results coming from the analysis of the datasets, necessary to inform the development of the Air Traffic Control model presented in chapter 1.4. The main tool of our analysis is the Navigation Point Network. This network is built using the navigation points crossed by aircraft in the dataset as nodes, connecting them if at least an aircraft has flown from one to another in a considered period of time. Since we have both flight plans and real trajectories, we built a planned and a real navigation point network. Topological changes from one to another are the result of the action of air traffic controllers so we could use them to study their action. The real navigation point network is more homogeneous with respect to the planned one indicating that controllers try to redistribute the aircraft from highly trafficked areas to lower ones. Moreover while the planned navigation point network has a grid-like structure, the real one has an increased number of long link. This is the result of the direct assignment procedure, aiming at reducing the traffic load within a sector shortening the trajectories of some aircraft flying inside it.

The STCAs data was automatically gathered by a tool developed by Eurocontrol. Since many nuisance events (i.e. false safety events recorded), we proposed a filtering technique based on the matching between STCAs and traffic data. This technique allowed us to identify the pair of aircraft involved in each events, since the data has been anonymized. The STCAs has been gathered within the Rome ACC, which is about the half of the whole Italian Airspace. Using the matching with the traffic data, we have been able to assign each STCA to a link of the planned navigation point network built within this area in order to study the relations between topological properties of the network and the occurrence of STCAs. We defined two criticality measures for the links of the network and, using them, other two criticality measures for the nodes of the network. Between these metrics, the “intensive” ones has been found to be power-law distributed and then well suited to classify the nodes or the links of the network according to their criticality. We
studied the correlations between the criticality measures, static network metrics related to
the topology or the traffic deployment in the network and some dynamical metrics related
to the dynamics taking place over the network defined within the ELSA project [101]. We
found that the most critical nodes were also the ones with large static network metrics,
while the only dynamical metrics that was large in this case was the one connected with
vertical movements. On the contrary nodes with large horizontal movements were not
critical and with low network metrics. This indicates that critical areas are managed by
ATC using mainly vertical deviation since they are less costly in terms of management
effort. Horizontal movements are instead used in safe areas in order to speed up the traf-
fic. Despite the interest of the method, the analysis has been performed with a small set
of STCAs and in a limited part of the airspace. Increasing the number of STCAs could
lead to clearer patterns, since some criticality measures were affected by large estimate
erors. On the other hand the fact that the analysis was limited just to an airspace means
that its results may not be universal: different airspaces may have completely different
management strategies.

In chapter 1.4 we introduced an air traffic control model based on a dynamics taking place
on a planned navigation point network. In our model aircraft fly according to a flight plan
that is a certain path over a navigation point network. Whenever an aircraft reaches one
of the navpoint in its flight plan, controllers look for conflicts on the next link which are
computed geometrically. If a conflict is spotted the controller tries to apply a rerouting to
the aircraft without generating other conflicts. When no options are available, the conflict
occurs and no redirection is applied. Several conflict resolution strategies have been de-
fined in 1.4.1, 1.4.1 and 1.4.4. Initially the model has been tested in a synthetic airspace
with periodic boundary condition in order to avoid boundary effects. Then we tested on
a more realistic airspace in which boundary effects have been eliminated. In both cases,
as the traffic density increases the model undergoes to a transition from a phase in which
all the conflicts are solved to a phase in which many conflicts are not solved anymore.
The order parameter of this transition is the number of unresolved conflicts that has
been found to grow with the traffic density following a power-law. The exponent of this
power-law depends on the protocol used for conflict resolution, i.e. which combination
of strategies are used. On the other hand the scaling law of the transition curve is just
airspace dependent. In fact, every curve scales as $N^{0.43}$ where $N$ is the number of nodes
in the navpoint network. We then proceeded to validate the model by simulating full
day schedules of flight in different days and in different national airspaces. Following the
analysis performed in chapter 1.3, we defined a realistic management protocol including: a strategy of conflict resolution based mainly on vertical changes, capacity constraints for the sector and direct assignment to speed up the traffic. All the parameters of the model have been measured from the dataset. The model has shown a good agreement with the measures performed with the dataset in almost every airspace with the exclusion of the Spanish one. Further investigations are needed in order to understand the discrepancies in this case, but it is possible to argue that they are due to particular regulations within the airspace. In every case the introduction of external disturbances has been crucial in order to reproduce correctly the distribution of the en-route delays produced by the ATC. On the other hand their introduction is irrelevant for every other measure.

In chapter 1.5 we introduce an adaptation of the Extremal Optimization Algorithm [19,109] to the problem of optimal trajectory planning. Our algorithm performs an optimization of a set of flight plans within a chosen airspace (the Greek one in our case) modulating from having the shortest possible trajectories to having conflict-free trajectories, depending on a parameter in the cost function. For every value of this parameter we found that the obtained planned navpoint network is different from the current one, since the traffic was more homogeneously distributed in the airspace. Moreover the trajectories were close to being geodesics also in the conflict free solution. The topological study of this new airspace is not sufficient in order to obtain a complete comparison with the current one. Thus we adapted the ATC model in order to test its efficiency under external disturbances and departure delays. The functioning of the current system is independent from the presence and the intensity of departure delays, while its performances are lowered by external disturbances. The optimal solutions are instead affected by both delays and external disturbances but their performances are always better than those of the current system. Surprisingly the trivial solution of the optimization algorithm where all the trajectories follow a straight line from the origin to the destination is the one that is less affected by both kinds of perturbations. In par. 1.2.1, we introduce the SESAR scenario in which aircraft will fly in a less structured airspace according to business trajectories, reflecting the needs of the owner of the aircraft. The reduction of fuel consumption is usually one of the major concerns of Airlines and straight trajectories are surely one of the best options if strong winds do not affect the system. Thus our study suggests that they are also well performing in terms of predictability (small en-route delays) and effort needed to be managed (small number of ATC actions). Despite the simplicity of the approach, we showed that our model is suited to study and test new future scenarios.
Chapter 2

Human Mobility in a Urban Environment

2.1 Introduction

Many different areas of research have been concerned in the study of human mobility [111–119]. In relatively recent years in fact the growth of urban areas into metropolitan cities requires a better understanding of the way in which human interact with the urban environment in order to improve the city planning process and to build more functional cities [2]. In fact, a more accurate planning could ease the human processes in the urban areas and make them more efficient, leading to a global improvement in the cities’ welfare. Other problems such as the understanding of epidemic spreading phenomena [120] are also linked to human movements in urban and extra-urban areas. Despite the fact that many means of transportation can be used for relatively short movements, car is still one of the most popular and used one.

Physics has already given many contribution in understanding the process of diffusion of individuals in a certain environment particularly in the field of car movements. In this approach urban traffic has been modeled as a collective motion of cars, meant as a many-body system of self-driven particles [73, 74]. These studies have been possible in past years thanks to sensors collecting measures on traffic in fixed spots along streets. In this sense the description of the phenomenon was “Eulerian”. The main role of physics was and is to derive “universal laws” of collective movements, both from the analysis of experimental data and from the development of realistic models. An interesting example
of how physical research has contributed in the development of tools capable of aiding the diffusion process of humans in a road network is the comprehensive traffic forecasting system in the German state of North Rhine-Westphalia [121], which is based on the well-know cellular automata model of car traffic developed by Nagel and Schreckenberg [79].

In recent years a great diffusion of Information and Communication Technologies has taken place [8,9]. This led to an increase of the available data concerning human behavior at many levels and social sciences have benefited from all this new experimental data [122]. The study of human mobility and in particular of car movement is not an exception: the spreading of Global Positioning Systems (GPS) makes now possible to dynamically track and record the position of a large number of private vehicles (floating car data). Therefore we are now able to study the same system previously analyzed by means of fixed sensors in a new “Lagrangian” approach.

In the more general field of human mobility, many studies have already been proposed in order to understand which patterns can be spotted with this new kind of data. These studies are usually called trajectory based [123] and focus on the statistical characterization of the diffusion process of individuals [111]. The first example of this kind of studies is [124], where using banknotes movements the authors have been able to reconstruct the movement of individuals. In particular they where able to analyze the distribution of the spatial ($\Delta r$) and temporal ($\Delta t$) distances of two successive point in space visited by an individual. The diffusion process that they found was anomalous since the distribution was a power-law $p(\Delta r) \sim \Delta r^{-(1+\beta)}$ with $\beta = 2$. This suggested that the human mobility could be described by a Lévy Flight [125], already used to describe the movements of other animal species [126]. Moreover the probability of remaining confined in a small spatial region for a certain time is again a power-law, which attenuates the superdiffusive process. The authors conclude that such diffusion process can be described by a continuous random walk with two parameters.

The availability of cell phones movements allows to sample the trajectories of the individuals directly, without the need to reconstruct them. In [111] one of these samples (called Cell Call Data Records) has been used in order to verify the results of [124]. In this case the jump distances $\Delta r$ are distributed according to a truncated power-law $p(\Delta r) \sim (\Delta r_o + \Delta r)^{-\beta} \exp \frac{-\Delta r}{k}$, with $\beta < 1$, compatible with the one generated by a truncated Lévy Flight. However the study of the temporal evolution of the gyration radius $r_g$, which measures the dispersion of the visited locations with respect to their barycenter, showed how this quantity saturates as the time increases, while with a truncated Lévy
2.1. INTRODUCTION

Flight it should have been grown. Most importantly, the study of the conditional probabilities \( p(\Delta r | r_g) \) showed how these distributions rescaled by \( r_g \), collapse into a single one of the form

\[
p(\Delta r | r_g) = r_g^{-\alpha} F \left( \frac{\Delta r}{r_g} \right),
\]

(2.1)

with \( \alpha > 1 \) and \( F(x) \sim x^{-\alpha} \) for \( x << 1 \) and rapidly decreasing for \( x >> 1 \). Thus the diffusion process can still be considered a Lévy flight within the gyration radius \( r_g \) and the distribution \( p(\Delta r) \) can be explained as the result of this confined stochastic process combined with the heterogeneity of the population expressed by \( p(r_g) \) (which can be described with a functional form for similar to \( p(\Delta r) \)). In [113] the results found in [111] has been extended with new observations and with the use of a Complex Networks approach. A universal mobility model is built using continuous random walks as in [111], but noting that successive jumps are not unrelated. In fact each individual has an Individual Mobility Network, i.e. the network of the sites that were visited which are connected if at least a travel between them has been performed. Before each jump there is a probability inversely proportional to the number of sites in the network that the individual would visit a new site. This site will be attached to the network with a preferential attachment rule [127]. The model is capable of reproducing the anomalies in the diffusion process as well as the saturation in the number of visited location. Another notable example of the use of individuals trajectories to derive general laws is the application of entropic measures presented in [114]. In this work the authors found that, if the observation period is long enough, the average predictability of an individual’s movements is around 93%. Thus an algorithm that has the past trajectory of an individual as input can predict its next movement with an accuracy of 93%.

These studies concerning general human mobility have also been extended to vehicular movements. For example in [116] (using GPS tracks) an Individual Mobility Network has been introduced for this kind of movements. Surprisingly the study of the distribution of the degree of the nodes in the network showed how this was a power-law with an exponent of \(-1.6\), while the model proposed in [111] lead to an exponent of \(-2\). In [116] other disagreements between measures performed with GPS tracks and Cell Call Data Records, such as differences in the distribution of stop times due to the fact that this quantities are better defined with GPS data. Moreover a study of the relations between the Individual Mobility Network and the Interaction Network (i.e. the network of individuals connected if they have visited the same location at the same time) is presented.

In [128] entropic measures similar to the ones presented in [114] are applied to vehicular
movements. These measures show how human activities can be separated in three categories depending on the time at which they occur.

This part of the thesis is devoted to the continuation of the studies introduced in the Master Thesis of Pierpaolo Mastroianni [129]. This work concerned many aspects of human mobility studied by means of the analysis GPS data, aiming at integrating the results found in [115–119]. Our contribution to this work is the development of numerical models capable of reproducing and explaining some of the most interesting results of the data analysis. This part of the thesis is structured as follows:

Section 2 In section 2 we will present the results of the analysis in [129] that we use later for the development of the model. After a brief presentation of the dataset that has been used, we will present results concerning the relation between the traveled space in the urban network and the travel time and the study of the daily dynamics of drivers dividing them according to their number of daily trips. Since all these results come from [130] we invite to read the reference for the missing part of the analysis that we will not present here.

Section 3 In section 3 we will introduce the model we developed in order to explain the results of the analysis in chapter 2. The model consist of a grid network with shortcuts in order to simulate the urban environment and a navigation algorithm used to sample the paths of the drivers over it. We will show how this model can reproduce qualitatively many results found in the analysis, both concerning the relation between space and time and the daily dynamics.

Conclusions We will review the results and discuss possible future developments of the work.

2.2 Analysis of GPS Tracks

In this section we will recall the main results of the master thesis [129]. These results will be the basis the simplified model of urban network and route choice we will present in section 2.3. The analysis performed in [129] can be summarized in two main areas: space-time relations, i.e. the relation between the travel time and the traveled space of single trips or sequence of trips, daily dynamics, i.e. the characterization of the daily activity of car drivers analyzing the evolution of the average quantities that define their travel. All the pictures in the following section come from [129] and [130]. The section is structured as follows: in 2.2.1 we will present the analyzed dataset, pointing out the
2.2. ANALYSIS OF GPS TRACKS

accuracy of the measures and how to deal with its systematic errors; in 2.2.2 we will present the studies concerning the relations between the traveled space and the travel time of the GPS tracks, showing how the distribution of the travel time at fixed length has a universal scaling property; in 2.2.3 we will present the analysis of the daily dynamics of the drivers, perform dividing them according to the number of trips performed in a day.

2.2.1 GPS Data

The analysis has been performed using data gathered by OCTOTELEMATICS [131] on behalf of an insurance company in the month of May 2011 in the of the Rome district. This data are recorded thanks to a device installed in a certain percentage of private cars in Italy (2.5% of the population in the whole Italy, 4% in the Rome district), named OBU (On Board Unit). OBU is equipped with a GPS locator, and accelerometer for crash revelation and a GSM-GPRS module to communicate recorded data to the OCTOTELEMATICS service center.

This device is able to collect data about the movements of the car in which it is installed. Specifically it can record the position of the car during its trip with a certain resolution as well as the start and stop points of the trip. The resolution depends on the kind of road that the car is traveling: the position is recorded every 30 s on freeways and every 2 km on regular roads. Each record contains the position of the car, the corresponding time and the status of the engine (i.e. if it has been turned on, turned off or if the car was moving at the time of the record). Since the quality of the measurements are highly affected by the quality of the signal, a measure of such quality is also recorded by the device. The error associated to the recording time is always negligible (around 1 s), while error on the position must be treated carefully. The Global Positioning System (GPS) can establish the position of the device using its relative distance from a certain number of satellites. Thus the precision of the determined position depends on the precisions of these distance measurements. There are two kinds of errors affecting such measurements. The first kind is systematic errors (time errors, tropospheric and ionospheric refraction, orbit errors). Systematic errors can be usually cured by means of differential (using the known position of a station on earth) or relative (determining the coordinates of the vector linking two receivers). The other kind of errors is observational errors (multipath, i.e. signals reflected or deviated by some obstacle that show a longer path, problems in the electronic components of the device, electromagnetic interferences, changes in the
antennas phase center) and can be fixed only with data filtering techniques. The quality of the signal is recorded by UBO counting the number of visible satellites. In the better case (when 3 or more satellites are visible) the error in the determination of the position of the car is around 10 m. In every case, the measures of the distances between consecutive recorded points and instantaneous speed are reliable despite the many error sources, since they are computed using GPS data every second (such data however is not stored by the device).

The analysis has been performed over 13527 vehicles in 20 working days (weekends have been excluded as well as the last two days of the months in order to have the same numbers of each day of the week). The vehicles are selected so that their start and stop positions have been recorded inside the Rome municipality, so they are likely to be Rome inhabitants. Meaningless trips shorter than 10 m have also been deleted. Two consecutive travels of the same car have also been merged together into a single one if the time difference between the end of the first (i.e. when the engine has been switched off) and the beginning of the second (when the engine has been turned on) is less than 5 min. This is a common procedure in literature [115, 118, 119] and different choices do not affect the results found in [129] since this kind of trips are just the 16% of the entire sample. Beside this concatenation other operations have been made in order to eliminate undesired sample issues. For example, in order to eliminate problems due to underground parking, if the time lapse between two consecutive points with engine on is more than 30 min, those points are respectively the end of a trip and the beginning of a new one. Other illogic sequences of points have been excluded, like sequences in which the engine was turned off in the first point, was on in the second and was turned on in the last one.

### 2.2.2 Space-Time Relation

The central quantity of the investigations performed in [129] is the conditional probability \( p(t|l) \), which represent the experimental probability that a trip with length \( l \) is performed in a time \( t \). According to our measurements this distribution can be described by a log-normal distribution [132] (figure 2.2). Despite the fact that the Kolmogorov-Smirnov test presented in [99] exclude that these distributions are exactly lognormal, such approximation has been useful in order to understand its scaling properties. The parameters of the lognormal distribution describing \( p(t|l) \) will be in general function of \( l \), so we can write

\[
p(t|l) = \frac{1}{\sqrt{(2\pi)\sigma(l)t}} \exp\left(-\frac{(\log t - \mu(l))^2}{2\sigma^2(l)}\right),
\]

(2.2)
2.2. ANALYSIS OF GPS TRACKS

Figure 2.1: (Left) Dependence of the experimental mean value \( \log t(l) \) of the logarithm of travel time \( t \) upon travel length \( l \) (red points). The solid blue line is the best fit logarithmic function (2.5). From the fit we obtain the value \( \alpha = 0.672(3) \). (Center) \( \sigma(l) \) calculated as a function of \( l \). It is clear that after \( l = 1 \text{ km} \) \( \sigma(l) \) begins to fluctuate around the constant value \( s \). (Right) Experimental average travel time \( \bar{t}(l) \) as a function of \( l \) (red points). The solid black line refers to the function in (2.8).

were \( \mu(l) \) is the average of the logarithm of the travel times of all the trips of length \( l \),

\[
\mu(l) = \frac{1}{N_l} \sum_{i=1}^{N_l} \log t_i(l)
\]

and \( \sigma(l) \) is the corresponding standard deviation of the logarithm

\[
\sigma(l) = \frac{1}{N_l} \sum_{i=1}^{N_l} (\log t_i(l) - \mu(l))^2.
\]

In equations (2.3) and (2.4) \( N_l \) is the total number of trips with length \( l \) and \( t_i(l) \) is a the travel time of the \( i^{th} \) trip of length \( l \). The average \( \mu(l) \) has found to be linearly growing with the logarithm of \( l \) as it is shown in figure 2.1 left panel. Thus its relation with \( l \) can be written as

\[
\mu(l) = \alpha \log \left( \frac{l}{l_0} \right),
\]

with \( \alpha = 0.672(3) \) and \( l_0 \approx 10^{-3} \text{ m} \). On the other hand \( \sigma(l) \) has a different behavior (figure 2.1 center panel ). After an initial decrease for small value of \( l \), \( \sigma(l) \) fluctuates around a constant value \( s = 0.41(3) \). Thus it is possible to write equation (2.2) with the
explicit dependence of $\mu(l)$ and $\sigma(l)$ from $l$:

$$p(t|l) = \frac{1}{\sqrt{(2\pi)s^2}} \exp\left(-\frac{(\log t - \alpha \log(l/l_0))^2}{2s^2}\right).$$  \hfill (2.6)

According to the properties of the lognormal distribution, we can relate $\mu(l)$ and $\sigma(l)$ to the average and the variance of the travel time:

$$\overline{t(l)} = e^{\mu(l) + \sigma^2(l)/2}$$

$$\text{Var}[t(l)] = e^{2\mu(l) + \sigma^2(l)}(e^{\sigma^2(l)} - 1).$$  \hfill (2.7)

Thus combining equation (2.7) with (2.5) and $\sigma(l) = s$, the average value of the travel time grows as a power of $l$ with exponent $\alpha$ (figure 2.1 right panel):

$$\overline{t(l)} = e^{\mu(l) + \sigma^2(l)/2} \sim l^\alpha.$$  \hfill (2.8)

The power-law growth of $\overline{t(l)}$ with $l$ indicates that the speed $v = \frac{l}{t(l)}$ grows on average with $l$ in a non-linear fashion:

$$\overline{v(l)} = \frac{l}{\overline{t(l)}} = \frac{l}{\overline{t(l)}} \sim l^{1-\alpha}.$$  \hfill (2.9)

Note that $1 - \alpha = 0.328(3)$, so as the trip length becomes longer the average speed also grows, indicating that the optimization of the path performed by car drivers improves as the path becomes longer. Combining the equations (2.5), (2.8) and $\sigma(l) = s$ we can write equation (2.6) in the form:

$$p(t|l) = \frac{1}{\sqrt{(2\pi)s^2}} \exp\left(-\frac{(\log t + \sigma^2(l)/2)^2}{2s^2}\right).$$  \hfill (2.10)

From equation (2.10) it is evident $p(t|l)$ depends on $l$ only through $t(l)$. This suggests that, with a proper redefinition of the variable $t$, it is possible to eliminate such dependence. In fact defining $\tau = \frac{t}{t(l)}$ we can write equation (2.10) in a universal form $p(\tau)$:

$$p(\tau) = \frac{1}{\sqrt{(2\pi)s^2}} \exp\left(-\frac{\left(\log \tau + \sigma^2/2\right)^2}{2s^2}\right).$$  \hfill (2.11)

This scaling property implies the existence of a universal mechanism that governs the fluctuations of travel times around its mean value for different travel length.
2.2. ANALYSIS OF GPS TRACKS

Figure 2.2: (Left) Experimental frequency distribution $p(t|l)$ for different $l$-values. Solid lines refers to the fitted lognormal distributions. (Right) Experimental frequency distribution of the scaled variable calculated at different value of $l$, as expressed in the key. The solid black line is the curve in Eq.10. The graph shows clearly the existence of the scaling property.

These relation has also been tested in different hour of the day in order to see if it depends on different traffic conditions. The dataset has been split in some time slots following [133]. These slots separate different times of the day (from 6AM to 10AM: morning slot, from 10AM to 4PM: early afternoon slot, from 4PM to 8PM: evening slot and from 8PM to 6AM of the next day: night slot) and the same measures are repeated for each one of them. It is clear that in this way it is possible to separate high traffic conditions during the daily hours from low traffic conditions during night time.

The conditional probability $p_h(t|l)$ for each time slot $h$ is again lognormal like (figure 2.3). It is evident from figure 2.3 that the distributions $p_h(t|l)$ during morning and evening are shifted to the left with respect to the remaining two distribution. This is the effect of the higher traffic that is present on the urban network in these parts of the day. Since $p_h(t|l)$ resembles, for every $h$, $p(t|l)$ it is possible to define a functional form similar to equation (2.2) by substituting $\mu(l)$ and $\sigma(l)$ with $\mu_h(l)$ and $\sigma_h(l)$, i.e. with the same quantities measured using only the track in the considered time slot $h$. Again, for every slot $h$, $\mu_h(l)$
grows logarithmically with $l$ while $\sigma_h(l)$ is asymptotically constant (figure 3): 

$$
\mu_h(l) = \alpha_h \log\left(\frac{l}{l_0}\right),
$$

$$
\sigma_h(l) = s_h.
$$

(2.12)

Therefore the functional relations obtained for the aggregated sample are robust against the separation of the sample in time slots. However, the value of the parameter $\alpha_h$ depends on the time slot (table 2.1). Higher value of $\alpha_h$ (and thus lower growth of the average velocity according to equation 2.9) can be found in the two most congested time slots during early morning and evening, while the minimum value is during the night. On the contrary the value of $s_h$ is quite constant in every time slot. All these features indicates that the universal behavior of $p_h(t|l)$ is independent from the considered time slot. Thus it is possible to define again $\tau_h = t/t(l)_h$ using equation (2.7) to write $t(l)_h$ as a function of $l$. In this way we can again rescale every $p_h(t|l)$ into a universal $p_h(\tau_h)$ (figure 2.5). Finally in [116] has been argued that the velocity of the cars should follow a Wiener process. All the values of $\alpha_h$ could be in agreement with this hypothesis, since they are close to the value $\frac{2}{3}$. However, if this hypothesis was true, we should observe that travel length $l$ should grow as a power of $t$ with exponent $\alpha^{-1} = \frac{3}{2}$. Figure 2.6 shows the average travel
2.2. ANALYSIS OF GPS TRACKS

Figure 2.4: (Left panel) Dependence of the parameter $\mu_h(l)$ on $l$ for different time slots $h$. The logarithmic scale of the $x$ axis suggest the functional relation (2.5) in the main text. (Right panel) Dependence of the parameter $\sigma_h(l)$ on $l$ for different time slots $h$. After 1 km the values of $\sigma_h(l)$ become constant.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$\alpha_h$</th>
<th>$s_h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 – 10</td>
<td>0.670(4)</td>
<td>0.40(9)</td>
</tr>
<tr>
<td>10 – 16</td>
<td>0.633(4)</td>
<td>0.40(5)</td>
</tr>
<tr>
<td>16 – 20</td>
<td>0.666(4)</td>
<td>0.37(3)</td>
</tr>
<tr>
<td>20 – 6</td>
<td>0.616(3)</td>
<td>0.39(7)</td>
</tr>
</tbody>
</table>

Table 2.1: Dependence of the value of $h$ and $h$ on the time slot, calculated by fitting experimental points with equations (2.5) and $\sigma(l) = s$.

length $l$ as a function of the travel time $t$. Since the travel length grows almost linearly with the travel time, it can be excluded that these observations are the results of a Wiener process. In section 2.3 we will see that this behavior is reproduced by a simplified model of urban network with a navigation process used to sample the paths over it.
CHAPTER 2.

Figure 2.5: Experimental frequency distribution of the scaled variable calculated on the time-split sample. It is clear that the curves tend to collapse on the same curve, regardless of the time separation.

Figure 2.6: Average travel length $l(t)$ as a function of travel time measured. Solid blue line is a linear function of travel time $t$. 
2.2.3 Daily Dynamics

The paths of the car drivers are highly regular and auto-similar so they can be easily predicted [114,128]. Recent studies showed how individuals tend to move in a limited number of “habitats” [134]. Thus the following results concern the characterization of the daily patterns of driver, dividing each trip according to the number $n$ of stops performed during a day. Note that a trip is defined as the travel performed by a driver from a point in our dataset in which the engine was turned on and the successive point in which it was turned off. Moreover, two consecutive trips are concatenated into a single one if the stopping time between the end of the first trip and the beginning of the second one is smaller than 5 min.

Parabolic Pattern of Trips

By dividing the car drivers by the number $n$ of trips performed in each day it is possible to define the quantity $l_k^n$ i.e. the distance traveled during the $k^{th}$ trips. We can then compute $\overline{l_k^n}$ by averaging over all the $k^{th}$ of the drivers with $n$ movements. This operation of average might hide the differences inside the sample, but as we will see, it makes regular patterns emerge from the data. If the way in which each driver choses the sequence of stops performed during a day would be random, the succession of $\overline{l_k^n}$ would not depend on $k$. Instead they follow the behavior in figure 2.7 panel (a). For every value of $n$, the succession of $\overline{l_k^n}$ strongly depends on $k$. Moreover the succession of $\overline{l_k^n}$ seems to be particularly regular. In fact, the succession is symmetric around $k = \frac{(n+1)}{2}$ when the minimum value of the succession is reached. Thus $\overline{l_k^n}$ decreases starting from $k = 0$ until $k = \frac{(n+1)}{2}$ is reached, then starts growing again until $k = n$. Note that $\overline{l_0^n}$ and $\overline{l_n^n}$ are the highest values of the succession and their values become more similar as $n$ grows. This suggest that for each $n$, $\overline{l_k^n}$ can be approximated by a parabola, whose coefficients depend on $n$:

$$p_n(k) = a_n k^2 + b_n k + c_n.$$  \hfill (2.13)

For every value of $n$, $p_n(k)$ is a parabola with symmetry axis orthogonal to the $k$ axis. Fitting $p_n(k)$ using the succession $\overline{l_k^n}$ with the corresponding value of $n$, the coefficients $a_n$, $b_n$ and $c_n$ has grows with $n$ in a power-law fashion as can be seen in figure 2.7 panel.
(b). Thus it is possible to write:

\[ a_n = An^{-\eta_a} \]
\[ b_n = Bn^{-\eta_b} \]
\[ c_n = Cn^{-\eta_c}. \]  

where \( A = 8.8(7) \text{Km}, \ B = 21(1) \text{Km}, \ C = 24.5(8) \text{Km}, \ \eta_a = 2.00(4), \ \eta_b = 1.33(4) \) and \( \eta_c = 0.55(2) \). Thus, combining equations (2.13) and (2.14),

\[ p_n(k) = An^{-\eta_a}k^2 + Bn^{-\eta_b}k + Cn^{-\eta_c}. \]  

Dividing each members of (2.15) by \( Cn^{-\eta_c} \), we get

\[ \frac{p_n(k)}{c_n} = \frac{A}{C}n^{-(\eta_a-\eta_c)}k^2 + \frac{B}{C}n^{-(\eta_b-\eta_c)}k + 1. \]  

Note that with the estimated values of the parameters,

\[ \eta_a - \eta_c = 1.45(4) \]
\[ \eta_b - \eta_c = 0.78(4). \]  

This indicates that, within the estimated errors, the following relation holds

\[ \eta_a - \eta_c = 2(\eta_b - \eta_c). \]  

The relation (2.18) allows to derive a scaling law for equation (2.15). In fact

\[ \frac{p_n(k)}{c_n} = \frac{A}{C}n^{-(\eta_a-\eta_c)}k^2 + \frac{B}{C}n^{-(\eta_b-\eta_c)}k + 1. \]

Then it is possible to define a new variable \( \hat{k} = n^{-(\eta_b-\eta_c)}k \), obtaining a universal expression for the parabola independent of \( n \)

\[ p(\hat{k}) = \frac{A}{C}\hat{k}^2 + \frac{B}{C}\hat{k} + 1. \]  

This indicates that rescaling all the values of \( \ln k \) by \( c_n \), all the data would collapse over the curve (2.20). Figure 2.8 shows the collapse of the data for all the \( \ln k \) successions. The variability of the data, combined with the parabolic approximation used, does not provide a very precise collapse but the indication of a universal scaling law still holds.
2.2. ANALYSIS OF GPS TRACKS

Figure 2.7: (a) Successions $l_n^k$ for some values of $n$. (b) Successions of the coefficients of the parabolic fit $p_n(k)$ for $n \geq 4$. Continuous lines are power law fit of the successions.

Similar properties can also be found for the succession of the euclidean distances between subsequent stops $d_n^k$ and for the succession of the travel times $t_n^k$. Their corresponding parabolas are shown in figure 2.9.

Despite the approximation of the parabolic fit, some general features can be identified for all the previously presented successions. Independently of the value of $n$, the first value of the succession with $k = 1$ is the largest one. Apart from $n = 2$ and $n = 3$, all the successions are initially decreasing until a certain value of $k$, say $k_n^*$, is reached.

---

\[ l_n^k \]

\[ a_n, b_n, c_n \]

---
Figure 2.8: Successions $\overline{l}_k^n$ divided by $c_n$ as functions of $\hat{k}$.

Around $k = k_n^\ast$ the successions reach a minimum, then for $k > k_n^\ast$ they start growing with $k$. As $n$ grows, the value of $k_n^\ast$ gets closer to the central value of the succession $k = \frac{n+1}{2}$. The dynamics of the individual is certainly related to the circadian rhythm. Thus the emergence of these parabolic pattern are probably related to the constraint of going back to the origin of the trip. In [129] in fact has been shown as the distance between the first and the last stop for each value of $n$ is on average much shorter than the distance between the initial position and the others. This distance is of the order of 1 km, indicating that despite the fact that a driver usually goes back to the origin in its last movement, the parking spot is close but does not coincide with the initial one. The fact that the distance between the beginning and the other stops was usually constant suggested two kind of possible dynamics for the choice of the succession of the stops: an orbital dynamics and a bipolar dynamics (figure 2.10). In the orbital dynamics the stops are disposed at a constant radius around the origin. In the other case the driver performs a long movement and then a series of shorter movements. These movements are so that the distances between the points they connect and the origin are quite constant so that they are displaced on average over a circular arc centered in the origin. The last movement is again a long one in order to get back to the origin.

In order to assess which one of these dynamics is dominant, an adimensional metric has been defined

$$\rho_n = \frac{R_n - d_{c,h}}{R_n + d_{c,h}},$$

(2.21)
where $d_{c,h}$ is the distance between the origin and the barycenter of the position of the central stops of the parabolas. $R_n$ is instead the radius of gyration on the central stops of the parabolas, which measures their dispersion with respect to the barycenter. If $\rho_n < 0$ then the bipolar dynamics is the dominating one, while the orbital dynamics dominates otherwise. It has been shows that, for $n > 5$, $\rho_n$ fluctuates around a value of 0.209(1). So when the successions $\overline{d_k}$ are more parabola-like the orbital dynamics is dominant. This result will be crucial in the definition of a model that could explain the emergence of the
Figure 2.10: Representations of an orbital (a) and bipolar (b) dynamics for a driver performing 4 stops. In (a) the radius of gyration $R_5$ is comparable to the distance between the origin and the barycenter $d_{c,h}$ since the points are rather scattered, thus leading to $\rho_5 > 0$. On the contrary in (b) the points are quite clustered around the barycenter and thus $\rho_5 < 0$.

parabolic structures that we will present in 2.3.3.

2.3 Grid Model with Shortcuts

The analysis so far pointed out as many interesting patterns emerge from the dynamics of the car drivers, both considering all the data aggregated together and dividing the drivers in categories based on the number of visited locations. Since such patterns are due to the complex interplay between the human activity and the urban environment, the only way to investigate further and to understand this mechanism is through modeling. Driven by some of the results of the analyses and by some well-known properties of the urban network, we will propose a model in order to explain how drivers move in cities and how the environment affect their behavior. We will see that this model is capable to reproduce many qualitative findings of the previous chapter.
2.3. GRID MODEL WITH SHORTCUTS

2.3.1 Definition of the Model

The results of the previous analysis emerge from the interplay between the dynamics of the car drivers on the network and the structure of the network itself. In particular equation (2.9) suggests that the drivers are able to move in the network in an efficient way and their knowledge of the environment is so that their optimization of the travel time improves as the path becomes longer. A first attempt to model their behavior is to assume that, to go from a selected origin to a certain destination on the urban network, each driver can always choose the path that minimize the travel time. This assumption allows us to focus on the structure of the network. Considering the behavior observed with equation (2.9), it is possible to argue that the presence of arterial roads plays a crucial role in the improvement of temporal optimization as the travel distance increases. In a certain way the function of this roads is to reduce the (temporal) distance between regions of the city that otherwise would be hard to reach in a reasonable time. So this roads resemble the shortcuts of small-world networks, with the difference that they can be traveled only in a finite time. Thus we model the urban network as a grid with shape $L \times L$. Each link of the grid has unit length and can be traveled with the same unit speed. The shortcuts are modeled as links connecting unconnected nodes in the grid. The length of each shortcut is assumed to be the euclidean distance between the connected nodes and each link can be traveled with a speed $v > 1$, that can be either a fixed value or taken randomly from a chosen distribution. Both the links of the grid and the shortcuts have an assigned travel time given by the ratio between the length of the link and its associated speed. This assumption on the structure of the road network is in accordance with the results of [29]. In particular the model reproduces correctly the effect of having a dimensionality larger than 2 when the temporal metric is considered in the computation of the reachable nodes. Following the results in [29], we measured the dimensionality of the network with a fixed number of shortcuts as the slope (in logarithmic scale) at the inflection point of the cumulative number of isochronous paths $N_t$, i.e. the number of paths reachable from the center of the network within a time $t$ (figure 2.11 right panel shows of example of such measure of dimension). The dimension of the network is presented in figure 2.11 left panel as a function of the number of shortcuts for some values of velocity on them. The dimension is larger than 2 with small values of $N_{shortcuts}$ and grows until it reaches a quite stable value with is larger than 2.8 and depending on $v$, similarly to what has been found for real urban networks [29].
Figure 2.11: (Left) Dimension of the network computed as the slope of $N_t$ (i.e. $N \sim t^\delta$) at its inflection point as a function of the number of shortcuts for a $100 \times 100$ grid. (Right) Cumulative number of isochronous path $N_t$ as a function of the time $t$ for a $250 \times 250$ grid with 100 shortcuts and shortcut velocity equal to 2. The blue line is the slope at the inflection point of the curve.

Navigation Algorithm

The scaling of equation (2.9) suggests that the drivers are able to move in the network in an efficient way and their knowledge of the environment is so that their optimization of the travel time improves as the path becomes longer. A first attempt to model their behavior is to assume that, to go from a selected origin $s$ to a certain destination $t$ on the urban network, each driver can always choose the path that minimize the travel time. This assumption allows us to focus on the structure of the network. If the number of shortcuts in the grid is 0 there is no difference between the spatial and temporal dimension. On the other hand if such number is larger than 0, space and time are not equivalent anymore since the shortcuts are traveled at higher speed. Thus for the computation of the shortest paths we weight each link on the grid with the travel time needed to go from a node to the other one that it connects. Thus a shortest-path is the one that minimizes the travel time needed to go from $s$ to $s'$. We will see with this simple choice the model can reproduce many of the findings of the analysis part. However some disagreements still remain indicating that some kind of noise must be introduced in the sampling process. As we will show in the following the variance of the logarithm of the travel time decreases for large $l$ if the paths are sampled using global shortest-paths (Fig. 2.13 center panel),
while instead they should be constant (Fig. 2.1). We defined a navigation algorithm in order to build the path from $s$ to $s'$ in a slightly different way with respect to the global optimum path. Starting from $s$, the navigation algorithm proceeds as follows:

- Assuming that the current node visited by the algorithm at its $i^{th}$ step is $n_i$, we chose an optimization distance $l_{optim}$ from a uniform distribution in $[3, l(n_i, s')]$, where $l(n_i, s')$ is the euclidean distance between $n_i$ and $s'$.

- The next visited node $n_{i+1}$ is chosen randomly between all the nodes whose distance from $n_i$ is less than $l_{optim}$. Moreover the angle between the lines connecting $n_i$ and $n_{i+1}$ and $s$ and $t$ is smaller than an assigned value $\alpha$ ($\alpha = 30^o$ in the following).

- If $t$ satisfies the conditions in the previous point, it is automatically chosen as $n_{i+1}$

- If $n_{i+1} = s'$ the process ends.

Once a sequence of nodes $\{s, n_1, n_2, \ldots, s'\}$ is made with the algorithm, the path connecting $s$ and $t$ is built by concatenating all the shortest paths connecting each $n_i$ with its successive node $n_{i+1}$. The choice of a new node with the navigation algorithm is shown in figure 2.12.

Figure 2.12: Picture of the algorithm used to chose the intermediate steps. The yellow circle sectors are the areas in which the driver can choose his next step.
2.3.2 Reproducing the Space-Time Relations

In 2.2.2 we have seen that the dependence on $l$ of conditional probability $p(t|l)$ can be eliminated rescaling the time variable using the average travel time at fixed length in equation (2.8). This property is due to the fact that the average of the logarithm of the travel time $\mu(l)$ grows with $l$ following a power-law while its fluctuations $\sigma(l)$ are constant with $l$ (figure 2.1).

Starting with the simplest navigation process, i.e. assuming that we assume that they always go from one point in the network to another one following the shortest path on the network, we can sample the possible paths on the network. This sample is performed by choosing a random starting node $s$ and a random origin node $t$. After we have collected a sample of paths over a grid model with a certain number of shortcuts with an assigned velocity $v$, we can check if the behavior of $\mu(l)$ and $\sigma(l)$ are reproduced. Figure 2.13 left and right panels shows these two quantities as a function of $l$ for a grid network with $N_{\text{shortcut}} = 100$ and $v = 3$. We can see that the power-law relation between $\mu(l)$ and sigma is correctly reproduce with an exponent $< 1$ similar to the one found with the dataset. On the other hand the behavior of $\sigma(l)$ does not agree with the one of the real case, since it clearly decreases as $l$ grows due to the finite size of the grid reducing the number available paths connecting distant locations. This suggest that while the short paths are more sensitive to fluctuations, the long ones are not and the distribution of the travel times defined in (2.11) becomes sharper for these path lengths. To recover the correct behavior of $\sigma(l)$ it is possible to argue that the drivers are certainly able to chose the shorter travel time between all the available, but the optimization is not performed globally but using instead intermediate steps. For this reason we have introduced the navigation algorithm presented in 2.3.1. Using the algorithm to construct the paths on the network, the fluctuations due to the short tranches of each path prevents $\sigma(l)$ from decreasing for large values of $l$ (Fig. 2.13 center panel), while preserving the power-law behavior of $\mu(l)$ of equation (2.5) (Fig. 2.13 left panel). Despite the fact that the distributions $p(t|l)$ can not be approximated with lognormal distributions (the Kolmogorov-Smirnov test performed on $p(t|l)$ for each value of $l$ rejects this hypothesis), the scaling property observed in 2.2.2 still holds (figure 2.14). The previous pictures concerned just the case of the grid network with $N_{\text{shortcut}} = 100$ and $v = 3$ but the relations found are valid for every value of these parameters. We have seen that the power-law behavior of $\mu(l)$ is present with both the optimization algorithms, but in the exponent is usually lower with the second one using the same number of shortcuts and velocity on them. The trends of the exponents of the
2.3. GRID MODEL WITH SHORTCUTS

Figure 2.13: (Left) Dependence of the mean value of the logarithm of travel time $t$ upon travel length $l$ (red points). The solid blue line is the best fit logarithmic function (2.5). From the fit we obtain the value $\alpha = 0.63$. (Center) Dependence of the mean value of the logarithm of travel time $t$ upon travel length $l$ (red points). The paths are sampled by choosing some intermediate stops between the origin and the destination nodes. The solid blue line is the best fit logarithmic function (2.5). From the fit we obtain the value $\alpha = 0.62$. (Right) $\sigma(l)$, calculated as in (2.5), as a function of $l$. Green points corresponds to paths sample using intermediate optimization steps between the origin and destination nodes of each path, red points correspond to the case in which these intermediate steps are not used. Each panel refers to measure performed on a $100 \times 100$ grid with $N_{\text{shortcut}} = 100$ and a fixed speed on each shortcut $v = 3.0$. 

power-law are presented in figure 2.14 for both the optimization processes. In both cases after an initial decrease of its value and the approach to a minimum, the exponent starts growing and eventually approaches the value of 1 for large values of $N_{\text{shortcuts}}$, when the shortcuts dominate respect to the links of the underlying grid. As a further validation of the model, we reproduced the sampling process also for the grid network with a realistic distribution of speed over its links and for the Rome Urban Network.
Figure 2.14: (Left) $p(t|l)$ distribution for some values of path length $l$. Solid lines refers to the fitted lognormal distributions. (Right) Distribution of the scaled variable calculated at different values of $l$. The solid black line is the curve in (2.11). All the paths are sample using intermediate steps on a $L \times L$ grid with $N_{\text{shortcut}} = 100$ and a fixed speed on each shortcut $v = 3.0$.

Figure 2.15: Exponents of the relation between the average logarithm of travel time and traveled distance for the grid model with shortcuts. In the left panel the paths have been sampled using the global shortest path between the starting and arrival point, while in the right panel the paths have been sampled considering random intermediate steps between the starting and arrival point. The picture refers to a $100 \times 100$ grid. Paths have been sampled using the global shortest path between the starting and arrival point.
Grid Model with Realistic Speed Distribution

The choice of the velocities assigned to the link of the grid network is completely arbitrary. It has been shown that the main ingredient needed to reproduce its just the presence of shortcuts faster than the regular links of the grid. At this point the model has two free parameters, say the number of shortcuts and the ratio between the velocity of the links of the grid and the velocity of the shortcuts, but using data regarding the real average velocity of a urban network it is possible to fix one the parameters as a function of the other. Figure 2.16 shows the distribution of the average velocities of the roads of the Rome Road Network \( p(v) \), where the velocities have been scaled so that the average speed is equal to 1 (adim.). Such distribution is bell-shaped with a large tail towards high speeds. Thus it is possible to argue the velocities in this tail should correspond to the shortcuts of the network that should be usually traveled at a speed higher than those of regular streets. Having fixed \( N_{\text{shortcut}} \) on a \( L \times L \) grid, we define the fraction of regular links \( f = \frac{2L(L-1)}{N_{\text{shortcut}}+2L(L-1)} \), where \( 2L(L-1) \) is the number of links in the grid. It is then possible to define \( v_p(f) \), i.e. the \( f^{th} \)-percentile of the distribution of the velocities as a function of \( N_{\text{shortcuts}} \) (figure 2.16). Considering each link of the grid we can assign to it a speed \( v \) randomly chosen from \( p(v) \) so that:

- \( v < v_p(f) \) if the link belongs to the grid,
- \( v \geq v_p(f) \) if the link is a shortcut.

In this way we obtain the desired speed distribution for the links of the network and we assure that all the shortcuts have a large speed. The results with this choice of velocity distribution confirms what we have found in the fixed velocities case. Figure 2.17 shows the average of the logarithm of the travel time as a function of the traveled distance, also in this case we have a power-law dependence of these quantities with an exponent smaller than 1. Moreover the fluctuations of the logarithm of the travel time shows a behavior which is more similar to the real case. When the velocities over the grid and the shortcuts are fixed, the fluctuations are constant and stable as the traveled distance \( l \) grows. Although for small values of \( l \) it is unlikely to travel through a shortcut and thus the regular Manhattan paths are the most used. This means that, just before the fluctuations become constant, there is a transient in which no power-law dependence between the \( \log(t) \) and \( l \) is observed and the fluctuations are growing from 0 to their final constant value. This behavior is quite different from what has been observed in real-world
Figure 2.16: (Left) $v_p(f)$ as a function of the fraction of regular slow links $f$ for the Rome Urban Network. (Right) Frequency distribution of the velocities $p(v)$ for the Rome Urban Network. The vertical line indicates the values of $v_p(f)$ for $N_{shortcuts} = 700$ in a $100 \times 100$ grid. The velocities of the shortcuts are taken from the filled part of the distribution above $v_p(f)$.

data (figure 2.1), since in that case fluctuations seems to grow as $l$ approaches 0, probably due to the noise induced by the different velocities of the streets and street regulations which affects short trips more than long ones. Having introduced stochastic velocities over the links of the grid partially reproduces this behavior. Figure 2.17 shows the difference between $\sigma(l)$ for the grid with fixed velocities and $v = 3$ over the shortcuts and the case with realistic velocity distribution. In the latter case, there is still a transient in which the fluctuations are not constant, but the fluctuations grow for small values of $l$ as in the real case showing a minimum between these values and the range in which they are constant. Since such minimum is not evident in real-world data, it is possibly due to the lack of regulations that should increase the level of noise in that range of travel length that have not been included in the model at this stage.
2.3. GRID MODEL WITH SHORTCUTS

Figure 2.17: (Left) Dependence of the mean value $\log(t(l))$ of the logarithm of travel time $t$ upon travel length $l$. These points have been obtained by sampling the paths using intermediate steps on a $100 \times 100$ grid with $N_{\text{shortcut}} = 100$ and a realistic velocity distribution. In this case we have $f = 0.995$ and $v_p(f) = 3.30$. (Right) $\sigma(l)$ for the same grid with realistic velocity distribution (red) and fixed velocities with $v = 3$ on the shortcuts (green).

Navigation Algorithm in the Rome Urban Network

Using a realistic velocity distribution increased the realism of the model and made possible to fix one of the two free parameters of the grid model. Since all the results of the simpler case with fixed velocities have been confirmed, we can study what happens if we apply the navigation algorithm in an even more realistic urban network. We applied the navigation algorithm in two similar urban networks of Rome, one coming from the combination of OpenStreetMap and Google Maps databases (for sake of simplicity we will refer to the network built with this data as the GoogleMaps Network) [135] and another coming from the TeleAtlas database [136]. The OpenStreetMap/GoogleMaps dataset is a list of roads whose extremities (crossroads) are indicated by their coordinates. Each road has an average traveling time assigned. The coordinates of the crossroads and their connections have been taken from OpenStreetMap. In order to compute the average traveling time of each road, the Google Directions API has been used to query Google Maps and extract this data. For more information about the construction of the network please refer to [135]. The Multi-Net TeleAtlas dataset that we used gives information about the structure of the roads in the form of a network whose nodes are crossroad and roads are the links connecting them. Each element of the network has a certain number of assigned
informations. The only two quantities that we are interested in are again the coordinates of the crossroad and the speed limit of the roads that we used to calculate the travel time on the links.

We proceeded to apply the navigation algorithm sampling some random origin and destination points over the network. For every couple of origin-destination nodes on the network, we built a path using the navigation algorithm used for the grid model. The results confirmed qualitatively the results observed for the previous cases. The average of the logarithm of the travel time $t$ grows with the length of the path $l$ as a power-law for both the Google Maps (fig 2.19) and the TeleAtlas networks (Fig. 2.18) although the exponent is higher than the one found with the data. The fluctuations of $\log(t)$ decrease with $l$ for small values of $l$ and then remain constant as $l$ grows, although their constant value is smaller than the one found in real data (Fig. 2.18 and 2.19). The distributions $p(t|l)$ have the same scaling property of the one found in real data (fig. 2.20). Moreover the Kolmogorov-Smirnov test does not exclude that these distributions are really lognormal (moreover a skewness test assured that such distributions cannot be normal), while the ones found with the grid model clearly are not. This suggest that further topological details must be introduced in the grid model in order to refine the shape of the distribution, such as speed correlations between adjacent links.

It is evident that the scaling property and the qualitative behavior of the distribution $p(t|l)$ is captured by our model, although the optimization process is less efficient than the one occurring in real life and the fluctuation are dumped in some way. Notice that the measures obtained with GPS data are aggregated at the end of the trip of each driver, meaning that during its travel the trajectory of the car had been affected by all the dynamical fluctuations of speed due to the traffic situation at the time the trip occurred. On the other hand, when sampling paths on a network like the GoogleMaps, these fluctuations are eliminated by the averaging process over the link of the network. In this way fluctuations due to traffic are averaged together in the average travel speed assigned to every road. A more predictive and accurate model should consider the interaction between different paths converging to the same roads at the same time and their effect on the traveling speed.
2.3. GRID MODEL WITH SHORTCUTS

Figure 2.18: (Left) Dependence of the mean value $\log(t(l))$ of the logarithm of travel time $t$ upon travel length $l$. These points have been obtain sampling the paths using intermediate steps on Rome Urban Network. (Left) $\sigma(l)$ for the Rome Urban Network. The paths of these figures are sampled using the navigation algorithm with an angle of $45^\circ$.

Figure 2.19: (Left) Dependence of the mean value $\log(t(l))$ of the logarithm of travel time $t$ upon travel length $l$. These points have been obtain sampling the paths using intermediate steps on Rome Urban Network (Google Maps). (Right) $\sigma(l)$ for the Rome Urban Network (Google Maps). The paths of these figures are sampled using the navigation algorithm with an angle of $30^\circ$ (red points) and using the shortest-path from the origin to the destination (green points).
Figure 2.20: (Left) Distributions $p(t|l)$ for some values of $l$ in the Rome Urban Network (Google Maps). Continuous lines are lognormal approximations of the distributions. (Right) Collapse of the same distribution using the scaling property derived in the main text. The black curve is the distribution of the scaled variable $\tau$. 
2.3. Reproducing Daily Dynamics Patterns

We have seen that the grid model with the navigation algorithm is able to reproduce many statistical features of the single trips aggregated together. In 2.2.3 we have seen that particular patterns emerge when we look at the way in which each driver choses the sequence of trips to be performed during the day. In particular emerged that, considering a driver that has to perform $n$ trips, the succession of the distances traveled between successive trips is initially decreasing and then starts increasing, so that the shortest trip can be found in the middle of the sequence. It has been shown that this pattern is linked with the circadian rhythm of drivers, since the last trip is used to go back to the origin. Moreover the sequence of stops is chosen according to an orbital dynamics, i.e. they are deployed at a constant radius around the origin. Thus we can use this information in order to build a model that can reproduce these patterns in a simple way.

We have seen that drivers seem to optimize every single trip according to the travel time rather than the traveled space. This suggests that, when a driver has to connect different locations during its daily dynamics, the total travel time in order to perform all the trips and go back to the initial point should be the cost to be minimized. Assuming that we want to reproduce a sequence of $n - 1$ trips performed by a driver on the grid model with shortcuts, we assume that the driver start from the node $i_{start}$. We can chose this node randomly from all the nodes in the grid, but we will see that its choice is fundamental in order to obtain the desired pattern. Then the algorithm to find the sequence proceeds as follows:

- Choose $n - 1$ distinct nodes as intermediate stops. These nodes are chosen randomly between all the nodes in the grid.
- We build a path connecting each couple of stops $i$ and $j$ (including $i_{start}$) using the navigation algorithm introduced in 2.3.1. Thus we will construct the matrix $t_{ij}$ of the travel times of the paths connecting every couple of stops.
- The sequence of stops $\{i_{start}, i_1, i_2, \ldots, i_{n-1}, i_{start}\}$ is then the one that minimizes the total travel time:
  \[
  T = \sum_{k=0}^{n-1} t_{i_k, i_{k+1}},
  \]  
  where $i_0 = i_n = i_{start}$.
- At the end of the procedure we will have the optimal sequences $d_k^n$, $t_k^n$ and $l_k^n$. 

2.3. GRID MODEL WITH SHORTCUTS
Note that this would be the Traveling Salesman Problem \cite{137} on the grid with shortcuts if the shortest path between two successive nodes would have been chosen instead of one build with the navigation algorithm. In the following we sampled more than 2000 paths for each value of \( n \) and \( N_{\text{shortcut}} \) in the grid model, so we can compute \( \bar{d}_n^k \), \( \bar{l}_n^k \) and \( \bar{t}_n^k \) as in 2.2.3. For each path, the shortcuts are reassigned over the grid. The speed on the links in the grid and in the shortcuts has been chose constant, with values of 1 and 2 respectively. Different choices lead to qualitatively similar results.

The first check that can be made is that the choice of \( i_{\text{start}} \) must not be random. Figure 2.21 shows the successions of \( \bar{d}_n^k \) for some values of \( n \) in grids with \( L = 100 \) and \( N_{\text{shortcuts}} = 0 \) and \( N_{\text{shortcuts}} = 100 \). The values of the succession as \( n \) increasing are quite constant or follows an irregular pattern, rather different from the one observed with the data. Figure 2.22 shows the same successions with \( i_{\text{start}} \) chosen randomly on the border of the grid. In this case it is evident that the successions show the desired behavior: an initial decrease towards a minimum and then an increase until a value similar to the initial one is reached. This indicates that the interplay between the optimization process and the confined geometry of the urban environment is crucial in order to understand how the patterns emerge. The “parabola-like” succession seems to be the results of an optimization process performed within a limited area, starting from a point close to the boundaries of such area. As reminded in the introduction, in \cite{111} has been shown how the mobility of each individual is characterized by a different gyration radius \( r_g \), so that its movements are confined within a distance \( r_g \) from its barycenter. Thus, it is possible to argue that in a real urban environment the area delimited by \( r_g \) play the role that the grid plays in our model, being the starting point of their daily dynamics close to its border. Note that, as we have seen in the data, also \( \bar{l}_n^k \) and \( \bar{t}_n^k \) have this structure (Figure 2.23).

In 2.2.3 has been shown that approximating the succession of \( \bar{d}_n^k \) with a parabola \( p_n(k) \) it is possible to derive a scaling law for the succession. Despite the fact that in our case the parabolic fit does not describe well the succession, we can still use it in order to see if there is a similar scaling also for the successions on the grid. Then we fit the curves in figure 2.23 with equation 2.13. In both cases we check that the coefficients \( p_n(k) \) are power-law decreasing with \( n \), so that equations (2.14) hold. Moreover also the relations between the coefficients (2.18) is still valid even though the approximation is worse for \( N_{\text{shortcuts}} = 100 \) (table 2.2 presents the values of the exponents of equations 2.14 for some values of \( N_{\text{shortcuts}} \)). Figure 2.25 shows the collapsed successions together with the universal parabola independent from \( n \). It is evident that, despite a scaling law seems to
2.3. GRID MODEL WITH SHORTCUTS

Figure 2.21: Succession of trip lengths $\overline{l_k}$ on the grid model with $N_{\text{shortcuts}} = 0$ (a) and $N_{\text{shortcuts}} = 100$ (b) when the origin of the trips is randomly chosen on the grid.

exist, the parabola does not describe well the curves since they have a smaller curvature.

Figure 2.22: Succession of trip lengths $\overline{l_k}$ on the grid model with $N_{\text{shortcuts}} = 0$ (a) and $N_{\text{shortcuts}} = 100$ (b) when the origin of the trips is randomly chosen on the border of the grid.
Figure 2.23: Succession of trip travel times $\overline{t}_k$ (a) and on the distances from the origin $\overline{d}_k$ (b) on the grid model with and $N_{shortcuts} = 100$ when the origin of the trips is randomly chosen on the border of the grid.

<table>
<thead>
<tr>
<th>$N_{shortcuts}$</th>
<th>$\eta_a$</th>
<th>$\eta_b$</th>
<th>$\eta_c$</th>
<th>$\eta_c - \eta_a$</th>
<th>$2(\eta_c - \eta_b)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.89(8)</td>
<td>1.76(8)</td>
<td>0.62(1)</td>
<td>-2.26(1)</td>
<td>-2.27(1)</td>
</tr>
<tr>
<td>50</td>
<td>2.91(7)</td>
<td>1.72(8)</td>
<td>0.58(2)</td>
<td>-2.34(9)</td>
<td>-2.3(2)</td>
</tr>
<tr>
<td>100</td>
<td>2.95(5)</td>
<td>1.74(5)</td>
<td>0.58(2)</td>
<td>-2.30(7)</td>
<td>-2.32(7)</td>
</tr>
</tbody>
</table>

Table 2.2: Values of the exponents of equations (2.14) for the grid model.
2.3. GRID MODEL WITH SHORTCUTS

Figure 2.24: Coefficients $a_n$, $b_n$ and $c_n$ of the parabolic fit for the succession $\ln k$ for some values of $n$ on the grid with $N_{\text{shortcuts}} = 0$ (a) and $N_{\text{shortcuts}} = 100$ (b). Continuous lines are power-law fits of the points.

Figure 2.25: Collapsed succession of trip lengths $\ln k$ on the grid model with $N_{\text{shortcuts}} = 0$ (a) and $N_{\text{shortcuts}} = 100$ (b) when the origin of the trips is randomly chosen on the border of the grid.
2.4 Conclusions and Perspectives

In this work we presented the development of a simple model in order to explain the results found analyzing GPS data in [129]. In chapter 1 we recalled the main results of such analysis. One of the most important results is the study of the conditional probability \( p(t|l) \), i.e. the probability that a driver performing a trip with length \( l \) will drive for a time \( t \). This probability can be described by a lognormal distribution, whose parameters depend on \( l \). It has been shown how the dependence on \( l \) of this distribution can be eliminated by rescaling the time with the average travel time at length \( l \), called \( \tau(l) \). Such scaling property is linked with the ability of driver to improve the optimization of their path as the length increases. In fact since \( \tau(l) \sim l^\alpha \) with \( \alpha < 1 \), the average velocity increases with \( l \).

In 2.2.3 has been shown that particular patterns emerge when the drivers are grouped together according to the number of trips performed during a day. In fact it has been shown that indicating with \( \overline{l}_n^k \) the average traveled distance of the \( k^{th} \) trip of all the drivers that has performed \( n \) trips in a day, such succession decrease initially until a minimum is reached around the middle point \( \hat{k} = \frac{n+1}{2} \) and the it starts increasing. Moreover for \( n \geq 4 \) the last and the first elements of the succession have close values. Approximating the successions with parabolas \( p_n(k) \) it is possible to find a scaling relation so that the dependence on \( n \) is eliminated and they all collapse together. This behavior has been found due to a particular choice of the deployment of the stops named orbital dynamics. According to this dynamics the stops are placed around a constant radius around the origin (figure 2.10). Also the successions of the travel time \( \overline{t}_n^k \) and of the distances from the origin \( \overline{d}_n^k \) follows a pattern similar to \( \overline{l}_n^k \).

In 2.3 we introduce our model of urban environment. Following the results of [29], the urban network is built as a grid whose links can be traveled at low speed. To simulate the presence of arterial roads, we introduced a certain number of shortcuts linking couple of nodes that were not connected in the original grid. These couples are chosen randomly and each link can be traveled at a speed higher than the ones of the grid (in the beginning the grid links can be traveled with speed 1 while the shortcuts with a fixed speed \( v > 1 \)). Initially we sampled the paths of the drivers assuming that each driver can perform a global optimization of its path using the travel time as cost. Despite the fact that with this method it is possible to reproduce the power-law relation between the average logarithm of the travel time and the traveled length of equation (2.5), the standard deviation
(\sigma(l)) of such quantity was decreasing with \(l\). \(\sigma(l)\) has been found constant with \(l\) in the data and this property is crucial in order to obtain the scaling of the distributions \(p(t|l)\). The introduction of the navigation algorithm in 2.3.1 to sample the paths reproduced the correct behavior of both \(\mu(l)\) and \(\sigma(l)\). According to this algorithm the driver does not perform a global optimization of their travel but use some milestones between the origin and destination of their trip, optimizing the path between these points. A scaling property similar to the one found with the data has also been observed for the \(p(t|l)\) measured on the grid model with the navigation algorithm. Although, the distribution is not well described by a lognormal distribution.

The navigation algorithm has also been applied to the grid model with a realistic distribution of speeds over the links and on the real Rome Urban Network, coming from the TeleAtlas database. In both cases the results were in good agreement with the ones found before.

In 2.3.3 we used the grid model and the navigation algorithm in order to reproduce the parabolic pattern of the successions of the movements. Fixing a origin node, we randomly placed \(n - 1\) stops on the grid. Then we connected each couple of stops (including the origin) with a path build with the navigation algorithm and found the sequence that the total travel time. This simple procedure reproduces pattern similar to those observed in 2.2.3, assuming that the origin point lays on the border of the grid. Thus these patterns are the result of the interplay between the optimization process of the total travel time performed by the drivers and the confined geometry in which they place their stops. The scaling relation of \(\frac{d_m}{d_k}\) is also reproduced.
Acknowledgments

I would like to express my gratitude to all the people that contributed to the development of this doctoral thesis and helped me during my research activity. First of all my supervisor Vittorio Loreto and Vito D.P. Servedio for the support given to me during these years.

Thanks to Simone Pozzi and Marco Ducci from Deep Blue s.r.l. for the support and the operative knowledges that have been essential in the development of the first part of the thesis.

Thanks to Carlo Libero, Gaetano Valenti from ENEA.

Thanks to Alina Sirbu for providing us the OpenStreetMap and Google Directions data of the city of Rome.

Special thanks to Pierpaolo Mastroianni and his interesting Master Thesis that was the starting point for the second part of this work.

Finally I would like to thank all the people that have not directly contributed to the work but still helped with other kinds of support. Pietro Gravino for the random discussions during our coffee breaks. Martina Pugliese for the patience during this years (I know sometimes I am annoying). Stefano Frontani and Francesco Cecconi for the logistic support during my travels. Valebot for the interesting discussions on Skype. Celeste Marzi for always being there.

Last but not least to all my family, whom I dedicate this work.
Bibliography


[121] North rhine-westphalia’s olsim traffic forecasting system. URL http://www.autobahn.nrw.de/.


