# Investigation into Air Traffic Complexity as a Driver of a Controller‘s Workload

Dissertation

to obtain the academic degree of

**Doctor of Science (Ph.D.)**

submitted to
the Faculty of Transportation and Traffic Sciences "Friedrich List"
of the Technische Universität Dresden


*by*

*M.Sc. Jelena DJOKIC*


*Dissertation Advisor:*

*Prof. Dr.-Ing. habil. Hartmut Fricke*

*June 2014*

# Untersuchung der Flugverkehrskomplexität als Treiber der Arbeitsbeanspruchung von Fluglotsen

Dissertation

zur Erlangung des akademischen Grades des

Doktoringenieur (Dr.-Ing.)

vorgelegt

der Fakultät Verkehrswissenschaften "Friedrich List"

der Technischen Universität Dresden

*von*

*M.Sc. Jelena DJOKIC*

*Betreuer:*

*Prof. Dr.-Ing. habil. Hartmut Fricke*

*Juni 2014*

*Gutachter:*

*Prof. Dr.-Ing. habil. Hartmut Fricke*

*Prof. Dr.-Ing. habil. Leon Urbas*

# Contents

# List of Figures

## List of Tables

## *Acknowledgements*

First and foremost I wish to thank my advisor, Professor Hartmut Fricke, for giving me an opportunity to pursue this work, who supported me and provided scientific advices throughout my research. I would also like to thank to Professor Leon Urbas for the valuable insights and also for taking the demanding role of the reviewer of my thesis. I am also very grateful to Bernd Lorenz, my advisor during the years I spent in Eurocontrol, CRDS, in Budapest. Bernd introduced me and taught me immensely statistical analysis and guided me through its application within my research. The time that I spent in Eurocontrol I cherish tremendously, as I was so fortunate to meet so many amazing persons with whom, not only that I worked with so much pleasure, but I also share so many unforgettable moments. My special acknowledgments go to Mustafa, Philippe, Andrew, Zoltan and Eniko. I also like to thank Eurocontrol CRDS who supported financially my thesis and gave me an opportunity to be a part of the research and development team, enabling not only the collection of data necessary for my research, but also a unique professional experience and the basis for my professional career.

I would like to express my infinite gratitude to Michael Schultz who has been supportive since the days I began working on the thesis. He has provided insightful discussions and scientific support throughout the years of my research, but also encouraged me enormously through the rough road to finish this thesis.
I am also very grateful to Stefano for his statistical advices and knowledge that was instrumental in helping me to carry through my research.

For the friendship and support that I have been surrounded by through years, I would like to thank Beatrice and Andrea who are always there for me, in good and bad, to help me with my personal challenges. Thank you for sharing these years with me.

I would like to express my heartfelt gratitude and appreciation to Paolo, my true partner in life, love and friendship, for all understanding and patience through my everyday moods, support and unwavering belief in me.

And the most important, I give my extraordinary thanks to my dad and my sister who supported me, encouraged me and always provide unconditional love and care. Thank you with all my heart!

*To my Mama*

# Abstract

*The thesis describes an investigation into Air Traffic Control (ATC) complexity as a contributory factor in changes of controllers' workload. It is considered that ATC complexity, together with equipment interface and procedural demands comprise the task demands imposed on the en-route controller to perform certain activities, which mediated by performance shaping factors create workload.*

*The data used to study this relationship came from ATC real-time simulations completed at EUROCONTROL CRDS in Budapest: recorded flown trajectories, communication performed by the controller (whether with other controllers or with the pilots), data entries related to flight data management, and instantaneous self-assessment ratings of workload provided by the controllers were used. The ATC complexity factors that have been consistently found to be important in the previous studies (related to aircraft density, flight attributes of each individual aircraft, aircraft conflicts and traffic disorder) and for which detailed calculation formula have been reported were selected for further analysis. Since the established set of factors resulted from multiple researches conducted in this field, it was assumed that some of these factors are correlated with one another, overlapping and possibly measuring similar concepts. Therefore, a reduction of the initial set of factors was performed by combining information contained within these factors into a smaller number of new artificial variables and by deleting statistically redundant portions of these factors prior to conducting further analysis. The Principal Component Analysis (PCA), which is the statistical method applied to achieve required reduction, resulted in the overall set of 6 complexity components, whose interpretations are driven by the factors that showed the strongest correlation with that component. In order to establish a link between ATC complexity and a controller's subjective workload, multiple regression analysis was performed, using the complexity components identified in the PCA as predictors of the workload ratings. In addition, some measures of controller's activity (data entries made by the controllers related to flight data management, cumulative duration of radio calls, i.e. frequency occupancy time, and average duration of single calls) were added to the analysis to test whether information about the controller's activity could be also useful for predicting workload, once the effect of complexity had been considered, and to verify whether the effect of complexity on workload could be mediated by the effect of complexity on the controller's activity. The analysis revealed that both ATC*

*complexity and the activities that the controller performs to deal with a demand imposed on him/her give a unique contribution to the prediction of workload ratings and therefore the workload of the controller is determined by both ATC complexity and controller's activities.*

*In addition, it was assumed that the workload is differently impacted by individual components of complexity, and further statistical analyses were performed to test this assumption. Understanding these differences could in fact facilitate comparison of the complexity levels of a single sector under different conditions, but also comparison of complexity levels of different sectors under same conditions. Firstly the changes in the workload and activities of the controllers under different conditions were investigated using analysis of variance. Subsequently, in order to be able to map these changes on the complexity components, it was necessary also to investigate into the changes that the complexity components undergo when observed under different conditions. The results revealed different behaviour of single complexity components when mapped on the changes recorded in the activities of the controller and workload, demonstrating that changes in controller's activities and perceived workload are driven by different complexity components in different sectors and under different operational conditions.*

*Shedding light on these contributors to the workload experienced by a controller can greatly facilitate the introduction of any change envisaged for the airspace under consideration. Namely, in the current structure, whenever new procedures or new working methods are subject to possible deployment, the identified complexity components could support the estimation of the impact that those changes would impose on the workload of the controller and further on decision making processes. Additionally, the complexity components are also applicable in the validation of the new concepts and new technologies to be introduced in the system when designing simulation scenarios against which new concepts would be assessed. As also demonstrated by the analysis, the comparison of different sectors, or even different sector designs within the same airspace, could be compared and contribute to the improvement of airspace design.*

# Zusammenfassung

*Die vorliegende Arbeit untersucht die Komplexität der Flugverkehrskontrolle (Air Traffic Control, ATC) als einen wesentlichen Einflussfaktor auf die Arbeitsbelastung des Radarlotsen. Die zentrale Annahme ist dabei, dass die Komplexität der ATC zusammen mit den Anforderungen aus den betrieblichen Rahmenbedingungen (technische Systemschnittstellen und Prozeduren) den Lotsen zu bestimmten Abläufen zwingen, welche die Arbeitsbelastung signifikant beeinflussen.*

*Für die durchgeführten Untersuchungen standen Daten von ATC-Echtzeitsimulationen von EUROCONTROL CRDS Budapest zur Verfügung, die folgende Informationen umfassen: abgeflogene Flugtrajektorien, Kommunikationsprotokolle der Lotsen (untereinander oder zwischen Lotse und Pilot), Daten aus dem flight-data Management und Daten aus der regelmäßigen Selbstbewertung der Lotsen bezüglich ihrer aktuell gefühlten Arbeitsbelastung. Die bereits in früheren Studien identifizierten Komplexitätsvariablen (insbesondere die lokale Flugzeugdichte, spezifische Flugzeugeigenschaften, Konfliktsituationen zwischen Flugzeugen und die Verkehrslage betreffend) sowie hierzu erarbeitete mathematische Vorschriften bilden die Grundlage für die weiterführenden, detaillierten Untersuchungen. Aufgrund der Vielzahl an Komplexitätsvariablen aus diversen wissenschaftlichen Quellen war davon auszugehen, dass Korrelationen unter den Variablen vorliegen. Aus diesem Grund wurden zunächst statistisch redundante Informationen der ursprünglich vorliegenden Variablen reduziert, sodass als Ergebnis neue voneinander unabhängige Faktoren klassifiziert werden konnten. Die hierfür verwendete Hauptkomponentenanalyse (Principal Component Analysis - PCA) führte zu sechs statistisch signifikanten Komplexitätsfaktoren, die anhand der höchsten Korrelation zur zugeordneten Komponente interpretiert wurden. Um die Verbindung zwischen der ATC Komplexität und der subjektiv empfundenen Arbeitsbelastung herzustellen, wurde eine multiple Regressionsanalyse zwischen den Komplexitätsfaktoren und den abgeleiteten Arbeitsbelastungszuständen durchgeführt. Zusätzlich lagen für die Analyse der Arbeitsbelastung auch Daten über die Arbeitsaufgaben des Lotsen vor (bspw. Dateneinträge des Lotsen, Gesamtlänge der Funkanweisungen, durchschnittliche Länge der Funkanweisungen), um zu untersuchen, inwieweit sich aus den aktuell durchgeführten Arbeitsaufgaben bei gegebener Verkehrsnachfrage eine verlässliche Vorhersage über die Arbeitsbelastung ableiten lässt. Die Analyse zur Vorhersage der Arbeitsbelastung konnte*

*zeigen, dass sowohl die ATC Komplexität als auch die aktuellen Arbeitsaufgaben einen individuellen und signifikanten Einfluss haben.*

*Weiterhin wurde unterstellt, dass die spezifischen Komplexitätsfaktoren einen unterschiedlichen Effekt auf die Arbeitsbelastung ausüben. Die Überprüfung dieser Annahme war ebenfalls Bestandteil der umfangreichen statistischen Untersuchungen. Tatsächlich könnte ein fundamentales Verständnis der Komplexitätsgrade den Vergleich einzelner Luftraumsektoren unter verschiedenen operativen Randbedingungen, als auch den Vergleich unterschiedlicher Luftraumsektoren mit vergleichbaren operativen Randbedingungen wesentlich erleichtern. Zuerst wurden die Veränderungen der Arbeitsbelastung und -die Tätigkeiten der Lotsen unter Verwendung einer Varianzanalyse untersucht. Um eine valide Zuordnung zu den Komplexitätsfaktoren sicherzustellen, war es ebenfalls notwendig, die Veränderungen dieser Faktoren und Tätigkeiten unter wechselnden Randbedingungen zu analysieren. Die Analysen zeigen hierbei unterschiedliche Resultate bezüglich der jeweiligen Komplexitätsfaktoren. So beeinflussen die verschiedenen Komplexitätsfaktoren die Handlungsabläufe der Lotsen und die wahrgenommene Arbeitsbelastung, jedoch in Abhängigkeit von den ausgewählten Sektoren und den betrieblichen Randbedingungen.*

*Unter Berücksichtigung dieser erarbeiteten Abhängigkeiten der Arbeitsbelastung des Lotsen können nun die Auswirkungen von Veränderungen im Luftraum zuverlässig bestimmt werden. Gerade in Bezug auf Veränderungen der gegenwärtigen Luftraumstruktur oder die Einführung neuer Prozeduren oder Arbeitsabläufe können die entwickelten Komplexitätsfaktoren bereits frühzeitig Aufschluss darüber geben, welche Konsequenzen solche Veränderungen auf die Arbeitsbelastung der Lotsen nach sich ziehen können und Entscheidungsprozesse unterstützen. Weiterhin sind die entwickelten Komplexitätsfaktoren als Grundlage für die Validierung neuer Konzepte und Technologien, gegebenenfalls unter Verwendung von entwickelten Simulationsszenarien, nutzbar.. Darüber hinaus können die Komplexitätsfaktoren für die Gegenüberstellung von verschiedenen Luftraumsektoren genutzt werden und zur Abwägung bzw. Optimierung von Entwürfen eines Luftraumdesigns dienen.*

# 1 Introduction

Over the past few decades there has been a huge growth in air passenger demand (STATFOR, 2009). Even though depressed levels were recorded during the financial crisis of 2008 and continued throughout 2009 (ICAO, 2010), the demand for air travel has rebounded and continued to increase since, and this tendency is likely to continue in the decades to come (EUROCONTROL, 2010; IATA, 2014). Therefore, air traffic management (ATM) is facing great challenges as it reaches its limits. Additionally, ATM is quite conservative when it comes to adopting new technologies, which can be attributed to safety considerations whenever there are new proposals or modifications to the existing equipment in place.

Nevertheless, not only is ageing technology a bottleneck for the predicted growth, but also existing ATM concepts need fundamental changes in order to cope with these increased demands. In addition, overall higher environmental awareness and the need for cost efficiency calls put additional pressure on the ATM world to evolve taking into consideration also these aspects.

Therefore, the modernization and changes in the ATM are necessary in order to increase its capacity while, if not elevating, at least maintaining existing safety levels.

Many ATM projects and programmes have been initiated in order to forestall all these impediments – reduced mobility, additional delays, and more frequent occurrences of safety issues, higher costs and pollution through $CO_2$ and noise emissions.

In Europe, the project that leads this research is SESAR - Single European Sky Air traffic Research (SESAR Consortium, 2008), while in USA that is NextGen - Next Generation Air Transport System (FAA, 2011). These two projects are running in parallel and even a preliminary agreement on their interoperability was reached (ICAO, 2011). Both of these projects are focusing on developments that will improve ATM performance when it comes to capacity, safety, environmental impacts, economy and security. The solutions are sought in new technologies as well as in new designs and concepts both in the airborne and on the ground side.

SESAR intends to evolve from an airspace to a trajectory based system while introducing new technologies and applying a new approach to airspace design and management (SESAR Consortium, 2008). The key element of this change is the information shared among all actors of the system in the synchronized way, enabling collaborative decision making processes.

Nevertheless, SESAR and NextGen have very different frameworks in which they should be incorporated when it comes to the airspace structures. European airspace (which this research focuses on) is fragmented and the systems and air navigation services that are supporting it are not sufficiently integrated to face challenges of traffic demands to come in the next decades. Therefore, besides SESAR new airspace structures are proposed to enable for improved provision of air navigation services and to encourage cooperation between ANSPs. Such airspace structures are Functional Airspace Blocks (FABs) on which SESAR's operational concept has been built. "Functional Airspace Block means an airspace block based on operational requirements and established regardless of State boundaries, where the provision of air navigation services and related functions are performance-driven and optimized with a view to introducing, in each functional airspace block, enhanced cooperation among air navigation service providers or, where appropriate, an integrated provider." (European Union, 2004).

Finally, there are many changes envisaged in the ATM system. However, the human element of this system should not be left behind. Even though all these improvements take place we have to look at the limitations put in front of us by the human capabilities. Tsonis (2006) argued that the "computer and display technologies matured to a stage where the presentation of the information is no longer primarily limited by technology's ability to process the information, but by the human's".

ATM is undergoing significant changes (new sectorisation procedures, advanced new generation displays, a range of potential automated aids, interface technology) that could involve fundamental changes for the role of the air traffic controller (ATCO). And in order "to support ATM systems development, and ensure that the human continues to perform with high reliability, a better understanding of the picture, how it is built, maintained, and lost, is essential" Kirwan et al. (1997).

And thus, it is of paramount importance to understand what makes the work of the controller difficult - the task demands imposed on the controller that he / she has to cope with when controlling the traffic. Following this, it is possible to identify how this work can be facilitated with the application of the advanced automation, new procedures and airspace designs.

## 1.1 Motivation

In order to understand how certain elements of the ATM system can be altered to achieve more efficient and safer air traffic control, this thesis focuses on the investigation into air traffic situations and their complexity as a driver of difficulty as perceived by a controller.
Namely, it is assumed that by the introduction of new technology the current work of the controller would be facilitated until the point where the capacity could be increased while maintaining the current safety level, or even improving it. But to be able to measure this positive impact, firstly we have to understand the current construct of the system.

Therefore, primarily, the aim of this work is to identify what in the existing ATM system is making the work of the controller difficult and why certain traffic situations are perceived by the controllers as more difficult than others. However, it is clear that to be able to capture air traffic control (ATC) complexity more accurately, besides simply counting the number of aircraft under control, it is necessary to take into consideration other important factors. Although the concept of complexity in ATC has been tackled in many researches until today (Delahaye and Puechmorel (2000), Chatterji and Sridhar (2001), Laudeman *et al.* (1998), Kopardekar and Magyarits (2003), Sridhar, Sheth and Grabbe (1998), Mou, Cho and Histon (2012)), it seems to be difficult to find a unique measurement of complexity. Therefore, the challenge of this work is to adopt and adapt earlier work so as to achieve the measurement of the complexity that would comprehensively cover the ATC complexity aspects of interest.

The second challenge motivating this work is how to measure the impact of the new technologies and procedures on the controller's work when comparing with those established and currently in use. We are looking into different parameters of the air traffic situations, i.e. ATC complexity elements, and analysing how their alterations are reflected into different levels of difficulty experienced by controllers. The aim is to anticipate the possible impact that new technology or new procedures may have on the controller. If the impact of modifications on a definite set of contributing parameters is in any certain way predictable, we can manipulate them and investigate their influence on the other elements of the system. This aspect can prove very useful when designing simulation scenarios against which new concepts could be assessed.

## 1.2 Outline of the thesis

This thesis focuses on the air traffic situation, its elements and the level of the ATC complexity that they constitute. Further, this research looks into the causal relationship between ATC complexity and the workload experienced by the controller when controlling the traffic.

It is organized as follows:

Chapter 2 provides the background of the thesis. It addresses the components of the ATC system and assesses more into depth how changes imposed on this system can be further investigated. It provides the literature review of the related issues, the elements and the definitions of the ATC complexity and its correlates in the context of the air traffic controllers' work, based on which the questions that need further research are formulated. Thus, the literature review not only addresses the previous researches into ATC complexity, but also on the controllers' task load and workload. The chapter concludes with the statement of the existing problem and hence the hypotheses on which the current work is focused.

In Chapter 3, the methodology that is developed to collect and provide evidence in the support of the research hypotheses is described together with the methods that would enable such an approach. Furthermore, the real-time simulations as well as the experiments used to obtain data are explained more into depth: the applied techniques and materials, participants of the experiments and data extraction procedures.

Chapter 4 addresses the analysis of the collected data and presents the results of these analyses, whereas the conclusions drawn from these results and how they are assessed against the research hypotheses are described in the Chapter 5. In this chapter possible practical applications of the results and recommendations for improvements and future work are also proposed.

# 2 Background

## 2.1 Air Traffic Control Elements

Airspace is divided into small units called sectors. Air traffic control sector is “a defined airspace region for which an associated controller (or controllers) has ATC responsibility” (EUROCONTROL, 2004). Each sector is commonly controlled by at least one, and most commonly by two air traffic controllers (executive and planning controller). The role of an air traffic controller is to assure the safe, orderly and expeditious flow of controlled aircraft between departure and destination points (Kirwan, Rodgers & Schaeffer, 2005). More specifically, air traffic controllers provide air traffic control service by preventing collisions (between aircraft, and on the manoeuvring area between aircraft and obstructions) and by expediting and maintaining an orderly flow of traffic in accordance with the procedures and rules of the air and air traffic services (ICAO, 2007).

At the same time, they are complying with letters of agreement defined between neighbouring air traffic control centres regarding specific points, speeds, flight levels, where the responsibility for controlling traffic is delegated from one facility to the other.

In performing these tasks, controllers use radar displays to follow the traffic movements. Radars are visualizing the positions of aircraft in two dimensions, while the altitude, speeds and other relevant information is contained in the tag label assigned to each aircraft.

Accordingly, air traffic controllers assess the situation and anticipate future positions of aircraft within their sector’s boundaries, but also in a certain part of the surrounding airspace. When needed, a controller takes actions to assure safe, orderly and expeditious traffic by changing an aircraft’s performance parameters (altitude, speed, or heading) to timely modify aircraft trajectories, and also supports the pilots in steering their tasks, e.g. climbing to a particular altitude to increase fuel efficiency, descending prior to arrival at the destination etc.

Furthermore, the environment in which the controller acts is dynamic and each action that he or she takes influences future ones. As Sperandio (1971) exemplified in the closed-loop model of ATCO’s workload, shown in Figure 1 below, actions performed in response to the task demand placed in front of the controller influence the task demand encountered in the future (feedback loop 2).

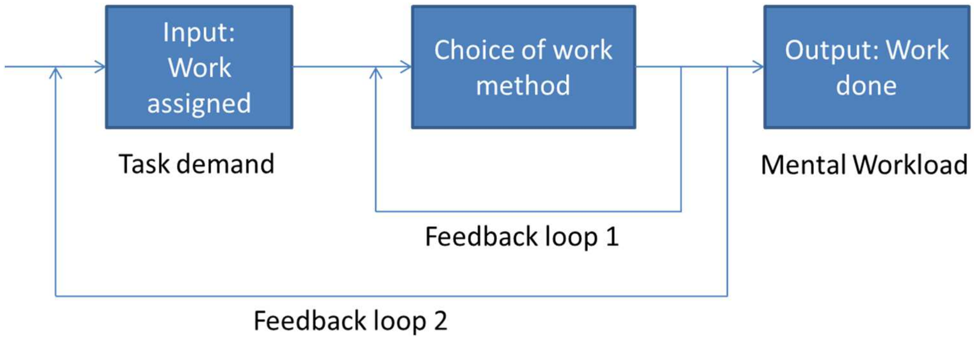



Figure 1. Closed-loop model of ATCO mental workload (Sperandio 1971)

On the other hand, the resulting workload also regulates the choice of working method that a controller will apply (feedback loop1). When the workload increases, the more economic strategies tend to be followed more often, but occasionally strategies appropriate for other levels of workload are also used (D'Arcy & Della Rocco, 2001). Controllers report that they become more conservative (which likely minimizes monitoring), do things faster and act earlier.

Nevertheless, not all controller tasks are observable. As defined by Pawlak (1996), there are four controller tasks while managing the ATC situation: monitoring, evaluating, planning and implementing the formulated plan. Specifically, out of these four tasks only one is observable, and that is the implementation process (Figure 2).

That is, that, the controller is monitoring the situation in the sector and is anticipating its possible evolvement. Based on this, he is planning his further set of actions and also foreseeing their impact on the traffic in the sector. Naturally, this process of planning results in a set of re-routes, vectors, speed adjustments, altitude changes, coordination with other controllers, and other actions that are then implemented through the use of various communication and data entry tasks. After implementation, the controller continues to monitor the situation to ensure conformance of the situation with the plan and to evaluate the effectiveness of the plan in resolving the conflicts (Pawlak1996).

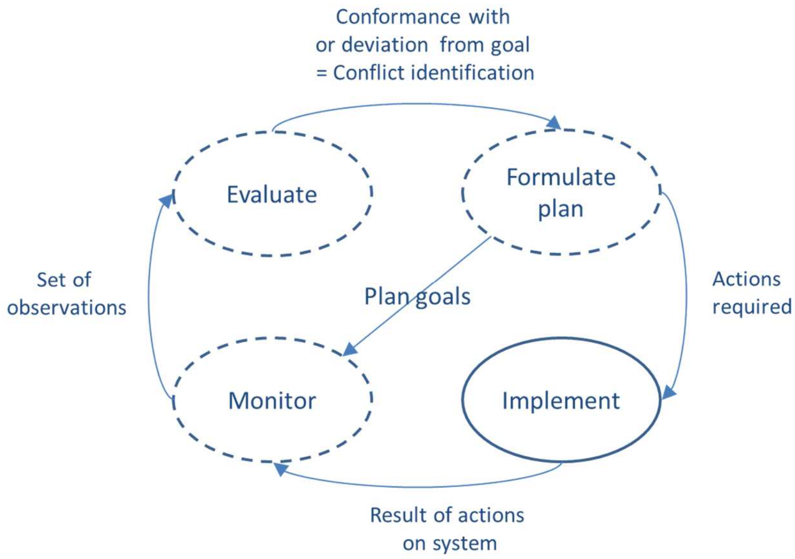



Figure 2. Mental and Physical Process Required in Air Traffic Control (Pawlak, 1996)

It means that by taking only objectively measurable tasks into consideration, it is possible to capture only one aspect of controller activity involved. However, this aspect of the controller’s activity is directly connected with changes made by the controller in the ATC situation, and therefore its significance was investigated in numerous studies.

The observable and measurable part of the process is referred to as “physical workload” by Pawlak *et al.*(1996), while the other activities in the process are referred to as “mental workload”. Similarly, Cardosi & Murphy (1995) whilst elaborating on measures that are corresponding to the observable activities of the controller emphasized that both the observable (objective) and perceived (subjective) aspects of demand on the controller need to be considered and that there is no absolute workload independent of skill and experience. Likewise, more recent studies and researches distinguish between objective and subjective workload (Hilburn and Jorna (2001), EUROCONTROL (2009)).

This process is very commonly in the literature presented through open-loop model, such as the one developed by Hilburn & Jorna (2001) shown on the Figure 3 below.

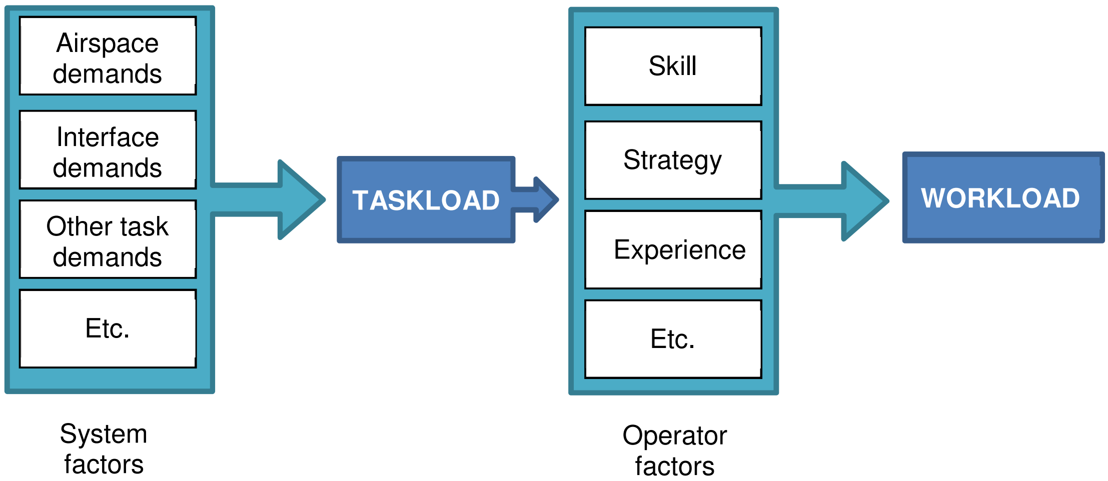



Figure 3. Model of ATC workload (Hilburn & Jorna 2001)

Therefore, it may be concluded that the objective workload or task load relates to task demands, while subjective or mental workload is induced by cognitive and physical task demand but it varies as a function of ability, skill, training and experience.

In line with these definitions, one should ponder that ATC complexity essentially contributes to a task demand imposed on the controller, i.e. task load. The activities that then controller perform in reaction to this demand - mediated by skill, experience, training, age, etc.- are translated into a workload perceived by air traffic controllers.

In order to better understand these elements contributing to the workload of the controller, as well as their interdependencies, the following sections are aimed at elaborating on them based on the survey of the researches conducted hitherto in the field: to begin with, workload (subjective or mental workload) is addressed, followed by the description of the controller's activities as a link between task demand and controller's workload (objective, behavioural workload or task load), and finally ATC complexity factors are tackled.

## 2.2 Workload

Even though workload has been investigated for decades already, a unique and widely accepted definition of workload has not been established. The operational definitions of workload from various fields continue to disagree about its sources, mechanisms, consequences and measurement (Huey & Wickens 1993). Cardosi & Murphy (1995) emphasize that it is imperative to define the term workload in context as there is no single agreed definition.

Several researchers in the air traffic management domain agree that workload is a result of such a complex interaction between the task demand and the way the controller actively manages the situation (e.g. Hilburn 2004; Loft *et al.* 2007; Majumdar *et al.*2004; Pawlak *et al.* 1996). Moreover, as previously noted, controllers, by performing certain activities, regulate the evolution of the task demands with the aim of keeping workload at an acceptable level (Sperandio 1971).

Workload is thought of as a construct that is not directly observable or measurable, but must be inferred, based on measures and observations of other elements (such as physical tasks) (Casali & Wierwille 1984; Stein 1993; Mogford *et al.* 1995).

As such, workload is judged to be very difficult to measure. Nevertheless, Cain (2007) provides an interesting list of criteria for workload measurement methods. Furthermore, the author notes that workload measurement techniques are typically organized into three broad categories:

1. Self-assessment or subjective rating scales;
2. Performance measures (including subdivisions of primary and secondary task measures);
3. Psychophysiological measures.

In line with this are also workload measures applied within the ATC domain. Hilburn and Jorna (2001) provided some examples of ATC workload measures that have been used in the past:

Table 1. ATC workload measures (Hilburn & Jorna 2001)

| |
|---|
| 1. Subjective |
| NASA TLX (Brookings & Wilson, 1994; Hooijer&Hilburn, 1996) |
| Air Traffic Workload Input Technique (ATWIT (Leighbody, Beck & Amato, 1992)) |
| Subject Matter Expert / Over-the-shoulder ratings (Schaeffer, 1991) |
| Instantaneous Self-Assessment (ISA) technique (Whittaker, 1995, Eurocontrol,1997) |
| 2. Behavioural |
| Number of control actions (Mogford, Murphy &Guttman, 1993) |
| Communications efficiency (Leplat, 1978; Geer, 1981) |
| Communication time, message length (Morrow, 1993) |

| |
|---|
| Flight data management (Cardosi& Murphy, 1995) |
| Inter-sector co-ordination (Cardosi& Murphy, 1995) |
| Decision and action frequency (Schmidt, 1976) |
| 3. Psychophysiological |
| EEG, EMG and EOG (Costa, 1993) |
| Heart rate measures (Brookings & Wilson, 1994; Laurig et al.,1971; Hooijer&Hilburn, 1996) |
| Eye blink rate (Stein, 1982; Brookings & Wilson, 1994) |
| Respiration (Brookings & Wilson, 1994) |
| Biochemical activity (Zeier, 1994; Costa, 1993) |
| Pupil diameter (Hilburn, Jorna&Parasuraman, 1995) |
| Eye scanning randomness ( entropy(Hilburn, Jorna&Parasuraman, 1995) |
| Visual fixation frequency (Stein, 1992; Hilburn, 1996) |

However, as also discussed before, there is no widely accepted taxonomy of the terms related to controller’s task load (objective workload) and workload (subjective or mental workload). Thus, it may be noticed that behavioural (performance) measures actually correspond to what in the previous section was referred to as controller’s activities, i.e. task load.

Therefore, further within this section, explanation will address only subjective and psychophysiological measures of workload, especially those that are widely accepted and frequently applied in the studies within the air traffic control domain. The performance or behavioural measures will be addressed in the following section (section 2.3) as related to the controller’s activities.

For each of these methods there are advantages and disadvantages that are driving the selection of the appropriate method in accordance with the predefined criteria. The selection criteria depends on the context of the study within which workload is measured. Likewise, EUROCONTROL (2003) listed several criteria in the review of workload measurements, analysis and interpretation methods conducted by EUROCONTROL team involved in the INTEGRA programme.[1] The aim of this review was to derive principles of workload measurement in human-in-the-loop simulations from experience in domains other than air traffic management. The authors firstly focused on the more general advantages of each of the proposed measures (subjective, performance-based and physiological/biochemical) and then further discussed more into detail how these measures respond to the predefined set of criteria (reliability, validity, sensitivity, diagnosticity, practicality and intrusiveness).

[1] For more information on INTEGRA programme see http://www.eurocontrol.int/integra

Subjective methods attempt to quantify the personal interpretations and judgements of their experienced demand. Therefore, they are mostly presented in the form of rating scales. The adequate representative of such a method is NASA Task Load Index (TLX). Hart and Staveland (1988) define the NASA Task Load Index (TLX) as "a workload rating scale that provides a sensitive summary of workload variations within and between tasks that is diagnostic with respect to the sources of workload and relatively insensitive to individual differences among subjects". Namely, NASA-TLX is a multi-dimensional rating scale that considers several dimensions of workload. These dimensions are reflecting the operator's experience of mental, physical and temporal demand, their own performance and the effort and frustration that they have in their jobs. The individual rating of each dimension facilitates the revealing of the source of a workload or performance problem. However, overall workload is then derived based on a weighted average of ratings of these dimensions. This measure of workload proved to be very reliable in many studies and also showed high sensitivity (Hart & Staveland 1988; Nygren 1991; Hill *et al.* 1992). Additionally, NASA TLX is an off-line measure, which takes several minutes to be completed. Even though it was developed more than 20 years ago, it is still very much relevant and, "it is being used as a benchmark against which the efficacy of other measures, theories, or models are judged" (Hart 2006).[2]

The ATWIT (the Air Traffic Workload Input Technique) is a tool specifically developed by the FAA Technical Centre for the air traffic control domain. It measures mental workload in "real-time" by presenting auditory and visual cues that prompt a controller to press one of seven buttons on the workload assessment keypad (WAK) within a specified amount of time to indicate the level of mental workload experienced at that moment (Stein 1985).

In some studies, workload is measured by applying more than one measurement method. Likewise, Manning (2000) conducted a study in which controllers used the ATWIT, the NASA TLX, and an estimate of the traffic sample's activity level (using a 5-point scale ranging from "Not at all busy" to "Very busy") to rate the amount of workload experienced. The authors give two main reasons for using these measures in parallel. The first reason is related to actuality of the real-time workload ratings provided shortly after the experience occurs. However, the authors also anticipated the negative impact of intrusiveness of this

[2] For more information on NASA TLX, as well as for free electronic and paper/pencil versions see http://human-factors.arc.nasa.gov/groups/TLX/index.html

method that may interfere with the performance of certain tasks and also increase the controller's perceived (mental) workload. On the other hand, the workload ratings obtained after a scenario is completed may be influenced by early or more recent events or the controller may forget to consider certain events altogether.

Another real-time "online" tool for subjective workload rating is ISA (Instantaneous Self-Assessment). ISA was developed by NATS (Whittaker, 1995) as a simple tool with which controller may opt for a rating perceived mental workload as on a scale from 1 to 5 (1 refers to very low experienced level of workload, while 5 represents very high workload rating) (Figure 4).


Figure 4. ISA workload assessment tool

Every 2 minutes, controllers are prompted (by a light signal) to score their ratings. As for ATWIT, ISA also has the advantage of workload assessments provided shortly after the experience occurred and therefore is not biased by other events that may occur before or after within the analysed traffic sample. Moreover, as stated by EUROCONTROL (2003), this method shows the lowest intrusiveness when compared with other subjective measures of workload besides minimal equipment required to apply it.

Beside those reported by Hilburn and Jorna (2001), another frequently used technique is SWAT (Subjective Workload Assessment Technique). Similar to NASA TLX, SWAT (Reid &

Nygren 1988) is a multidimensional ratings tool which was more frequently used in aircrew studies than in air traffic control studies. Dimensions covered by this tool are: time load, mental effort, and psychological stress, each with three categorical levels (low, medium and high). The time load dimension depends on the availability of spare time and the overlap of task activities, while mental effort load is an indicator of the amount of attention or mental demands that are required to accomplish a task, independent of the number of subtasks or time limitations. The third dimension, psychological stress load refers to conditions that produce confusion, frustration, and/or anxiety during task performance and, therefore, make task accomplishment seem more difficult.

Even though it proved to be very sensitive and reliable in comparison with NASA TLX (Colle & Reid 1998), SWAT was not as well accepted as NASA TLX by users as the effort needed to complete SWAT measures is much higher. This is because subjects are asked to perform a card sorting procedure (27 cards) - beginning with the card that represents the lowest mental workload and ending with the card that represents the highest workload. And only then subjects were asked to provide ratings for particular tasks or events. Additionally, the analysis of these ratings also requires significant effort. However, Colle & Reid (1998) and De Eggemeier (1983) consider SWAT scale as useful in estimating changes in mental workload, while Boyd (1983) concluded that the three dimensions “lack subjective orthogonality” (e.g. high levels of time-load will also artificially inflate the level in the mental workload category).

Another group of workload measures are psychophysiological[3] measures such as eye movements, heart rates, electroencephalography and respiration. The biggest advantage of psychophysiological measures when compared to the subjective measures is that it gives a continuous and unbiased objective measurement of the operator’s state. In the past, these types of measurements were more intrusive and very costly. However, as technology lately has advanced very much, today they may be used with minimal intrusion on operator’s activities even though they are still very costly. In addition, these types of measurements produce a large volume of data that requires sophisticated and very costly analysis.

Above and beyond this, there is still no clearly defined link between objective psychophysiological measures and workload, as there are many factors (not workload

[3] Psychophysiology is the study of relations between psychological manipulations and resulting physiological responses, measured in the living organism, to promote understanding of the relation between mental and bodily processes (Andreassi, 2007).

related) that may influence changes in psychophysiological measures. For example, due to large individual differences between controllers, an increase in traffic load may also affect a change in the heart rate for one controller, while for the other this may not be observed or recorded (Cardosi & Marphy 1995).

Furthermore, eye movements are affected by other factors rather than only by changes in workload level. Namely, fatigue and ambient illumination can greatly influence eye movement and therefore it is very difficult to distinguish to which extent eye movements are reflecting changes in workload level only. Although many of these potential problems can be minimized simply through experimental design (Hilburn & Jorna 2001), “...their independent use as a predictor of workload levels seems quite limited...” and “formal, coupled models relating various psychophysiological measures and workload need to be developed” (Cain 2007).

## 2.3 Controller activity – link between task demands and controller’s workload

The performed controller’s activities, as explained in section 2.1 above, present the reaction of the controller to the task demand imposed on him. In certain way, those activities are presenting the link between the task demand and the workload perceived by the controller (mental or subjective workload). Moreover, as Sirevaag et al. (1993) stated, the measures related to the behaviour of the operator (i.e. controller) are ideal in that they provide an indication of both operator and system performance. That is, changes in the system performance (e.g. more traffic within the sector) directly drive changes in the activities performed by the controller (e.g. the number of radio/telephony communications).

Within ATC domain, these measures mainly fall into two categories (see behavioural measures listed in Table 1, section 2.2):

1. measures related to the communication performed by the controller (whether with other controllers or with the pilots) and
2. measures that are reflecting data entries related to flight data management.

The category of communication measures predominantly obtains measures that are related to the duration and number of the communication conducted by the controller. Such communication measures are: total time spent communicating during a traffic sample, time required for individual communication, average time needed for each transmission and total number of transmissions during a considered traffic sample. However, there are measures that are also taking into consideration the contents of the transmissions.

The second category of measures, measures that reflect controller data entries related to the flight data management, are inputs made by the controller referring to assignments of aircraft performance within sector boundaries (e.g. vertical rate, number of cleared flight levels, number instructions for headings changes, speed instructions, etc.).

Manning, Fox and Pfleiderer (2003) conducted a study on the relationship between measures of air traffic controller voice communications, task load, and traffic complexity. Communication measures were based on transmissions obtained from voice tapes associated with the traffic sample. The measures included the total number of transmissions, total time spent communicating during a traffic sample, and time required for individual transmissions (for all speakers). These transmissions were also categorized by the content (address, courtesy, instructional clearances, frequency changes, advisory/remark, request, readback /acknowledgement and non-codable). During the transmission it was possible that more than one message content occurred.

The task load measures taken into consideration were POWER measures (Performance and Objective Workload Evaluation; Mills, Pfleiderer, & Manning, 2002). POWER measures include the numbers of controlled aircraft, handoffs made and accepted, altitude changes, controller data entries and data entry errors, variations in aircraft headings, speeds, altitudes and average time aircraft were under control. In all, 23 POWER measures were taken into account. In this context, task load measures refer to routinely recorded ATC data that describe both aircraft and controller activities.

The main hypothesis was that these measures would be significantly correlated with each other and with activity, since as the traffic situation gets busier, more communications should occur. But, a second hypothesis forecasted that these measures of communication would not make unique contribution to the prediction of complexity, over and above the contribution provided by the POWER task load measures. Eventually, it was found that neither task load (POWER) measures alone, nor the communication measures alone statistically predict complexity ratings by themselves, contrary to the full model containing all variables.

In the same way, Rantanen, Maynard and Özhan (2005) examined the impact of sector characteristics and aircraft count on air traffic control communications and workload.

Similarly, when Casali and Wierwille (1983) investigated into any correlation between 16 workload measures and communication, they found that of 16 workload measures, half were sensitive to communication. Even more, the results suggested that communication load also affected subjective judgments (i.e. Cooper-Harper ratings), performance (i.e. response time and errors), and physiological measures (i.e. pupil diameter).

Laudeman *et al.* (1998) validated the dynamic density equation in an operational environment. Observations of air traffic controller activity at the radar position of an en route sector were taken into consideration as an acceptable independent measure of controller workload. A set of representative activities was selected, which included radio communication activities and radar scope related activities. Observers were sitting behind and to the side of the R-side controller while recording activity observations on laptop computers. The Activity Catalogue Tool (ACT) was used to collect observations. The ACT output files included time stamps that could be synchronized with dynamic density file time stamps and activity labels. A count of activity events recorded during each two minute sample interval was used as the controller workload measure. Even though it does not really answer objective data, still this data can be used in similar way as a controller's subjective response for validation of the importance of objective measures.

Besides observations of ATC activities, Histon and Hansman (2002) used interviews with controllers to elicit key complexity factors. The radar display was observed directly and radio communications between controllers and pilots were monitored using an extra headset. At the same time, controllers often explained a set of control actions and the reasoning for performing them. Furthermore, in order to examine how structure reduces complexity associated with implementing commands, they performed an analysis of the commands that the controller provided to the pilots. For this purpose data was gathered from an internet website broadcasting live pilot-controller and controller-pilot communications for an en-route sector in the Atlanta Centre ARTCC[4]. A series of buttons were provided where each button corresponded to a different communication event. When the audio recordings were played back, the appropriate button was selected each time a communication event was heard, with the time of the event and aircraft involved also recorded also.

Manning (2000), in order to evaluate ATC performance, used two measures developed by Bruskiewicz *et al.* (2000). These are the Over-the-Shoulder (OTS) rating forms (used to

[4] www.atcmonitor.com

evaluate controller performance across broad dimensions) and the Behaviour and Event Checklist (BEC) (used to record specific mistakes made by the controller during the simulation exercises). However, Rantanen and Nunes (2003) found that the OTS method is “labour intensive, time-consuming, and expensive. Moreover, a human evaluator may not be able to provide sufficiently accurate quantitative data for research purposes, due to the limitations of human observation capabilities. The latter is the case particularly in observation of simultaneous events.” Elaborating on many of ATC’s performance measures, the authors propose a taxonomy of ATC performance measures that from their point of view represents an emphatically systematic and comprehensive approach to the measurement problem in ATC.

Nevertheless, this proposed taxonomy, however valid, is not widely recognised and accepted. As evident from the numerous studies conducted in the field, there are many different taxonomies as well as approaches applied to measure controller’s activities depending on the context of the studies in which they were investigated.

## 2.4 ATC complexity

ATC complexity has been the subject of a significant number of studies since research into it started long ago in the 1960s (e.g. Arad 1964). Yet despite many complexity factors having been proposed, up to now a comprehensive and generally accepted set of measures has still not been defined.

ATC complexity was predominantly investigated in relation with other factors rather than a stand-alone independent measure. Most frequently, the causal relationship between complexity factors and controllers’ workload was addressed. Many researchers were seeking a measure that would accurately capture this relationship between air traffic complexity, the controller’s workload and how these relate to safety. Even more, as Loft *et al.* (2007) remarked, systematic comparison among these studies is very much more cumbersome for two main reasons. The first reason is the presence of a wide variety of research methodologies reported, and the second reason is the presence of a wide variety of workload measures used.

A straightforward determinant of air traffic control complexity is simply the number of aircraft for which the controller is responsible in a specified time and sector (Manning *et al.* 2001; Kopardekar & Magyarits 2003). This measure is referred to as the “sector load”.

Predicting sector load and avoiding sector overload is the basic tool upon which current traffic flow management is built. However, the level of difficulty experienced by the controllers depends on additional factors beyond the number of aircraft present in a sector (Sridhar, Seth & Grabbe 1998).

Therefore, many studies that followed, relied significantly on so-called “traffic density” (e.g. Chatterji & Sridhar 2001; Hill *et al.* 2005; Kopardekar & Magyarits 2003; Sridhar, Seth & Grabbe 1998, Albasman & Hu 2012). This stands for a count of aircraft in a volume of airspace, taking into account also their relative distance and distribution within the airspace boundaries under consideration.

As a furtherance and expansion of this work, Laudeman *et al.* (1998) defined a new metric called dynamic density (DD) with a hypothesis that it would better resemble an air traffic controller’s workload than measures based only on traffic density. In fact, this newly defined concept captured both traffic density and traffic complexity (a measure of complexity in a volume of airspace). Additionally, it was assumed that the complexity factors together will better correspond to controller workload than if they were considered individually. The initial intention was first to identify the most relevant complexity factors and, based on those factors, to delineate a weighted linear dynamic density as their function. The traffic factors included in the equation were selected based on informal interviews with experts in the field - current and former air traffic controllers. The complexity factors identified here characterise the traffic present in the sector:

- dynamic factors that captured changes such as aircraft speed or heading
- aircraft density factors that captured the variability in the distribution of aircraft in the sector
- conflict factors that captured predictions of aircraft conflicts.

The traffic factor weightings were computed using multiple regression analysis, but also at the same time subjective weights were collected from survey data. Consequently, the correlations of observed air traffic controller activity with dynamic density values were computed with both regression-weighted and subjectively-weighted traffic complexity factor values. The set of complexity factors appeared to contribute substantially to the variance in controller activity accounted for by dynamic density. Results suggested that dynamic density equation using factor weights computed from the multiple regression analysis was the strongest predictor of air traffic controller’s workload.

Sridhar, Seth and Grabbe (1998) further used this measure and studied how well dynamic density can be predicted up to a specified period of time in advance (up to 20 minutes) using information about future positions and speeds of aircraft by taking into consideration radar tracks, flight plans, aircraft dynamics models and weather data. Results suggested that the availability of inter-centre data, i.e. data with a 250 nm range outside the centre airspace (and most specifically of aircraft intent), can help to extend this analysis for larger prediction intervals. The authors recommended further improvement in accuracy by introducing wind estimates, reduced radar tracker errors, better aircraft models, effects of structural characteristics like airway intersections, as well as other dynamic flow events such weather (Sridhar, Seth and Grabbe 1998).

As some other authors also conducted researches in order to define the list of complexity factors, the question was which of these factors can be considered as the most relevant. Moreover, Masalonis, Callaham and Wanke (2003) assessed four studies (Kopardekar 2000; Chatterji & Sridhar 2001; Laudeman*et al.* 1998 and Wyndemere 1996) to determine which metrics provided unique contributions to the prediction of subjective complexity and to see the extent to which the same complexity factors fared related to subjective workload in different airspaces.

Similarly, Kopardekar and Magyarits (2003) took into account four of the same researches: WJTHC/Titan Systems Metrics (Kopardekar 2000), NASA Metric 1 (Chatterji & Sridhar 2001), NASA Metric 2 (Laudeman *et al.* 1998) and Metron Aviation Metric (Wyndemere 1996). Their intention was to compare these DD metrics and to select the most relevant DD.

In order to provide a better overview of these factors, they are summarized in Table 2 and initially grouped by the categories proposed by Laudemanet al. (1998).The factors identified by both Chatterji and Sridhar (2001) and Wyndemere (1996) fit into these categories. On the other hand, Kopardekar (2000) suggests additional factors that by their nature do not correspond to any of the groups identified by Laudeman et al. (1998). In addition, they define the categories of their own (i.e. coordination task load index and qualitative factors categories as shown in the table below). At the same time, Kopardekar (2000) does not define any factors that are reflecting changes in aircraft speed, heading or altitude (i.e. dynamic factors in Table 2).

Table 2.Complexity factors identified by Laudeman et al. (1998), Wyndemere (1996), Chatterji and Sridhar (2001) and Kopardekar (2000)

| | Laudeman *et al.* (1998) | Wyndemere (1996) | Chatterji and Sridhar (2001) | Kopardekar (2000) |
|---|---|---|---|---|
| **Dynamic factors**<br><br>(factors that captured changes such as aircraft speed and heading) | -Heading Change (HC) - The number of aircraft that made a heading change of greater than 15 degrees during a sample interval of 2min<br>-Altitude Change (AC) - The number of aircraft that made an altitude change of greater than 750 feet during a sample interval of 2min<br>-Speed Change (SC) - The number of aircraft that had a computed airspeed change of greater than 10 knots or 0.02 Mach during a sample interval of 2min | -Count of number of bearing changes above a threshold value with the sector<br>-Count of number of altitude changes above a threshold value with the sector<br>-Count of number of aircraft within a threshold distance of a sector boundary (e.g. 10 miles)<br>-The squared difference between the heading of each aircraft in a sector and the direction of the major axis of the sector, weighted by the sector aspect ratio | -Variance of speed<br>-Ratio of standard deviation of speed to average speed | |
| **Aircraft density factors**<br><br>(factors that captured the variability in the distribution of aircraft in the airspace) | -Minimum Distance 0–5 n. mi. (MD 5) – The number of aircraft with a Euclidean distance of 0–5 n. mi. to the closest other aircraft at the end of each 2min sample interval.<br>-Minimum Distance 5–10 n. mi. (MD 10) – The number of aircraft with a Euclidean distance of 5–10 n. mi. to the closest other aircraft at the end of each 2min sample interval, excluding conflict aircraft | -Aircraft count within a sector<br>-Aircraft count divided by the usable volume of sector airspace. | -Number of aircraft<br>-Number of climbing aircraft<br>-Number of cruising aircraft<br>-Number of descending aircraft | -Aircraft density 1 - number of aircraft/occupied volume of airspace<br>-Aircraft density 2 - number of aircraft/sector Volume<br>-Sector volume<br>-Square of aircraft count |

| | | | | |
|---|---|---|---|---|
| **Conflict factors**<br><br>(factors that captured predictions of aircraft conflicts) | -Conflict Predicted 0–25 n. mi. (CP 25) – The number of aircraft predicted to be in conflict with another aircraft whose lateral distance at the end of each 2min sample interval was 0–25 n. mi.<br>-Conflict Predicted 25–40 n. mi. (CP 40) – The number of aircraft predicted to be in conflict with another aircraft whose lateral distance at the end of each 2min sample interval is 25–40 n. mi.<br>-Conflict Predicted 40–70 n. mi. (CP 70) – The number of aircraft predicted to be in conflict with another aircraft whose lateral distance at the end of each 2min sample interval is 40–70 n. mi. | -Number of aircraft with predicted separation less than a threshold value (e.g.8 miles) at a particular time<br>-The angle of converge between aircraft in a conflict situation<br>-Count of number of other aircraft in close proximity to a potential conflict situation (e.g. within 10 miles laterally and 2000 feet vertically)<br>-Count of predicted conflicts within a threshold distance of a sector boundary (e.g.10 miles) | -Horizontal proximity metric 1<br>-Vertical proximity metric 1<br>-Horizontal proximity measure 2<br>-Vertical proximity measure 2<br>-Horizontal proximity measure 3<br>-Vertical proximity measure 3<br>-Time-to-go to conflict measure 1<br>-Time-to-go to conflict measure 2<br>-Time-to-go to conflict measure 3<br>-Conflict resolution difficulty based on crossing angle | -Convergence recognition index – measure of the difficulty of detecting converging aircraft with shallow angles<br>-Separation criticality index - proximity of conflicting aircraft with respect to their separation minima<br>-Degrees of freedom index – based on manoeuvre options in a conflict situation |
| **Coordination task load index** | | | | -Coordination task load index 1 - based on aircraft distance from the sector boundary prior to hand-off<br>-Coordination task load index 2 – different formula based on the same principle as CTI1 |
| **Qualitative variables** | | | | - categorical variables such as facility and sector types (i.e. high/low) |

Firstly, Kopardekar and Magyarits (2003) performed regression analysis to determine the relationships between different DD variables and complexity ratings. Additionally, a unified metric that contained all variables from all metrics was developed. After a detailed correlation analysis between different metrics and subjective complexity ratings provided by controllers, it was determined that the best results in all conditions were those that included the unified DD metric. Since it consisted of several variables, the researchers used a Principal Component Analysis (PCA) to analyse the relationship among these variables. It was found that variables are significantly correlated with each other and 12 potential factors were extracted.

Also, the prediction of a unified DD metric for a look-ahead-timeframe of up to 120 minutes was investigated by adding a so called “look-ahead” variable into the regression equation. Consequently, this model with “look-ahead” time appeared to better predict the complexity ratings than the instantaneous model. However, it was only slightly better than the predicted aircraft count, possibly due to the inherent inaccuracy in the method of aircraft count prediction rather than the model itself.

Kopardekar and Magyarits (2003) suggested that the further developing and testing of these findings with techniques such as neural networks, genetic algorithms and non-linear regression in the simulation environment would contribute to the accuracy of the variables and their weights.

With similar reasoning, Gianazza and Guittet (2006) suspected non-linear interactions between complexity factors when investigating into dynamic re-sectorisation. They evaluated air traffic complexity metrics by using recorded radar data as an input for a unified list of complexity factors provided by Kopardekar (2001) and as suggested by Kopardekar and Magyarits (2003). Gianazza and Guittet (2006) performed neural networks analysis rather than linear regression to find the relationship between complexity factors and actual sector configuration. It was found that only four of the considered complexity factors were significantly related to the decision to split or merge sectors.

Similarly, Martin *et al.* (2006) investigated into the prediction of workload again using neural networks. To train neural networks, data was used that was calculated using recorded sector data following the approach proposed by Chatterji and Sridhar (2001).

The study conducted by the Federal Aviation Authority FAA (Baart 2001) compared previously identified DD metrics, but using RAMS (the Reorganized Air traffic Control Mathematical Simulator). These DD metrics were: FAA DD metric by Magyarits and Kopardekar (2000), NASA DD Metric by Chatterji and Sridhar (1997) and DD metric developed for RAMS (Baart 2001). Basically, human–in-the-loop simulation workload ratings were compared with DD calculations obtained from an algorithm which uses output from RAMS.

Even though DD metrics give insight into the complexity of traffic, some researchers highlighted that ATC situation cannot be considered as complex or less complex without taking into consideration the subjective perception of the controller about the complexity of the traffic. The hypothesis is that controllers use structure based abstractions to reduce cognitive difficulty and therefore the complexity of ATC situations.

In that manner, Delahaye and Puechmorel (2000) emphasized the importance of intrinsic traffic disorder when investigating complexity. According to their work, it is possible to identify some high density zones and clusters of traffic with strong disorder. Therefore the goal of their work was to extend work done previously by Wyndemere (1996), Laudeman *et al.* (1998) and Sridhar, Seth and Grabbe (1998) by introducing two new classes of indicators: The first one uses geometrical properties in order to build a new complexity coordinate system in which sector complexity evolution through time is represented (Figure 5).

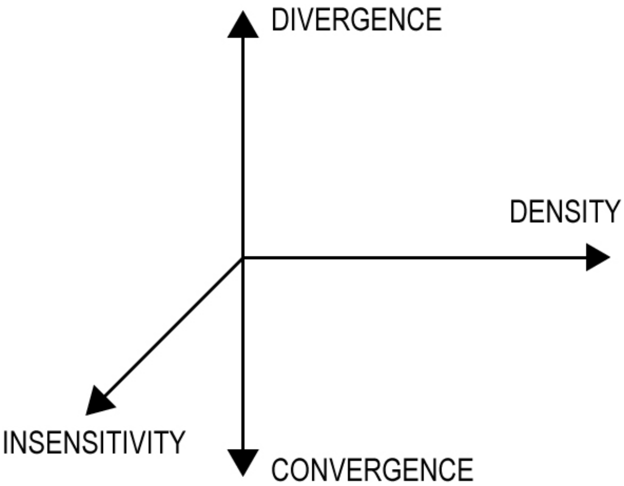



Figure 5. Complexity coordinate system (Delahaye &Puechmorel 2000)

In this new coordinate system, the density axis is a function of relative aircraft distance and their relative speed to each other.

The divergence/convergence axis represents the level of "disorder", where divergence between two aircraft measures how fast they move away from each other and convergence how fast they move toward each other. And the final axis reflects the sensitivity of the conflict duration with the speed and heading modifications. Delahaye and Puechmorel (2000) argue "that a convergent situation with a high sensitivity is better than a convergent situation with a low sensitivity, the induced complexity will be higher in the later than in the former." Therefore, in order to have a homogeneous coordinate system, the final axis represents so-called insensitivity.

Further Delahaye and Puechmorel (2000) suggest that also the evolution of the set of aircraft as a whole should be considered, since the perceived complexity often arises from the observation of the history of the traffic. Thus, they modelled the history of air traffic as the evolution of a hidden dynamical system such that aircraft correspond to pointwise observations. To measure the intrinsic complexity of such dynamical system Delahaye and Puechmorel (2000) use the topological entropy. Accordingly, the second class of indicators describes a complexity indicator based on the dynamic systems theory using the concept of topological entropy (Kolmogorov entropy) which produce a good measure of the actual disorder of the speed vector in an air sector even if the number of aircraft is small (Delahaye and Puechmorel 2000).

Histon *et al.* (2002) argued that cognitive complexity is an attribute of the controllers' working mental model and is related to the scope and details of the factors being considered as well as any simplifying abstractions employed by the controller. Histon *et al.* (2001, 2002) and Histon and Hansman (2002) relied heavily on the controllers' subjective opinions, and based on their focused interviews with them they listed the key factors influencing complexity. Even though no attempt was made in ranking them, they were classified into three categories: airspace factors, traffic factors and operational constraints.

Thus, when the set of complexity factors was identified, Histon *et al.* (2001) and Histon and Hansman (2002) studied it more deeply, in order to reveal structure-based abstractions that simplify the cognitive complexity of an air traffic controller's task. Actually, the underlying structure was identified as relevant in many of the factors. Moreover, based on the analysis of the field observations three key structure –based abstractions are suggested:

- standard flows (based on explicit structural elements in the airspace such as navigational aids, airways and documented standardized procedures-including ingress and egress points; and common practices, or standardized, but unpublished patterns of operation)

- groupings (associating aircraft linked by common properties-proximity, speeds etc. e.g. a wide distribution of speeds makes the process of projecting future positions more difficult than if all aircraft fly at a uniform speed)
- critical points (based on known "hot-spots", or locations where controllers know to expect potential conflicts, e.g. in the form of crossing or merge points of flows).

To explore the influence of structure on cognitive complexity in more depth, the authors used a model of Situational Awareness in Human Decision Making developed by Endsley (1994). And so, in spite of some differences, a generalized ATC process model was proposed, which included each of the following three levels of a controller's situational awareness: Decision process and performance of actions as well as function of structure-based abstractions in these processes (Histon *et al.*2001,2002; Davison *et al.* (2003); Reynolds *et al.*2002 and Histon and Hansman 2002).

While these authors conducted studies in order to identify relevant complexity factors, some other authors used these factors in their work for further investigation, or just listed factors they considered relevant according to the desired results. Therefore, when comparing the performances of the European and American centres, Flynn, Leleu and Zerrouki (2003) emphasized many factors that contribute to complexity, but for their study they considered only the following categories:

- General traffic measures - number of flights, kilometres controlled, flight hours controlled.
- Airspace measures – dimensions, volume, sector characteristics.
- Traffic pattern characteristics – spatial distribution, vertical movements of flights, traffic flows.

Pawlak *et al.* (1996) worked on the framework for developing and evaluating a model of perceived complexity of an air traffic situation. Therefore, one of the phases of their study focused on the identification of complexity factors. Pawlak *et al.* (1996) relied on their own experience and expertise in the field, and held a number of meetings designed to identify a set of initial complexity factors. Input in these meetings included reviews of existing studies and various ATC manuals. To validate these factors, controllers were asked to think aloud as they made a decision about dealing with the traffic during simulations. Also, after each scenario, they were asked to rate how difficult it was to control the traffic, considering both safety and efficiency.

As previously mentioned, research work conducted by Wyndemere (1996) also included a number of meetings designed to identify a set of initial complexity factors that might be useful for a measurement of air traffic complexity.

Mogford *et al.* (1994) conducted a study with the intention of identifying factors that contribute to airspace complexity. In this attempt, work was divided into two phases: phase 1 – factor identification phase, created a list of candidate sector complexity factors; phase 2 – factor selection; the list of factors that was used by the controllers to evaluate all of the sectors on factor complexity ratings, and then the resulting complexity ratings were analysed in their ability to account for the overall sector complexity. Moreover, Mogford *et al.* (1994) and Mogford, Murphy and Guttman (1999) used both rating scales and interviews as a direct data collection in comparison with the results obtained from indirect data collection. Controllers firstly rated their familiarity with the sectors and then made comparisons in complexity between all possible pairs of sectors. After the similarity judgments were collected, controllers were asked to list factors that they thought contributed to sector complexity. A revised list of 16 factors was finally created using the calculated correlations between overall complexity and complexity ratings of each factor in addition to conducting a further multiple regression analysis to determine which of the final factors were most salient in accounting for overall sector complexity. The resulted list of complexity factors is shown in Table 3 below:

Table 3. Complexity factors as identified by Mogford *et al.* (1994)

| **Complexity factors as identified by Mogford *et al.* (1994)** | |
|---|---|
| 1 | Number of climbing or descending aircraft |
| 2 | Degree of aircraft mix (e.g., visual-flight rules, instrument-flight rules, props, turboprops, jets) |
| 3 | Number of intersecting aircraft flight paths |
| 4 | Number of multiple functions the controller must perform (e.g., approach control, terminal feeder, en route, and in-trail spacing) |
| 5 | Number of required procedures to be performed |
| 6 | Number of military flights |
| 7 | Frequency of contacts (coordination) or interface with other entities (e.g. adjacent sectors, approach controls, centres, or military facilities) |
| 8 | Extent to which the controller is affected by airline hubbing or major terminal/airport traffic |

| | |
|---|---|
| 9 | Extent to which weather-related factors affect ATC |
| 10 | Number of complex aircraft routings |
| 11 | Extent to which the controller's work is affected by restricted areas, warning areas, and military operating areas and their associated activities |
| 12 | Size of sector airspace |
| 13 | Requirement for longitudinal sequencing and spacing |
| 14 | Adequacy and reliability of radio and radar coverage |
| 15 | Amount of radio-frequency congestion |
| 16 | Average volume of traffic |

In the previously mentioned studies complexity and factors affecting this complexity were investigated in relationship with other factors influencing controller's cognition, workload, performance and activity.

On the contrary, some authors used complexity measures as criteria in sector classification and sector comparison. These papers were found to be interesting from the aspect of the selection of the appropriate factors for sector comparison.

Accordingly, Christien and Benkouar (2003) described a classification process that identifies groups of sectors sharing similar complexity indicators levels. They used the hypothesis stated by Chaboud *et al.* (2000), that sectors with similar complexity indicator levels would have a similar macroscopic workload model. Also, authors reasoned that this classification process allows improvement of the quality of the sectors' workload evaluation and capacity estimates within the airspace.

Manning, Fox and Pfleiderer (2003) applied a method where controllers were prompted every two minutes during each traffic sample to provide complexity ratings on a 1-7 point scale using an electronic keypad. Alike, both Kopardekar and Magyarits (2003) and Pfleiderer (2005) used SATORI (Systematic Air Traffic Operations Initiative) to replay traffic samples, for which controllers rated complexity every two minutes. SATORI is a research tool that uses routinely recorded ATC computer and voice data to recreate and display air traffic control operational incidents in the same way that they appeared on the controller's radar screen.

Moreover, Chaboud *et al.* (2000) combined two different approaches while addressing the issue of measuring the level of service through analysing air traffic complexity. One approach is adopted by Eurocontrol Experimental Centre (EEC) and estimates the production process, especially the Air Traffic Service (ATS) workload involved in delivering the service or output. The other approach is adopted by the National Air Traffic Service Limited of the UK (NATS), which estimates ATS output i.e. the provided service.

In many previously cited researches where complexity in ATC was investigated, one of the crucial aspects that was taken into consideration is the controller’s perception about the ATC situation. Even if the number of aircraft remains the same in certain situations, the controller’s perception about the complexity of the situation might change. Therefore, besides objectively measurable complexity factors, the subjective assessments of the controllers were taken into consideration. These subjective responses were obtained either through interviews, questionnaires or other subjective complexity assessment techniques (e.g. ratings on scales).
For that reason, to examine factors that contribute to these changes in complexity level, it is of utmost importance to examine the correlation between complexity factors and the task load and workload measures of controllers as they provide an insight into controller’s perception of complexity.

## 2.5 ATC Validations

Many ongoing European projects are focusing on developments that will improve ATC performance when it comes to capacity, safety, environmental impacts, economy and security. For example, SESAR Programme aims at enabling a 3-fold increase in capacity which will also reduce delays both on the ground and in the air; improving safety by a factor of 10; enabling a 10 % reduction in the effects flights have on the environment and providing ATM services to the airspace users at a cost of at least 50% less (SESAR, 2006).
The solutions are sought in new technologies as well as in new designs and concepts both on the ground side and in the airborne, such as ALICIA[5] (All Condition Operations and Innovative Cockpit Infrastructure) and ACROSS[6] (Advanced Cockpit for Reduction Of Stress and Workload), both co-funded by European Commission under the Seventh Framework Programme.

---

[5] http://www.alicia-project.eu/
[6] http://www.across-fp7.eu/

However, as emphasized in SESAR (2008), but equally valid in other projects, the human element remains central in ensuring expected benefits (in safety, environment, cost efficiency and capacity).
That is, the introduction of new technologies and concepts would undoubtedly impact the current work of controllers. Performing the validation activities with the controllers ensures that before any modifications are made to the existing ATC elements, or new concepts, new tools, new working methods or a new Human Machine Interface (HMI) are introduced into the current ATC system, possible issues and negative impacts on controllers’ performance are timely identified and addressed.

There are a wide range of validation settings, techniques, methods, and tools in support to validation activities. However, their selection depends on the level of the maturity of the technology or concept under consideration. The level of the maturity of the technology or concept ”facilitates the settings of appropriate validation objectives, the choice of evaluation techniques, shows how concept validation interfaces with product development and indicates where requirements should be determined” (EUROCONTROL, 2010b). The European Operational Concept Validation Methodology (E-OCVM) identifies 6 levels of concept maturity (V0-V5) ranging from the outset of the idea based on the ATM needs, to the final implementation of the concept (EUROCONTROL, 2010b). The later phases of the concept lifecycle (Pre-Operational and Operational maturity phases) indicate that the concept is mature enough to be transformed into industrial products ready for implementation. By these stages, all the identified issues and possible negative impact on the controller are addressed and resolved. Therefore, the negative impact on the workload of the controller should be tackled before, while performing the V0-V4 validation activities.
The validation settings used in these stages of the concept lifecycle are initially low-fidelity mock-ups/prototypes, followed by more mature and further developed mock-ups, prototypes, simulators and operational trials (EUROCONTROL, 2010b).

Mock-ups and prototypes are usually used in very early stages of the concept design to acquire feedback from users about designs and design ideas. “Mock-ups are 'very early prototypes' made of cardboard or otherwise low-fidelity materials. The user, aided by the designer, may test the mock-up (imagining that it works) and thus provide valuable feedback about functionality/usability/understanding of the basic design idea/etc.” (Interaction Design

Foundation, 2004). Prototypes are used to make assumptions about technical aspects in order to avoid system engineering which can be costly and lengthy (EUROCONTROL, 2010b). They are considered "as a concrete representation of part or all of an interactive system. A prototype is a tangible artifact, not an abstract description that requires interpretation. Designers, as well as managers, developers, customers and end-users, can use these artifacts to envision and reflect upon the final system." (Beaudouin-Lafon & Mackay, 2002)

"Simulation is an engineering method used to predict a system's performance and to assess working methods, layouts, risk identification, and training needs. The aim of the simulations is to produce an environment that represents pertinent features of the realistic environment in which the system will operate, to be able to predict the systems operational performance" (EUROCONTROL 2011).

Simulations can address many aspects of the ATM system, such as: airspace configuration, air/ground surveillance and communication equipment, flight data systems, controller working methods and procedures, controller Working Position (CWP) and console, traffic levels corresponding to current or forecast levels, and new ATM concepts for general use etc. The resources required to conduct a simulation depend largely on the nature of the sought results and realism of the simulation. Therefore, they could range from very simple ones, such as pen and paper simulations, to those that require more sophisticated apparatuses: hardware, software, defined functions and tasks, operating procedures, and suitable subjects.

There are two different types of simulations:

- model-based simulations (also known as fast-time simulations) and
- real-time simulations (i.e. human-in-the-loop simulations).

The model-based simulation is a general term for analytical, macroscopic and fast-time simulator tools. Another frequently used name is fast-time simulation.

Model-based simulations are generally done without human observers or controllers and can be referred to as computer simulations. Controller behaviour and decision making are defined by the rule base of the model-based simulation tool.

The simulator is configured to realistically represent the required ATC system. Airspace and traffic are processed mathematically within a computer program which models controller tasks. Results are statistics for traffic load/flow plus an assessment of controller workload for all tasks including conflict detection and resolution.

The advantage of model-based simulations is that by using them it is possible to rapidly evaluate a large number of combinations of airspace and traffic, as well as simulating either very large or very small geographical areas with only a few resources. On the other hand, the weakness of this type of simulation is that the controller interaction with the problem under study is very limited.

Real-time simulations (RTS) are also referred to as human-in-the-loop simulations. Simulators used in real-time simulations are configured to represent as realistically as possible both the new concept but also the context in which this novelty should be implemented.

In the course of a real-time simulation, generally qualified controllers participate in a series of exercises designed to evaluate either controlled airspace, procedures to be applied or the ATC system. Nevertheless, depending on the concept that is validated, participation of the qualified controllers is not always required, but participation of the pseudo-ATCOs may meet the needs of the validation. The roles of pseudo-ATCOs may be taken by subject matter experts or other personnel with sufficient knowledge on the concept under validation. Also pseudo-pilots participate in this type of simulation to provide an essential element of realistic radio/telephony (R/T) and data link (DL) communication. In many cases the RTS is used to verify the results of model based studies. Because of realism that they provide, real-time simulations require a large amount of resources and therefore are rather costly. Nevertheless, RTS is frequently used for the validation before new operational concepts are introduced.

Figure 6. Controller Working Position in the simulation environment

The conduct of the real-time simulation consists of several important steps. The type of simulation depends on its focus- be it, for example, a new concept, a new procedure or a new tool to be introduced. In accordance with this, the tasks to be investigated are determined plus the required level of realism needed.

As the simulation represents the imitation of the operation of a real-world process or system over time, the simulation involves the generation of an artificial history of the system and the observation of that artificial history. This aims to draw inferences concerning the operating characteristics of the real system that is represented. A simulation can be used to investigate a wide variety of "what if" questions about the real-world system. One of the greatest advantages of using simulations is that once a valid simulation model is developed, it is feasible to better understand the interactions among the variables that make up such a complex system. Diagnosing problems and gaining insight into the importance of these variables increases understanding of their important effects on the performance of the overall system (Banks 1998).

Therefore, one of the most important steps in conducting the simulation is the identification of the parameters to be recorded and the recording processes (e.g. performance data from a part task simulation) for subsequent analysis that will allow the investigator to assess workload, predict system performance, evaluate alternatives, evaluate layout and procedures, identify training needs, and identify potential risks.

So, it is of the utmost importance, once the objective of the simulation is defined, to set up an experimental plan for the simulations, be they high and low level objectives, associated hypotheses, and which methods and measurements to carry out in order to obtain relevant evidence to support those hypotheses (EUROCONTROL, 2010b).

At the same time some other factors should be taken into consideration as well. These include the validity of stimuli, realism of control responses, stimulus-response relationships between displays and controls, task complexity, temporal representation, environmental conditions, situational pay-off (should the simulator be built to represent the task or provide a safe area for simulation?), social environment, and the control the investigator has over the simulation conditions.

The major advantage of the real-time simulations is that in a large scale system (e.g. aircraft and ATM networks) they provide an economical and significantly safer alternative to live trials on operational systems. Nevertheless, the activities assessed in the simulations may not be fully representative of those in the actual system and also during the simulations large quantities of data might be collected that will need filtering and careful analysis. Therefore, full simulation studies that aim to high realism are extremely expensive.

## 2.6 Statement of the problem

Numerous studies have focused on the investigation into ATC complexity. The majority have emphasized their significant influence on the controller's workload, and consequently on ATC safety. Different approaches and even different sets of relevant complexity factors were identified, but a unique and widely accepted set of complexity factors could not be defined, yet.

Overall, air traffic complexity was considered in correlation with and not independently from two main aspects: cognitive factors, such as the degradation in the controllers' workload and performance factors - controllers' mistakes, slip-ups and lapses. Additionally, it was shown that there is a strong influence of changes in workload level to the proposed sets of complexity indicators. Therefore, their use value is potentially high in the prediction of the controllers' workload.

Nevertheless, not much work was focused on identifying the individual influence of each complexity factor independently because proposed sets of complexity factors were investigated frequently as drivers of changes in workload level as a whole. However, what drives complexity, and further-on workload, in one airspace sector, can be trivial in other, where complexity is driven by a different complexity factor. Yet despite this, overall complexity levels for both airspace sectors may be assessed as equal. Thus, an accurate comparison of complexity across facilities is difficult if it is not possible to distinguish which complexity factor is more or less present in the considered airspace sectors.

Furthermore, when new concepts in ATC are envisaged, first it should be proven their application is not jeopardizing existing system functionality before being employed. For this reason, often either real-time or model-based simulations are used (see section 2.5). To build a simulation that would adequately address validation objectives, it is necessary to understand the interactions among all the relevant variables impacted by the new concept. However, as shown in Figure 7, it is assumed that interface and equipment demands, procedural demands and above all, ATC complexity represent the task demand imposed on the controller to perform certain activities. Those activities further on mediated by performance shaping factors (such as skill, age, training etc.) are translated into workload experienced by the controller.

Gaining insight into the importance of these variables increases understanding of their important effects on the performance and workload of the controller.

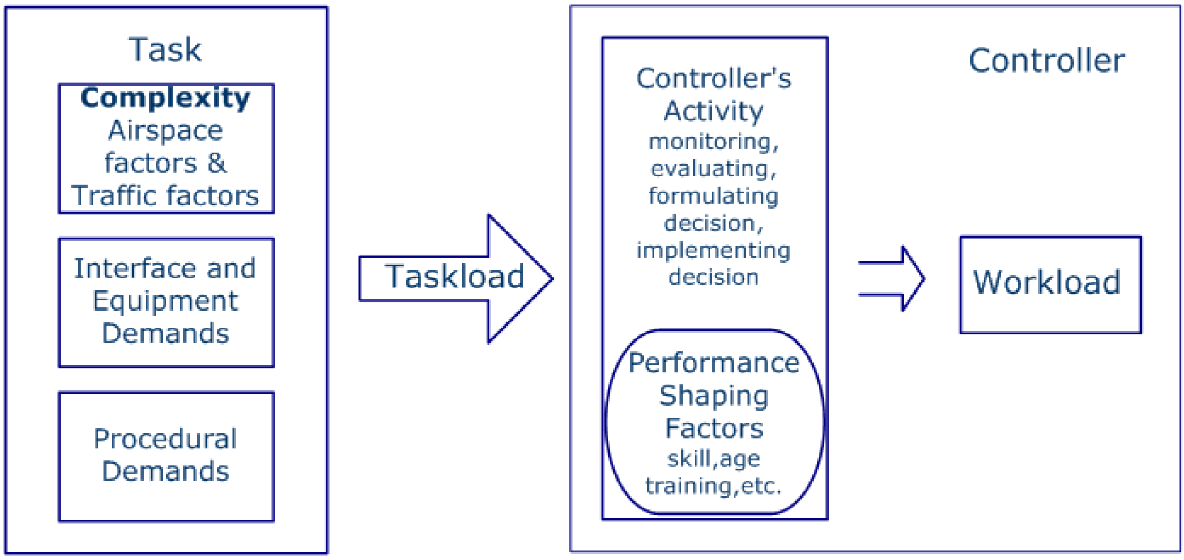



Figure 7. Simplified scheme of the relationship between ATC complexity and workload, adapted from Hilburn (2004)

Additionally, if the impact of modifications of some variables is in certain way predictable, we can manipulate a definite set of parameters and investigate its influence on the controller's workload. This aspect can prove very useful when designing validation scenarios against which new concepts would be assessed.

## 2.7 Scope, overview and applications of the research

The focus of this thesis is **the measurement of ATC complexity and the impact it creates onto the workload of air traffic controllers**.

Based on the identified scope, the aim of the research is twofold:

Firstly, it aims to identify a complexity measure that would adequately correspond to the controller's workload that can be applicable in different ATC sectors and under different simulated conditions (e.g. different traffic loads).

Therefore, **the first hypothesis** of the research is:

**Using the objectively recorded data it is possible to measure air traffic complexity as a predictor of the controller's workload.**

Secondly, the research aims to go beyond straightforward prediction of workload. That is, if a set of complexity factors as a whole performs well as a workload predictor, it is assumed that there are several complexity components (driven by limited number of related complexity factors) that can account for and explain overall complexity of the considered ATC situation and additionally that not all of these components correspond to the workload in the same way.

Hence, **the second hypothesis** is:

**The workload is differently impacted by individual components of complexity. Understanding these differences can facilitate comparison of the complexity levels of a single sector under different conditions, but also comparison of complexity levels of different sectors under same conditions.**

However, it should be noted that different sector units (area control centre, approach control unit or aerodrome control tower) are characterized with different airspace design, the traffic flows and airway structure are also differently organized, and the operational procedures of the controllers differ for these ATC units (ICAO, 2007). Consequently, it is assumed that the complexity associated to such sectors will also differ.

Therefore, in order to have consistent and adequate comparison of the measured complexity levels in different sectors, and not to be biased by the predefined elementary discrepancies of the different sector types, the scope of this thesis is an investigation only into ***area or en-route sectors***.

Although challenging and promising, there remains as an opportunity for future research to investigate into the complexity of other sector types (e.g. approach and terminal area), as proposed by Vogel, Schelbert, Fricke and Kistan (2013).

The results of the research originating from these hypotheses may have wide range of applications. Several of these applications are addressing:

- the improvement and the design of the airspace structure.
- the validation of the new concepts, procedures and technologies to be introduced in the current system.
- support air traffic management in the decision-making processes.

In order to address air traffic control complexity as the main driver of the workload experienced by controllers and also to understand how different components of complexity correspond to workload and different operational conditions, an adequate approach was defined.

# 3 Approach, Methods and Experimental Scenarios

This chapter presents the approach applied in order to investigate into ATC complexity factors and their impact on the workload of ATCOs. Also it addresses the methods and experimental scenarios used for the collection of data as evidence in support of the research hypotheses.

## 3.1 Approach of the research

With the purpose of identifying an adequate approach to investigate into ATC complexity and its causal relationship to the workload of the controller, the starting point was an in-depth examination of the defined research hypotheses.
Since the hypotheses are defined as quite generic and high level, they are broken into the research objectives that are more tangible and therefore assessable.

The first hypothesis can be deconstructed into the following research objectives:

- **Objective 1:** Identify objectively assessable complexity measures
- **Objective 2:** Evaluate the relationship between identified complexity measures and controllers' performance measures and subjective assessment of the controllers' workload
- **Objective 3:** Assess the predictive power of the identified complexity measures on the controllers' performance measures and workload when applied under different experimental conditions

The objectives drawn out from the second hypothesis are:

- **Objective 4:** Assess the impact of different experimental conditions on the performance measures and workload of the controller
- **Objective 5:** Assess the impact of different experimental conditions on the complexity measures

- **Objective 6:** Compare the effects of different conditions on the controllers' performance and workload measures with the effects on the complexity measures, and identify those complexity measures that are potential drivers of these effects.

In order to achieve all of the stated objectives, they should be addressed in the order as listed above.
This approach initially enables recognition of objectively measurable complexity factors relevant in the context of the research, continues with the investigation into how they are impacting the changes in controllers' performance and experienced workload, and concludes with the analysis of how all these factors are actually driven by different conditions.

They draw the outline of the approach and the steps to be followed and therefore they are here explained sequentially together with the methods intended for their accomplishment:

**Objective 1: Identify objectively assessable complexity measures**

To be able to capture more accurately ATC complexity, it is evident that besides taking into consideration the simple number of aircraft in the sector, it is necessary to consider other important factors. The significant element of the ATC complexity is the involvement of each individual aircraft and its flight characteristics that relate to instantaneous changes of the state of the aircraft, such as changes in altitude, heading or speed.
However, looking at each aircraft independently it is not sufficient to comprehend the overall situation and therefore to have an insight into the overall situational complexity. Besides this, the number of aircraft with the transitioning related to speed, altitude or direction that is not negligible plays an important role in the changes of the level of complexity.
Further, the interactions between all pairs of aircraft are substantially contributing to the overall ATC complexity.
As discussed earlier, interactions between aircraft are not considered only in terms of potential conflicts, but furthermore include the pattern of how aircraft converge or diverge from each other and, as defined by Delahaye and Puechmorel (2000), the degree of disorder among aircraft, i.e. the variability in headings and speeds of aircraft.

Having all this in mind, to create a comprehensive list of complexity factors the following aspects (categories) should be considered:

- aircraft density (concentration of aircraft in the measure of space and their count).
- flight attributes of each individual aircraft (considering also the count of those in the process of transitioning - changing speed, direction, altitude).
- aircraft conflicts (distance between aircraft, speed with which they are moving to/from each other, etc.)
- traffic disorder (discrepancy in their speeds and headings).

Nevertheless, as described in the section 2.4, lots of research was conducted in this field and also lots of objectively measurable factors were defined. Therefore, in a certain way Objective 1 of the research has been already completed by other authors in numerous previous researches. However, in the context of the present thesis, this is true only if looked at independently from the other requirements and the subsequent objectives of this research.

Therefore, as it is hardly possible, but also redundant, to proceed with the identification of completely new complexity factors, the aim of the initial step of this approach is to identify and select by relevance plus to adopt and combine factors identified by other authors in the past.
The selection of the factors relies on two important criteria: 1 – the selected complexity factors should concern / fall into one of the identified categories defined above; 2 – the list of selected factors should be limited in their number, so that once all previously defined aspects of complexity important for the next phases of this research are potentially covered, the list should be concluded in order not to continue indefinitely. The list of the complexity factors selected to comply with these criteria is further addressed in the section ahead (3.3.4 ATC Complexity factors).

Following this, the predefined set of complexity factors should be examined for possible redundancy. That is, since the established set of factors resulted from multiple researches conducted in this field, it is likely that some of these factors are overlapping and are correlated with one another, given that, to certain extent, they are measuring similar concepts. Actually the intention is to perform a reduction by combining information contained within these factors into a smaller number of new artificial variables and by deleting statistically redundant portions of these factors prior to conducting further analysis.

To evaluate such relationships between selected factors many statistical methods may be employed, both linear and non-linear. However, within the scope of this research, it was decided one of the linear data reduction techniques. This is reasoned out with two main arguments: firstly, one of the purposes of this research is to identify a limited set of complexity measures that would be easily reproduced or calculated when applied on a new set of data. Therefore, a more applicable method to achieve this is through straightforward linear modelling rather than probing for non-linear models where the complexity factors identified by different authors for different purposes would be forced into one model and therefore would be difficult to interpret and recalculate when applied on the new set of data. The second reason, that actually relies on the former one, is that the identified set of complexity factors, within the scope of this work, when merged and combined should retain the key information related to the complexity areas of interest (complexity categories listed above) and that the redundant and marginally significant elements would be disregarded.

In this light, the suitable method to be applied is Principal Component Analysis (PCA) that is also commonly used for these purposes. PCA is a statistical method for expressing the data in such a way as to highlight their similarities and differences, while at the same time reducing the number of variables, by creating a smaller number of artificial variables (called principal components) that nonetheless retains the majority of the information. These principal components may then be used as predictor or criterion variables in subsequent analyses and in the context of the current research they are referred to as complexity components.

General concepts and objectives of PCA are further detailed in the following section addressing the methods used in this research (3.2.1 Principal Component Analysis (PCA)).

**Objective 2: Evaluate the relationship between identified complexity measures and controllers' performance measures and subjective assessment of the controllers' workload**

Once the complexity components are defined, further analysis is performed to investigate into their relationships with controllers' performance and workload measures.

For this step, similar as for the objective one a number of statistical methods, can be used considering whether this causal relationship is either linear or non-linear. Yet again, the use of one of the linear methods was opted for. In general, non-linear regression is used when the prediction cannot be described with a linear equation. However, standard linear regression is able to fit many non-straight line relationships. This is done by creating variables with squares, taking the natural log of a variable or any other algebra function on a variable the algorithm is supported to solve for curved fits.

Linear regression has another important advantage, specifically in the convergence of the algorithms to calculate the coefficients. Non-linear methods rely on the iterative procedures to identify the relationship between dependent and independent (predictor) variables. The use of iterative procedures requires the starting values for the unknown parameters before the optimization is initiated. The starting values must be reasonably close to the as yet unknown parameter estimates or the optimization procedure may not converge. For example, the neural network, one of the commonly used non-linear methods, needs a large training sample, and the resulting neural networks cannot be retrained. If the data is added sequentially, this is almost impossible to add to an already existing network.

As a result, in order to understand the relationships between the complexity components and the performance and workload of controllers, and to develop a predictive model, a multiple-regression analysis has been applied.

General concepts and objectives of multiple-regression analysis are further detailed in the following subchapter addressing the methods used in this research (3.2.2 Multiple regression analysis).

**<u>Objective 3: Assess the predictive power of the identified complexity measures on the controllers’ performance measures and workload when applied under different experimental conditions</u>**

The question that arises is as to whether those complexity components will also prove relevant under different conditions, in different airspace and sectors considered, and also when different subjects are observed. Therefore, the approach is further developed in order to validate the complexity components’ application and their predictive power when applied in different airspace and under changed conditions. Therefore, in this step of the approach due heed is paid to test previously obtained data on the new dataset where it would applied on different sectors, operational conditions and procedures, and also with the participation of different subjects.

A detailed description of the experiment exercises and materials used, as well as data collection methods, are presented in the section 3.3 Experimental scenarios and data collection.

The objectives 4 and 5 are explained here together because within this approach they are addressed simultaneously applying the same method but on the different set of variables (performance and workload of the controller and complexity components).

Also, a description of objective 6 is incorporated within this description as it results from the comparative analyses of the findings obtained for objectives 4 and 5.

**Objective 4: Assess the impact of different experimental conditions on the performance measures and workload of the controller**

**Objective 5: Assess the impact of different experimental conditions on the complexity measures**

**Objective 6: Compare the effects of different conditions on the controllers' performance and workload measures with the effects on the complexity measures, and identify those complexity measures that are potential drivers of these effects.**

Referring to the hypothesis based on which these objectives have been extracted, the intention is to demonstrate that changes in the activities of the controller and workload are driven by different factors (complexity components) in different sectors and under different operational conditions.

Therefore, the first step would be to see how the performance and the workload of the controllers change when the conditions under which they operate are changing. Subsequently, in order to be able to map these changes on the complexity components, it is necessary also to investigate into the changes that the complexity components undergo when observed under different conditions. Therefore, controllers' performance measures and workload that are confirmed to be related to the construct (for which there are theoretical grounds for expecting it to be related) are independently examined. This is in regards to different airspace sectors' structure and different traffic loads, as well as their evolution through time during analysed traffic samples, followed by the same analysis performed on the complexity components.

Therefore, a statistical method that could assess the impact of a set of experimental conditions (sector, traffic load, and time interval into exercise) on the various workload and controller performance metrics, as well as on the complexity components, was sought to address these objectives.

To these aims, a particularly suitable statistical method to use is the Analysis Of Variance (ANOVA). ANOVA is, just like multiple-regression, from a mathematical statistical point of view, a form of General Linear Model (GLM), which can be used to test the effects of one or more experimental factors (i.e. categorical independent variables), and of their interactions, on a single dependent variable. An advantage of ANOVA is that it is very easy to model the effect not only of between-subjects factors, but also of within-subject factors, that arise in the case that the same subjects are tested multiple times in different conditions. In this case, we speak of a repeated measures ANOVA, in that the GLM is used to model dependent variables measured at multiple times using analysis of variance. In the ANOVA/GLM framework, moreover, it is possible to include covariates to check for possible confounds that could not be addressed in the design of the experiment.

Further descriptions of the concept and objectives of the GLM and in particular ANOVA are provided in the subchapter 3.2.3 General Linear Model (GLM).

For a better overview of the proposed research approach, on Figure 8 below is depicted the simplified representation of this approach starting from the research hypotheses, through associated objectives and finally methods used to obtain and analyse data as evidence to support these hypotheses.


Figure 8. Scheme of the research approach (hypotheses, objectives and applied methods)

## 3.2 Methods

In this section more into details are addressed statistical methods used for the analysis of the data obtained in the real-time simulations.

As described in the previous section (section 3.1), once when the relevant complexity factors have been selected, their similarities and differences have been looked at in order reduce their number, while at the same time retaining the majority of the information that they account for (Objective 1). To achieve this objective, Principal Component Analysis (PCA) is performed. The following section (3.2.1) provides more details on this statistical method.
With the aim of revealing the predictive power of the reduced set of complexity components (obtained through PCA) on controllers' performance measures and workload (objective 2, see section 3.1), multiple regression analysis was opted for. Therefore, more detailed description of this method is addressed in the section 3.2.2.
Finally, the method used to assess the impact of a set of experimental conditions (sector, traffic load, and time interval into exercise) on the complexity components, as well as on the workload and controller performance measures (objectives 4, 5 and 6), General Linear Model (GLM) is explained (see section 3.2.3).

### 3.2.1 Principal Component Analysis (PCA)

Principal component analysis is a variable reduction procedure that is useful when data is obtained for a large number of variables (in the present study those are complexity factors). Principal component analysis is a large-sample procedure, which is also in line with the amount of data that can be collected during the real-time simulation. The aim of PCA is to reduce the large number of complexity factors into a smaller number of principal components (artificial variables) that will account for most of the variance in the observed variables. The resulting principal components may then be used in subsequent analyses.
These new artificial variables (principal components) present linear combinations of observed variables are weighted in such a way that the resulting components account for a maximal amount of variance in the data set:

$$C_1 = b_{11}(X) + b_{12}(X_2) + \ldots b_1p(X) \quad (1)$$

where

$C_1$ = the score on principal component 1 (the first component extracted)

$b_{1p}$ = the regression coefficient (or weight) for observed variable p, as used in creating principal component 1

$X_p$ = the score on observed variable p (transformed so that it has a mean of zero and a variance of one).

When a new data set is obtained, component scores for this new data set can be calculated using the regression coefficients (weights) resulting from the principal component analysis performed on the previously observed data set.

The number of components extracted in a principal component analysis is equal to the number of observed variables being analysed. Nevertheless, only the first few components account for significant amounts of variance (eigenvalues), as each new component is accounting for progressively smaller and smaller amounts of variance. Therefore, only the first few components are retained for the subsequent analyses (such as in multiple regression analyses that is the following analysis within defined methodology).

There are many criteria to determine the number of components that should be considered as significant. One of them, the most commonly used, is the Kaiser criterion (Kaiser 1960). According to this criterion only components with the eigenvalue greater than 1 are retained for further analysis, while those components whose eigenvalue are less than 1 are excluded as insignificant. Namely, each variable contributes one unit of variance to the total variance in the data set. Therefore, any component with eigenvalue greater than 1.00 accounts for a greater amount of variance than had been contributed by one variable, while a component with an eigenvalue less than 1.00 accounts for less variance than had been contributed by one variable.

Since each of the resulting components is composed of more than one observed variable, the interpretation of component construct is hindered by the meaning of each contributing variable. To facilitate this interpretation these component solutions are linearly transformed (rotated).

An orthogonal solution is one in which the components remain uncorrelated, while the solution that results in correlated components is oblique. For that reason, orthogonal solutions are also rather less complicated to interpret and therefore used in the context of this study.

Further interpretation of components is determined by inspecting the correlations between each observed variables and the components, correlations known as factors/component *loadings* – identifying the variables that demonstrate high loadings for a given component, and determining what these variables have in common. The most commonly used criterion for considering loading as "high" is if its value is greater than 0.40. Generally when some variable has a loading higher than 0.40 on more than one component, that variable is excluded from further analysis since it is not a pure measure of only one construct.

### 3.2.2 Multiple regression analysis

Multiple regression is a flexible method of data analysis that may be appropriate whenever a quantitative variable (the dependent or criterion variable) is to be examined in relationship to (typically) several other factors (expressed as independent or predictor variables) (Cohen et al 2003).

A multiple regression equation for predicting Y can be expressed a follows:

$$Y' = A + B_1X_1 + B_2X_2 + \ldots + B_kX_k \quad (2)$$

The values for A and the $B_{1 \ldots k}$ are determined mathematically to minimize the sum of squared deviations between the predicted Y′ and the actual Y scores and $X_1$, $X_2$, $X_3$,..., $X_k$ a set of independent variables or potential predictor variables.

B-coefficients, which are unstandardized original values, show the net effect in Y which is associated with one unit change in X, while the other variables are kept constant.

Because beta-coefficients are standardized values of B-coefficients, they are more convenient for comparing the "effects" of different variables within equations, having different unit of measures.

If the b-coefficient is significant, determined by applying the t-test to the ratio of the coefficient to its standard error, then the beta-coefficient is significant.

The correlation between Y′ and the actual Y value is also called the multiple correlation coefficient, $R_{y.12...k}$, or simply R. Thus, R provides a measure of how well Y can be predicted from the set of X scores. In ordinary least squares regression, the squared value of the multiple correlation coefficient, $R^2$, also known as *coefficient of determination*, is a measure of the goodness-of-fit of the model to the data, as it quantifies the proportion of the variance in the outcome that is explained by the model (i.e. by a specific linear combination of predictors).

An especially useful application of multiple regression analysis is to determine whether a set of variables (Set B) contributes to the prediction of Y beyond the contribution of a prior set (Set A). The statistic of interest here is the change in R squared, that is the difference between the R squared of the model containing both sets of variables ($R^2_{Y.AB}$) and the R squared of the model including only the first set ($R^2_{Y.A}$).

The F value associated with a multiple regression equation tests for the significance of the multiple R for the entire equation.

It is possible to have a significant R and some variables that are not significant. These probably should be removed from the equation, for they are not adding any appreciable explanation -- but there may be theoretical reasons for leaving them in.

Different types of multiple regression are distinguished by the method for entering the independent variables into the analysis:

- standard (or simultaneous) multiple regression: all of the independent variables are entered into the analysis at the same step.
- hierarchical (or sequential) multiple regression: the independent variables are entered in an order prescribed by the analyst, usually according to theoretical reasons.
- stepwise (or statistical) multiple regression: the independent variables are entered according to their statistical contribution in explaining the variance in the dependent variable.

Stepwise regression is designed to find the most parsimonious set of predictors that are most effective in predicting the dependent variable. Variables are added to the regression equation one at a time, using the statistical criterion of maximizing the $R^2$ of the included variables. Starting from a model in which the outcome is modelled by a constant, at each stepwise regression step all the possible predictors are considered and the one that is explaining the

most variance is selected and entered into the model. The procedure is then repeated adding further variables according to their relative contribution to explanative power of the model.
The process of adding more variables stops when all of the available variables have been included or when it is not possible to make a statistically significant improvement in $R^2$ using any of the variables not yet included. Since variables will not be added to the regression equation unless they make a statistically significant addition to the analysis, all of the independent variables selected for inclusion will have a statistically significant relationship to the dependent variable.

### 3.2.3 General Linear Model (GLM)

GLM repeated measures (Keppel & Wickens, 2004) is a procedure used to model dependent variables measured at multiple times using analysis of variance (ANOVA). A repeated measure is a term used when the same participants take part in all conditions of an experiment. Experimental conditions are the levels of the factors of the ANOVA, i.e. independent variables, whereas the metrics are the dependent variables. Each dependent variable is represented by as many variables as there are measurement times for n time periods. Predictor variables may be categorical factors or continuous covariates. Normally factors define subgroups in the population and covariates are conceived as control variables. The GLM repeated measures model can test the main effects on repeated measures of between-subjects (grouping) factors, the main effects of within-subjects factors like measurement times, interaction effects between factors, covariate effects and effects of interactions between covariates and between-subjects factors.
Since the term "repeated measures" ANOVA refers to the fact that measurements across experimental conditions were taken from the same subjects (in this case, air traffic controllers), this ANOVA is also called a "within-subject" ANOVA as opposed to a "between-subject" ANOVA, in which different subjects are tested in the different experimental conditions.
Key assumptions are linear relationships, normal distribution of the dependents, homogeneity of the variances between the groups, and fixed effects. In repeated measures ANOVA a further assumption is what is known as sphericity, the condition where the variances of the differences between all combinations of related groups (levels) are equal, Interaction effects are modelled by default, as in other analyses of variance procedures.

## 3.3 Experimental scenarios and data collection

As the focus of this thesis is the controller's perception of difficulty of the instantaneous traffic situation, instantaneous subjective workload assessment values were sought.

Whereas the real-time simulations ensure the participation of the controllers (and therefore subjective assessments of the workload plus performance measures (see section 2.5)) the participation of the human observers or controllers is not foreseen within the scope of the fast-time (or model-based) simulation that essentially represent computer simulations that are based on the mathematical models. Therefore, in order to acquire data needed to pursue the research, human-in-the-loop simulations (i.e. real-time simulations) have been opted for as the resource.

This study and associated collection of data were conducted in EUROCONTROL Central European Air Traffic Services (CEATS) Research, Development and Simulation Centre (CRDS) that in period from year 2001 to 2011 operated in Budapest, Hungary.

A brief description of the CRDS simulation environment is provided in the following section (3.3.1 Simulation environment).

Further, as explained in a previous section (section 3.1), to proceed with the sufficient amount of data for further statistical analysis the precondition is to have two independent datasets. More specifically, in order to verify the validity of the results when applied on the different sectors and under different experimental conditions (foremost objective 3) data collected during two different real-time simulations was required.

That is, within the first phase of the research a measure of complexity is sought that could be applicable in different airspace sectors. Therefore, a study of different sectors from the same type (en-route sectors) is needed where the differences in the sector structure are evident (in conformance with the hypothesis 1). This is rather than to use the same airspace where re-sectorisation or changes in its design are introduced, as it could bring bias into the results.

As a result, the real-time simulation used for the collection of data to address objective 1 and 2 is the LINK2000+ Small Scale Real Time Simulation 2 experiment (LINK 2000+ SSRTS2). The aim of this simulation was to develop and validate new principles of task delegation between the planning and executive ATCO with the aim to best accommodate the Controller-Pilot Data-Link Communication (CPDLC) capability in three different en-route sectors. A more detailed explanation of this simulation and data extraction procedures applied is provided in the section 3.3.2 Real-Time Simulation 1 (LINK2000+).

However, to test the appropriateness of the identified complexity measure on other types of sector and also to test for their performance when applied only on one sector under different conditions, it was necessary to opt for a different set of data. The focus in this case was one sector, but with the profound changes introduced – not only changes in the work of the controllers, but also changes in the sector design. The experiment that enabled for such a set of data was the simulation addressing the validation of a package of changes proposed by the Irish Aviation Authority (IAA) to improve air traffic services in the Shannon low-level airspace (below FL245). Further description of this simulation and data extraction procedures applied is provided in the sub-chapter 3.3.3 Real-Time Simulation 2 (IAA).

### 3.3.1 Simulation environment

The simulations carried out in the CRDS aimed to assist civil aviation authorities and air navigation service providers in the validation of their operational and technical choices in line with the European Operational Concept Validation Model (E-OCVM) for operability, safety, capacity, environment and economics.

The Centre performed both real-time and model-based simulations. Model based simulations (MBS) were employed to investigate issues like dynamic re-sectorisation, validation of new procedures, route network development, economical impact, environmental analysis and future airspace concepts by applying modelling and analysis tools (Figure 9).

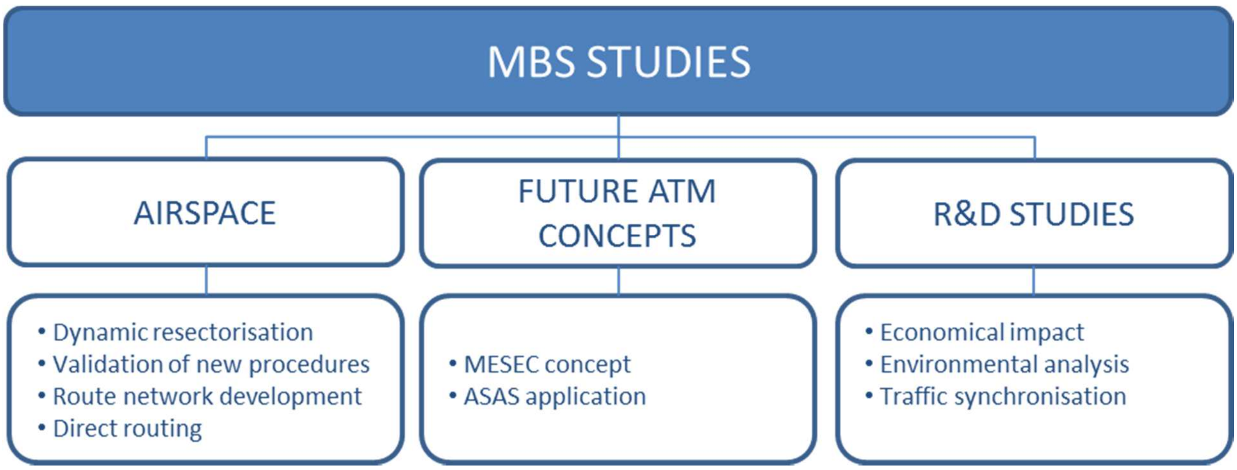



Figure 9. MBS studies’ topics

The RTS aimed to evaluate a specific airspace, procedures or an ATC system (Figure 10) and frequently represented continuance of the validation process that was commenced by MBS.

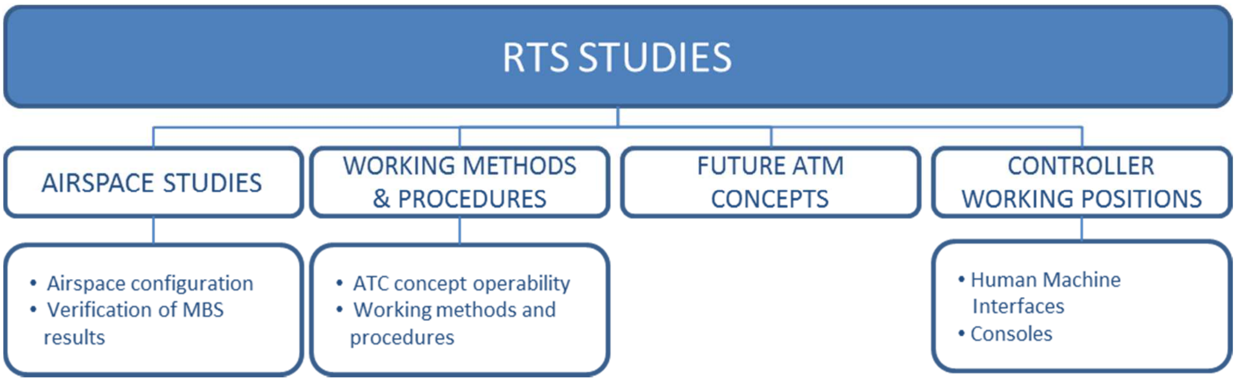



Figure 10. RTS studies' topics

The Centre was equipped with the state-of-the-art EUROCONTROL Simulation Capability and Platform that consisted of 26 controllers' working positions (CWP) (Figure 11). At the same time, in order to provide the required level of realism regarding radio/telephony (R/T) and data link communication, beside CWP, the simulation equipment consisted also of several pseudo-pilots' positions that were placed in a separated room.

Figure 11. CRDS simulation room

In compliance with E-OCVM and the EUROCONTROL Safety Regulatory Requirements (ESARRs), the CRDS simulation team conducted assessments of both human factors (HF) and safety for each validation simulation with the objectives of validating new concepts, new tools, new working methods or a new Human Machine Interface (HMI) and to identify potential hazards and risks and, further to this, to design and assess possible risk mitigation strategies.

### 3.3.2 Real-Time Simulation 1 (LINK2000+)

In order to obtain relevant values, data was recorded during the two-week LINK2000+ Small Scale Real Time Simulation 2 experiment (LINK 2000+ SSRTS2). As previously noted, the aim of this simulation was to develop and validate new principles of task delegation between the planning and executive controller with the aim to best accommodate the Controller-Pilot Data-Link Communication (CPDLC) capability in an en-route environment (EUROCONTROL 2007; Schuen-Medwed, Lorenz & Oze 2007). The simulation involved three different sectors of the Central European Air Traffic Services (CEATS) airspace (see Figure 12). The simulated measured area consists of parts of the Austrian and Hungarian airspace from FL 285 to FL 460 without a vertical split.

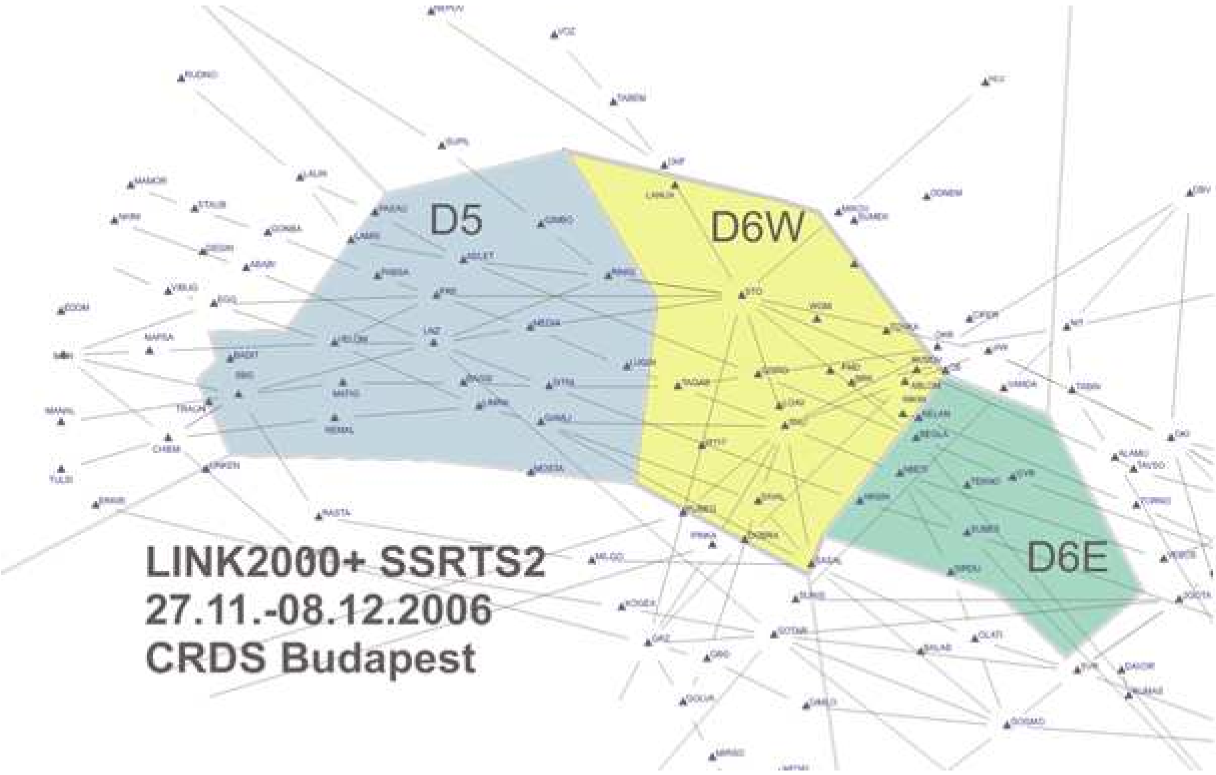



Figure 12. Map of the simulated area

The data used for the present study is data obtained for the two busiest sectors simulated (D5 and D6W). Namely, within those two sectors the average number of controlled aircraft on certain routes during a measured exercise goes beyond 30 in sector D6W and beyond 55 in sector D5, while in the third sector (D6E) that number does not go above 25 as shown in the Figure 13.

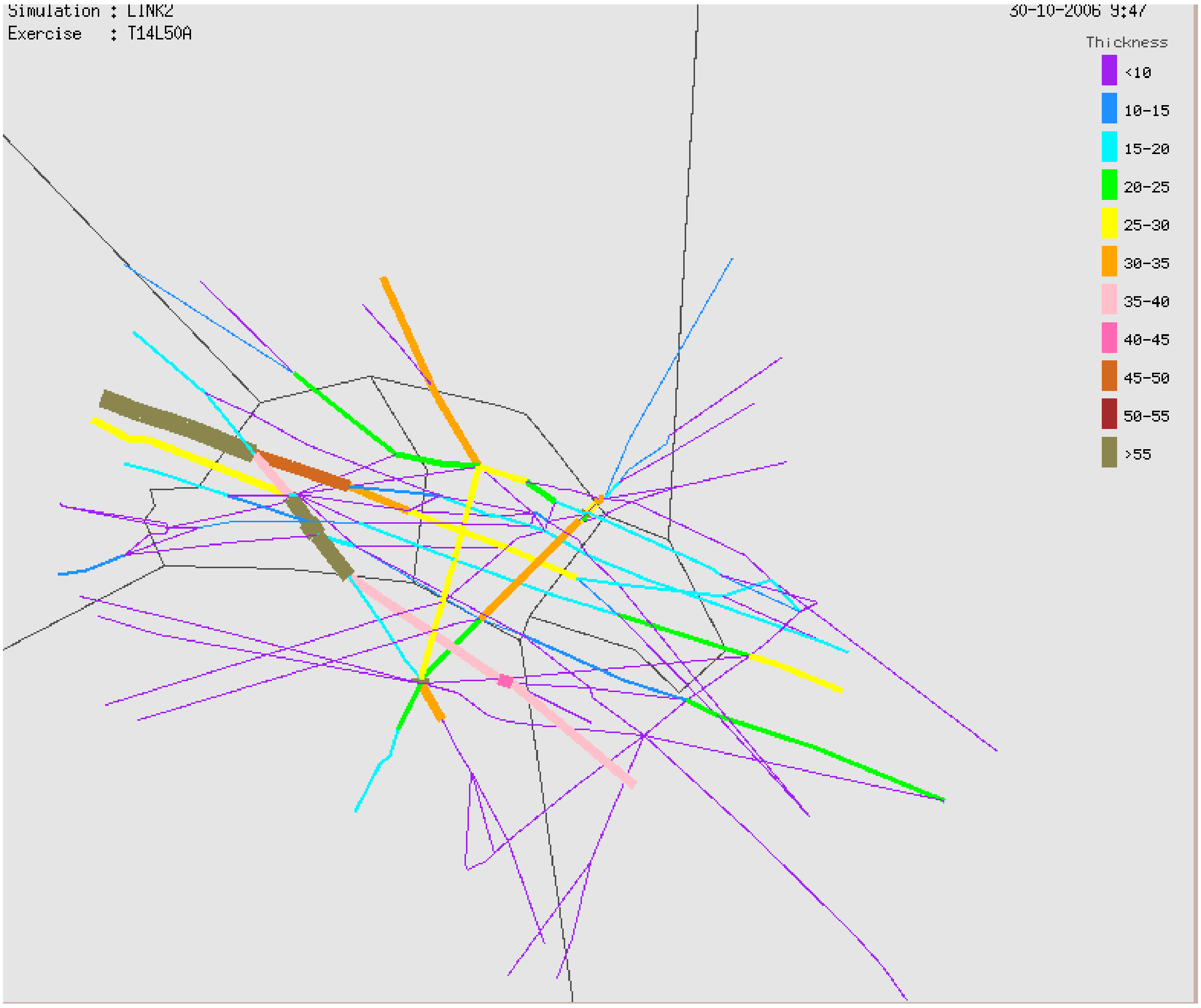



Figure 13. The traffic load of the traffic on the routes of the sectors under consideration

During the simulation different variables were manipulated: data link equipage (20 and 50%), traffic load (baseline - from year 2009 and increased traffic – traffic anticipated for year 2014.) and two different working methods that differ in the task related to the use of CPDLC (rigid and flexible). All this resulted in 8 different simulated conditions, depicted in Table 4.

Table 4. Simulated conditions based on the manipulation of different variables

| Traffic sample | 2009 | | | | 2014 | | | |
|---|---|---|---|---|---|---|---|---|
| DL equipage | 20% | | 50% | | 20% | | 50% | |
| Working method | Rigid | Flexible | Rigid | Flexible | Rigid | Flexible | Rigid | Flexible |
| Sim. condition | SC1 | SC2 | SC3 | SC4 | SC5 | SC6 | SC7 | SC8 |

Participants and data extraction procedure

The data used for the statistical analysis was derived from 5 participants working on two of the considered sectors. Each controller completed 8 exercises overall (refer to each of 8 simulated conditions) of 1 hour and 20 minutes, from which 1-hour recordings were extracted for analysis. Scores were derived for every 2 minutes, resulting in 30 measurements per exercise. This data was obtained for each indicator (ATC complexity measures, controllers' activity and workload measures). The overall dataset comprised 5 (controllers) x 8 (exercises) x 30 (time segments) = 1.200 measurements for each indicator. Prior inspection of the data set revealed that in 58 time segments (4.8%) of the data was missing. Therefore, subsequently reported results related to the first three objectives of this thesis are based on the measurements obtained in 1.142 time segments.

### 3.3.3 Real-Time Simulation 2 (IAA)

The data for the analysis was obtained during the real time simulation dedicated to the validation of a package of changes proposed by the Irish Aviation Authority (IAA) to improve air traffic services in the Shannon low-level airspace (below FL245).

The simulation consisted of a series of exercises that aimed at investigating the operational feasibility, efficiency and benefits of proposed modifications.

These modifications include the optimized sectorisation (sector split) of Shannon low-level sector (SHLOW) into Shannon low-level North (LONO) and Shannon low-level South (LOSO), introduction of Precision Area Navigation (P-RNAV) procedures in the new Shannon and Cork TMAs, establishment of unidirectional routes and other routes and introduction of Single Person Operation (SPO).

Shannon low-level airspace (SHLOW) is currently managed by a single sector composed of an executive controller (EC) and a Planning Controller (PC). It is proposed that sectors resulted from the split of this sector (LONO and LOSO) are manned by a single controller each at ranges between 60 and 90 nautical miles rather than a range of approximately 150 nautical miles as in the current sectorisation (see Figure 14).

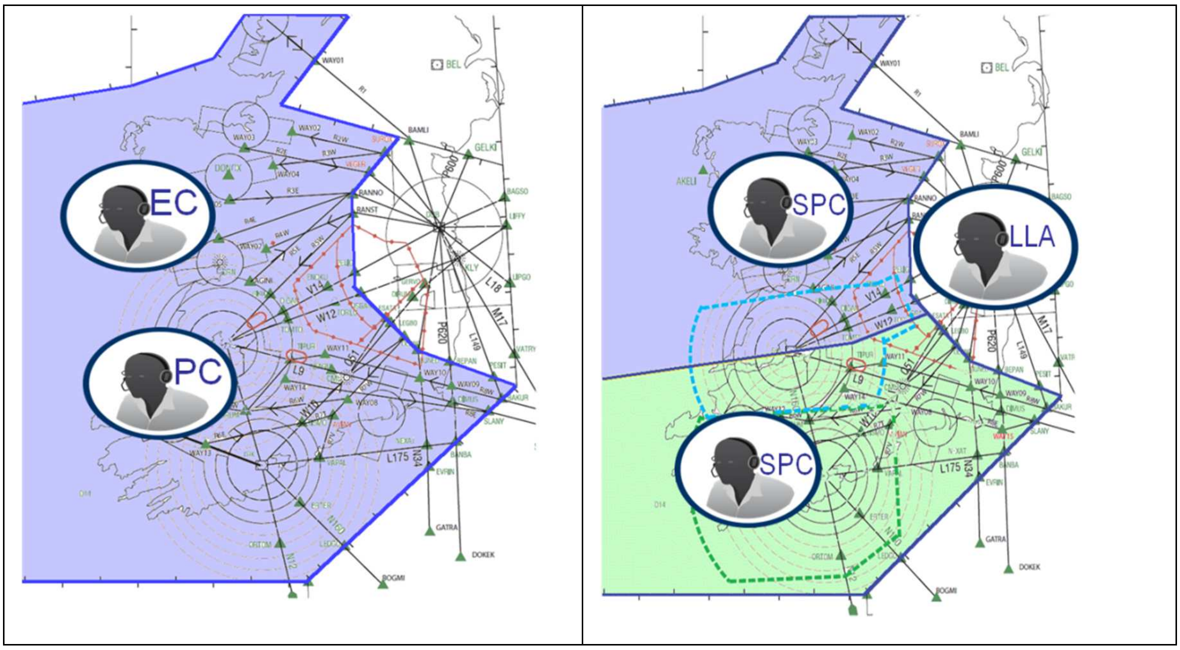



Figure 14. Split of Shannon low-level sector (SHLOW) into Shannon low-level North (LONO) and Shannon low-level South (LOSO)

Traffic load

In order to validate and exploit benefits gained through the introduced package of changes, traffic volume varied in both the current and new sectorisation.

The traffic sample for the IAA RTS was based on the real (24Hour) traffic sample over two days: the 28th and 30th of June 2007. The samples were chosen for the simulation to represent different traffic patterns and thus verify the impact of different traffic patterns in the new airspace organisation being analysed.

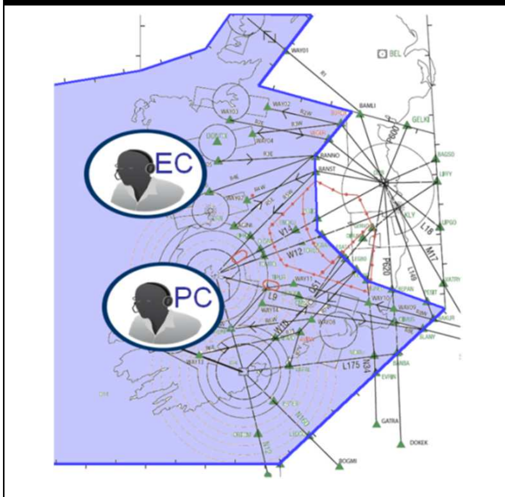
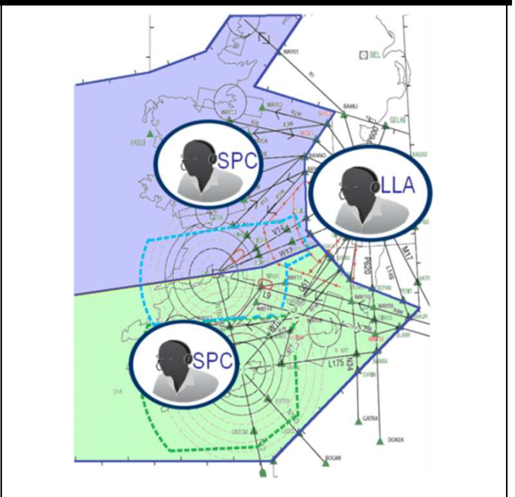
The two hour period with the highest traffic load was selected for baseline. The traffic increase of 20% and 40% was calculated by the Flight Increase Processor Software (FIPS) methodology developed at the EUROCONTROL Experimental Centre (EEC), Brétigny-sur-Orge. The traffic increase of 20% was applied to the current environment, while an increase of 40% was considered for the new environment taking into consideration the envisaged benefits gained by new sectorisation and concept of operations:

- the higher accuracy of P-RNAV procedures to enable shorter and more direct routes with simple connections to the en-route structure
- the establishment of unidirectional routes intended to reduce the number of potential conflict points and
- the introduction of SPO can be regarded as a means to achieve the same or even higher level of controller productivity as in the current management of the Shannon low-level airspace.

<u>Experimental conditions</u>

The combination of these experimental variables reflecting changes in airspace structure and traffic volume resulted in four different conditions:

Table 5. Simulated conditions based on the manipulation of different variables

<table>
<tr><th colspan="2">SHLOW</th><th colspan="2">LONO & LOSO</th></tr>
<tr><td colspan="2">
</td><td colspan="2">
</td></tr>
<tr><td>peak 2007 traffic</td><td>20% increased traffic</td><td>peak 2007 traffic</td><td>40% increased traffic</td></tr>
<tr><td>Sim. condition A</td><td>Sim. Condition A120</td><td>Sim. Condition B</td><td>Sim. Condition B140</td></tr>
</table>

| | |
|---|---|
| **Sim. condition A** | Current concept of operation: SHLOW, peak 2007 traffic |
| **Sim. condition A120** | Current concept of operation: SHLOW, 20% increased traffic, |
| **Sim. condition B** | New concept of operation: LONO and LOSO, peak 2007 traffic |
| **Sim. condition B140** | New concept of operation: LONO and LOSO, 40% increased traffic |

Participants and data extraction procedure

Out of 11 participants that were involved in the IAARTS1 four licensed controllers of the Shannon ACC were working on the Shannon sectors under consideration.

Each of the four controllers of Shannon's low-level sectors experienced all four experimental conditions:

- conditions A and A120 at least once on the EC position (SHLOW-EC)
- conditions B and B140 at least once on both SPO positions (SPC-North, SPC-South).

It means that each controller had to complete overall 6 exercises to fulfil the requirement of working on each of the measured positions.

The total duration of each exercise was 1 hour and 20 minutes, from which 48 minutes of recordings were extracted for the analysis. Scores were derived for every 2 minutes, resulting in 24 measurements per exercise. The overall dataset comprised 4 (controllers) x 6 (exercises) x 24 (time segments) = 576 measurements for each indicator.

### 3.3.4 ATC Complexity factors

For the purpose of the present study, a list of complexity factors was selected to represent different aspects of complexity:

1. aircraft density
2. flight attributes of each individual aircraft
3. aircraft conflicts
4. traffic disorder.

The factors selected to correspond to these aspects of complexity are those that have been consistently found to be important in the previous studies and for which detailed calculation formula have been reported. The factors were partially elicited from work done by Delahaye and Puechmorel (2000), Chatterji and Sridhar (2001), Laudeman *et al.* (1998), Kopardekar and Magyarits (2003), Chatton (2001), Gianazza and Guittet (2006), Kopardekar (2000), Kopardekar *et al.* (2009), Martin *et al.* (2006) and Sridhar, Sheth and Grabbe (1998).

The selected overall set of 24 complexity factors is presented in Table 6. It is out of the scope of this thesis to describe here all 24 factors in detail. For a more thorough review of the listed factors readers are referred to the indicated source literature. Nonetheless, the formula for calculation of the selected factors can be found in Annex A.

As an input for their calculations were used the recorded flown trajectories. Moreover, the values used are those related to the instantaneous position of the aircraft (latitude, longitude, altitude), as well as data reflecting movements of the aircraft within the sector (climb or descent rate, heading and speed). Customized application CALAN (“Calculation Analysis”) was developed using MATLAB software to calculate the complexity factors using these recorded values for chosen time steps.

Table 6. The list of complexity factors selected from the literature for the further analysis

| | | Complexity Factors | References |
|---|---|---|---|
| 1 | Aircraft density | number of aircraft | Chatterji & Sridhar (2001); Kopardekar & Magyarits (2003); Sridhar, Sheth & Grabbe (1998); Kopardekar (2000) |
| 2 | | density indicator | Delahaye & Puechmorel (2000) ; Chatton (2001) ; Gianazza & Guittet (2006) |
| 3 | Flight attributes | number of climbing aircraft | Chatterji & Sridhar(2001); Kopardekar & Magyarits (2003); Kopardekar et al. (2009) |
| 4 | | number of descending aircraft | Chatterji & Sridhar (2001); Kopardekar & Magyarits (2003); Kopardekar et al. (2009) |
| 5 | | number of aircraft with heading change >15˚ | Laudeman et al. (1998); Kopardekar & Magyarits (2003); Kopardekar et al. (2009) |
| 6 | | number of aircraft with the speed change >10 knots | Laudeman et al. (1998); Kopardekar & Magyarits (2003); Kopardekar et al. (2009) |
| 7 | Aircraft conflicts | number of a/c with lat. distance between 0-25nm and vert. separation < 2000ft above 29000ft | Laudeman et al. (1998); Kopardekar & Magyarits (2003); Kopardekar et al. (2009) |
| 8-10 | | horizontal proximity measure 1,2,3 | Chatterji & Sridhar (2001); Kopardekar & Magyarits (2003); Kopardekar (2000); Martin et al. (2006) ; Gianazza & Guittet (2006) |
| 11-13 | | vertical proximity measure 1,2,3 | Chatterji & Sridhar (2001); Kopardekar & Magyarits (2003); Kopardekar (2000); Martin et al. (2006) ; Gianazza & Guittet (2006) |
| 14 | | time-to-go to conflict measure | Chatterji & Sridhar (2001); Kopardekar & Magyarits (2003); Kopardekar (2000); Martin et al. (2006) |
| 15 | | divergence between pairs of aircraft | Delahaye & Puechmorel (2000); Chatton (2001); Gianazza & Guittet (2006) |
| 16 | | convergence between pairs of aircraft | Delahaye & Puechmorel (2000); Chatton (2001); Gianazza & Guittet (2006) |
| 17,18 | | sensitivity indicator (a/c converging; a/c diverging) | Delahaye & Puechmorel (2000); Chatton (2001); Gianazza & Guittet (2006) |
| 19,20 | | insensitivity indicator (a/c converging; a/c diverging) | Delahaye & Puechmorel (2000); Chatton (2001); Gianazza & Guittet (2006) |
| 21 | Traffic disorder | variance of ground speed | Chatterji & Sridhar (2001); Kopardekar & Magyarits (2003); Gianazza & Guittet (2006); Martin et al. (2006) |
| 22 | | ratio of standard deviation of speed to average speed | Chatterji & Sridhar (2001); Kopardekar & Magyarits (2003); Gianazza & Guittet (2006); Martin et al. (2006) |
| 23 | | variability in headings (track_disorder) | Delahaye & Puechmorel (2000); Chatton (2001); Gianazza & Guittet (2006) |
| 24 | | variability in speed (speed_disorder) | Delahaye & Puechmorel (2000); Chatton (2001); Gianazza & Guittet (2006) |

### 3.3.5 Controllers' activity measures

The measures of controller activity as a link between task demands imposed on him/her and the workload (see section 2.3) mainly fall into two categories:

1. measures related to the communication performed by the controller (whether with other controllers or with the pilots) and
2. measures that are reflecting data entries related to flight data management.

Thus, in order to adequately address these measures, during the real-time simulations it was sought to obtain the data that reflect those measures.

In CRDS, the recordings obtained during the simulation, except for the flown trajectories, also include every system input on the human machine interface (HMI) made by the controller, that is in line with the second category of measure listed above (data entries related to flight data management).

These inputs refer to assignments of vertical rate, exit flight levels/planned entry levels, cleared flight levels, headings, speed instructions, and direct clearances. These were summed across each 2-minute time step and across each input, resulting in only one measure named Actions_SUM.

Furthermore, when it comes to the communication performed by the controller, the recordings of radio/telephony (R/T) communication of the controller were made (duration and number of voice communications, but without the content).

The cumulative duration of radio calls (= frequency occupancy time per 2-minute time step; in further text also referred to as R/T occupancy time) was calculated as well as the average duration of single calls.

Altogether, three measures of the controller's activity obtained for every 2-minute time steps were used:

- Actions_SUM,
- Frequency (or R/T) Occupancy Time and
- Average Radio Duration.

### 3.3.6 Workload measures

To collect workload measures during the simulation the Instantaneous Self-Assessment (ISA) technique as operator-subjective metric was applied, where the air traffic controller gives subjective ratings of workload. This tool was developed by the UK NATS and offers a 5 point rating scale (see section 2.2). During the simulation exercise, the participants were prompted at regular intervals (every 2 minutes) by a flashing light to give a rating from 1 (Very Low) to 5 (Very High) in order to describe perceived workload at that moment (see Table 7). The controllers used the ISA box that was placed at the CWP that was equipped with five buttons labelled by the above-mentioned rating categories.

Table 7. ISA ratings definition

| ISA DEFINITION | | |
|---|---|---|
| Rating | Self Assessment of Workload | |
| **5** | Very High | A very high workload, any additional task would push you into overload. |
| **4** | High | High workload level that leaves very little spare capacity. |
| **3** | Comfortable | Enough work for the job to be interesting and challenging. Some spare capacity. |
| **2** | Relaxed | Not quite enough traffic to fully occupy you. |
| **1** | Under-utilised | You have very little to do and few, if any, aircraft on the frequency. |

Even though the controllers provide their subjective assessment of workload only on every 2 minute intervals, these assessments nevertheless reflect their workload perceived for the overall period of 2 minutes and not only for the specific moment when the assessment is recorded.

Instead, the complexity factors related data and controllers' activity measurements recorded at specific time steps account only for those instants when the recordings are made.

Therefore, this issue should be taken into consideration when performing the analysis. That is, the traffic during the period of 2 minutes might undergo significant changes and as a result, the recordings on these specific time steps could not adequately correspond to the workload assessed by the controllers.

In order to test this assumption, the predictive power of the complexity components and controller's activity measures recorded on more frequent time steps (5 seconds) and then averaged over 2 minute intervals should be compared with the predictive power of the same measures recorded only on 2 minute intervals.

If the predictive power remains alike, the conclusion may be drawn that the traffic does not undergo such a significant developments during the 2 minute interval to require the recordings of complexity and controller's activity measures on more frequent time steps.

# 4 Analyses and Results

## 4.1 Real-Time Simulation 1 (LINK2000+)

### 4.1.1 Principal Component Analysis

In a first analysis step, a PCA on all 24 complexity factors was computed in order to search for a potentially reduced number of uncorrelated predictor variables for the subsequent computation of regression models (see section 3.2.1 for a more in depth explanation of the ratio of using this statistical technique).

Principal components having an eigenvalue > 1 were extracted and subsequently rotated using the VARIMAX method. This analysis resulted in the extraction of 8 orthogonal (i.e. uncorrelated) principal components that accounted for 67.26 % of the total variance in the factors.

Table 8 displays these components sorted by the sizes of their eigenvalues and along with the percentage of variance they account for.

Table 8. Results of the Principal Component Analysis

| Components | Eigenvalue | % of Variance | Cum. % of Variance |
|---|---|---|---|
| Component 1 | 4.94 | 20.56 | 20.56 |
| Component 2 | 3.44 | 14.34 | 34.91 |
| Component 3 | 1.78 | 7.40 | 42.30 |
| Component 4 | 1.45 | 6.04 | 48.34 |
| Component 5 | 1.34 | 5.60 | 53.94 |
| Component 6 | 1.13 | 4.71 | 58.65 |
| Component 7 | 1.04 | 4.32 | 62.97 |
| Component 8 | 1.03 | 4.29 | 67.26 |

Table 9 shows the rotated component matrix which contains the loadings of the 24 complexity factors on the eight extracted principal components. Weak loading values below 0.40 are suppressed to better visualize the pattern.

Table 9. Rotated Component Matrix

| Complexity factors | Component 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| variance of ground speed | 0.884 | | | | | | | |
| ratio of standard deviation of speed to average speed | 0.845 | | | | | | | |
| divergence between pairs of aircraft | 0.787 | | | | | | | |
| convergence between pairs of aircraft | 0.785 | | | | | | | |
| variability in speed (speed disorder) | 0.703 | | 0.452 | | | | | |
| number of aircraft | | 0.816 | | | | | | |
| horizontal proximity measure #2 | | 0.672 | 0.431 | | | | | |
| variability in headings (track disorder) | | -0.657 | | | | | | |
| number of climbing aircraft | | 0.643 | | | | | | |
| number of aircraft with heading change greater than 15˚ | | 0.442 | | | | | | |
| horizontal proximity measure #1 | | | 0.894 | | | | | |
| density indicator | | | 0.815 | | | | | |
| number of descending aircraft | | | | 0.785 | | | | |
| number of a/c with the speed change greater than 10kt | | | | 0.732 | | | | |
| vertical proximity measure #1 | | | | -0.569 | | | | |
| number of aircraft with lateral distance between 0-25nm and vertical separation less than 2000ft above 29000ft | | | | 0.559 | | | | |
| sensitivity indicator (a/c diverging) | | | | | 0.772 | | | |
| sensitivity indicator (a/c converging) | | | | | 0.751 | | | |
| time-to-go to conflict measure | | | | | 0.623 | | | |
| insensitivity indicator (a/c converging) | | | | | | 0.723 | | |
| insensitivity indicator (a/c diverging) | | | | | | 0.686 | | |
| vertical proximity measure #3 | | | | | | | 0.849 | |
| vertical proximity measure #2 | | | | | | | 0.540 | |
| horizontal proximity measure #3 | | | | | | | | 0.908 |

By the inspection of Table 9 the following component meanings could be derived. Note, that the loading of a given metric on a given component is equivalent to the correlation between that factor and that component. Therefore, by-and-large the factor with the highest loading guides the interpretation of the component.

Comp.1 – ***ground speed variance and divergence/convergence***: strongly related to the variance of the ground speed (0.884) and the ratio of the standard deviation to the mean ground speed (0.845). Also, the strong correlation with divergence and convergence factors (0.787 and 0.785 respectively) was recognised, which is in compliance with speed significance, as divergence / convergence factors actually measure how fast aircraft are moving toward/from each other.

Comp. 2 – ***aircraft count***: this component has the strongest correlation with the number of the aircraft in the sector (0.816)

Comp. 3 – ***horizontal proximity***: this component can be considered as addition to the previous one, as it shows high correlation with the horizontal distance between aircraft taking into consideration the aircraft count - horizontal proximity measure: 0.894 and density_mean: 0.815 .

Together these two components (Comp. 2 and Comp. 3) can be considered as representatives of so-called sector density.

Comp. 4 – ***aircraft vertical transitioning***: highly correlated to the number of descending aircraft (0.785) as well as speed change related to this vertical evolution (0.732)

Comp.5 – ***conflict sensitivity***: this component is loaded highly by both sensitivity indicators (Sd+(i):0.772 and Sd-(i): 0.751). Sensitivity is related to the gradient of the relative distance between aircraft. This indicator measures the change in terms of relative distance in response to changes in speed and heading of the involved aircraft. If sensitivity is high only small changes in heading and speed imply a high impact on relative distance. This is the case, e.g. when two aircraft are heading towards each other. The sensitivity indicators are designed to set a weight on potential conflicts that are difficult to solve. Note that a situation with high sensitivity is easier to resolve for the controller than one with a low sensitivity (Delahaye & Puechmorel 2000).

Comp.6 – ***insensitivity***: This component is strongly related to the insensitivity indicators both for convergence and divergence of the aircraft (insen_c: 0.723 and insen_d: 0.686). It is not simply an analogue with the opposite direction to the previous component. High insensitivity is given for a pair of aircraft when the degree of convergence is high while the sensitivity for convergence is low.

Comp.7 – ***vertical separation***: high correlation with the measure of the vertical separation of aircraft in close horizontal proximity defines this component (0.849)

Comp.8 – ***horizontal separation***: analogously to the previous component, this component is defined based on the correlation with the measure of horizontal separation of the aircraft in close vertical proximity (0.908).

The PCA yielded 8 component scores for each two-minute interval, however, as shown in Table 9, two variables (variability in speed (speed disorder) and horizontal proximity measure #2) have loading higher than 0.40 on more than one component. Thus it was decided to re-run the PCA excluding these metrics. In addition, from the second PCA the factors that comprised components 7 (vertical proximity measure #3 and #2) and 8 (horizontal proximity measure #3) were excluded. The reason for excluding these variables was twofold. On the one hand, it was considered that having more than one variable coding horizontal and vertical proximity does not bring added value. On the other hand, even though the eigenvalues for components 7 and 8 satisfied Keiser criterion (i.e. were higher than 1), they were only minimally higher than 1.

The second PCA was thus performed on the initial set complexity factor, excluding the following factors:

- variability in speed (speed disorder)
- horizontal proximity measure #2
- vertical proximity measure #2
- horizontal proximity measure #3
- vertical proximity measure #3

As in the first analysis, the criterion for choosing the number of components was that of having eigenvalue > 1. The components that were extracted were further rotated using the VARIMAX method. This second PCA analysis resulted in the extraction of only 6 principal components that accounted for 65.9 % of the total variance in the metrics. Table 10 displays these components sorted by the sizes of their eigenvalues and along with the percentage of variance they account for.

Table 10. Results of the Principal Component Analysis (PCA 2)

| Components | Eigenvalue | % of Variance | Cum. % of Variance |
|---|---|---|---|
| Component 1 | 3.99 | 20.99 | 20.99 |
| Component 2 | 3.38 | 17.80 | 38.79 |
| Component 3 | 1.43 | 7.54 | 46.34 |
| Component 4 | 1.34 | 7.05 | 53.38 |
| Component 5 | 1.28 | 6.75 | 60.13 |
| Component 6 | 1.10 | 5.77 | 65.90 |

In Table 11 the rotated component matrix is shown, which contains the loadings of the 19 considered complexity factors on the six extracted principal components. Again, loading below 0.40 are suppressed to better visualize the pattern. As shown in the table, this time none of the metrics considered have loading higher than 0.4 on more than one component. Moreover, the 6 components that were extracted match perfectly the first 6 components that were extracted in the initial PCA, and for each of them the loadings of the relative factors are similar to what previously found.

Table 11. Rotated Component Matrix (PCA 2)

| Complexity factors | Component | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| variance of ground speed | 0.908 | | | | | |
| ratio of std of speed to average speed | 0.875 | | | | | |
| convergence between pairs of aircraft | 0.775 | | | | | |
| divergence between pairs of aircraft | 0.743 | | | | | |
| number of aircraft | | 0.794 | | | | |
| number of climbing aircraft | | 0.707 | | | | |
| variability in headings (track disorder) | | -0.705 | | | | |
| number of aircraft with heading change greater than 15˚ | | 0.482 | | | | |
| number of descending aircraft | | | 0.798 | | | |
| number of a/c with the speed change greater than 10kt | | | 0.740 | | | |
| vertical proximity measure 1 | | | -0.562 | | | |
| number of aircraft with lateral distance between 0-25nm and vertical separation less than 2000ft above 29000ft | | | 0.547 | | | |
| horizontal proximity measure 1 | | | | 0.898 | | |
| density indicator | | | | 0.836 | | |
| sensitivity indicator (a/c diverging) | | | | | 0.778 | |
| sensitivity indicator (a/c converging) | | | | | 0.746 | |
| time-to-go to conflict measure | | | | | 0.619 | |
| insensitivity indicator (a/c converging) | | | | | | 0.717 |
| insensitivity indicator (a/c diverging) | | | | | | 0.662 |

### 4.1.2 Multiple Regression Analyses

In order to assess the effectiveness of predicting ISA workload ratings on the basis of the extracted principal components, multiple regression models were computed (see section 3.2.2).

Instead of applying the stepwise linear regression using all the complexity components and the controllers' activity measures (Number of instructions given, Frequency (or R/T) Occupancy Time, and Average Radio Duration, see section 0) as predictors of workload, a hierarchical strategy was used. In the first step, only the 6 complexity components' scores were used to predict ISA workload ratings. In the second step the 3 activity measures were entered into the regression equation, and the model was re-fitted to the data. This was done in order to assess the contribution of ATC complexity components in relation to the controller activity metrics. The activity measures were inserted in the model after the complexity components because it is reasonable to assume that complexity can have a casual influence the controller activity, while the clearly the opposite could not hold. Table 12 contains the global statistics of the two models. As it can be seen the model containing only complexity components yielded a multiple R of 0.36 corresponding to $R^2$ of 0.13. Using only the complexity components we can thus account for 13% of the variance of the ISA workload ratings.

The increase in $R^2$ gained by adding the controller activity measures to the regression model (step 2) was statistically significant, although quite small (3%) in terms of predictive power (as measured by the percentage of variance explained). The second model thus was able to account for 16% of the total variance in the ISA ratings. Therefore it can be concluded that once information about complexity has been used to predict controllers' workload, prediction is only slightly (although significantly) improved by taking into account also information about their activity. On the other hand, given that the increase in $R^2$ was significant, it implies that activity measures give a unique contribution to the prediction. In other words, they are able to explain a fraction of the variance in ISA ratings that is not accounted by the complexity factors alone.

Table 12. Comparison of alternative multiple regression models for prediction of ISA

| Regression equation containing | mult. R | $R^2$ | $R^2$ change | F change | df | Sig. F change |
|---|---|---|---|---|---|---|
| **complexity components** | 0.36 | 0.13 | 0.13 | 27.89 | 6 , 1133 | p< 0.001 |
| **complexity components and controller's activity measures** | 0.40 | 0.16 | 0.03 | 23.91 | 3 , 1130 | p< 0.001 |

A final multiple regression model was computed using a traditional stepwise linear regression approach in order to identify those predictors that are significant for the workload prediction. This model is referred to as the optimised model as all insignificant variables have been removed. The parameter statistics of this model are given in Table 13. The model consists of 8 parameters (Comp.1 – Comp.6, Frequency Occupancy Time and Average Radio Duration). The corresponding parameters of the excluded predictors are not reported in the Table 13.

Table 13. Parameter statistics of the optimised model for the prediction of ISA workload ratings.

| | B | Std. Error | Beta | t | Sig. |
|---|---|---|---|---|---|
| **ground speed variance and divergence/convergence** | 0.066 | .020 | .090 | 3.298 | p< 0.05 |
| **aircraft count** | 0.107 | .021 | .146 | 5.161 | p< 0.001 |
| **aircraft vertical transitioning** | 0.089 | .020 | .121 | 4.423 | p< 0.001 |
| **horizontal proximity** | -0.092 | .020 | -.125 | -4.584 | p< 0.001 |
| **conflict sensitivity** | -0.144 | .021 | -.196 | -7.003 | p< 0.001 |
| **insensitivity** | 0.072 | .020 | .098 | 3.607 | p< 0.001 |
| **Frequency Occupancy Time** | 0.012 | .003 | .143 | 4.701 | p< 0.001 |
| **Average Radio Duration** | -0.184 | .033 | -.165 | -5.531 | p< 0.001 |

The stepwise regression analysis revealed that all the 6 complexity components were significant predictors of ISA ratings. The components that showed the strongest effect on ISA ratings are Comp. 2 and Comp. 5 which consider *aircraft count* and *conflict sensitivity*. The latter component had a negative effect of workload. This means that the that when sensitivity of the conflict increased, the workload ratings of the controller decreased, which is consistent with previous studies (Delahaye & Puechmorel 2000). Conversely, the number of aircraft had a positive effect on workload.

Two of the activity measures considered also remained in the final model as significant predictors of workload. These were *Frequency Occupancy Time* and *Average Radio Communication Duration* representing the communication load also remained in the model, while the number of actions performed was excluded. Frequency Occupancy Time had a positive effect on controller's perceived workload (i.e. when overall frequency occupancy time is increased, the workload rating was also higher), while average communication duration had a negative effect.

To sum up, the results of multiple regression analyses suggest that subjective workload hinges on other aspects not only of ATC complexity but also on the communication load of the controller. Both the total frequency occupancy time and the average radio duration significantly correlate with ISA workload ratings. These task demand factors are more closely linked to how the controllers interact with the traffic demand. The finding that the average time for an individual communication is negatively related to workload (see also Manning *et al.* 2001) reflects already some kind of an active adaptation on behalf of the controller to cope with increased task load. Thus the controller reduces the amount of time that she/he spends on a single communication as the situation gets busier.

Additionally, it was found that subjective controller workload as measured by the ISA ratings depends on additional factors rather than only on aircraft count. This is in agreement with a number of other studies (e.g. Delahaye & Puechmorel 2000; Laudeman *et al.* 1998; Gianazza & Guittet 2006).

With regard to the aim of using complexity measures in the design of traffic samples in ATC real-time simulations as well as parameters of comparison between different scenarios and sectors, the present findings (complexity components) are further employed as the measures of the complexity. Those components (as previously identified and interpreted) are:

1. ground speed variance and divergence/convergence
2. aircraft count
3. aircraft vertical transitioning
4. horizontal proximity
5. conflict sensitivity
6. insensitivity.

## 4.2 Real-Time Simulation 2 (IAA RTS1)

### 4.2.1 Complexity components

In order to calculate complexity components’ scores based on the previous findings, standardized regression coefficients resulting from the principal component analysis were used (Table 14). For each time step from which data was collected, the components were calculated by taking the linear function of the 19 complexity factors’ (see section 4.1.1) standardized scores multiplied by the coefficients.

Table 14. Component Score Coefficient Matrix.

| Complexity factors | Component | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| | Coefficients | | | | | |
| number of aircraft | .012 | .301 | -.011 | -.032 | -.039 | .033 |
| number of descending aircraft | -.128 | -.101 | .468 | .016 | .027 | -.116 |
| number of climbing aircraft | .090 | .335 | -.098 | .072 | -.074 | .035 |
| number of aircraft with heading change greater than 15˚ | .143 | .286 | -.053 | .129 | -.096 | -.320 |
| number of a/c with the speed change greater than 10kt | -.014 | -.044 | .414 | .056 | .076 | -.074 |
| number of aircraft with lat. distance 0-25nm and vert. sep. less than 2000ft above 29000ft | -.056 | .029 | .295 | .141 | .041 | .182 |
| horizontal proximity measure 1 | -.006 | .182 | .063 | .569 | -.018 | .004 |
| vertical proximity measure 1 | -.004 | .001 | -.278 | .042 | -.021 | .057 |
| time-to-go to conflict measure | -.008 | -.139 | .054 | -.139 | .398 | .141 |
| variance of ground speed | .361 | .140 | -.023 | -.056 | .067 | .021 |
| ratio of std. of speed to average speed | .333 | .121 | .009 | -.078 | .019 | .021 |
| density indicator | -.019 | -.028 | .013 | .450 | -.082 | .025 |
| variability in headings (track disorder) | -.003 | -.320 | -.095 | -.131 | -.214 | -.092 |
| divergence between pairs of aircraft | .253 | -.017 | -.049 | .101 | .027 | -.034 |
| convergence between pairs of aircraft | .264 | -.034 | -.049 | .043 | .018 | -.057 |
| sensitivity indicator (a/c diverging) | .070 | .179 | -.011 | -.029 | .476 | -.032 |
| sensitivity indicator (a/c converging) | .036 | -.008 | .042 | .053 | .425 | -.064 |
| insensitivity indicator (a/c diverging) | -.034 | .020 | -.020 | .029 | .030 | .588 |
| insensitivity indicator (a/c converging) | -.012 | -.024 | -.072 | -.016 | .026 | .549 |

### 4.2.2 Multiple regression analysis

Validation of the statistical association between identified complexity components and workload ISA ratings was conducted by performing multiple regression analysis with complexity components as predictors of the workload. Just as previously done with the data gathered in the first simulation, a hierarchical regression strategy was used. First it was aimed at predicting the workload ratings from the complexity components. In a second step, the 3 activity metrics were added to the best fitting model containing only complexity components. Again, this was done in order to assess the contribution of identified ATC complexity components in relation to the controller activity metrics.

As performed in the previously reported analyses (section 4.1.2), first a multiple regression model was fitted using only the 6 complexity components as predictors. The fitted model resulted in a multiple R of 0.67 corresponding to $R^2$ of 0.45. Using these 6 components, thus, the model was able to account for (i.e. predict) 45% of variance of the ISA workload ratings. The regression coefficient for the second complexity component (i.e. the one relative to the aircraft count), however, was not statistically significant. The most important complexity components for predicting workload ratings were component 5 (i.e. conflict sensitivity) and 4 (i.e. horizontal proximity), which together were able to account for 43% of the variance. The other 3 components were thus only able to explain about an additional 2% of the variance, and the least important of them (i.e. the one with the smaller effect on workload) was component 6 (i.e. insensitivity).

Given that component 2 was not significant, it was removed from the model and the analysis was again performed. Not surprisingly, the revised model was still able to account for 45% of the variance in the ISA ratings, and all the predictors had a significant effect. Then the controller activity measures were added to the model, to check whether they could contribute to increase the predictive power. The change in $R^2$ gained by adding these new predictors was statistically significant, and the new model was able to account for 54% of the total variance in the ISA ratings (see Table 15 for a statistics summary). However, once the three controller activity measures were inserted into the equation, complexity component 3 (i.e. ground speed variance) was no longer significant.

Overall, the results confirm that both sources of information, ATC complexity and controller activity, are useful to predict workload, each source being able to explain a fraction of the variance in the ISA ratings. The fact that the addition to the model of the activity measures resulted in a statistically significant increase in $R^2$ shows that these measures give a unique contribution to the explanatory power of the model, beyond the one already given by the complexity components.

Table 15. Comparison of alternative multiple regression models for prediction of ISA (2min)

| Regression equation containing | mult. R | $R^2$ | $R^2$ change | F change | df | Sig. F change |
|---|---|---|---|---|---|---|
| complexity components | .67 | .45 | - | 93.59 | 5 , 570 | p< 0.001 |
| complexity components and controller's activity measures | .74 | .54 | .09 | 37.41 | 3 , 567 | p< 0.001 |

In further regression models also the complexity components were used to predict the different activity measures, in order to see how much of the effect of complexity on workload could be mediated by their effect on the controllers' activity. The best fitting model for Frequency occupancy time was able to account for 22% of the variance with 4 complexity components (namely, components 5, 4, 3, and 1). Similarly, 24% of the variance in the number of actions was accounted for by a 4-components model (comprising components 5, 4, 1, and 2, with the latter one interestingly having a significant negative correlation coefficient). The model that performed the poorest was the one trying to predict the average duration of radio communication, which was only able to account for 4% of the variance in the outcome, with only two significant predictors (components 5 and 3). Overall, this seems to suggest that the effect of the complexity components on the subjective workload ratings is only partially mediated by the effect they have on the activity.

The question that arose during the analysis is related to the time-steps on which data related to complexity should be collected. Controllers are assessing their instantaneous subjective workload in time steps of 2 minutes, but these assessments should reflect their workload for the whole of that period and not only for the moment when the assessment is recorded.

Therefore, the complexity data recorded only at those moments could not provide sufficiently comprehensive data for the overall 2 minute intervals. To test this assumption, data was recorded on more frequent time steps (5 seconds) and then averaged over 2 minute intervals. Furthermore, they were entered first into a multiple regression model as the predictors of ISA workload, and in the second equation of the model, controllers' activity measures were added at the same increased recorded frequency (5 seconds).

The results are shown in Table 16. As can be seen, the predictive power of complexity components calculated for shorter intervals and then averaged over 2 minutes time-steps did not increase when compared to complexity components calculated only for the moment when ISA workload ratings are recorded (see Table 15). In accordance with the discussion provided in section 3.3.6, in further analysis are considered data collected on 2 minute time steps as sufficiently comprehensive.

Table 16. Comparison of alternative multiple regression models for prediction of ISA (5 sec)

| Regression equation containing | mult. R | $R^2$ | $R^2$ change | F change | df | Sig. F change |
|---|---|---|---|---|---|---|
| complexity components | .67 | .46 | .46 | 79.12 | 6 , 569 | p< 0.001 |
| complexity components and controller's activity measures | .74 | .55 | .09 | 37.72 | 9 , 566 | p< 0.001 |

### 4.2.3 Analysis of variance (repeated measures)

In the following sub-sections the results of a series of analyses of variance (ANOVA) will be presented. The purpose of these analyses is to assess the impact of a set of experimental factors (sector, traffic load, and time interval into exercise) on the various workload metrics (ISA rating, R/T occupancy, controller intervention actions).

Different en-route sectors are considered as different experimental conditions since one of the assumptions was that the measurements would behave differently depending on the characteristics of the airspace in question. The measurements for different traffic loads (baseline and increased traffic load) were analysed to understand how an increase of the traffic load influences changes of the measurement topics.

Except when explicitly noted, the data was analysed in 2-way, repeated measures ANOVA. This means that the impact of the experimental variables was assessed concurrently in a 3 (sector) by 2 (traffic load) ANOVA.

The 3 levels of the factor 'sector' are given by the three sectors: Shannon Low Merged (SHLOW), Shannon Low North (LONO) and Shannon Low South (LOSO). The two levels of the factor 'traffic load' were current peak traffic (baseline) and increased traffic load.

Separate ANOVAs were computed for each metric (hence the term univariate, i.e. the ANOVA considers only one dependent variable). The term repeated measures ANOVA refers to the fact that measurements across experimental conditions were taken from the same subjects (the four air traffic controllers). Therefore, this ANOVA is also called within-subjects ANOVA as opposed to a between-subjects ANOVA.

During the statistical analysis mean values are compared and a test of significance is performed. Statistically significant results guided the interpretation of the operational relevance. For an easier understanding, the results of the statistical analysis are here presented in graphical form.

In the following sections the results of three ANOVAs are presented. These were performed on:

1. ISA workload ratings,
2. R/T frequency occupancy times,
3. Amount of controller intervention activity

All the experimental variables (sector split and traffic increase) can be expected to have a sizable impact on these metrics.

The results of the ANOVA also provide some insight into more differential patterns of changes. For example: Is the effect of increasing traffic on ISA workload, R/T occupancy times or controller intervention activity uniform across all three sectors, or does the effect vary between sectors?

These questions are addressed by character, size and significance of the interaction effect between the experimental factors; in the first case, this is the 'traffic load' by 'sector' interaction. Thus, the ANOVA effect analysis provides a kind of signature of the workload impact of the imposed change.

In a second series of ANOVAs, the same experimental conditions are investigated as to their impact on the complexity component scores. Result patterns of both series of ANOVAs are then compared. Complexity component scores that display similar effect patterns to those of the workload indicators may be considered as potential drivers or at least mediating factors in the generation of the workload effects.

#### 4.2.3.1 Workload (ISA ratings)

The ANOVA revealed a significant main effect of all experimental factors on the average ISA workload rating.

First, as expected, the increase from the current to increased traffic load caused the average ISA workload to increase resulting in a significant main effect of 'traffic load' ($F(1;3) = 20.83$; $p = 0.02$). A main effect considers only the effect of a single factor (in this case traffic) and average the dependent variable (ISA ratings) across the levels of the other factor (sector), meaning that the average workload for low traffic (averaged across sectors) is different from the average workload when traffic load is high.

The break-down of average ISA workload ratings for two different traffic load conditions, averaged across different sectors and controllers, is shown in Figure 15. Additionally, the average workload rating for the higher traffic load is slightly above 3 (“Fair”)-meaning that some of controller’s recorded ratings were either 4 (“High”) or 5 (“very high”) thus resulting in an overall average value of above 3 (“Fair) (even if it is not statistically higher than 3).

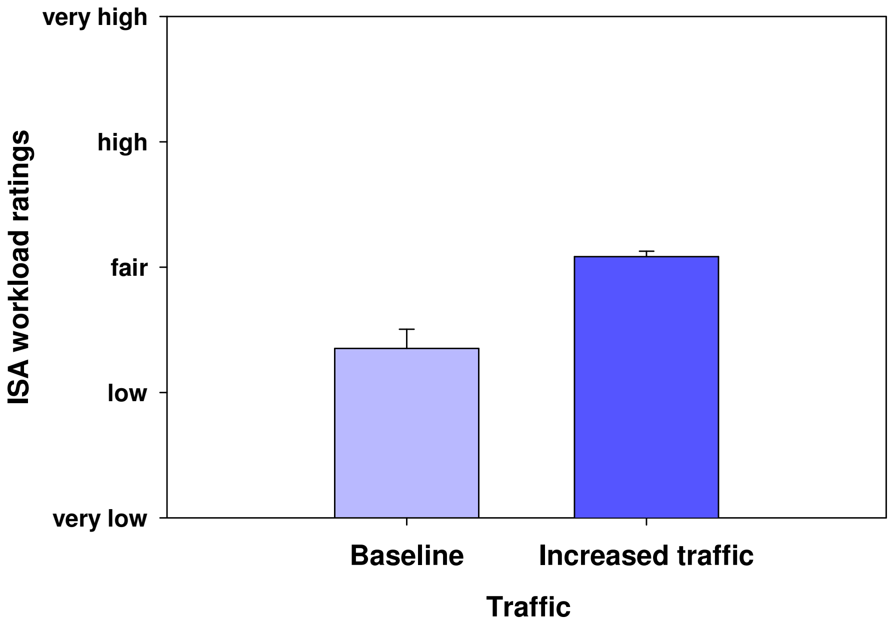



Figure 15. Average ISA workload ratings as a function of traffic, averaged across sectors and controllers (error bars represent standard errors of the means)

Second, average ISA workload ratings differed between sectors causing a highly significant main effect of ‘sector’ ($F(2;6) = 60.9$; $p < 0.001$).
As illustrated in Figure 16 the ISA workload was rated highest when controllers operated the merged sector (SHLOW), averaging across traffic load. Further tests revealed that the workload ratings in SHLOW sector were significantly higher than the average workload ratings in the two subsectors ($p<0.0001$). Moreover, a post-hoc comparison revealed that the average ISA workload was rated higher in the South (LOSO) as compared to the North sector ($p=0.047$), although this difference was not significant when the Bonferroni correction for type-II error was applied.

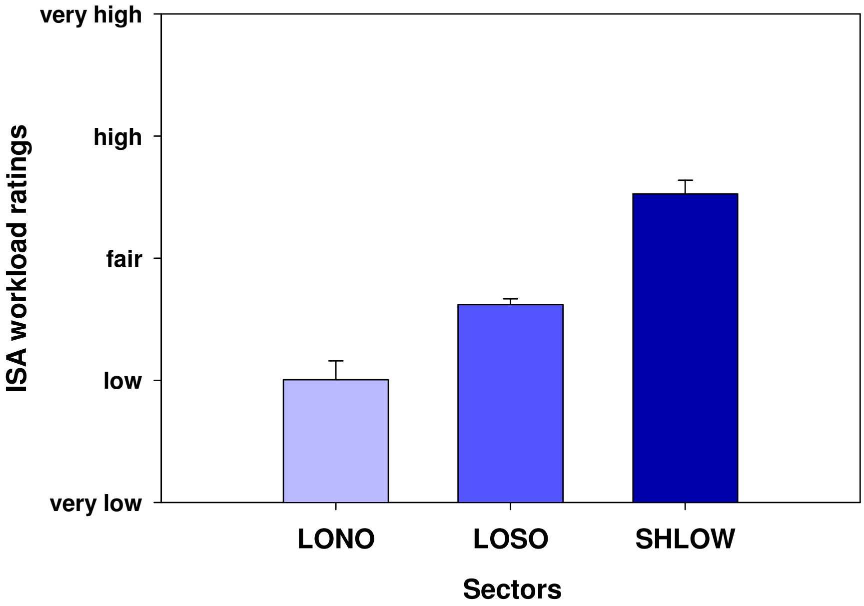



Figure 16. Average ISA workload ratings over the sectors, averaged across traffic loads and controllers (error bars represent standard errors)

The ANOVA further revealed a significant 'sector' by 'traffic load' interaction effect (F(2;6) = 26.7; p = 0.001) suggesting that the effect of increasing the traffic was different between sectors. This is illustrated in Figure 17.

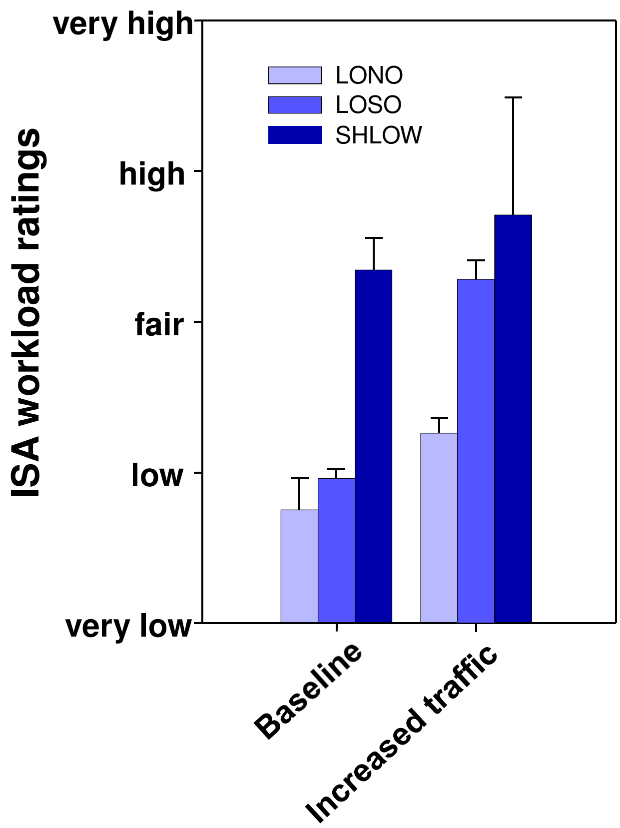



Figure 17. ‘Sector’ by ‘traffic load’ interaction effect on ISA ratings (error bars represent standard errors)

As can be derived from Figure 17 the average ISA workload did not differ much between the North and the South sector when operated in baseline traffic condition. Post-hoc pairwise comparisons indeed failed to reveal a significant difference between the ISA ratings in the two sectors. When operated at an increased traffic load, however, the average ISA rating increased much more in the South (LOSO) as compared to the North (LONO), although the difference was only marginally significant ($p=0.055$). Moreover, the ISA ratings in the SHLOW sector were only significantly higher than in the LONO sector ($p<0.01$), but not higher than in the LOSO one. This pattern is interesting as it shows that an increased traffic load can have different effects dependent on characteristics of the sector, which points to increased complexity created in the South as compared to the North sector. It is also interesting to note that in the SHLOW sector the ANOVA did not find a significant effect of traffic, showing that the workload in the increased traffic conditions was not significantly higher than the workload in the baseline conditions.

#### 4.2.3.2 Controllers' activity measures

In order to reveal whether the controllers' activity measures mirrored the same effect as the workload ratings for different sectors and under different traffic conditions, further analyses of variance were conducted on the frequency (R/T) occupancy time and the measure reflecting the actions performed by the controllers (see section 0).

- Frequency (R/T) Occupancy Time

Each R/T event (controller listening and talking to pilot via the R/T frequency) and its duration were recorded. The duration of R/T events was accumulated across the measured exercise time of one hour and transformed into a percentage score of R/T occupancy time relative to the measured exercise time. Thus, a 20% R/T occupancy time meant that the ATCO was listening or talking to aircraft in 12 min (20%) of the 60 min measured exercise time. Concerned about possible violations of the assumption of the ANOVA due to the nature of this type of measure (with proportion data the mean and the variance tend to be related, and one of the assumption of the ANOVA is that they are independent) all the analyses were also conducted on the arcsin-square-root transformed data, as recommended by Keppel & Wickens (2004). This transformation has the virtue to rescale the data in a way such as the variance and the mean are less dependent, and thus correct the violation of the assumptions. Given that the results of the significance tests did not change, only the results relative to the untransformed data are reported here.

A result illustration for different traffic levels is shown in Figure 18. As was anticipated, overall frequency occupancy time resulted in a significant main effect of traffic load ($F(1;3) = 150.859$; $p = 0.001$), in line with the workload ratings (see Figure 15).

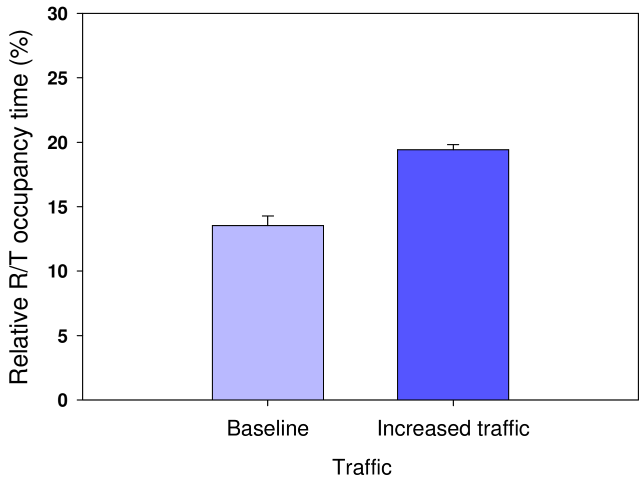



Figure 18. Average frequency (R/T) occupancy time as a function of traffic, averaged across sectors and controllers (error bars represent standard errors)

Similarly to what found about workload, the ANOVA revealed a significant main effect of sector ($F(2;6) = 75.69$; $p < 0.001$). As shown in Figure 19 the pattern of R/T occupancy across sectors is consistent with the pattern found for workload ratings (see Figure 16): communication load was significantly higher in the SHLOW sector than in the other two sectors ($F(1,39=132.02)$; $p<0.01$). Again, these results could be said to have been expected, due to the fact that, with the change of the volume of airspace under control, also the number of aircraft with which the communication is performed is changed. Therefore, the biggest sector correspondingly requires the highest communication load. The results, however, conversely to what was found for the ISA ratings, did not show significant differences between the measure of communication load in the two subsectors.

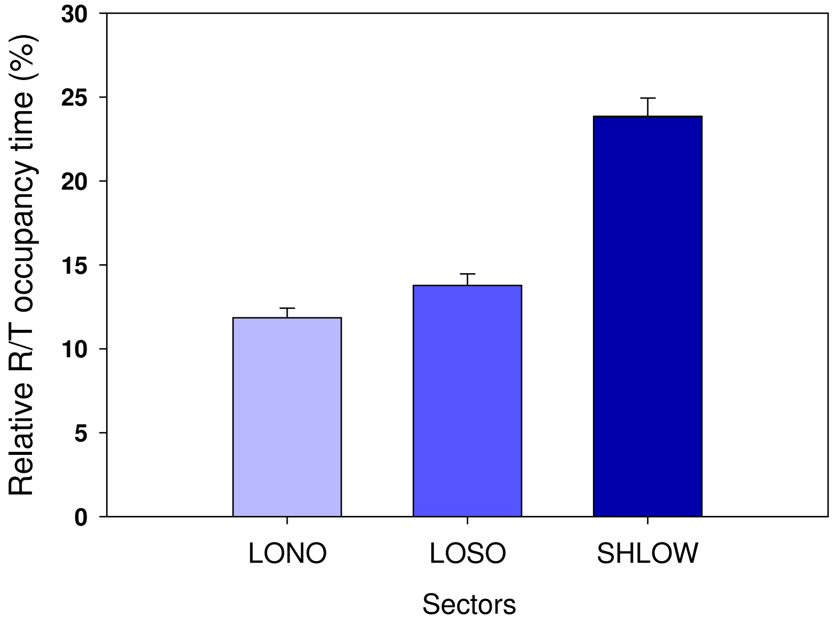



Figure 19. Average frequency (R/T) occupancy time over the sectors, averaged across traffic loads and controllers (error bars represent standard errors)

When looked at the differences between sectors separately for the baseline and the increased traffic condition (i.e. the simple effects of sector on R/T occupancy at low and high traffic), following the significant “traffic by sector” interaction ($F(2;6)=75.71$; $p<0.001$), different patterns of communication load were found across the sectors. Communication load for sector SHLOW, similarly as for workload, does not increase significantly with the increase of traffic volume (Figure 20). This increase is more significant for new sectors LONO and LOSO. Also with the increased traffic load, the pattern obtained for different sectors is analogous to the pattern obtained for the workload ratings for the same conditions (increased traffic load): lower communication load in the LONO sector than in the other two sectors, but no significant differences between LOSO and SHLOW. Additionally, even though workload ratings for LOSO are higher than workload ratings for LONO under baseline conditions (see Figure 20), at least numerically more communication is recorded for LONO than for LOSO ($F(1,2)=25.24$; $p<0.05$).

However, here again it can be noted that communication load in the LOSO sector increased with traffic load much more than it did for the LONO sector. This means that with the same increase of the traffic for LONO and LOSO, still more communication is required in the

LOSO sector than in LONO. Looking into complexity factors may reveal the cause of this effect.

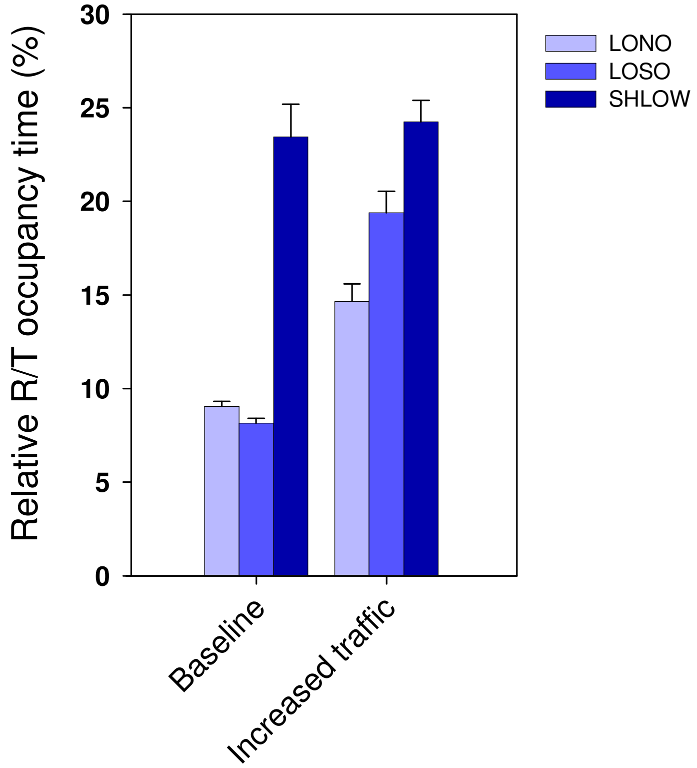



Figure 20. ‘Sector’ by ‘traffic load’ interaction effect on frequency (R/T) occupancy time (error bars represent standard errors)

- Amount of controllers’ intervention activities

In order to reveal the effect of increased traffic and sectors’ configuration on the number of activities performed by the controller, further ANOVA tests were performed taking into consideration all inputs made by the executive controller recorded during the simulation. As in the case of the analysis of the other activity measure, concerns about possible violations of the assumptions drove us to conduct all the analyses were also over transformed data. In this case the transformation applied was the square-root of the activities’ count for each time interval as suggested by Keppel & Wickens (2004). The results of the test were the same as the ones conducted on the untransformed data, which will be thus reported in the following paragraphs.

These inputs refer to those such as assignments of vertical rate, exit flight levels/planned entry levels, cleared flight levels, headings, speed instructions, and direct clearances (see section "*Controller activity measures*" in section 3.3 Experimental scenarios and data collection).

In the light of the previously reported analyses (those over ISA ratings and R/T occupancy), we several effects were expected to be significant. Namely, it was expected that with the higher number of aircraft under control, the controllers' inputs would also increase. As shown on Figure 21, indeed this expectation was confirmed by the results: the number of controller's inputs in the Increased traffic condition was significantly higher than the number of input in the baseline condition (F(1;3) = 46.213; p = 0.007).

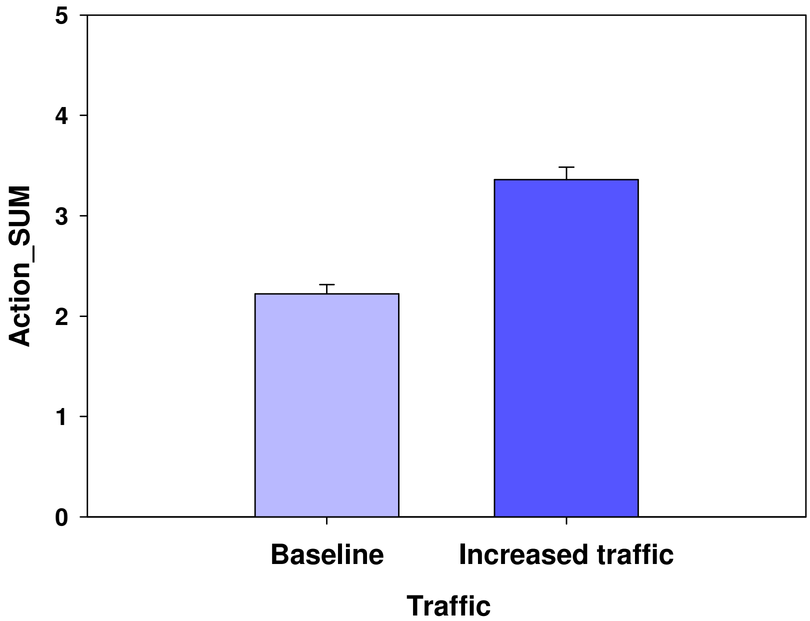



Figure 21. Average number of controller's actions as function of traffic, averaged across sectors and controllers (error bars represent standard errors)

Moreover, the analyses also revealed a significant main effect of sector on the average number of inputs provided by controllers (F(2;6) = 31.651; p = 0.001). As it can be seen in Figure 22, the average number of input across the different sectors follows the same pattern shown by ISA workload ratings and R/T occupancy time.

Further statistical tests revealed that the average number of actions performed in the LONO sector was lower than in the both the LOSO (p<0.05) and the SHLOW sectors (p<0.01).

However, the difference in the number of action in the LOSO and in the SHLOW sectors was only marginally significant (p=0.05).

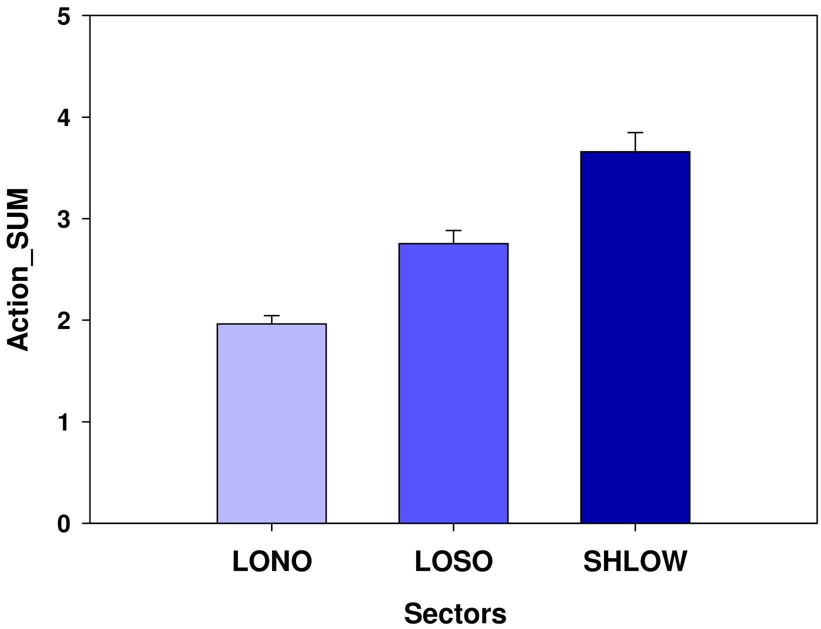



Figure 22. Average number of controller's actions over the sectors (error bars represent standard errors)

The analysis of the number of inputs made by controllers revealed also a highly significant interaction effect of 'sector' by 'traffic' ($F(df=2;6) = 20.757$; $p = 0.002$), showing that the effect of traffic load on the number of actions performed was not the same in the three sectors, or that, conversely, the pattern of the means (of the number of actions performed) across the different sectors in the baseline condition was different from the pattern found in the increase traffic condition. Namely, as shown in Figure 23 below, in the baseline conditions the number of actions was higher in the SHLOW than in the other sectors ($p<0.01$), but no difference was found between the reduced sectors. Conversely, with an increased traffic load no difference was found between the merged Shannon Low (SHLOW) and the South sector (LOSO), and in both sectors more actions ($p<0.01$) were recorded than in the North one (LONO), but no differences were found between LOSO and SHLOW. As for the ISA workload ratings, the same question may arise: what makes LOSO have more difficulty in coping with increased traffic and therefore to increase extensively the number of controllers' inputs when compared to other two sectors?

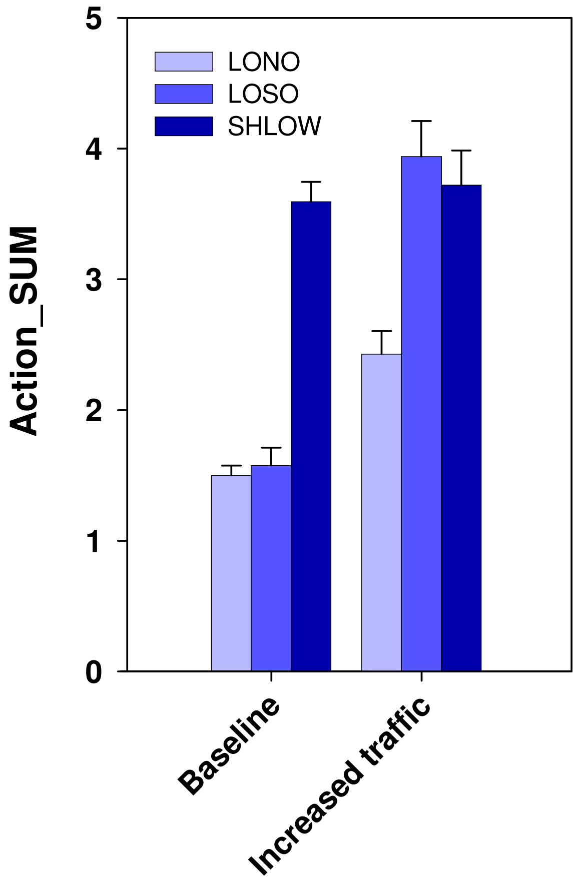



Figure 23. ‘Sector’ by ‘traffic load’ interaction effect on the number of controller’s actions (error bars represent standard errors)

From these findings it is evident that an increase in traffic load resulted in an increase in all of the variables considered - ISA workload, R/T occupancy time and amount of controller intervention activity.

At the same time, for all the three variables a similar pattern of means was found across the different sectors. For each measure, in fact, the highest values were found in the Shannon Low sector (SHLOW), while the lowest values were found in the Shannon Low North (LONO).

However, it is interesting to note that the effect of increasing traffic on ISA workload ratings, R/T occupancy time and controller intervention activity when observed across sectors individually is not uniform, but varies considerably, as witnessed by the significant “sector” by “traffic” interactions that were found in the analyses for each measure.

On the one hand, in fact, in the merged Shannon Low sector the values recorded under baseline conditions do not differ much from those recorded with the traffic load increment. In the Shannon Low South (LOSO), instead, we always found strong effect of traffic load on each of the measured variables, for which the average values recorded in the increased traffic condition were more than twice as big as those recorded in the baseline. It is also interesting to notice that in this sector (LOSO), in the baseline traffic condition the lowest R/T frequency occupancy time was recorded (compared to the other sectors), while when traffic increased these recordings surpassed recordings in LONO sector. Also with increased traffic, the interventions conducted by the controller propagated so much even to go beyond those activities recorded for SHLOW sector. Questions that arise here are: why is the driver of this big intensification of the R/T communications only in this sector and not in other two sectors? What is the characteristic of this sector or its traffic that influences such an increase of R/T coordination?

On the other hand, in Shannon Low North (LONO) sector, on each of the measured variables a significant effect of traffic was found, and the averaged values recorded in the increased traffic condition were higher than those recorded in the baseline.

#### 4.2.3.3 Complexity Components

- Complexity Component 1: **ground speed variance and divergence/convergence**

The complexity component, whose interpretation is guided by the speed of the aircraft and divergence/convergence interaction among the aircraft in the sector, was significantly influenced by the traffic load ($F(1;3) = 27.494$; $p = 0.014$) as shown in Figure 24.

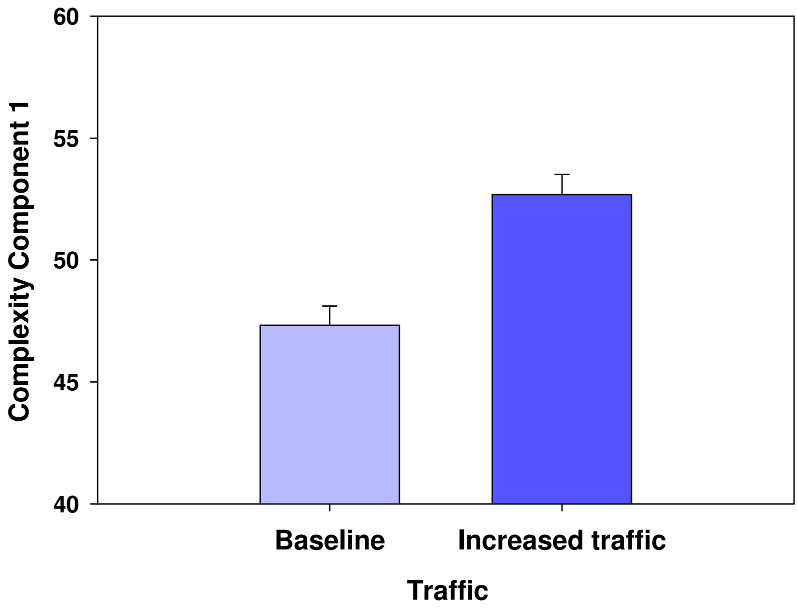



Figure 24. Average Complexity Component 1 (ground speed variance and divergence/convergence) value as function of traffic load, averaged across sectors and controllers (error bars represent standard errors)

Also, the ANOVA revealed a significant effect of sector on this complexity component ($F(2;6) = 27.323$; $p = 0.001$), as can be seen in Figure 25. The pattern of values assumed by this component across the sectors does not mirror the one recorded for the average values of the ISA workload ratings, the R/T occupancy time or the number of controller's actions.

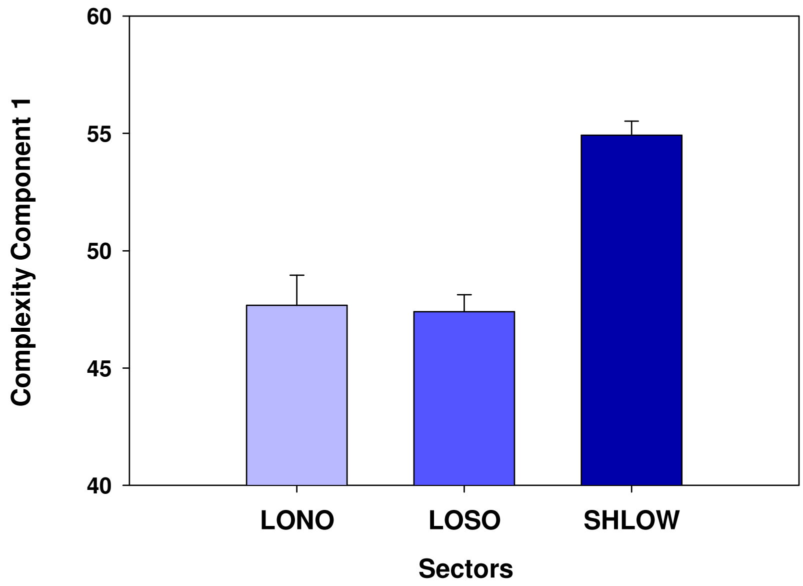



Figure 25. Average Complexity Component 1(ground speed variance and divergence/convergence) value over the sectors, averaged across traffic load and controllers (error bars represent standard errors)

Pairwise comparisons (see Table 17), in fact, showed no differences between the LONO and the LOSO sectors in the average values of this component, and, at the same time, the values recorded in the SHLOW were significantly higher than the values recorded in the other two sectors ($p<.01$).

Table 17. Pairwise Comparisons of the Complexity Component 1 (ground speed variance and divergence/ convergence) value over the sectors, averaged across traffic loads and controllers

| Pairwise Comparisons | | | | | | |
|---|---|---|---|---|---|---|
| **(I) sector** | (J) sector | Mean Difference (I-J) | Std. Error | Sig | 95% Confidence Interval for Difference | |
| | | | | | Lower Bound | Upper Bound |
| **LONO** | LOSO | .277 | 1.150 | .825 | -3.384 | 3.939 |
| | SHLOW | -7.242 | 1.163 | .008 | -10.945 | -3.540 |
| **LOSO** | LONO | -.277 | 1.150 | .825 | -3.939 | 3.384 |
| | SHLOW | -7.520 | 1.147 | .007 | -11.169 | -3.870 |
| **SHLOW** | LONO | 7.242 | 1.163 | .008 | 3.540 | 10.945 |
| | LOSO | 7.520 | 1.147 | .007 | 3.870 | 11.169 |

Here again a significant 'sector' by 'traffic load' interaction effect can be noted ($F(2;6) = 8.973$; $p = 0.016$): under baseline conditions, the lowest values of this component were found in the LOSO sector, while when the traffic is increased the lowest values were recorded in the LONO sector (Figure 26). However, neither in the baseline condition nor in the increased traffic condition significant differences were found between the values recorded for this complexity component in the LONO and LOSO, but only between the values recorded in the SHLOW and those recorded in the other sectors, which were significantly lower ($p<0.01$). In both the LONO and in the LOSO sectors, however, a significant effect of traffic was found on this component, and the values recorded in the increased traffic condition were significantly higher than the values recorded in the baseline, although its significance was only marginal for the LONO sector. No effect of traffic was instead found on this complexity component in the SHLOW sector.

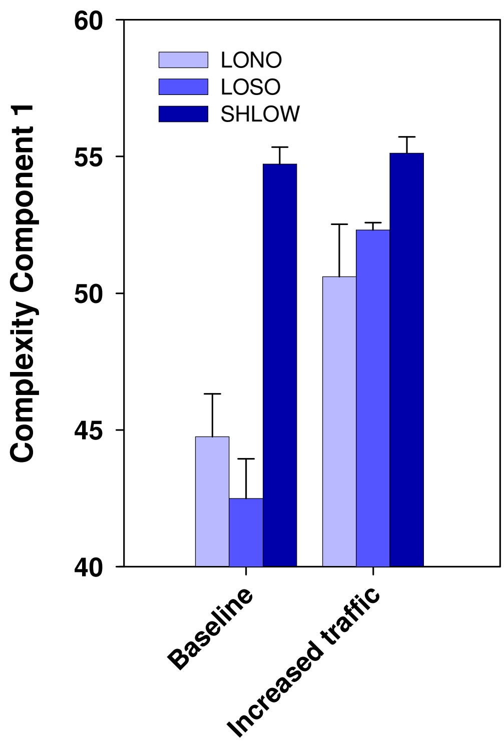



Figure 26. ‘Sector’ by ‘traffic load’ interaction effect on Complexity Component 1 (ground speed variance and divergence/convergence) (error bars represent standard errors)

- Complexity Component 2: **aircraft count**

Also for the second complexity component, the one whose interpretation is driven by the highest loading of the number of the aircraft in the sector, the ANOVA revealed significant main effects both of the traffic load ($F(df=1;3) = 11.282$; $p = 0.044$) and of the sector $F(df=2;6) = 94.358$; $p < 0.001$). The effect of traffic is shown in Figure 27, where can be seen that the average value of this component recorded in the increased traffic condition was significantly higher than the value recorded in the baseline.

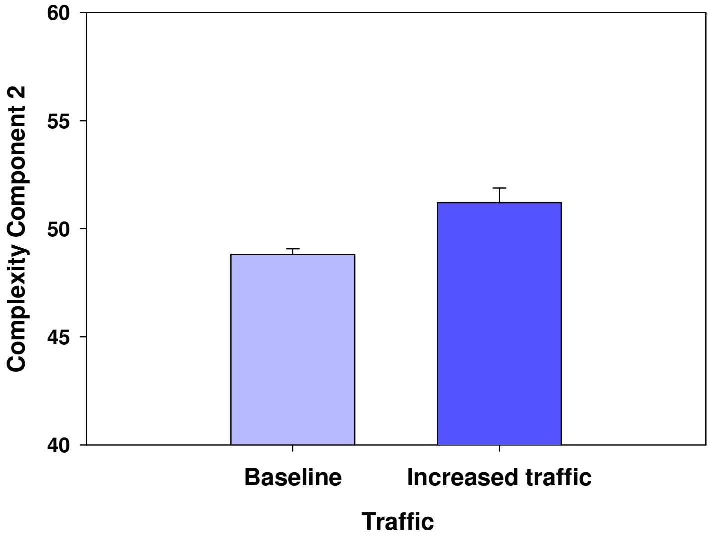



Figure 27. Average Complexity Component 2 (aircraft count) value as a function of traffic, averaged across sectors and controllers (error bars represent standard errors)

On the other hand, as shown in Figure 28, the average value of this component was significantly higher in the SHLOW sector than in the other two sectors, while no differences were found between LONO and LOSO (the results of the pairwise comparisons are reported in Table 18).

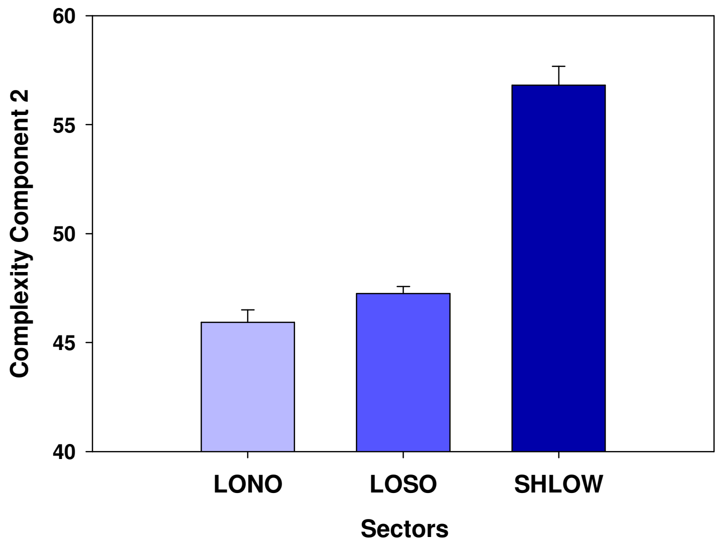



Figure 28. Average Complexity Component 2 (aircraft count) value over the sectors, averaged across traffic loads and controllers (error bars represent standard errors)

Table 18. Pairwise Comparisons of the Complexity Component 2 (aircraft count) values over the sectors, averaged across traffic loads and controllers

| Pairwise Comparisons | | | | | | |
|---|---|---|---|---|---|---|
| **(I) sector** | (J) sector | Mean Difference (I-J) | Std. Error | Sig | 95% Confidence Interval for Difference | |
| | | | | | Lower Bound | Upper Bound |
| **LONO** | LOSO | -1.321 | .604 | .116 | -3.242 | .600 |
| | SHLOW | -10.882 | .898 | .001 | -13.740 | -8.023 |
| **LOSO** | LONO | 1.321 | .604 | .116 | -.600 | 3.242 |
| | SHLOW | -9.561 | 1.035 | .003 | -12.854 | -6.267 |
| **SHLOW** | LONO | 10.882 | .898 | .001 | 8.023 | 13.740 |
| | LOSO | 9.561 | 1.035 | .003 | 6.267 | 12.854 |

The interaction effect of ‘sector’ by ‘traffic load’ was also again significant, showing that the effect of increased traffic on this component was different between sectors ($F(2;6) = 7.238$; $p = 0.025$). This is also depicted on the Figure 29 below. As it can be seen in the figure, only in the LOSO sector the increase of traffic was followed by a (statistically) significant increase in the average value of this complexity component. Looking at the differences between the average values of this component across the sectors showed first of all that both in the baseline and in the increased traffic condition the complexity was higher in the SHLOW than in all the other sectors ($p<.01$). In the baseline condition, however, the average complexity value recorded in the LONO was not significantly different to the one recorded in the LOSO. In the increased traffic condition, conversely, the complexity recorded in the LOSO sector was significantly higher than the one recorded in LONO, and significantly lower than the one recorded in the SHLOW.

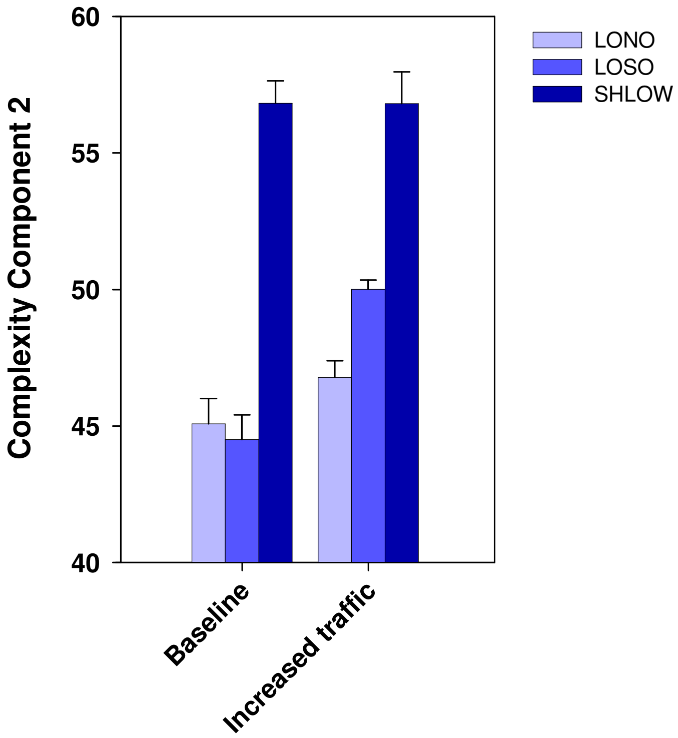



Figure 29. 'Sector' by 'traffic load' interaction effect on Complexity Component 2 (aircraft count) (error bars represent standard errors)

- Complexity Component 3: **aircraft vertical transitioning**

The analysis of complexity component 3, the one mainly related to aircraft vertical transitioning, yielded very interesting results. The ANOVA showed a significant effect of both traffic ($F(1;3) = 34.066$; $p = 0.010$) and sector ($F(2;6) = 278.386$; $p < 0.001$) on this component. As it can be seen in Figure 30 the average value of the component was higher in the increased traffic than in the baseline. As it can be seen in Figure 31, however, the pattern of the average values of this component across the sector was different from the pattern that was found for the other components described so far.

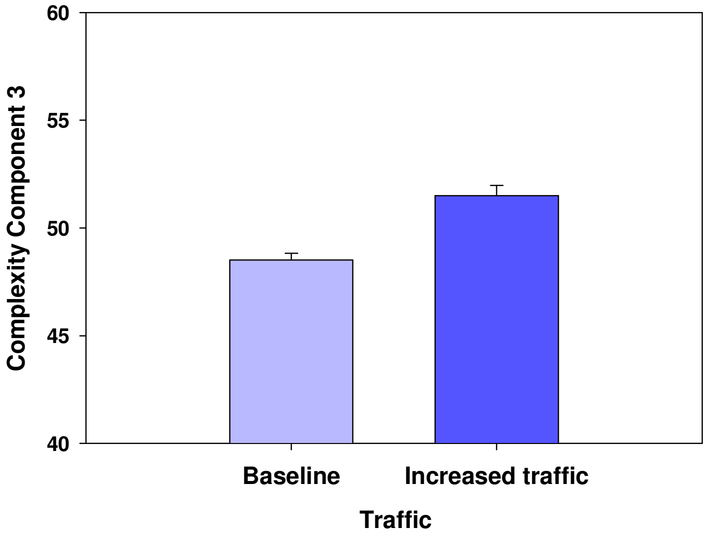



Figure 30. Average Complexity Component 3 (aircraft vertical transitioning) value as function of traffic, averaged across sectors and controllers (error bars represent standard errors)

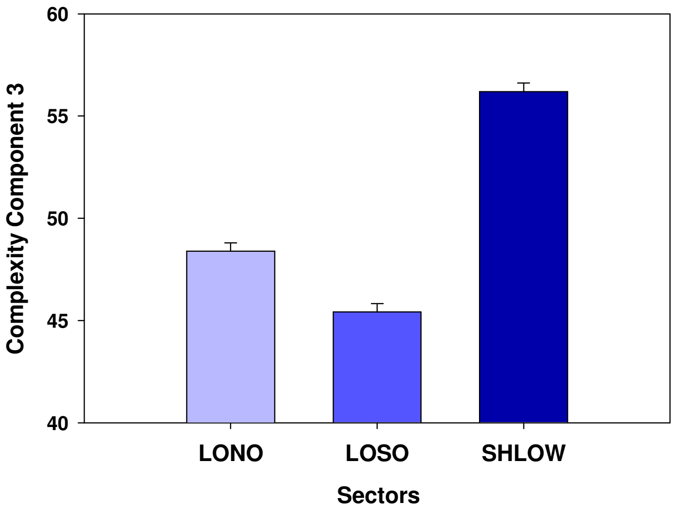



Figure 31. Average Complexity Component 3 (aircraft vertical transitioning) value over the sectors, averaged across traffic loads and controllers (error bars represent standard errors)

As can be seen, the highest values of this component were recorded for SHLOW, as for other variables previously described. However, when compared to other variables, here the smallest values are recorded for sector LOSO, and not for sector LONO. Indeed pairwise comparisons (see Table 19 below) showed that the scores for the LOSO sectors were significantly lower than both the scores recorded in the LONO sector ($p<0.05$) and those found in the SHLOW sector ($p<0.01$). One of the possible explanations for this may lie in the fact that in the sector Shannon North many aircraft decrease significantly the flight level in order to achieve necessary altitude for the safe landing on the airports that actually lie below Shannon South low-level sectors. Therefore, once when aircraft enter the LOSO sector they are already flying at sufficiently low flight levels and therefore not many vertical movements are required, but more sequencing of the incoming traffic for their transfer into terminal area.

Table 19. Pairwise Comparisons of the Complexity Component 3 (aircraft vertical transitioning) values over the sectors, averaged across traffic loads and controllers

| Pairwise Comparisons | | | | | | |
|---|---|---|---|---|---|---|
| **(I) sector** | (J) sector | Mean Difference (I-J) | Std. Error | Sig | 95% Confidence Interval for Difference | |
| | | | | | Lower Bound | Upper Bound |
| **LONO** | LOSO | 2.962 | .556 | .013 | 1.193 | 4.731 |
| | SHLOW | -7.799 | .346 | .000 | -8.899 | -6.699 |
| **LOSO** | LONO | -2.962 | .556 | .013 | -4.731 | -1.193 |
| | SHLOW | -10.761 | .487 | .000 | -12.312 | -9.210 |
| **SHLOW** | LONO | 7.799 | .346 | .000 | 6.699 | 8.899 |
| | LOSO | 10.761 | .487 | .000 | 9.210 | 12.312 |

It was also interesting to look at the 'sector' by 'traffic load' interaction effect, i.e. the way effect of increasing traffic vary across the different sectors (Figure 32). The ANOVA here revealed again a significant interaction effect ($F(2;6) = 14.356$; $p = 0.005$). Both in the LONO and in the LOSO sector, an increase of traffic caused a significant increase in this complexity component, while no effect of traffic load was found in the SHLOW. In the LONO sector, however, this increment is smaller when compared with the increment recorded in LOSO. Indeed when we looked at the difference between LOSO and LONO in the increased traffic condition, the difference was only marginally significant ($p<0.1$).

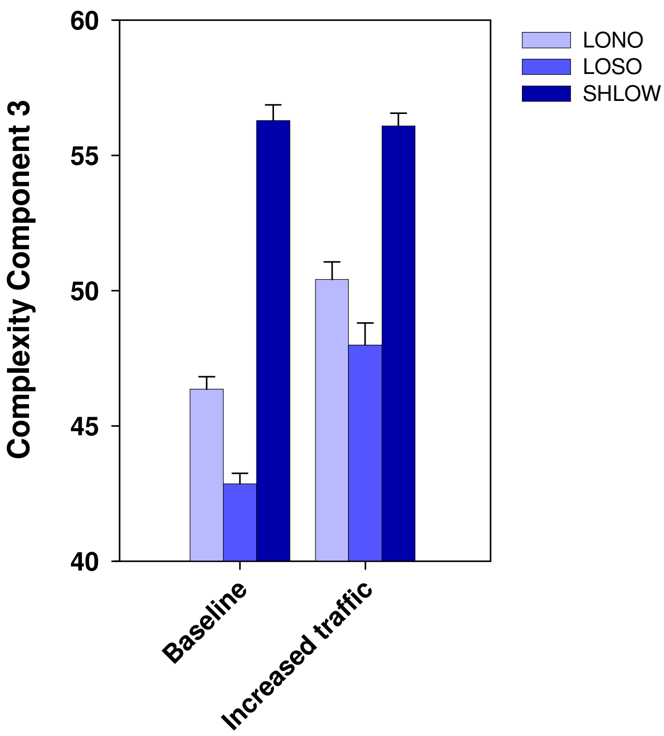



Figure 32. ‘Sector’ by ‘traffic load’ interaction effect on Complexity Component 3 (aircraft vertical transitioning) (error bars represent standard errors)

- Complexity Component 4: horizontal proximity

The complexity component reflecting the horizontal proximity of the aircraft within the controlled airspace was also significantly affected by traffic load level ($F(1;3) = 54.997$; $p = 0.005$), and its average value was higher in the increased traffic condition than in the baseline, as it can be seen in Figure 33. Significant differences in the average values of this component were also found between the different sectors ($F(2;6) = 32.710$; $p = 0.001$). This is shown in Figure 34, where it can be seen that the average complexity recorded in the LONO was lower than the average complexity recorded in the LOSO, and that, in turn, complexity in the LOSO was lower than in SHLOW. As it can be verified in Table 20, pairwise comparisons for this component showed that the average values recorded in each sector was significantly different from the average values recorded in the other sectors. This particular patter of means is very similar to the one found for the average ISA ratings across the sectors.

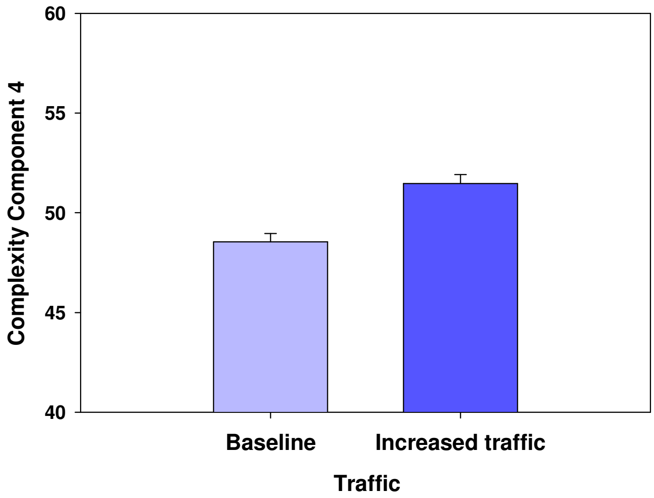



Figure 33. Average Complexity Component 4 (horizontal proximity) value as function of traffic, averaged across sectors and controllers (error bars represent standard errors).

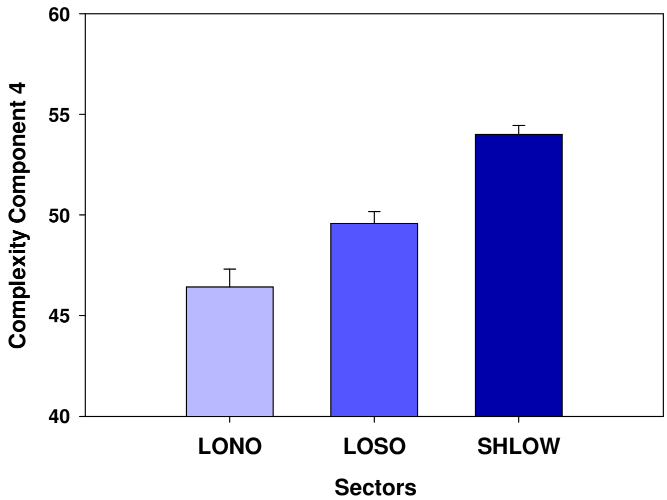



Figure 34. Average Complexity Component 4 (horizontal proximity) value over the sectors, averaged across traffic loads and controllers (error bars represent standard errors)

Table 20. Pairwise Comparisons of the Complexity Component 4 (horizontal proximity) values over the sectors, averaged across traffic loads and controllers

| Pairwise Comparisons | | | | | | |
|---|---|---|---|---|---|---|
| **(I) sector** | (J) sector | Mean Difference (I-J) | Std. Error | Sig | 95% Confidence Interval for Difference | |
| | | | | | Lower Bound | Upper Bound |
| **LONO** | LOSO | -3.156 | .891 | .038 | -5.993 | -.319 |
| | SHLOW | -7.572 | 1.017 | .005 | -10.809 | -4.334 |
| **LOSO** | LONO | 3.156 | .891 | .038 | .319 | 5.993 |
| | SHLOW | -4.415 | .908 | .017 | -7.304 | -1.527 |
| **SHLOW** | LONO | 7.572 | 1.017 | .005 | 4.334 | 10.809 |
| | LOSO | 4.415 | .908 | .017 | 1.527 | 7.304 |

For the horizontal proximity complexity component the analysis also found a significant 'sector' by 'traffic load' interaction effect ($F(2;6) = 12.990$; $p = 0.007$). In Figure 35 below the large impact of increasing traffic load on this component in sector LOSO can be seen, while an impact of traffic load was absent in the other two sectors, and not only in SHLOW. The analysis also showed that in the baseline there was no significant difference between average value of this component in LONO and in LOSO, that the values recorded in both these sectors were significantly lower than SHLOW ($p<0.01$). In the increased traffic conditions, instead, no differences were found between LOSO and SHLOW, and both had recorded significantly higher scores than LONO ($p<0.001$).

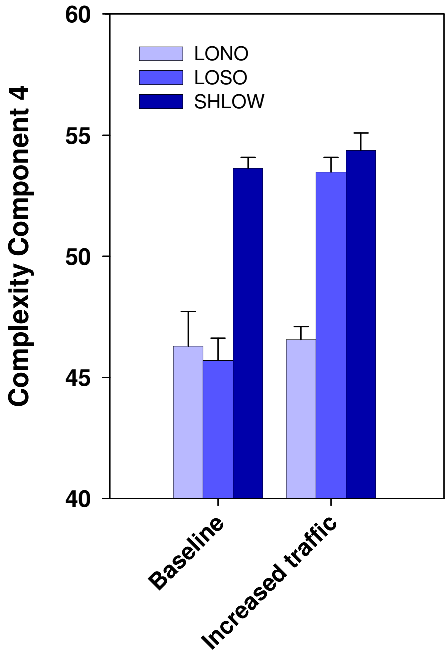



Figure 35. ‘Sector’ by ‘traffic load’ interaction effect on Complexity Component 4 (horizontal proximity) (error bars represent standard errors)

- Complexity Component 5: **conflict sensitivity**

For the complexity component whose meaning is largely driven by the conflict sensitivity, ANOVA also revealed significant effect of ‘traffic load’ ($F(1;3) = 67.572$; $p = 0.004$). As can be seen in Figure 36, also for this component the average value recorded in the increased traffic condition was significantly higher than the average value recorded in the baseline.

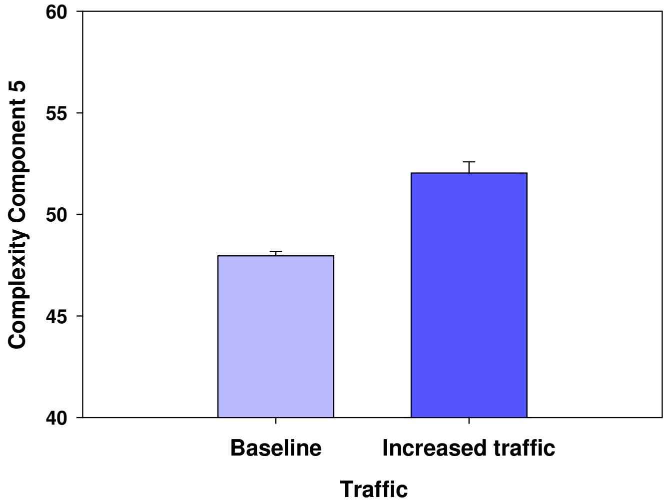



Figure 36. Average Complexity Component 5 (conflict sensitivity) value as function of traffic, averaged across sectors and controllers (error bars represent standard errors)

Moreover, ANOVA also revealed a highly significant main effect of 'sector' ($F(2;6) = 131.516$; $p < 0.001$), meaning that the average value of conflict sensitivity changes across the sectors. This is depicted in the Figure 37 below. As it can be seen in the figure, the average value of this component was higher in SHLOW than in the other sectors. No significant differences were found between the average value in LONO and in LOSO (Table 20).

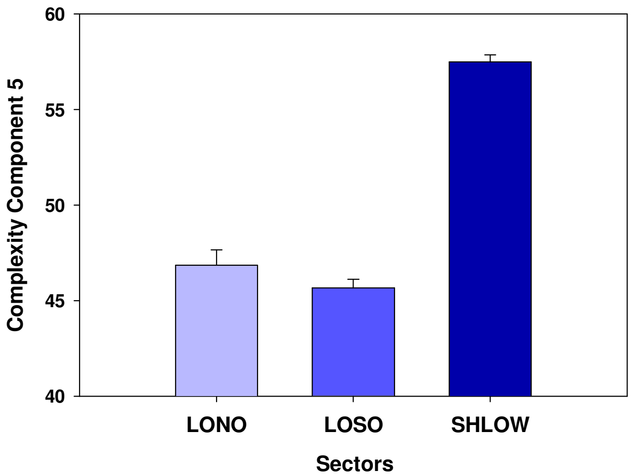



Figure 37. Average Complexity Component 5 (conflict sensitivity) value over the sectors, averaged across traffic loads and controllers (error bars represent standard errors)

Table 21 Pairwise Comparisons of the Complexity Component 5 (conflict sensitivity) values over the sectors, averaged across traffic loads and controllers

| Pairwise Comparisons | | | | | | |
|---|---|---|---|---|---|---|
| **(I) sector** | (J) sector | Mean Difference (I-J) | Std. Error | Sig | 95% Confidence Interval for Difference | |
| | | | | | Lower Bound | Upper Bound |
| **LONO** | LOSO | 1.185 | .629 | .156 | -.817 | 3.187 |
| | SHLOW | -10.637 | 1.096 | .002 | -14.125 | -7.150 |
| **LOSO** | LONO | -1.185 | .629 | .156 | -3.187 | .817 |
| | SHLOW | -11.822 | .581 | .000 | -13.671 | -9.974 |
| **SHLOW** | LONO | 10.637 | 1.096 | .002 | 7.150 | 14.125 |
| | LOSO | 11.822 | .581 | .000 | 9.974 | 13.671 |

The ‘sector’ by ‘traffic load’ interaction effect was also once again statistically significant ($F(2;6) = 6.017$; $p = 0.037$), and the analysis of the simple effects showed that in both LONO and in LOSO the average values of this complexity component were higher in the increased traffic than in the baseline (Figure 38). Conversely, no significant effect of traffic on this component was found in SHLOW. Moreover, in the baseline conditions the difference between LONO and LOSO was marginally significant ($p<.1$), while in the increased traffic conditions no differences were found between these sectors.

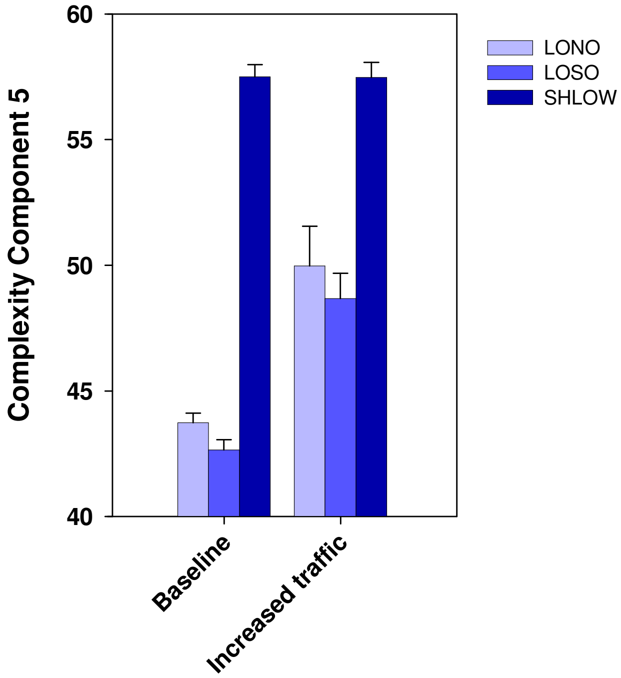



Figure 38. 'Sector' by 'traffic load' interaction effect on Complexity Component 5 (conflict sensitivity) (error bars represent standard errors)

- Complexity Component 6: **insensitivity**

For this component ANOVA did not reveal any statistical significant effect of sector or traffic load, nor of their interaction. And therefore further detailed interpretation of the results was found unnecessary. However, the graphs depicting the obtained results for the consistency with other reported variables are provided below (Figure 39, 40 and 41).

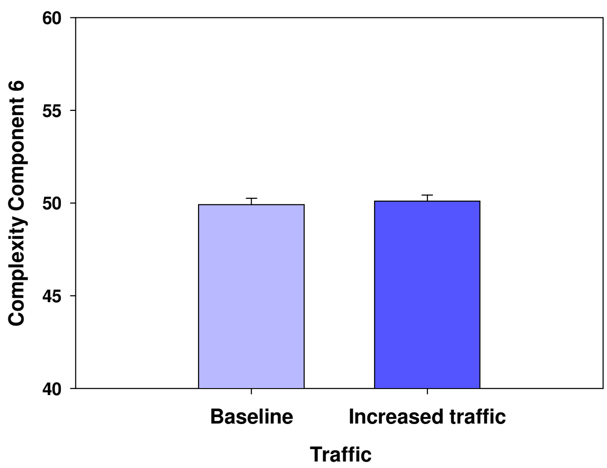



Figure 39. Average Complexity Component 6 (insensitivity) value as function of traffic, averaged across sectors and controllers (error bars represent standard errors)

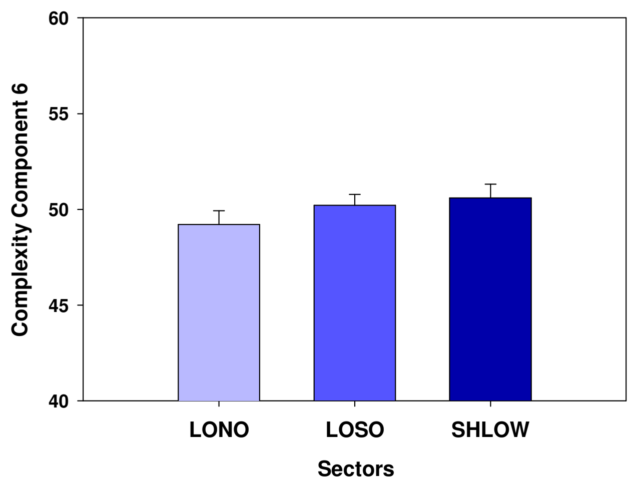



Figure 40. Average Complexity Component 6 (insensitivity) value over the sectors, averaged across traffic loads and controllers (error bars represent standard errors)

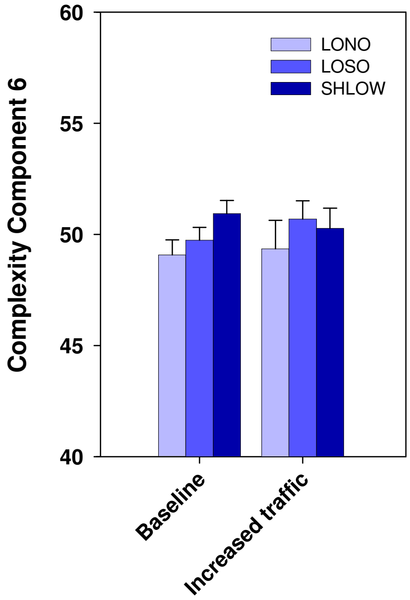



Figure 41. ‘Sector’ by ‘traffic load’ interaction effect on Complexity Component 6 (insensitivity) (error bars represent standard errors)

- The ANOVA conclusions

Based on the ANOVA findings for the complexity components, differential patterns of changes are detected for three considered sectors.

For all the first 5 components, the analyses found that in the baseline condition complexity was significantly higher in the merged Shannon Low sector (SHLOW) than in the other sectors. The same pattern was found in the analyses of the ISA workload ratings, of the frequency occupancy time and of the amount of intervention activities of the controllers. In the increased traffic condition, on the contrary, not all the components were found to follow this pattern, showing that the effect of traffic on the complexity in the different sector was not the same for all the components. In other words, increasing traffic did not increase in all the sectors also the complexity. The only sector, in which the effect of traffic was the same for all the components, as well as for the activity measures, was SHLOW, where basically increasing traffic did not have any effect.

For CC1 (ground speed variance and convergence/divergence), CC3 (aircraft vertical transitioning) and CC5 (conflict sensitivity) the analyses showed that increasing traffic increased complexity in both LONO and LOSO. For components CC2 (aircraft count) and CC4 (horizontal proximity), instead, the effect of traffic was only found in LOSO. The latter was also the pattern that was found for the three activity measure, for which a significant difference between baseline and increased traffic was only found in the LOSO sector.

The effect of traffic in the LOSO sector was thus found for all the 5 components. In LONO, instead, only three components, related to ground speed, aircraft vertical transitioning and conflict sensitivity, were affected by traffic load.

In the following tables are summarized the main findings of the ANOVAs that have been conducted, relative to the nature of the effects of the factors considered in the analyses and of their interactions on the dependent variables. More precisely, the first table (Table 22) reports the simple effects of traffic load changes separately in the different sectors.

As can be seen, for both the workload measures and the complexity components, two general patterns can be identified. On one hand, in the SHLOW sector the increase in traffic load did not significantly affect either workload or complexity. On the other hand, increased traffic always increased workload and complexity levels in the LOSO sector. As for the LONO sector, instead, a rise in the traffic also increased the measures of workload (subjective and objective), and some of the complexity components, but not all. No significant differences were in fact found in this sector between the baseline and the increased traffic conditions in the complexity components related to the aircrafts counts (component 2) and to the horizontal proximity (component 4).

Table 22. Effect of traffic load changes on the variables across the sectors

| | Effect of traffic load change across the sectors | | |
|---|---|---|---|
| **Parameter (measure)** | LONO | LOSO | SHLOW |
| **ISA** | Baseline < Increased | Baseline < Increased | Baseline ≈ Increased |
| **R/T** | Baseline < Increased | Baseline < Increased | Baseline ≈ Increased |
| **AS** | Baseline < Increased | Baseline < Increased | Baseline ≈ Increased |
| **CC1** | (Baseline < Increased) | Baseline < Increased | Baseline ≈ Increased |
| **CC2** | Baseline ≈ Increased | Baseline < Increased | Baseline ≈ Increased |
| **CC3** | Baseline < Increased | Baseline < Increased | Baseline ≈ Increased |
| **CC4** | Baseline ≈ Increased | Baseline < Increased | Baseline ≈ Increased |
| **CC5** | Baseline < Increased | Baseline < Increased | Baseline ≈ Increased |

A more complex picture emerges when we look at the data from a different angle, analysing the interaction in terms of the differences between the sectors separately for the baseline and the increased traffic conditions. As can be seen in Table 23 there are again two general patterns of differences that seem to occur frequently. On the one hand, in the baseline conditions the most frequent pattern is the one in which the dependent variable is higher in the SHLOW sector than in the other two sectors (on average), and no differences are found between them. This is the pattern we find for the ISA workload ratings, for the number of actions performed, as well as for the complexity components 1, 2 and 4. An additional difference, however, is found between the subsectors in the R/T variable, with less radio communication being performed in LONO than in LOSO. More interestingly, for the remaining two components (components 3 and 5) the opposite trend of development is found when the subsectors are compared: in these cases, in fact, complexity seems higher in the LONO than in the LOSO subsector.

The more frequent pattern in the increased traffic condition is the one that was found for the ISA ratings, as well as for AS and for complexity factor 4: for all of these variables significantly lower values were recorded in the LONO sector than in the other two sectors, that however were not statistically different from each other (LOSO ≈ SHLOW).

For other variables, namely components 1 and 5, however, the pattern is different, in that no differences can be found between the subsectors that on average show lower complexity than SHLOW. For R/T and component 2, instead, values were recorded in each sector that were significantly different from the ones recorded in the other sectors, with a general trend of increasing values in the following order: LONO < LOSO < SHLOW. Finally, for component 3, the trend of complexity between the subsectors is reversed, as we found significantly higher values in LONO than in LOSO.

Table 23. Difference of the variables among sectors under different traffic load

| | Difference between sectors under different traffic load | |
|---|---|---|
| **parameter** | Baseline | Increased |
| **ISA** | (LONO≈LOSO)<SHLOW | LONO<(LOSO≈SHLOW) |
| **R/T** | LONO<LOSO<SHLOW | LONO<SHLOW |
| | | LONO≤LOSO |
| | | LOSO≤SHLOW |
| **AS** | (LONO≈LOSO)<SHLOW | LONO<(LOSO≈SHLOW) |
| **CC1** | (LONO≈LOSO)<SHLOW | (LONO≈LOSO)<SHLOW |
| **CC2** | (LONO≈LOSO)<SHLOW | LONO<LOSO<SHLOW |
| **CC3** | LOSO<LONO<SHLOW | LOSO(<)LONO<SHLOW |
| **CC4** | (LONO≈LOSO)<SHLOW | LONO<(LOSO≈SHLOW) |
| **CC5** | LOSO(<)LONO<SHLOW | (LONO≈LOSO)<SHLOW |

# 5 Summary, Conclusions and Recommendations

In order to address the existing problems in the ATM system (such as reduced mobility, delays, the more frequent safety occurrences issues, higher costs and pollution through $CO_2$, noise emissions, etc.) the ATM system is subject to numerous changes and modernization in general, both already on-going but also envisaged in the next decades. These changes comprise the introduction of new technologies, new design concepts, new procedures and operating methods, even the introduction of new operator roles. All these changes will inevitably have an impact on the controller and the way he/she currently performs the work.

The main idea motivating this work has been the possibility to assess and / or predict the impact that new technologies and procedures may imply for the controller’s work when comparing with those established and currently in use. Additionally, if the impact of these changes could be anticipated for certain variables of the current system, by their manipulation it would also be possible to investigate the impact that new technologies and procedures would create onto other elements of the system. Gaining insight into the importance of these variables moreover increases understanding of their effects on the performance of the overall system and significantly facilitates the diagnosis of problems that could emerge.

Therefore, the focus of this work was the identification of different parameters of the air traffic situations, i.e. ATC complexity components, and the impact that their variations induce on the work of the air traffic controllers resulting into different workload levels that they experience while organizing the aircraft that are following flight plan routes within boundaries of airspace under his/her responsibility.

Lots of research has been conducted in the field of the complexity where the contributing factors were considered as drivers of changes in workload levels as a whole. The challenge of this work was to investigate into the identification of the individual influence of each single complexity component independently, with the reasoning that the overall complexity of different air traffic sectors is led by different complexity features. The comparison of complexity across facilities cannot be considered as completely accurate if one cannot distinguish which complexity component is more (or less) present in one airspace when compared to another one, although the overall complexity of both sectors may be assessed as equal. Namely, while one component of complexity may affect by a large degree the overall complexity of that specific sector and consequently workload of the controller, on the other hand the same component may be proven as trivial in the other sector at issue.

Therefore, to start with, this research aimed at identifying complexity measures that would together adequately correspond to the controller's workload in different airspace sectors and under different conditions simulated. The hypothesis, for which the first phase of this work was conducted to put to test, states that it is possible to measure air traffic complexity as a predictor of the controller's workload under different conditions using the objectively recorded data.

It should be however emphasized that the scope of this work encompasses only en-route type of sectors. Namely, the formula based on which the complexity measures were defined was developed considering traffic characteristics of en-route sectors, but also the distinctive features of the activities that en-route controllers apply within their working methods and procedures. Nevertheless, the methodology applied in the present work to derive the complexity measurement adequate for the en-route sector could be used for guiding the identification of the complexity measurement for other sectors associated to different air traffic control units (approach and TMA, as in Vogel *et al.*, 2013).

The ATC complexity factors that have been consistently found to be important in the previous studies accounted for the following characteristics of the en-route traffic (see section 3.3.4):

- aircraft density (concentration of aircraft in the measure of space and their count)
- flight attributes of each individual aircraft (considering also the count of those in the process of transitioning - changing speed, direction, altitude),
- aircraft conflicts (distance between aircraft, speed with which they are moving to/from each other, etc.) and
- traffic disorder (discrepancy in their speeds and headings).

Hence, to create a list of complexity factors that would comprehensively cover the ATC complexity, the challenge and the first objective of this work was to adopt and adapt from prior research conducted in this field the of the complexity factors. Those complexity factors that have been consistently found to be important in the previous studies and for which detailed calculation formula have been reported were selected for further analysis. This activity resulted in the initial list of 24 complexity factors that are then treated by the PCA. Namely, since the established set of factors resulted from multiple researches conducted in this field, it was assumed that some of these factors are correlated with one another, overlapping and possibly measuring similar concepts. In order to ensure that the statistically redundant portions of these factors are removed, but that at the same time the information that they contain is preserved by combining information contained within these factors into a smaller number of new artificial variables, the PCA was performed on the overall set of 24 complexity factors. This analysis was performed based on the data recorded during the real-time simulation addressing en-route airspace where a new concept was introduced (i.e. CPDLC, see section 3.3.2). This PCA resulted in the 6 complexity components, whose interpretations are driven by the factors that showed the strongest correlation with that component (see section 4.1.1 for more detailed description of the analysis and the interpretation of each component):

- ground speed variance and divergence/convergence
- aircraft count
- horizontal proximity
- aircraft vertical transitioning
- conflict sensitivity
- insensitivity.

Subsequently, with the aim of establishing a link between ATC complexity and a controller's subjective workload, complexity components identified are related to workload measures. The multiple regression analysis was conducted using the instantaneous self-assessment ratings provided by the controllers as measures of workload and complexity components as predictors. Besides the complexity components, it was decided to look at the controllers' performance measures (inputs made by the controller, cumulative duration of radio calls, i.e. frequency occupancy time, and average duration of single calls) and their correlation with workload measures. The aim was to investigate the relationships between complexity, performance and subjective workload, to test whether information about the controller's activity could be useful for predicting workload, once the effect of complexity had been considered, and to verify whether the effect of complexity on workload could be mediated by the effect of complexity on the controller's activity.

The analysis revealed that both of these sources of information (about complexity and about activity) give a unique contribution to the prediction of ISA workload ratings and therefore the workload of the controller is determined by both ATC complexity and the activities that the controller performs to deal with a demand imposed on him/her. In addition, the results revealed the single contribution of each complexity component and controller's activity measure in the prediction of workload: those complexity components that played the most significant role in the prediction of workload are horizontal proximity between aircraft in the airspace and the sensitivity to the conflict. On the other side, the controllers' activity measure that did not confirm the correlation with the workload ratings was the one that characterizes the inputs made by the controllers, demonstrating that only communication related measurements play a significant role in the prediction of workload, once complexity has been taken into account. Further, while frequency (R/T) occupancy time was directly correlated with workload (the more time spent on frequency, the higher is workload), the average duration of single communication was negatively affected, i.e. in the situation in which controllers perceive higher workload, they tend to spend less time on a single call (these results are provided in the section 4.1.2). This kind of behaviour can be attributed to the active adaptation of the controllers to the increased demand in order to maintain the workload at an acceptable level. These findings are compliant with the findings of Manning *et al.* (2001) and also with the closed-loop model of ATCO's workload developed by Sperandio (1971) explaining that the actions performed in response to the task demand placed in front of the controller influence the task demand encountered in the future.

The regression analysis showed that the set of 6 complexity components extracted by PCA could be used to predict of the controller's workload for that en-route sector. However, this does not prove that the same complexity components would be equally applicable as predictors of workload in different sectors. Therefore, we performed a second regression analysis using data about complexity and workload in different en-route sectors. The data were recorded in another real-time simulation focusing particularly on the design of the sectors (IAA RTS 1, see section 3.3.3). The complexity components were calculated based on the previous findings, using the coefficients from the PCA performed on the first set of data (obtained during the first real-time simulation addressing CPDLC described in section 3.3.2). The multiple regression analyses performed on the new set of data confirmed the predictive power of complexity components, revealing even stronger correlation with the workload ratings. As in the first regression analysis, the components which were found to have a stronger effect on workload were those related to horizontal proximity and conflict sensitivity. Overall, these results, together with the ones of the first regression, confirmed the first hypothesis of the research.

Interestingly, however, once controller's activity measures were added to the regression model, only four (out of six) complexity components were still able to give a significant contribution to the prediction of workload. These results led to the possibility that the remaining two complexity components (the ones related to aircraft count and to ground speed variance) are strongly related to and impact the controller's activity measures, and therefore, once these measures are in considered in the model the components become redundant. Alternatively, it might be that the excluded complexity components are not playing a significant role in the prediction of workload in this particular airspace – which, in fact, leads to the assumption that the different complexity components correspond differently to the workload experienced by the controller in different airspace sectors.

Moreover, the results of the analysis revealed that complexity components are not as successful for the prediction of controller's activity measures as they are for the prediction of workload. This confirms previous findings that both complexity and controllers' activity measures have a unique contribution to workload ratings, and also confirms that the complexity components which are important in one air traffic sector may be trivial in the other one.

The impact of different experimental conditions on the measures of controller's activity and workload, as well as on complexity components was then assessed with further statistical analyses. The aim was to compare the effects that different conditions impose on these measures and identify those complexity components that mirror the changes in the controller's activity measures and workload measures. These analyses correspond to a test of the second hypotheses of this work that states that different components of complexity correspond differently to the workload experienced by the controller, and that understanding these differences can facilitate comparison of the complexity levels of a single sector under different conditions, but also comparison of complexity levels of different sectors under same conditions.

To collect the evidence in support of this hypothesis, the analysis of variance was performed by assessing the impact of a set of experimental factors (sector and traffic load) on the workload and controller's activity measures (ISA rating, frequency (R/T) occupancy time and controller intervention actions) on one side, and on the complexity components on the other side (more details on the results of ANOVAs are provided within the section 4.2.3). Result patterns of both series of ANOVAs were then compared. It was assumed that the complexity component scores that display similar effect patterns to those of the workload indicators may be considered as potential drivers or at least mediating factors in the generation of the workload effects in different sectors. Three different sectors were considered, and two different traffic loads (baseline and increased traffic load), in order to be able to compare the average values of the dependent variable across sectors and in different traffic conditions.

The results showed that the effect of the increasing traffic on workload, communication load (frequency (R/T) occupancy) and controller intervention activity is identical: for all three variables significantly higher values were recorded in the increased traffic load condition than in the baseline condition. The effect of traffic, however, was not the same in the sectors. Within the largest sector considered no significant effect of traffic load was found on any of the variables considered (workload, communication load and controller intervention activity, but also complexity components).

These findings are consistent with the results of the multiple regression analysis for the communication load (i.e. frequency occupancy time) and could be attributed to the behaviour demonstrated in the Sperandio's closed-loop model of ATCO's (Sperandio 1971). Similarly Loft et al. (2007) argued that air traffic controllers are in the loop with air traffic events, reacting to the consequences of his/her own proactive behaviour, where the link between task demands and the workload perceived by the controller is basically connected to the way in which the controller manages his/her own resources. This means that when the workload increases, controllers tend to use the more economic (time saving) strategies more often, they become more conservative, do things faster and act earlier. Actually, we can assume that for the controllers the workload is already quite high even under baseline conditions, and therefore with higher traffic load they are using more "economic" strategies to cope with increased number of aircraft under control. Here "economic" strategies would be those where less R/T communication is needed, less input required and, therefore, all three recorded variables are kept at a tolerable level for the controller.

The other two sectors also demonstrated different effects of the increased traffic: while in one of the sectors the higher traffic load resulted in the doubling of the recorded values for workload and the controller's activity measures, in the other sector, the recorded values mirrored the increment of the traffic.

Similarly, the results showed that traffic had a different impact on the complexity components in those two sectors. While in both sectors traffic seemed to affect aircraft vertical transitioning and conflict sensitivity, the horizontal proximity did not change under different traffic loads in one of the sectors. Furthermore, while all the complexity components were affected by the increment of the traffic in one sector, this was not the case for the other sector where the average values of some of the complexity components remained the same level for both levels of traffic load. Interestingly, one of the complexity components, that is conflict insensitivity, did not record any significant effect of any of the experimental conditions and, therefore, was excluded from further analysis as a trivial aspect of the complexity for this particular airspace.

Overall, the results of these analyses seem to bring evidence in support of our hypothesis, confirming that different components of complexity correspond differently to the workload experienced by the controller, and that understanding these differences can facilitate the understanding of differences in the complexity of different airspace sectors and also for the same sectors under different conditions.

Having an insight into these contributors to the workload experienced by a controller can greatly facilitate the introduction of any change envisaged for the airspace under consideration. Namely, in the current structure, whenever new procedures or new working methods are subject to possible deployment, the identified complexity components could support the estimation of the impact that those changes would impose on the workload of the controller and further on decision making processes. Additionally, the complexity components are also applicable in the validation of the new concepts and new technologies to be introduced in the system when designing simulation scenarios against which new concepts would be assessed. As also demonstrated by the analysis, the complexity of different sectors, or even different sector designs for the same airspace, could be compared and contribute to the improvement of airspace design.

Nevertheless, there are certain shortcomings of the findings obtained within this research that would require improvements and further studies. The first limitation of the current results is that they address the workload as the final outcome of the demand imposed on the controller, but without providing more light onto the relationship between controller's activity measures (other than communication load) and complexity components. Gaining an insight into this aspect of the contribution to the overall workload (as controller activity is mediating factor between complexity and the workload) would increase greatly the accuracy of the results and the precision in the prediction of the level of workload assessed by the controllers. Thus, it remains as a challenging opportunity for future research to further develop and improve the complexity measures identified here to allow for the adequate identification of the correlation with the controller's activity measures. Additionally, the methodology developed here to derive the complexity components relevant for the en-route sectors could be used for guiding the identification of the complexity measure for the sectors of other air traffic control units (approach and TMA).

# Annex A - List of Complexity Factors

CALAN (Calculation Analysis) Output data:

1. number of the aircraft in the sector
2. number of the descending aircraft in the sector
3. number of the climbing aircraft in the sector
4. number of aircraft on the same route through the sector
5. number of aircraft with the heading change greater than 15°
6. number of aircraft with the speed change greater than 10 knots
7. horizontal proximity measure for the pair of aircraft *i* and *j*
8. vertical proximity measure for the pair of aircraft *i* and *j*
9. 3D Euclidean distance between aircraft *i* and *j*
10. number of aircraft with 3D Euclidean distance between 0-5nm
11. number of aircraft with 3D Euclidean distance between 5-10nm
12. number of aircraft with lateral distance between 0-25nm and vertical separation less than 2000ft/1000ft above/below 29000ft (FL290)
13. number of aircraft with lateral distance between 25-40nm and vertical separation less than 2000ft/1000ft above/below 29000ft (FL290)
14. number of aircraft with lateral distance between 40-70nm and vertical separation less than 2000ft/1000ft above/below 29000ft (FL290)
15. The measure of complexity associated with the mean weighted horizontal separation distance is defined as:

$$C_5 = \frac{N}{\sum_{1 \le i \le N} \left( \frac{\sum_{1 \le j \le N} W_{ij} d_{ij}}{\sum_{1 \le j \le N} W_{ij}} \right)};$$

where $N$ is the number of aircraft within the sector airspace, $d_{ij}$ is the horizontal separation distance between the two aircraft *i* and *j* (in nautical miles) and $W_{ij}$ is the associated weighting factor.

The rationale for using the inverse of the mean weighted distance is that decreasing mean distance results from reduced separation between neighboring aircraft.

$W_{ij}$ is defined as:

$$W_{ij} = \frac{[j \neq i]}{d_{ij}^2 + S_h^2 h_{ij}^2 + [j = i]}$$

where $h_{ij}$ is the vertical separation distance between the two aircraft $i$ and $j$ (in feet) and $S_h$ is the scaling factor for making the altitude separation distance comparable to the horizontal separation distance. At altitudes above (below) 29,000ft the horizontal separation minimum is 5nm and the vertical separation is 2000ft (1000ft) therefore the scaling factor $S_h$ = 5/2000 ($S_h$ = 5/1000)can be used. The purpose of weighting is to reduce the contribution of aircraft that are further away horizontally and vertically from $i^{th}$ aircraft under the consideration. Thus, the expression in equation (1) inside the parenthesis describes the weighted average horizontal separation between the $i^{th}$ aircraft and the neighbouring aircraft with the bias towards the neighbouring aircraft.

The logical operator [ ] in equation (2) takes a value of unity when the enclosed logical expression is true, otherwise it takes a value of zero.

16. The vertical proximity complexity measure:

$$C_6 = \frac{N}{\sum_{1 \leq i \leq N} \left( \frac{\sum_{1 \leq j \leq N} W_{ij} h_{ij}}{\sum_{1 \leq j \leq N} W_{ij}} \right)}$$

17. The complexity measure as the inverse of the average minimum horizontal separation between two aircraft is defined as:

$$C_7 = \frac{\sum_{1 \leq i \leq N} [j \in J_i]}{\sum_{1 \leq i \leq N} \min_{j \in J_i} \{d_{ij}\}};$$

where $J_i$ is set of aircraft that are within a $\Delta h$ vertical neighbourhood about the aircraft $i$:

$$J_i := \{ j \mid h_i - \Delta h / 2 \leq h_{ij} \leq h_i + \Delta h / 2;\ j \neq i \}$$

the numerator counts the number of aircraft for which at least one other aircraft is found within its altitude neighbourhood.

18. The complexity measure related to the average minimum vertical separation is defined as:

$$C_8 = \frac{\sum_{1\le i\le N} [j \in K_i]}{\sum_{1\le i\le N} \min_{j\in K_i}\{h_{ij}\}} ;$$

where a horizontal neighbourhood of radius *r* around aircraft *i* can be defined as:

$$K := \{ j \mid d_{ij} \le r; j \ne i \}$$

19. The measure that is based on the minimum separation for a pair of aircraft within the group for horizontal separation of aircraft within an altitude band:

$$C_9 = \frac{1}{\min_{1\le i\le N}\left\{ \min_{j\in J_i}\{d_{ij}\} \right\}}$$

20. The measure that is based on the minimum separation for a pair of aircraft within the group for vertical separation of aircraft within an altitude band:

$$C_{10} = \frac{1}{\min_{1\le i\le N}\left\{ \min_{j\in K_i}\{h_{ij}\} \right\}}$$

21. The variance of ground speed is defined as:

$$\sigma_{vg}^2 = \frac{\sum_{1\le i\le N} \left(V_i - \overline{V}\right)^2}{(N-1)} ;$$

where the mean ground speed is:

$$\overline{V} = \frac{\sum_{1\le i\le N} V_i}{N}$$

Low variance indicates less performance variation between the aircraft.

22. A complexity measure based on the variance and the mean of the groundspeed can be developed as the ratio of standard deviation to mean of the groundspeed (contrast ratio):

$$C_{15} = \frac{\sigma_{vg}}{\overline{V}}$$

23. if $s_{ij}$ is the distance between the *i* and *j* aircraft pair and $s_{xij}$ , $s_{yij}$ and $s_{hij}$ are Cartesian components of $\vec{s}_{ij}$ with respect to the reference frame attached to the *i* aircraft, then the range rate $\dot{s}_{ij}$ is given as:

$$\dot{s}_{ij} = \frac{\left(s_{xij} \cdot V_{xij} + s_{yij} \cdot V_{yij} + s_{hij} \cdot V_{hij}\right)}{s_{ij}}$$

The time-to –go to conflict $t_{ij}$ is : $t_{ij} = -\frac{s_{ij}}{\dot{s}_{ij}}$

And it is positive if the closing rate is negative. A negative closing rate indicates that the aircraft pair is moving closer to each other.

24. The weighting function $f\left(d_{ij}\right) = \frac{e^{-\alpha\left(d_{ij}\right)^2} + e^{-\beta d_{ij}}}{2}$ ; where α=0.002 and β=0.01 and $d_{ij}$ is expressed in nautical miles.

25. Density measure is defined as:

$$Dens = \sum_{j=1}^{N} f(d_{ij})$$

26. The variability in headings is defined as:

$$track\_disorder(i) = \sum_{j \neq i} \left|\theta_j - \theta_i\right| f\left(d_{ij}\right)$$

27. The variability in speed is defined as:

$$speed\_disorder(i) = \sum_{j \neq i} \left|\upsilon_j - \upsilon_i\right| f\left(d_{ij}\right)$$

28. The global divergence of the aircraft *i* is the weighted sum of all the divergence between the pairs of aircraft:

$$Div(i) = \sum_{\substack{j=1 \\ j \neq i}}^{N} 1_{R^+}(v_{ij}) \cdot |v_{ij}| \cdot f(d_{ij})$$

where R is a neighborhood distance and $1_{R^+}$ is indicator function of $R^+$, same as $1_{R^-}$ is indicator function of $R^-$.

29. The global convergence of the aircraft *i* is the weighted sum of all the convergence between the pairs of aircraft:

$$Conv(i) = \sum_{\substack{j=1 \\ j \neq i}}^{N} 1_{R^-}(v_{ij}) \cdot |v_{ij}| \cdot f(d_{ij})$$

30. if $\left\| \vec{\nabla} v_{ij} \right\|$ is measure of the change in term of relative distance when small modification is applied to the speeds and the headings of the aircraft involved.

$$\left\| \vec{\nabla} v_{ij} \right\| = \left\| \begin{pmatrix} \dfrac{\partial \left\| \overrightarrow{d_{ij}} \right\|}{\partial v j} \\ \dfrac{\partial \left\| \overrightarrow{d_{ij}} \right\|}{\partial v_i} \\ \dfrac{\partial \left\| \overrightarrow{d_{ij}} \right\|}{\partial \theta_j} \\ \dfrac{\partial \left\| \overrightarrow{d_{ij}} \right\|}{\partial \theta_i} \end{pmatrix} \right\|$$

$$\left\| \vec{\nabla} v_{ij} \right\| = \left\| \begin{pmatrix} a_x \sin(\theta_j) + a_y \cos(\theta_j) \\ -(a_x \sin(\theta_{ij}) + a_y \cos(\theta_{ij})) \\ v_j(-a_x \cos(\theta_j) + a_y \sin(\theta_j)) \\ v_i(a_x \cos(\theta_i) - a_y \sin(\theta_i)) \end{pmatrix} \right\|$$

and : $$a_x = \frac{\Delta x}{\sqrt{\Delta_x^2 + \Delta_y^2}} \qquad a_y = \frac{\Delta y}{\sqrt{\Delta_x^2 + \Delta_y^2}}$$

$$\Delta_x = x_j - x_i \quad \Delta_y = y_j - y_1$$

31. then the “sensitivity” indicators designed to set a weight on potential conflicts that are difficult to solve are:

$$Sd_+(i) = \sum_{\substack{j=1 \\ j \neq i}}^{N} 1_{R^+}(\upsilon_{ij}) \cdot \left\| \vec{\nabla} \upsilon_{ij} \right\| \cdot f(d_{ij})$$

32. and:

$$Sd_-(i) = \sum_{\substack{j=1 \\ j \neq i}}^{N} 1_{R^-}(\upsilon_{ij}) \cdot \left\| \vec{\nabla} \upsilon_{ij} \right\| \cdot f(d_{ij})$$

A situation with a high “sensitivity” is easier to resolve for the air traffic controller that one with a low “sensitivity”.  As these indicators “increase” with number of aircraft, it is unclear whether they actually are “complexity” or “simplicity” indicators. Therefore, the authors defined pair of indicators:

33. $ISd_+(i) = \dfrac{Div^2(i)}{Sd_+(i)}$

34. $ISd_-(i) = \dfrac{Conv^2(i)}{Sd_-(i)}$

# References

Albasman, M. and Hu, J. (2012). An Approach to Air Traffic Density Estimation and Its Application in Aircraft Trajectory Planning. Published in: Chinese Control and Decision Conference (CCDC), 2012 Taiyuan, China: 706-711,

Arad, B.A. (1964). The controller load and sector design. Journal of Air Traffic Control, May, 12-31.

Andreassi, J. L. (2007). Psychophysiology: Human behavior and physiological response. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Baart, D. (2001). Evaluation of Dynamic Density (DD) Metrics using RAMS. DOT/FAA/CT-TN, Atlantic City, NJ: FAA William J. Hughes Technical Center, ACT-540.

Banks, J. (1998). Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice, John Wiley and Sons, New York, NY.

Beaudouin-Lafon, M. and Mackay, W.E. (2002) Prototyping Development and Tools. In J.A. Jacko and A. Sears (Eds), Handbook of Human-Computer Interaction. New York: Lawrence Erlbaum Associates, pages 1006-1031. (Revised edition 2007)

Boyd, S. P. (1983, October 10-14, 1983). Assessing the validity of SWAT as a workload measurement instrument. Paper presented at the Proceedings of the Human Factors Society- 27th Annual Meeting, Norfolk, Virginia.

Brookings, J.B. and Wilson, G.F. (1994). Physiological and workload changes during a simulated air traffic control task. In Proceedings of the Human Factors and Ergonomic Society, 38th Annual Meeting.

Bruskiewicz, K.T., Hedge, J.W., Manning, C.A., and Mogilka, H.J. (2000, January). The development of a high fidelity performance measure for air traffic controllers. In C.A.

Manning (Ed.), Measuring air traffic controller performance in a highfidelity simulation. (Report No. DOT/FAA/AM-00/2). Washington, DC: FAA Office of Aviation Medicine.

Cain, B. (2007). A review of the mental workload literature.Technical Report, Defence Research and Development Canada Toronto.

Cardosi, K. M and Murphy, E.D. (1995). Human factors in the design and evaluation of air traffic control systems (DOT/FAA/RD-95/3). Cambridge, Massachusetts: U.S. Department of Transportation, Volpe Research Center.

Casali, J.G. and Wierwille, W.W. (1983). A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. Human Factors 25(6): 623-641.

Casali, J.G. and Wierwille, W.W. (1984). On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues. Ergonomics 27(10): 1033-1050.

Chaboud, T., Hunter, R., Hustache, J. C., Mahlich, S. and Tullett, P. (2000). Air Traffic Complexity: Potential Impacts on Workload and Costs, EEC Note No. 11/00, Project GEN-4-E2, Eurocontrol Experimental Centre, Bretigny, France

Chatterji, G. B., Shridhar, B. (1997). Measures of Airspace Complexity, Preliminary Draft and Unpublished Work, NASA Ames Research Center

Chatterji, G.B. and Sridhar, B. (2001). Measures for air traffic controller workload prediction. First AIAA Aircraft Technology, Integration and Operations Forum. Los Angeles, California.

Christien, R., Benkouar, A., Chaboud, T. and Loubieres, P. (2003). Air traffic complexity indicators and ATC sectors classification.Presented at ATM 2003.Budapest, 23-27 June, 2003.

Cohen, J., Cohen, P., West, S.G. and Aiken, L.S. (2003). Applied multiple regression/correlation analysis in the behavioural sciences (Third Edition). Mahwah, NJ: Erlbaum

Colle, H.A. and Reid, G.B. (1998). Context effects in subjective mental workload ratings. Human Factors, 40 (4), 591-600.15

Costa, G. (1993). Evaluation of workload in air traffic controllers. Ergonomics, 36(9), 1111-1120.

D'Arcy, J. F., and Della Rocco, P. S. (2001). Air traffic control specialist decision making and strategic planning – a field survey (DOT/FAA/CT-TN01/05). Atlantic City International Airport: Federal Aviation Administration William J. Hughes Technical Center.

Davison, H. J., Histon, J. M., Ragnarsdottir, M. D., Major, L. M., and Hansman, R. J. (2003). Impact of operating context on the use of structure in air traffic controller cognitive processes. Paper presented at the 5th Eurocontrol/ FAA ATM R&D Seminar, Budapest, Hungary.

Delahaye, D. and Puechmorel, S. (2000). Air traffic complexity: Towards intrinsic metrics. Presented at the 3rd FAA/EUROCONTROL ATM RandD Seminar. Naples Italy, 13-16 June.

Edwards, E. (1972). Man and machine: Systems for safety. In Proc. of British Airline Pilots Associations Technical Symposium, (pp. 21-36). British Airline Pilots Associations, London

Endsley, M. R. (1994). Situation awareness in dynamic human decision making: Theory. In R. D. Gilson, D. J. Garland, and J. M. Koonce (Eds.), Situational Awareness in Complex Systems. Daytona Beach, FL: Embry Riddle Aeronautical University Press.20

EUROCONTROL (1997). PD/1 final report annex A: experimental design and methods. Report
PHARE/NATS/PD1-10.2/SSR;1.1. Brussels, Belgium: Eurocontrol.

EUROCONTROL (2003), Review of Workload Measurement, Analysis and Interpretation

Methods, CARE-Integra-TRS-130-02-WP2 -1-0

EUROCONTROL (2004), EUROCONTROL ATM Lexicon

EUROCONTROL (2007). LINK2000+ small scale real-time simulation No.2 (LINK2000+ SSRTS2), final report (CRDS Note No.26), EUROCONTROL

EUROCONTROL (2009). CREDOS – Human Factors Case Report (D4-10), EUROCONTROL

EUROCONTROL (2010a). EUROCONTROL Long-Term Forecast: IFR Flight Movements 2010-2030, Document Identifier: CND/STATFOR Doc415, EUROCONTROL

EUROCONTROL (2010b).European Operational Concept Validation Methodology E-OCVM, 3rd Edition, February 2010,download from: http://www.eurocontrol.int/valfor/public/standard_page/OCVMSupport.html

EUROCONTROL (2011) http://www.eurocontrol.int/hifa/public/standard_page/Hifa_HifaMethods_EvalRTSim.html>

European Union (2004). REGULATION (EC) No 549/2004 OF THE EUROPEAN PARLIAMENTAND OF THE COUNCIL of 10 March 2004 laying down the framework for the creation of the single European sky, Official Journal of the European Union

Farmer, E. and Brownson, A. (2003). Review of workload measurement, analysis and interpretation methods, CARE-Integra-TRS-130-02-WP2, under the EUROCONTROL INTEGRA programme

Federal Aviation Administration (2011). Next Generation Implementation Plan, Federal Aviation Administration. March 2011. Retrieved April 16, 2012.

Flynn, G. M., Leleu, C. and Zerrouki , L. (2003). Traffic complexity indicators and sector typology analysis of U.S. and European centres. Technical report, Eurocontrol

Gaillard, A.W.K. (1993). Comparing the concepts of mental load and stress. Ergonomics 36(9): 991-1005.

Gaillard, A.W.K. and Kramer, A.F. (1999). Theoretical and methodological issues in psychophysiological research. Engineering psychophysiology: issues and applications. R.W. Backs and W. Boucsein. Mahwah, NJ., Lawrence Erlbaum Associates, Inc.: 31-58.

Gianazza, D., Guittet, K. (2006), "Evaluation of air traffic complexity metrics using neural networks and sector status", 2nd International Conference on Research in Air Transportation, Belgrade, Serbia.

Hart, S.G. and Stavelend, L.E. (1988). Development of the NASA-TLX (Task Load Index): results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.) Human mental workload.Holland: Elsevier.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 Years Later. Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, 904-908. Santa Monica: HFES.

Hilburn, B. and Jorna, P. (2001).Workload and Air Traffic Control. In PA Hancock and PA Desmond (Eds.) Stress, Workload and Fatigue: Theory, Research and Practice. Hillsdale, New Jersey, USA: Erlbaum)

Hilburn, B., (1996). The impact of ATC decision aiding automation on controller workload and human machine system performance.NLR Technical Publication TP96999. Amsterdam, Netherlands: NLR.32

Hilburn, B., Jorna, P.G.A.M. and Parasuraman, R (1995). The effect of advanced ATC strategic decision aiding automation on mental workload and monitoring performance: an empirical investigation in simulated Dutch airspace. In Proceedings of the Eighth International Symposium on Aviation Psychology, April 1995, Columbus, Ohio: Ohio State University.

Hilburn, B. (2004). Cognitive Complexity In Air Traffic Control – A Literature Review. Note No. 04/04, Eurocontrol Experimental Centre, Brétigny, France.

S. G. Hill, H. P. Iavecchia, J. C. Byers, A. C. Bittner, A. L. Zaklad, R. E. Christ (1992). Comparison of Four Subjective Workload Rating Scales, Human Factors: The Journal of the Human Factors and Ergonomics Society (August 1992), pp. 429-439.

Hill, J. C. , Archibald, J. K., Stirling, W. C. and Frost, R. L. (2005). A multi-agent system architecture for distributed air traffic control. In AIAA Guidance, Navigation and Control Conference, San Francisco, CA, AIAA-2005-6049.

Histon, J.M. and Hansman, R.J. (2002). The impact of structure on cognitive complexity in air traffic control (Report No. ICAT-2002-4)", MIT International Center for Air Transportation, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA

Histon, J., Aigoin, G., Delahaye, D., Hansman, R.J., and Puechmorel, S., (2001). Introducing Structural Considerations Into Complexity Metrics, Presented at the 2nd USA/EUROPE ATM RandD Seminar, Santa Fe

Histon, J. M., Hansman, R. J., Gottlieb, B., Kleinwaks, H., Yenson, S., Delahaye, D., et al. (2002, October). Structural considerations and cognitive complexity in air traffic control. Paper presented at the IEEE/AIAA Digital Avionics System Conference, Irvine, CA.

Hooijer, J.S. and Hilburn, B.G. (1996). Evaluation of a label oriented HMI for tactical datalink communication in ATC. NLR Technical Publication TP 96676 L. Amsterdam, the Netherlands: National Aerospace Laboratory NLR

Huey, F.M. and Wickens, C.D. (1993). Workload transition: Implications for individual and team
performance. Washington, DC, National Academy Press.

IATA - International Air Transport Association (2014). Air Passenger Market Analysis - December 2013. IATA Economics

ICAO - INTERNATIONAL CIVIL AVIATION ORGANIZATION (1989). Human factors digest no 1. Fundamental human factors concepts. Circular 216-AN/131. ICAO (Montreal, Canada)

ICAO - INTERNATIONAL CIVIL AVIATION ORGANIZATION (2007). Air Traffic Management-Procedures for Air Navigation Services (ATM-PANS, DOC 4444 ATM/501), Fifteenth Edition, ICAO (Montreal, Canada)

ICAO - INTERNATIONAL CIVIL AVIATION ORGANIZATION (2010). Environmental Report-Aviation and Climate Change Produced by the Environment Branch of the International Civil Aviation Organization (ICAO) in collaboration with FCM Communications Inc., Montreal, Canada

ICAO - INTERNATIONAL CIVIL AVIATION ORGANIZATION (2011). The Framework for Global Harmonization. Working Document for Global Air Navigation Industry Symposium (GANIS). ICAO Aviation System Block Upgrades

Interaction Design Foundation (2004). Mock-ups. Retrieved 30 April 2014 from http://www.interaction-design.org/encyclopedia/mock-ups.html

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Keppel, G., Wickens, T.D. (2004). Design and Analysis: A Researcher's Handbook (4th Edition). Pearson Prentice Hall, New Jersey.

Kirwan, B., Evans, A., Donohoe, L., Kilner, A., Lamoureux, T., Atkinson, T. and MacKendrick, H. (1997). Human Factors in the ATM System Design Life Cycle. FAA/Eurocontrol ATM R&D Seminar, 16 – 20 June, 1997, Paris, France.

Kirwan, B., Rodgers, M & Schaeffer, D. (2005). Human Factors Impacts in Air Traffic Management. Aldershot, UK: Ashgate.

Kopardekar, P. and Magyarits, S., (2003). Measurement and prediction of dynamic density.In Proceedings of 5th USA/Europe Air Traffic Management RandD Seminar. Budapest, Hungary

Kopardekar, P. (2000). Dynamic Density – A Review of Proposed Variables. FAA NAS Advanced Concepts Branch ACT-540: FAA.44

Kopardekar, P., Schwartz, A., Magyarits, S. And Rhodes, J., Airspace Complexity Measurement: An Air Traffic Control Simulation Analysis, International Journal of Industrial Engineering, 16(1), 61-70, 2009

Laudeman, I., Shelden, S., Branstrom, R. and Brasil, C. (1998). Dynamic Density: An Air Traffic Management Metric. Report NAS/TM-1998-112226. Ames Research Center, Moffett Field, California: NASA.45

Laurig, W.; Luczac, H.; Philipp, U. (1971) Ermittlung der Pulsfrequenzarrhythmiebeikörperlicher Arbeit, InternationaleZeitschriftfürAngewandtePhysiologie, 30, 40-51.

Leighbody, G., Beck, J., and Amato, T. (1992). An operational evaluation of air traffic controller workload in a simulated en route environment. 37th Annual Air Traffic Control Association Conference Proceedings (pp. 122-130). Arlington, Virginia: Air Traffic Control Association.

Leplat, J. (1978). The factors affecting workload. Ergonomics, 21, 143-149.

Loft S., Sanderson P., Neal A., Mooij M. (2007). Modeling and Predicting Mental Workload in En Route Air Traffic Control: Critical Review and Broader Implications. Human Factors, Vol. 49, No. 3, pp.376-399.

Magyarits, S., and Kopardekar, P. (2000). Dynamic density: The measure of air traffic control sector complexity. Atlantic City, NJ: NAS Advanced Concepts Branch, FAA William J. Hughes Technical Center.

Majumdar, A, Ochieng, W.Y. McAuley G, Lenzi, J.M. and Lepadatu C. (2004). The Factors Affecting Airspace Capacity in Europe: A Cross-Sectional Time-Series Analysis Using Simulated Controller Workload. Journal of Navigation, 57(3), 385-405.

Manning, C. and Stein, E. (2005). Measuring Air Traffic Controller Performance in the 21st Century. In B. Kirwan, M. Rodger, and D. Schaefer (Eds.) Human Factors Impacts in Air Traffic Management (pp. 283-316). Aldershot: Ashgate.

Manning, C. A. (2000). Measuring Air Traffic controller performance in a High-Fidelity Simulation (DOT/FAA/AM-00/2). Washington, DC: FAA Office of Aviation Medicine.

Manning, C.A., Mills, S.H., Fox, C., Pfleiderer, E., and Mogilka, H.J. (2001). Investigating the validity of Performance and Objective Workload Evaluation Research (POWER). (Report No. DOT/FAA/AM-01/10). Washington, DC: FAA Office of Aerospace Medicine.

Manning, C., Fox, C., Pfleiderer, E. (2003). Relationships between Measures of Air Traffic Controller Voice Communications, Task load, and Traffic Complexity. Presented at the 5th USA/EUROPE ATM RandD Seminar, Budapest, Hungary

Martin, L., Kozon, T., Verma, S. and Lozito, S. (2006). Using Neural Networks to Explore Air Traffic Controller Workload, NASA Ames Research Center Moffett Field, CA

Masalonis, A., Callaham, M., and Wanke, C. (2003). Dynamic Density and Complexity Metrics for Realtime Traffic Flow Management.Presented at ATM 2003.Budapest, 23-27 June, 2003.

Mills, S.H., Pfleiderer, E.M., and Manning, C.A. (2002). POWER: Objective Activity and Task load Assessment in En Route Air Traffic Control. (Report No. DOT/FAA/AM-02/2). Washington, DC: FAA Office of Aerospace Medicine

Mogford, R.H., Murphy, E.D., Roske-Hofstrand, R.J., Yastrop, G. and Guttman, J.A. (1994). Research techniques for documenting cognitive processes in air traffic control: sector complexity and decision making (Report DOT/FAA/CT-TN94/3). Pleasantville, New Jersey: CTA Incorporated.

Mogford, R. H., Guttman, J. A., Morrow, S. L., and Kopardekar, P. (1995). The complexity construct in air traffic control: A review and synthesis of the literature (DOT/FAA/-CT TN95/22). Atlantic City, NJ: FAA.

Mogford, R.H., Murphy, E.D., Guttman, J.A. (1999). Using knowledge exploration tools to study airspace complexity in air traffic control. The International Journal of Aviation Psychology 4(1) 29-45

Mogford, R.H., Murphy, E.D. and Guttman, J.A. (1993). The use of direct and indirect techniques to study airspace complexity factors in air traffic control. Proceedings of the First Mid-Atlantic Human Factors Conference (pp 196-202). Norfolk, Virginia: Old Dominion University.

Morrow, D. (1993). Collaboration in Controller-Pilot Communication.Reprinted in Kanki, B. (ed.) & V. Prinzo, (coed.): A-29–A-44.

Mou, PPY., Cho, A. and Histon, J. M. (2012). Managing Flight Pattern Visualization Complexity. Conference Paper. Retrieved from http://moo-mou.com/data/paper.pdf

Nygren, T.E. (1991). "Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload." Human Factors 33(1): 17-33.

O'Donnell, R. D. and Eggemeier, F. T. (1986). Workload assessment methodology. In K.R. Koff and J. Thomas (Eds.) Handbook of Perception and Human Performance, 2, Cognitive processes and performance, Wiley, New York.

Pavard, B. and Dugdale, J. (2006). The contribution of complexity theory to the study of socio-technical systems.In Unifying Themes in Complex Systems. Editors: Ali A. Minai and Yaneer Bar-Yam. Publishers: Springer. Book Chapter.

Pawlak, W.S., Brinton, C.S., Crouch, K. and Lancaster, K.M. (1996). A framework for the evaluation of air traffic control complexity. Presented at the AIAA national Conference.

Pfleiderer, E. (2005), Relationship of the aircraft mix index with Performance and Objective Workload Evaluation Research Measures and controllers' subjective complexity ratings (DOT/FAA/AM-05/16), Washington, DC:FAA

Rantanen, E. M., and Nunes, A. (2003). Taxonomies of measures in air traffic control research. Proceedings of the 12th International Symposium on Aviation Psychology (pp. 977-980). April 14-17, 2003, Dayton, OH.

Rantanen, E. M., Maynard P.W. and Özhan, D. (2005). The impact of sector characteristics and aircraft count on air traffic control communications and workload. Presented at the 13th International Symposium on Aviation Psychology, Dayton, OH

Reid, G.B., and Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock and N. Meshkati (Eds.), Human mental workload (pp. 185–218). Amsterdam: Elsevier.

Reynolds,T.G, Histon, J.M., Davison, H, and Hansman, R.J. (2002). Structure, Intent and Conformance Monitoring in ATC, Workshop on Advanced Technologies and their Impact on Air Traffic Management in the 21st Century (ATM 2002), Capri, Italy

Schaffer, M. (1991). The use of video for analysis of air traffic controller workload.Journal of Air Traffic Control, January - March, 35-37, 40.

Schmidt, D. K. (1976). On modeling ATC workload and sector capacity. Journal of Aircraft, 13(7), 531-537.

Schuen-Medwed, R., Lorenz, B. and Oze, S., 2007. Empowerment of the planning controller in CPDLC environment, In Proceedings of the Human Factors and Ergonomics Society, Europe Chapter Annual Meeting., Braunschweig, Germany

SESAR Consortium (2006). The Performance Target, SESAR Definition Phase- Deliverable 2, Issued by the SESAR Consortium for the SESAR Definition Phase Project co-funded by the European Commission and EUROCONTROL

SESAR Consortium (2008). SESAR Master Plan, SESAR Definition Phase- Deliverable 5, Issued by the SESAR Consortium for the SESAR Definition Phase Project co-funded by the European Commission and EUROCONTROL

Sirevaag, E. J., Kramer, A. F., Wickens, C. D., Reisweber, M., Strayer, D. L., & Grenell, J. F. (1993). Assessment of Pilot Performance and Mental Workload in Rotary Wing Aircraft. Ergonomics, 36(9), 1121-1140.

Sperandio, J.C. (1971). Variations of operators' strategies and regulating effects on workload. Ergonomics, 14, 571-577.

Sridhar, B., Seth, K. S., Grabbe, S. (1998). Airspace Complexity and its Application in Air Traffic Management, 2nd USA/Europe Air Traffic Management RandD Seminar

STATFOR, the EUROCONTROL Statistics and Forecast Service (2009). Mitigating the Challenges for Air Transport 2030, Publication reference number: 09/07/15-21

Stein, E. S. (1993). Workload for operators in complex person-machine systems.In M. Pelegrin and W. M. Hollister (Eds). Concise Encyclopedia of Aeronautics and Space Systems, Oxford England: Pergamon Press.

Stein, E. S. (1982). How hard is it to fly an airplane? Paper presented at the Annual Convention of the Eastern Psychological Association, Baltimore, MD.

Stein, E.S. (1985). Air traffic controller workload: An examination of workload probe. (Report No. DOT/FAA/CT-TN84/24). Atlantic City, NJ: Federal Aviation Administration Technical Center 2

Stein, E.S. (1992). Air traffic control visual scanning (FAA Technical Report DOT/FAA/CT-TN92/16). Washington, DC: US Department of Transportation.

Vogel, M., Schelbert, K., Fricke, H. and Kistan, T. (2013). Analysis of Airspace Complexity Factors' Capability to Predict Workload and Safety Levels in the TMA (Explorative Factor and Regression Analyses of Radar and Simulated Trajectory Data ), Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013)

Whittaker, R.A. (1995). Computer assistance for en-route (CAER) trials programme: future system 1 summary report. CS Report 9529. London: Civil Aviation Authority National Air Traffic Services.

www.atcmonitor.com

Wyndemere (1996). An Evaluation of Air Traffic Control Complexity, Final report contract NAS2-14284. Boulder, Colorado USA: Wyndemere.

Zeier, H. (1994). Workload and Psycho-physiological Stress Reactions in Air Traffic Controllers, Ergonomics, 37 (3): 525-39.

***Declaration of Authorship***

Except where reference is made in the text of this dissertation, this dissertation contains no material published elsewhere or extracted in whole or in part from a dissertation presented by me for another degree or diploma. No other person’s work has been used without due acknowledgement in the main text of the dissertation. This dissertation has not been submitted for the award of any other degree or diploma in any other tertiary institution.

The dissertation was written under supervision of Prof. Dr.-Ing. habil. Hartmut Fricke, Chair of Air Transport Logistics and Aviation of Dresden University of Technology.

I accept the doctorate regulations of the Faculty of Transportation and Traffic Sciences "Friedrich List" of Technische Universität Dresden.

***Versicherung***

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen wörtlich oder sinngemäß übernommenen Gedanken sind als solche kenntlich gemacht. Ich erkläre ferner, dass ich die vorliegende Arbeit an keiner anderen Stelle als Prüfungsarbeit eingereicht habe oder einreichen werde.

Die vorliegende Dissertation wurde an der Professur für Technologie und Logistik des Luftverkehrs der Fakultät Verkehrswissenschaften "Friedrich List" an der Technischen Universität Dresden unter der wissenschaftlichen Betreuung von Herrn Prof. Dr.-Ing. habil. Hartmut Fricke angefertigt.

Ich erkenne die Promotionsordnung an.

*Jelena Djokic*

Juni 2014.