EUROPEAN ORGANISATION
FOR THE SAFETY OF AIR NAVIGATION

EUROCONTROL

EUROCONTROL EXPERIMENTAL CENTRE

COGNITIVE COMPLEXITY IN AIR TRAFFIC CONTROL
A LITERATURE REVIEW

EEC Note No. 04/04

Project COCA - COmplexity and CApacity

Issued: March 2004

REPORT DOCUMENTATION PAGE

Reference: EEC Note No. 04/04	**Security Classification:** Unclassified
Originator: **EEC/ N**etwork **C**apacity and **D**emand	**Originator (Corporate Author) Name/Location:** EUROCONTROL Experimental Centre Centre de Bois des Bordes B.P.15 F - 91222 Brétigny-sur-Orge CEDEX FRANCE Telephone : +33 (0)1 69 88 75 00
Sponsor:	**Sponsor (Contract Authority) Name/Location:** EUROCONTROL Agency Rue de la Fusée, 96 B -1130 BRUXELLES Telephone : +32 2 729 9011

TITLE:

COGNITIVE COMPLEXITY IN AIR TRAFFIC CONTROL –
A LITERATURE REVIEW

Author	Date	Pages	Figures	Tables	Annexes	References
B. Hilburn (STASYS)	03/04	vi+69	7	1	3	213

Contact	Project	Task No. Sponsor	Period
Geraldine M. Flynn +33 1 69 88 76 12	COCA – Complexity and Capacity	NCD-R-CY	2003/2004

Distribution Statement:
(a) Controlled by: Head of NCD
(b) Special Limitations: None
(c) Copy to NTIS: ~~YES~~ / NO

Descriptors (keywords):

ATC - Complexity - Workload - Traffic Density - Taskload - Cognitive - Model - Human Factors

Abstract:
This report reviews literature into air traffic control complexity. This work was carried out in the context of the Complexity and Capacity (COCA) project. This work reviewed past research (both theoretical and empirical) into ATC complexity, and its relationship to controller workload. In reviewing the forty-plus year history of work into ATC complexity, this effort identified:

- The major theoretical views concerning ATC complexity;
- Candidate complexity factors;
- Data collection methods for identifying, refining and validating a model of ATC cognitive complexity.

On the basis of this review, a functional model of ATC cognitive complexity is proposed that can help guide the next phase of the COCA work.
The overriding conclusion from this review was that despite the breadth and depth of previous work done into identifying ATC complexity factors, a good deal of work remains. Nobody, it seems, has yet managed to construct a valid and reliable model of ATC complexity that [1] moves substantially beyond the predictive value of simple traffic density alone, and [2] is sufficiently context-free. Further, it is proposed that COCA explore the development of non-linear techniques to refine and develop its model of cognitive complexity.

TABLE OF CONTENTS

GLOSSARY

ACC	Area Control Centre
ATC	Air Traffic Control
ATFM	Air Traffic Flow Management
ATM	Air Traffic Management
ATWIT	Air Traffic Workload Input Technique
CAASD	Center for Advanced Aviation System Development
CANAC	Computer Assisted National ATC Centre (in Brussels)
CCM	Cognitive Control Model
CDI	Control Difficulty Index
CDM	Collaborative Decision Making
CFMU	Central Flow Management Unit
COCA	Complexity and Capacity
CRI	Convergence Recognition Index
DEL	Dynamic Elements of Load
EAM	European Airspace Model
EATMP	EUROCONTROL Air Traffic Management Programme
EEC	EUROCONTROL Experimental Centre
EEG	Electro-Encephalogram
EMG	Electro-Magnetogram
EOG	Electro-Oculogram
ERP	Event related Potential
GSR	Galvanic Skin Response
ICAO	International Civil Aviation Organisation
IMC	Instrument Meteorological Conditions
IPME	Integrated Performance Modelling Environment
ISA	Instantaneous Self Assessment
LOA	Letter of Agreement
MEG	Magneto-Encephalography
MIDAS	Man-Machine Integrated Design and Analysis System
MDS	Multidimensional Scaling
NCD	Network Capacity and Demand
NDM	Naturalistic Decision Making
OE	Operational Error
PCA	Principal Components Analysis
POSWAT	Pilot Objective/Subjective Workload Assessment Technique
POWER	Performance and Objective Workload Evaluation Research
PSF	Performance Shaping Factor
RT	Radio Telephony
SWAT	Subjective Workload Assessment Technique
TFM	Traffic Flow Management
TLI	Task Load Index
TLX	Task Load Index (NASA)
RAMS	Reorganised ATC Mathematical Simulator
RVSM	Reduced Vertical Separation Minimum
WOODSTOCK	Wide Object-Oriented Data Standard Traffic Observable Complexity Knowledge

EXECUTIVE SUMMARY

Controller workload is likely to remain the single greatest functional limitation on the capacity of the ATM system). One of the key factors contributing to controller workload is air traffic *complexity.* It is thought that improved measures of ATC complexity could benefit, for instance, evaluation of ATM productivity, benchmarking cost effectiveness, assessment of the impact of new tools and procedures, and airspace redesign. To date, research into air traffic complexity has tended to overlook cognitive aspects of the controller's task. EUROCONTROL's Network Capacity and Demand (NCD) Business Area has undertaken the Complexity and Capacity (COCA) project, in part to address this shortcoming.

The chief goal of COCA is to identify, develop and evaluate factors related to air traffic control complexity, and to validate and test complexity factors and identify those linked with controller workload and sector capacities. Within this broad goal, COCA has specifically set out to:

- Build a model of air traffic complexity;
- Apply this model to a comparison of complexity factors;
- Build a model of sector capacity;

This paper is a review of literature into cognitive complexity in Air Traffic Control (ATC). This work reviewed past research (both theoretical and empirical) into ATC complexity, and its relationship to controller workload. In reviewing the forty-plus year history of work into ATC complexity, this effort identified:

- The major theoretical views concerning ATC complexity;
- Candidate complexity factors;
- Data collection methods for identifying, refining and validating a model of ATC cognitive complexity.

On the basis of this review, a functional model of ATC cognitive complexity is proposed that can help guide the next phase of the COCA work.

It is clear that the relationship between complexity and workload is an indirect one that is highly mediated by the influence of many individual characteristics. Whilst this poses obvious difficulties for ever fully capturing the notion of cognitive complexity mathematically, it would be overly pessimistic to conclude that the human factor must remain an unknown. In fact, the literature on human factors makes it clear that there are known aspects of human cognitive functioning (regarding attention and decision making, memory, and perception) that can/could be incorporated into a predictive model of cognitive complexity.

The overriding conclusion from this review was that despite the breadth and depth of previous work done into identifying ATC complexity factors, a good deal of work remains. Nobody, it seems, has yet managed to construct a valid and reliable model of ATC complexity that [1] moves substantially beyond the predictive value of simple traffic density alone, and [2] is sufficiently context-free. Further, it is proposed that COCA explore the development of non-linear techniques to refine and develop its model of cognitive complexity.

1. OVERVIEW

Controller workload is likely to remain the single greatest functional limitation on the capacity of the ATM system (Majumdar & Polak, 2001). Given predicted traffic increases, as well as corresponding developments in ATC procedures and technologies, it is increasingly necessary to understand the abilities of controllers and to identify the “safe” limits of workload. One of the key factors contributing to these limits is air traffic *complexity.* It is thought that improved measures of ATC complexity could benefit, for instance, evaluation of ATM productivity, benchmarking cost effectiveness, assessment of the impact of new tools and procedures (e.g. Collaborative Decision Making, or CDM), and airspace (re)design (Mills, 1998). To this end, EUROCONTROL’s Network Capacity and Demand (NCD) Business Area has undertaken the Complexity and Capacity (COCA) project.

1.1 The COCA Project

The chief goal of COCA is to identify, develop and evaluate factors related to air traffic control complexity, and to validate and test complexity factors and identify those linked with controller workload and sector capacities. Within this broad goal, COCA has specifically set out to:

- Build a model of air traffic complexity;
- Apply this model to a comparison of complexity factors;
- Apply this model to airspace design;
- Build a model of sector capacity;
- Apply this model to sector classification; and
- Apply this model to ATFM validation of airspace design.

To date, research into air traffic complexity has tended to overlook cognitive aspects of the controller’s task. In the context of the COCA project, EUROCONTROL has recognised the need to incorporate such cognitive aspects of air traffic complexity. The aim of the current activity is to ensure a model of traffic complexity that both adequately captures those cognitive aspects, and is compatible with EEC’s WOODSTOCK and COLA modelling tools.

1.2 The Current Literature Review

As the first step in this activity, a literature review was recently conducted into ATC cognitive complexity, with an eye toward:

- Documenting current theoretical views on the role of complexity in ATC operations;
- Cataloguing and evaluating past research into cognitive complexity indicators for ATC; and
- On the basis of the above, to provide a work plan for capturing cognitive aspects in the project’s developmental Complexity Index.

1.2.1 Organisation and scope of this report

This report is organised into SEVEN chapters. The remaining chapters shall, roughly in turn, set out to:

- Review the concept of complexity as it has been applied in various domains;
- Summarise theoretical views on complexity and the related notion of workload, from the perspectives of human factors, information processing and cognitive psychology;
- Provide an overview of past research, and summarise ATC complexity indicators that have been used to date (along with case studies of several research projects);
- To summarise lessons from the literature regarding both data collection and data analysis methods that can be applied within COCA;
- Based on the preceding review, to propose a functional model of cognitive complexity that will facilitate the project's ongoing modelling work; and
- To summarise lessons learnt and draw specific conclusions for the COCA project.

Annex A to this report provides a tabular overview of complexity indicators by project. This covers indicators both actually used and merely suggested. Further, Annex A does not claim to cite all studies in which a given indicator was used/mentioned. In some cases, this was either not clear from the literature source, or was insufficiently elaborated/defined. Where available, primary references and brief methodological notes are provided. Annex B provides a similar tabular overview of data collection methods that seem potentially useful for COCA. Annex C presents a summary of potential workload indicators.

1.2.2 Literature sources

This review relied on a combination of government and contractor technical reports, scientific journal articles, book chapters and operational reviews. Primary sources were used as much as possible. Relevant references were drawn primarily from the fields of ATM research, human factors and cognitive psychology, and computer modelling. Though focused on cognitive complexity, this review necessarily borrowed heavily from theoretical and empirical work in the related field of workload assessment. The majority (but certainly not all) of the identified literature on ATC complexity factors originates from the US, primarily from NASA, FAA and affiliated contractors.

A number of seminal reports and articles on the subject of ATC complexity have been produced in recent years, including several in-depth reviews of complexity factor literature. This report relies heavily on those works, most notably, the work of Mogford and colleagues in the US (Mogford et al, 1993; Mogford et al, 1994; Mogford et al, 1995; Rodgers et al., 1998); Work by NASA into the concept of "Dynamic Density" (Laudeman, 1998; Kopardekar 2000); Airspace complexity work by the Wyndemere Corporation (Wyndemere, 1996; Pawlak et al. 1996); and Majumdar & Ochieng's (2000) study of ATC workload factors. Whilst this report does not claim to be an exhaustive review, it is thought to capture and synthesise the major theoretical, empirical and operational perspectives on ATC complexity.

2. COMPLEXITY DEFINED

2.1 What is Complexity?

The Merriam Webster dictionary defines complexity (2) as the state of being "hard to separate, analyse, or solve…," and that would seem to agree with most people's intuition. Pelz et al (2001) suggested that the terms complex and difficult are colloquially synonymous. Other cited synonyms have included *complicated, intricate, difficult* and *involved*.

Cilliers (1998), in a fascinating scientific and philosophical analysis of complexity, distinguishes between "complex" and "complicated" systems. According to Cilliers, if a system consisting of a huge number of parts or elements (and thereforee complicated) can be given a complete description, it is not by definition "complex." By this reasoning a supercomputer or jumbo jet is complicated, but not complex. In a complex system, interaction between elements of the system is such that the nature of the whole cannot be determined by analysis of some subset. Cilliers (1998) cites human brains, natural language and social systems as examples of complex systems, and identifies the following characteristics that define such systems:

- A large number of elements whose interaction defies analysis by traditional mathematical means;
- Dynamic interaction between elements, that involves transfer of energy and/or information;
- Redundancy that permits some subset of the system to carry out the function of the whole;
- Localised autonomy and lack of information sharing between all elements;
- Non-linear interactions between elements, which makes it possible for small perturbations to have large effects.

One could be forgiven for reading the above characteristics, and assuming that Cilliers (1998) was referring to ATC in particular! In fact, one realises that ATC is not entirely unique as a complex human-machine system. Other examples of complex systems include Emergency Management, C^3I systems (Worm, 2001), Nuclear Power Generation, and Maritime systems (Perrow, 1999), a group that Fields et al. (1998) termed "distributed cognitive systems."

Dörner (1995, in Schaefer, 2001) added *opacity* to Cilliers' (1998) list of complex system characteristics—that is, not all system variables can be directly observed. A controller, for instance, may ask pilots for performance data on their aircraft, but must still guess about the influence of weather (cf. Schaefer, 2001).

In the literature on ATC complexity, surprisingly few definitions of "complexity" appear to have been given (Schmeidler & D'Aanzo, 1994), presumably because the authors assume common knowledge. Meckiff et al (1998) defined complexity as a "…measure of the difficulty that a particular traffic situation will present to an air traffic controller…" Meckiff et al. (1998) went on the describe workload as

> *"…a function of three elements, firstly, the geometrical nature of the air traffic; secondly, the operational procedures and practices used to*

handle the traffic and thirdly, the characteristics and behaviour of individual controllers (experience, orderliness etc.)…."

It is this third element (which, of the three, seems closest to the notion of cognitive complexity) that has thus far proven most difficult to mathematically formalise. Mogford et al. (1995) defined complexity as "a multidimensional concept that includes static sector characteristics and dynamic traffic patterns," and noted that the concept is subjectively defined by controllers.

2.2 System Engineering Approaches to Complexity

In natural systems, the concept of complexity is closely tied to the notion of "entropy" which has been used to describe the state of "disorder" in a system (as in the second law of thermodynamics, which states that entropy of a system can only increase, as a system naturally evolves in energy state toward randomness). The notions of system complexity and entropy have also been applied more generally to man-made systems, in particular the analysis of information transfer systems. This breakthrough owes much to the work of Shannon (1949), who developed a mathematical theory of information transfer that formed the basis for *information theory*. According to this view, information content (literally, the system's ability to reduce uncertainty) can be mathematically described in terms of its entropy (expressed in binary units, or bits, of information). Within such a system, a completely predictable message would be said to have zero entropy (randomness), and therefore zero complexity. In a sense, this view captures the nature of latent information within a system, as do thermodynamic views of energy state. According to information theory, a complex system is one in which randomness (and therefore uncertainty) is high. The quantification of this concept is easy to grasp, even at the level of an individual information transfer (i.e. message). Selecting the letter "A" from a possible 26 letters transfers 5 bits of information, namely the maximum number of binary split half decisions ("is the chosen letter in the upper or lower half?") one would need to determine a randomly-chosen letter from the entire alphabet. This concept clearly underlies the working of current day computers, and has also been a useful one in the evaluation of man machine systems (Johannsen and Rouse, 1979) to evaluate, for instance, team communications, and human eye scan behaviour (Harris, Glover & Spady, 1986). The notions of entropy and information theory have even been used to evaluate information transfer in honeybee and ant behaviour (Michelson, 1993). In ATC, the notion of entropy has been applied to the predictability of traffic arrival location (Mehadhebi 1996, cited in Christien et al., 2003) and also the general dispersion of traffic (Histon & Hansman, 2002).

Over the years, various other system engineering and mathematical approaches (e.g., control theory, utility theory, and queuing theory) have been used to model behaviour in complex human-machine systems. In general, such approaches have attempted high level descriptions of the task, to *normatively*[1] model human behaviour. Curry and Gai (1976), for instance, modelled the human as a failure detection system that employs a Kalman filter to eliminate uncertainty due to perceptual noise. Optimal control models (cf. Sheridan and Ferrell, 1974) liken the human to a state estimator who must perceive a signal in against background noise (again, perceptual), generate a control output, compare the output to a desired goal state, etc.

[1] *Normative* models (which describe optimal behaviour) are distinguished from *descriptive* models (which attempt to capture actual behaviour).

Of the various system engineering efforts at modelling human-machine interaction, at least two have focused on ATC. Schmidt (1978) used queuing theory to analyse ATC workload. Unlike some other system engineering approaches, queuing theory does not view the operator as some comparator between current and desired state (a sort of glorified thermostat). Rather, it defines performance in terms of task completion times. On the basis of operational ATC data (collected from Los Angeles en route controllers), Schmidt derived a mathematical expression that showed a reasonable ability to predict average delay and server occupancy as a function of demand.

Hunt and Zellweger (1987) suggested that the ATC system can be described in terms of closed loop control theory, with a planning function involving a set of goals to define the future state of the system, a controlling function to identify deviations and formulate actions needed to return the system to normal, and a communication function within the controlling function that conveys these actions to the aircraft. Finally, a data management function is employed to collect, record and distribute information. The model does not appear to have been empirically evaluated.

Despite the successes of past system engineering approaches in modelling human-machine interaction, there are some remaining shortcomings. Humans are not, after all, a set of servomotors and filters and comparators that seek to optimise behaviour in some normative way. They are subject to biases and limitations in terms of perception, decision making and attention that can not always be captured a priori. As Johannsen and Rouse (1976) noted, such system engineering attempts to analyse human behaviour in complex systems have tended to disregard: The task environment per se; Human adaptability to fluctuating task demands; and human limitations in analytic thinking. Rouse (1980) noted that the difficulties that have beset system engineering approaches to human-machine interaction have tended to fall into one of three broad categories:

- *Measurement difficulties*— can take several forms. First (and as noted several times throughout this report) it is not possible to infer inner process on the basis of observable behaviour. Second, measures might be noisy due to non-repeatability, the time varying nature of human behaviour, or individual differences. As a result, modelling attempts to capture such unobservable processes as decision making or planning have often been forced to rely on average or aggregate measures of behaviour, thereby losing precision.

- *Non-uniqueness of constrained optimality*—powerful modelling techniques are best at modelling normative, or optimised, behaviour. It is clear from evidence in various fields (including economics (Kahneman, Slovic & Tversky, 1982) and emergency management (Klein, 1993) that humans do not “optimise” but “satisfice”—that is, they tend to follow a utilitarian approach that settles for a solution that is “good enough” but not necessarily the best available. In multi-stage decision making tasks (in which each decision point is dependent on previous decisions) the sequence is therefore hard to predict.

- *Pervasiveness of task environment*—an analytic modelling of human behaviour can only be relevant if all task environments are common (Simon, 1969). Rouse (1980) noted that human behaviour is difficult to predict in the absence of a context.

After roughly half a century of analytic modelling, and system engineering approaches to human-machine interaction, many of the same criticisms seem to persist: Niessen et al (1999) argued that quantitative attempts to capture human cognition still largely fail to adequately capture the cognitive parameters of coping with high workload. It is in response to this recognised shortcoming that the COCA project explicitly set out to address the notion of cognitive complexity.

3. COMPLEXITY IN ATC

The consensus view among the ATC research and operational communities is that complexity drives controller workload, which in turn is thought to ultimately limit sector capacity[2] (Christien, Benkouar & Chaboud, 2003, Majumdar and Ochieng, 2000). Whereas Mogford et al (1995) claim that "ATC complexity generates workload," Athenes et al (2002) noted that "the functional relationship between the two is largely unknown." In any event, research into ATC complexity has been inextricably intertwined with the notion of workload, as discussed in the next chapter. Given the apparent consensus on the definition of "Complexity" in ATC, one might be surprised to learn that research into the factors underlying ATC complexity has run a long, and still inconclusive, course. In fact, the earliest clear research reference to ATC complexity and associated factors dates back nearly 40 years (cf. Arad, 1964), nearly to the beginning of the ATC era itself.

Traditionally, traffic density has been the single factor most associated with complexity. However, it is increasingly clear that density by itself is an insufficient indicator of the difficulty a controller faces. Anecdotal evidence suggests that controllers increasingly speak not of the difficulty of a given traffic density, but of the associated traffic complexity (Kirwan et al., 2001). Past attempts to assess complexity have generally relied on geometric relationships between aircraft (Histon, 2000), or on observable physical activity (Pawlak et al., 1996). Increasingly, it is being recognised that complexity factors can interact (Fracker & Davis, 1990) in non-linear ways (Majumdar & Ochieng, 2000; Athenes et al, 2002), and that individual differences between controllers can mean that different controllers respond differently to the same constellation of complexity factors (Mogford et al., 1994). These are among the considerations that seem to be driving the search for a better way to describe and predict complexity as it affects the controller.

Traffic complexity seems to have been studied in the context of five main areas:

- The occurrence of operational errors or incidents (Stein, 1985; Grossberg, 1989; Mogford et al, 1995; Breitler et al, 1996; Rodgers et al., 1998; Rodgers & Nye, 1993; Gosling, 2002);
- Controller workload (Arad, 1964;Grossberg, 1989; Redding, 1992; Athènes, S., Chaboud et al., 2000);
- Conflict risk (Knecht, Smith & Hancock, 1996; Schmidt, 1976; Arad, 1964) and
- Controller decision making (Mogford et al., 1994);
- The design of decision support and flight planning tools (Leal de Matos, 1998; Schaefer et al., 2001; Masalonis et al., 2003).

[2] Interestingly, the implicit intervening factor between workload and capacity—namely *controller error*—has thus far shown little relationship to traffic complexity (Stager & Hameluck, 1990; Gosling et al., 2002), although the data may be inadequate in this area (Rodgers, et al., 1998).

Whereas the goals, methods and theoretical bases often differed across the areas mentioned above, there was substantial overlap in how complexity was defined and measured across studies. To help identify potential complexity factors, a wide net was therefore cast across literature from these various areas.

Addressing the complexity literature presented some differences in terminology in how complexity is defined, measured and used. For the sake of coherence, the remainder of this report will adopt the following conventions:

- CONSTRUCT- is a high level characteristic of the system (e.g. Complexity, Safety, etc.);
- FACTOR- is a qualitative expression of that indicator (e.g., Traffic Density); sometimes referred to as “METRIC”;
- DEFINITION— also known as an *operationalisation*, this is specification of how the construct will be concretely measured and expressed (e.g., aircraft per sector volume);
- UNIT—is the quantitative expression of that definition (e.g., number of aircraft per km^3).

3.1 Correlates of ATC Complexity

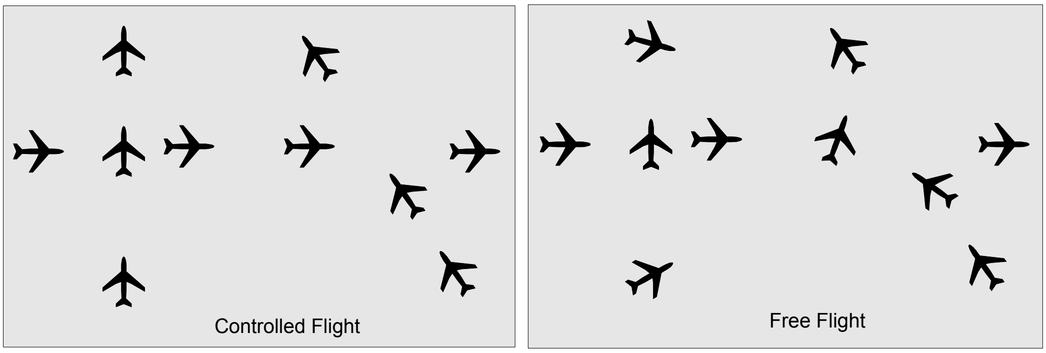
3.1.1 Traffic density

It seems from the literature that no single traffic characteristic has been as cited, studied and evaluated as has traffic density (whether it is termed traffic count, density, or traffic load) in terms of its influence on complexity and controller workload. The body of literature seems at the same time to praise the concept of traffic density (as the best available indicator of complexity), and to criticise it (mainly on the theoretical grounds that it does not capture the richness of what controllers find complex (Kirwan et al., 2001; Mogford et al., 1995; Athenes et al., 2002)).

This ambivalence is built on a long history of experimental and operational evidence. Davis (1963) showed that both communication time and manual performance time increase with traffic density. Arad (1964) derived a formula for controller workload based on observational data that related the number of aircraft (together with sector size, average traffic speed, rules of separation (in NM), and tightness of organisation) directly to controller load. Hurst & Rose (1978) found that density accounted for 53% of variance seen in activity and workload in enroute controllers.

A simple diagram (below) illustrates how complexity can vary independent of traffic density. This diagram compares flight with (left) and without airway route structure (right (van Gent et al, 1997)). Ten aircraft are presented in the same locations in each diagram.

In the figure on the right, headings have been altered for four of the ten aircraft (assume, for the sake of simplicity, that all aircraft are level at the same altitude). It is obvious at a glance that the task of monitoring for conflicts and predicting where such conflicts might occur, is made more difficult if traffic flow is made only slightly less organised. This figure, by the way, also highlights one of the major potential cognitive hurdles to be overcome if mature free flight is to be achieved, namely, the removal of current day route structure would greatly increase the complexity of the controller's task and would, in particular, make it much more difficult to detect conflicts.



Traffic flows with route structure (left) and without (right). After van Gent et al. (1997).

It has been noted that traffic density is not only an important driver of complexity, but also correlates well with conflict rate. (EUROCONTROL, 2002b). It does not, however, seem to correlate highly with the number or extent of altitude transitions (Chaboud et al, 2000). Further, a single indicator of density may not accurately capture the traffic pattern over time. Traffic volumes that fluctuate wildly over time (say, low routine traffic punctuated with one or two pronounced rushes per day) are more likely to generate conflicts (and appear complex to the controller) than is a sector of uniform traffic flow. (Chaboud, 2000).

Density itself can be defined in different ways. It can, for instance, be the average number of aircraft present in a fixed airspace over some defined period of time (Hilburn, 1996); It can also be the average density encountered by each flight (Chaboud et al, 2000).

Ideally, a complexity indicator should apply independent of such factors as equipment sophistication, traffic volume, or size of the controlled airspace (Chaboud et al, 2000). As several have suggested, however, this is seldom the case (Koros et al., 2003; Kirwan et al., 2001).

3.1.2 Incidents / operational errors

Several studies have focused on the role of air traffic complexity in ATC errors (alternatively referred to as Incidents, Losses of Separation, AirProxes, or—as they are somewhat benignly termed in the US—Operational Errors). Surprisingly, the operational evidence does not convincingly make the case to link the two. Grossberg (1989) found a relationship between complexity (as defined by FAA Order 7210.46) and errors in Chicago enroute airspace. Reviews of ATC incidents in Canada (Stager & Hameluck, 1990; Stager, 1991) concluded that most errors occurred during low or moderate traffic load and normal traffic complexity. Similar data have emerged from studies of US ATC operational errors (Kinney, 1977; Redding, 1992). In Europe, it should be noted, the 2000 mid-air collision over Germany occurred on a clear, quiet night.

Of the various complexity indicators evaluated by Breitler et al. (1996), most showed very low correlations with operational errors. The three that showed significant correlations were: Number of Letters of Agreement (LOAs) associated with a sector (r=.162); Number of airports and fixes in a sector (r=.219); and subjectively-rated "ease of transitioning" (r=.231). Rodgers & Nye (1993) classified the severity of errors, and found no correlation between either the number of aircraft or traffic complexity and error incidence. Kirwan et al (2001) reviewed the occurrence of UK airspace conflicts for 1997, and found that traffic volume (or rate of increase), communications, and procedures were the factors most cited. Somewhat surprisingly, weather and airspace design were not cited in the incident reports as causal factors (though the authors note potential reporting biases).

Rodgers et al. (1998), noting the lack of convincing operational evidence tying complexity and error, argued that there <u>is</u> a logical link between the two, and speculated that the data are to blame for the weak confirmatory evidence. In fairness to this view, it does seem that the prevalence of light traffic conditions might skew some of the data..That is, it is likely for much of the world that light air traffic prevails (perhaps with pronounced rush periods) and that high traffic might only appear for, say, 10% of each day. For this reason, it might be more appropriate to control for this pattern statistically when examining error. It is not clear that the research community has fully considered this option.

3.2 The Cognitive Nature of the ATC Task

The ATC system has two often-cited, and rather self-evident, goals. First, to ensure adequate separation between airborne aircraft, and second to expeditiously move aircraft through a fixed airspace. In a sense, beyond adherence to the ICAO mandated separation standards of 5 nm and 1000 feet (under RVSM), there are very few other constraints on how a controller should handle air traffic. Gosling (1987; 2002) noted that ATC problems are often complex and ill-defined, in that ATC represents a large *solution space*– that is, ATC accommodates any number of successful strategies within the basic system constraints (Cardosi & Murphy, 1995). ATC represents a probabilistic environment (Leroux, 1992), in which improper behaviour does not necessarily lead to a negative outcome (Reason, 1988). Because many different strategies can be used to reach the same acceptable outcome, criterion measures of ATC performance have proven elusive (Hopkin, 1980; Stein,1987).

Danaher (1980) noted that, in carrying out the functions of ATC, the controller must perform a variety of tasks. These include:

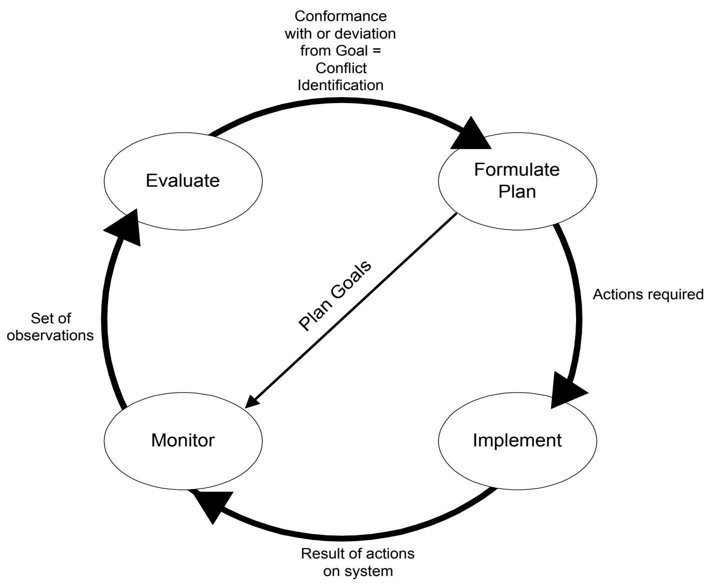
- Observing aircraft (either directly or via computer-generated displays);
- Operating display controls;
- Making data entries;
- Processing and updating flight progress information;
- Communicating with both aircraft and ground-based agents;
- Co-ordinating with co-workers; and
- Selecting/ revising plans and strategies.

A more detailed list of controller tasks was provided by Seamster (1993), who conducted a cognitive task of en route ATC:

- Maintain situation awareness—i.e. maintain understanding of current and projected positions of aircraft in the sector to determine events that require attention;
- Develop and revise sector control plan;
- Resolve aircraft conflict;
- Reroute aircraft;
- Manage arrivals;
- Manage departures;
- Manage overflights;
- Receive handoff;
- Receive pointout;
- Initiate handoff.

These ATC tasks have obvious implications for such aspects of human performance as visual perception (Day, 1994), monitoring (Thackray & Touchstone, 1989; Thackray, 1991), planning (Layton, Smith & McCoy, 1994), decision making (Amaldi, 1994), and memory (Stein, 1991).

While many other authors over the years have provided similar analyses of the tasks underlying ATC, a simple and more recent one is presented by Pawlak et al. (1996) who combined four major controller activities into a functional schematic, as shown below.



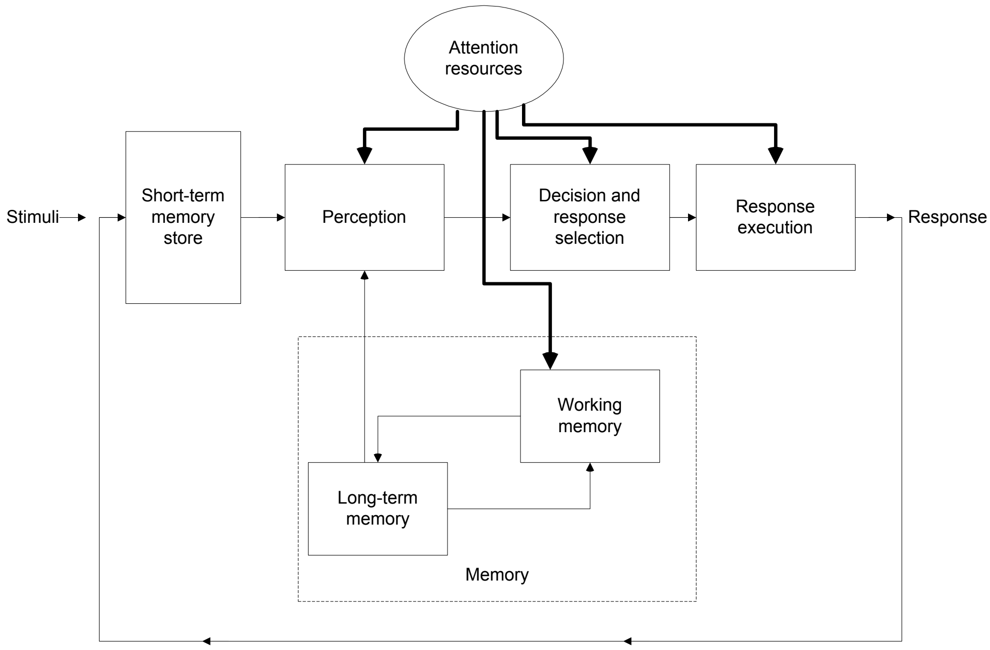
A model of the mental and physical processes required in ATC (after Pawlak et al., 1996)

Pawlak et al. (1996) defined four types of general tasks that controllers must perform. Of these, only the Implementation processes are observable (though the authors are careful to note that this is not always so—planned co-ordination can, in fact, be a form of implementation without observable action). Pawlak et al (1996) note that it is the other three processes—planning, monitoring, and evaluating—that combine to create mental effort for the controller. Smith et al 1992 (cited in Newman et al., 1993) similarly identified four types of cognitive processes in ATC: Perception; Planning; Control (i.e. selection of next behaviour) and Execution.

3.2.1 An information processing view of the ATC task

Notice that all approaches to modelling human-machine interaction assume a sort of analogy, whether it is the human as a optimal feedback device (Baron. & Levison, 1977), or as a pattern recogniser (Curry and Gai, 1976), or as a time-shared computer (Schmidt, 1976), etc. Whilst each approach has its utility, and has been applied in various applications, the model that seems currently favoured for ATC from a human factors perspective is that of the controller as an *information processor* (Mogford et al, 1994; Pawlak et al, 1996, Wickens et al., 1997). This seems a useful one to adopt, as it permits an inherent link to the cognitive processes that the developmental model of cognitive complexity is trying to incorporate. This section will briefly review some of the work into the information processing model of human machine interaction, focusing on ATC.

It has been useful (for both research and system engineering) to conceive of human cognition in terms of information processing steps, somewhat akin to the current-day computer. This model distinguishes the sequential steps of *Input* (the stimuli of, say, sound or vision), *Processing* (perception and decision processes) and *Output* (response execution). Overseeing these processes is the role of attention (cf. section 4.3), which can be deployed as required.



A model of human information processing (after Wickens, 1980).

This generic model, which is built up on decades of work dating back to the 1950s, has proven useful because it allows human cognition to be deconstructed in a way that permits the underlying processes (e.g., decision making, memory) to be studied individually. It provides the basis for a number of more specific models, such as the Cognitive Control Model (CCM, cf. Schaefer, 2001) that explore a part of the Input-Processing-Output chain.

3.2.2 The role of decision making in cognitive complexity

No review of the literature on cognitive complexity would be complete without considering the role of human decision making and, in particular, what is known about how humans recognise patterns, and make decisions, under demanding real world conditions.

Perhaps the most salient aspects of the ATC job are the highly skilled, yet routine and stereotyped nature of most required actions. Ironically, these are just the combination of factors that can lead to certain decision making biases. Decision making is generally defined as that process occurring midway between the acquisition/exchange of information and a subsequent action (Nagel, 1988). Decisions made *under risk* (when outcomes are associated with known probabilities) are often distinguished from those made *under uncertainty* (when outcome probabilities are unknown). It is assumed that most real-world decisions are made under uncertainty. A pilot electing to continue an instrument flight in deteriorating weather, for instance, is unaware of the a priori probability of a successful outcome (i.e., safe landing) for exactly such an aircraft, weather pattern, mental state of pilot, etc.

Normative (i.e., optimum performance) models of decision making maintain that a decision maker will act rationally (minimising loss, maximising profit) in goal seeking. Clearly, humans do not always act so rationally. For instance, human limitations in statistical estimation can introduce systematic decision making errors. From their observations of systematic biases in human decision making, Tversky & Kahneman (1974)[3] inferred the use of several major "heuristics," or strategies that lessen the cognitive burden of the decision making task, by narrowing the range of contingencies to be examined. Heuristics are "rules of thumb" that humans develop with experience, and which guide them in both problem recognition and decision making. In daily life, such individual heuristics normally serve each of us well. Because heuristics can colour our expectations, however, there are occasions when they can introduce decision making biases and sub-optimal strategies. Some of these sub-optimal decision making strategies include:

- *Bias against vagueness in subjective probabilities* (Baron, 1988)– people will consistently predict an event of known (albeit low) probability over one whose probability is unknown;
- *Biased risk aversion* (Tversky & Kahneman, 1974)– people are risk aversive for gains, yet risk taking for losses;
- *Memory availability bias* (Tversky & Kahneman, 1974)– memory vividness drives subjective probability;
- *Unwarranted confidence* (Tversky & Kahneman, 1974)– people show an unwillingness to reconsider options;
- *Inability to extrapolate growth functions* (Wagenaar & Segaria, 1975)– people underestimate growth in an increasing function;
- *Rejoice and regret* (Bell, 1982)– people consider both subjective probability and their anticipated reaction.

Further, empirical data suggest that decision making can also be impaired by stress (Broadbent, 1971; Hockey, 1986), which can lead to premature closure, non-systematic scanning, and temporal narrowing (Keinan, 1987). Decisions that rely on

[3] Kahneman was awarded the 2002 Nobel Prize in Economics for the application of their joint work to the field of economics (Tversky died in 1996, and the Nobel prize is not awarded posthumously).

working memory (as opposed to, say, retrieval from long term memory) are especially error-prone: Planning is likely to be shallower (Johannsen & Rouse, 1991) and consideration of alternatives more limited (Sheridan, 1987; Halford, Wilson & Phillips, 1998) under such conditions.

Unfortunately, domain experts seem just as susceptible (and in some cases, more so) to decision making biases that can colour how a situation is perceived, or data are estimated, or how thoroughly alternatives are considered. There has been extensive research overt the last 20 or so years into the notion of Aeronautical Decision Making (ADM (Jensen, 1992)). ADM describes the host of techniques available to the flightdeck crew (or single pilot) to minimise and mitigate the effect of judgement errors (e.g. continuing VFR flight into IMC (cf. Goh & Wiegmann 2002)). Techniques include risk assessment, stress management, interpersonal crew co-ordination and communication, etc.

The now nearly defunct traditional view of decision making was that humans acted rationally so as to optimise utility (the field grew out of, and borrowed much of its terminology from, the field of economics) by fully identifying the problem, considering alternatives, and acting so as to maximise gain. Increasingly (Dietsch 2001) emphasis is being placed in the notion of Naturalistic Decision making (NDM), which addresses decision making in real world conditions that are often characterised by risk, dynamic changing settings, feedback, shifting goals, and ambiguous information. In practice, this means that experts tend to base decisions on intuition (Dreyfus & Dreyfus, 1986) and quick recognition of a situation (Klein, 1993). For the expert, selection of action is then trivial. Plans need to be modified only if the situation later makes it clear that the original assessment was in error. Observational data from naturalistic settings (e.g., fire fighting (Klein, 1993)) suggest that experts spend more time deliberating over recognition of a given situation than they do on response selection, and are reluctant to reverse their initial assessment. Indeed, anecdotal evidence suggests that quick situational recognition, and confidence in judgements, are two characteristics highly valued in air traffic controller selection and training assessment. Unfortunately, there are times when hasty decisions and unflappable self-confidence can lead to disastrously unanticipated outcomes.

Bisseret (1981) found that air traffic controllers were more willing to detect a conflict and issue a corrective command), as the difficulty of predicting the future state of an ATC system increased. This finding has potential implications for advanced controller tools (e.g., MTCD), which might extend the time horizon for trajectory prediction and conflict detection capabilities beyond that of the human. As Bisseret's (1981) data would suggest, controllers in such future systems might be overly sceptical of machine-supplied strategic advice.

In summary, it is useful to bear in mind that cognitive complexity in ATC relies heavily on controllers' perception and recognition of the pattern ("ah, this is going to be a hard pattern"). Research into decision making biases strongly suggest that "normative" models of complexity overlook these decision making biases. Whereas these biases might be systematic in nature (e.g., controller's last traffic pattern might unduly colour recognition of a new pattern), their effect is likely very idiosyncratic in practice (e.g., each controller has a unique history).

Given the close relationship between ATC complexity and workload, it seems useful to now review some of the literature on the factors underlying workload, in particular the relationship between complexity, taskload and workload.

4. WORKLOAD IN ATC

This section reviews what is currently known about mental workload, particularly from the field of ATC. One of the few areas of almost universal agreement in the literature on ATC complexity is that complexity plays a major role in driving controller workload. Workload in ATC is generally mental, as opposed to physical, in nature. The following section will therefore briefly review what is known about mental workload.

4.1 Mental Workload

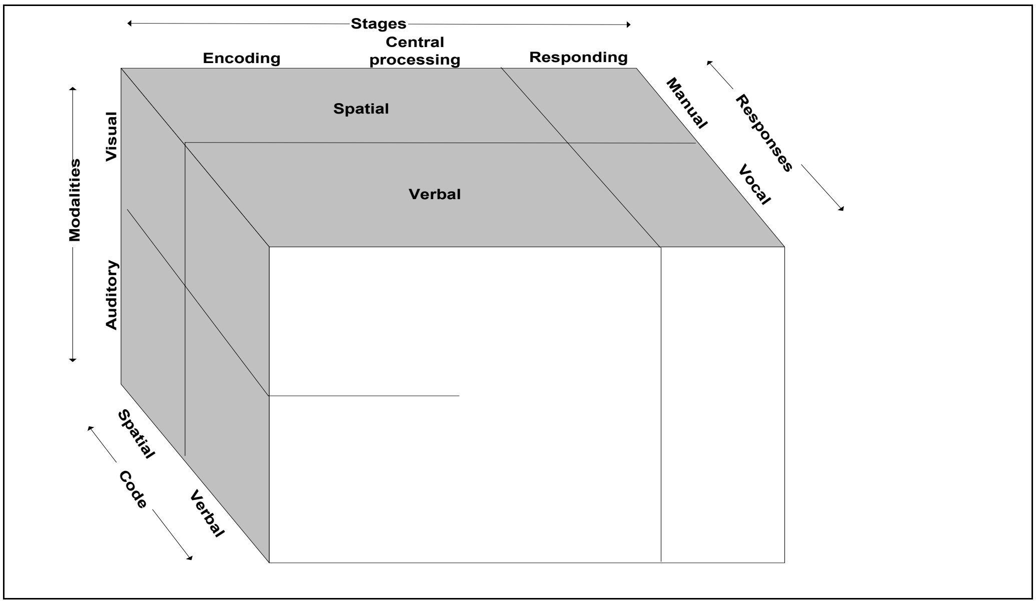
Interest in defining and developing metrics of mental workload has grown dramatically since the mid 1970s (Sanders & McCormick, 1987; Wickens, 1980). Most attempts to define mental workload have grown by way of analogy out of the concept of physical workload (Meshkati, Hancock & Rahimi, 1990).

The lack of a clear definition is reflected in the disagreement over appropriate metrics of mental workload. It seems generally agreed that mental workload is not a unitary, but a multi-dimensional concept (Leplat, 1978; Moray, 1979; Kramer, 1991), that taps both the difficulty of a task and the effort (both physical and mental) brought to bear (Gopher & Donchin, 1986). It therefore represents an interaction between task and operator, that can vary for different task-operator combinations (Leplat, 1978). Such factors as time pressure, noise, stress, and distraction can all influence the 'human costs' of performing a given task (Hancock & Chignell, 1988; Jorna, 1993). Aptitude, skill, experience, operating behaviours, and personality traits have all been cited as determinants of subjective workload in ATC (Bisseret, 1971; Sperandio, 1978). Clearly, the same given task might represent a reasonable amount of workload for an experienced operator, yet overtax a novice. The distinction is generally made between taskload (the objective demands of a task) and workload (the subjective demand experienced in the performance of a task).

Inherent in the notion of mental workload has been the concept that the human operator has a limited capacity to process information. Information processing models of the 1950s grew out of the field of communications engineering. Experiments into “dichotic listening” by Colin Cherry in the early 1950s demonstrated the difficulty humans have in dividing attention. In what has been termed the “cocktail party phenomenon,” people are able to attend to only one source of information, unless some salient stimulus— such as their own name—is broadcast. The notion of *channel capacity* was adopted to explain limitations of the information processing system.

Based on mounting evidence, theories of attention have been significantly refined over the years. The 1970s saw the emergence of “resource models” of attention, which postulate that all cognitive processes demand resources that are available only in limited supply. Whereas channel theories had assumed a structural bottleneck on information processing, resource models held that limitations were functional, in the form of attention or effort (Sanders, 1979). If task demands exceed available resources, performance declines. Conversely, if task demands fall short of supply, then the amount of residual resource provides a measure of spare mental capacity.

Perhaps the most widely accepted current model of human attention is the Multiple Resources model, as proposed by Wickens (1980). According to this model, tasks differ on the basis of demands they place in terms of: input modality (visual versus auditory); data input code (spatial /verbal) stage of processing (encoding/central/response); and response type (manual versus verbal) characteristics. This model is diagrammed below.



The Multiple Resources model of human attention (after Wickens, 1980).

According to this multiple resource model, the degree of interference between two tasks can be characterised by the tasks' compatibility along the dimensions outlined above. This model has proven useful in predicting what sorts of tasks (do they rely on spatial or verbal information? Do they require verbal or manual response? Is information auditory or visual?) will interfere with one another. Further, workload in this model is the demand that tasks place on a limited supply of attentional resource.

4.2 ATC Task Load Factors

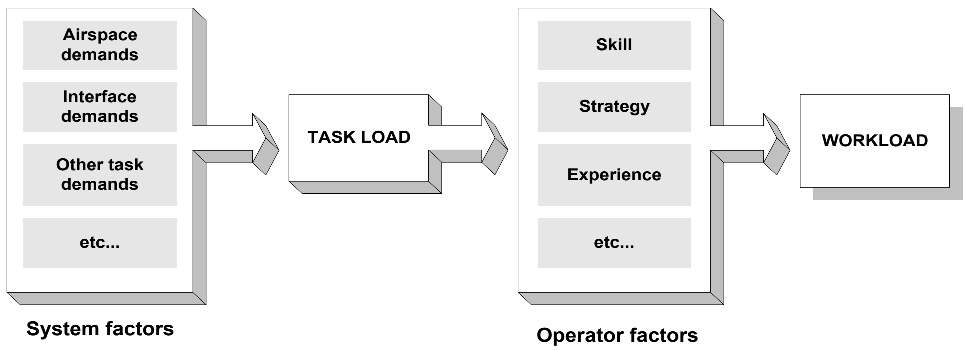
Task load (i.e., the demand imposed by the ATC task itself) is generally distinguished from workload (i.e., the controller's subjective experience of that demand). A number of studies have attempted to identify traffic-related workload factors for ATC. Of many prospective task load indices, the number-of-aircraft-under-control (i.e., traffic load) has shown the clearest predictive relationship to workload measures (Hurst & Rose, 1978; Stein, 1985). As with complexity, traffic density does not appear to fully capture workload. Some of the other (airspace-related) ATC task load factors include:

- Sector flow organisation (Arad, 1964);
- Number of traffic problems (Kalsbeek, 1976);
- Number of flight altitude transitions (Cardosi & Murphy, 1995);
- Mean airspeed (Hurst & Rose, 1978);
- Sector area (Arad, 1964);
- Mean aircraft separation (Arad, 1964);
- Aircraft mix (as it relates to differences in aircraft performance envelopes);
- Variations in directions of flight (Wyndemere, 1996);
- Proximity of aircraft and potential conflicts to sector boundaries (Wyndemere, 1996); and
- Weather (Scott, Dargue & Goka, 1991; Mogford et al., 1994).

Airspace factors are clearly not the only contributors to ATC task load. Such other considerations as the ATC position (e.g., oceanic versus terminal (Wickens, Mavor & McGee, 1997)), and the controller interface (including both the visual display and the data entry system) are critical in determining a controller's task load. They do not always do so, however, in predictable or beneficial ways. As research from other domains has demonstrated, a system's interface itself can impose additional task demands. For instance, automated tools can have the unintended effect of raising task load (Selcon, 1990; Kirlik, 1993). The potential for such situations appears increasingly likely as more sophisticated "advisory" types of decision aids emerge within ATC. By presenting the controller the additional tasks of (1) considering the system's advice, and (2) comparing the system's solutions to those he/she must continue to generate (if he/she is to remain "in the loop"), such decision aiding automation might paradoxically force an additional task upon the controller (Hilburn, Jorna & Parasuraman, 1995), or lead the controller to feel "driven" by the system (Whitfield, Ball & Ord, 1980).

4.3 ATC Operator Factors

The link between ATC task load and workload is a causative (albeit indirect) one, that is influenced by a number of internal factors. In the past, attempts to assess ATC workload have sometimes equated measures of direct task performance (such as time to perform discrete ATC tasks) with workload. Such observable workload (Cardosi & Murphy, 1995), however, provides only a partial picture of the workload experienced by a controller. For instance, a controller's observable performance cannot always convey the <u>cognitive</u> task demands—such as planning, decision making, and monitoring—imposed by ATC. Factors such as skill, training, experience, fatigue and other "stressors" all mediate the relationship between task demands and the workload experienced by a controller. Further, strategy plays an especially important part in determining a controller's workload. Notice that within very basic system demands, there are few constraints on how a controller should handle air traffic (Cardosi & Murphy, 1995; Leroux, 1992; Reason, 1988). As a result, the system can accommodate various control strategies without suffering a negative outcome (i.e., a loss of separation or, worse, a collision). How a controller chooses to prioritise tasks, or the compensatory strategies used to respond to workload fluctuations (e.g., shedding or deferring tasks, and deciding which tasks to handle first, or becoming more cautious in bad weather), all influence the controller's workload (Koros et al., 2003). The figure below (after Hilburn & Jorna, 2001) depicts a simplified schematic of the relationship between ATC task load and controller workload. Many other factors (e.g., time pressure, motivation, effort) are omitted from this figure. The literature review revealed a number of compensatory cognitive strategies that controllers use in accommodating changes in complexity and workload, and these are presented later in section 4.8.

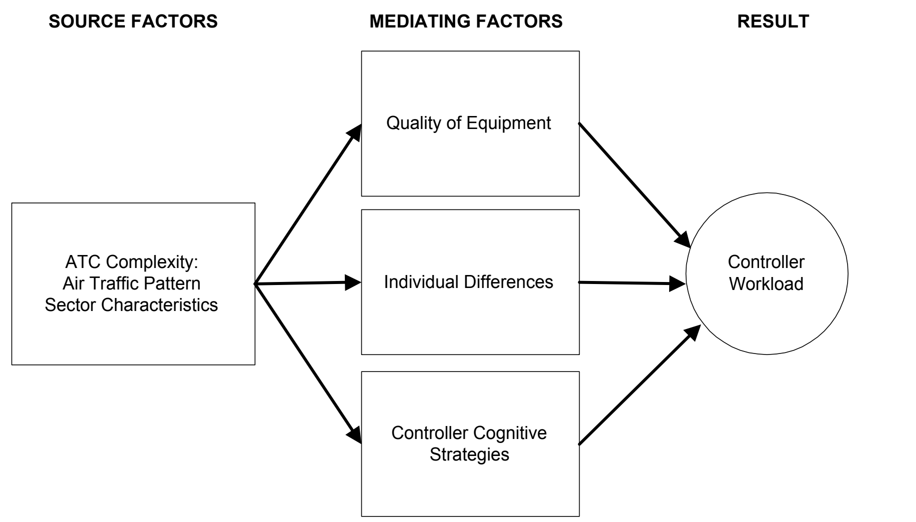


A model of ATC workload (after Hilburn & Jorna, 2001).

5. THE RELATIONSHIP BETWEEN ATC COMPLEXITY AND WORKLOAD

Over the years, various studies have shown a strong relationship between complexity factors and controller workload (Hurst & Rose, 1978; Stein, 1985; Grossberg, 1989; Laudeman et al., 1998). Mogford et al. (1995) reviewed a number of studies examining the effects of ATC complexity on workload and performance. Their resulting model of ATC workload links “source factors” (analogous to *the task load* factors of Hilburn & Jorna’s (2001) model presented in section 4.4.1.), through mediating factors (the *operator factors* of Hilburn and Jorna (2001)) to resulting workload. As shown below, Mogford et al. (1995) considered ATC complexity a task load factor (along with traffic and sector characteristics).

Not all researchers are sanguine that the relationship between complexity and perceived workload can ever be adequately expressed mathematically. Delahaye and Puechmorel (2000) suggest that the “complexity” of controller workload prevents its use as a concept. Their work, instead, set out to define a complexity indicator that specifically disregards controller workload.



The relationship between ATC complexity and workload (after Mogford et al. (1995).

5.1 Controller Compensatory Strategies for Handling Complexity

As noted, there are many operator factors that can influence the transformation from complexity to perceived workload. These can be loosely grouped in two groups: controller characteristics, and compensatory strategies. In trying to refine a model of cognitive complexity, both must be considered. Examples of the former would include skill, which can obviously mediate the perception of workload by shifting tasks from controlled to automatic processing (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977), as when a car driver ultimately learns to shift a manual transmission automatically (in the cognitive sense). Age is another characteristic that might mediate the workload function. At first glance, it would seem feasible to empirically evaluate the influence of controller characteristics such as age, years of experience etc., and to incorporate this information into the model. Likely to be more challenging, however, is the process of capturing and embedding in a predictive complexity formula the second group—namely, the strategies that controllers seem to use to respond to fluctuating task conditions, by changing the way in which they perceive or organise information. Such strategies are likely to be dynamic (Davison & Hansman, 2002), and idiosyncratic (Mogford et al., 1997).

Histon (2002) appears to have doe the most work (or at least brought the most attention to) the issue of how controllers respond to complexity through cognitive strategies.

5.2 Complexity and Workload: Some Distinctions

Several researchers have distinguished between characteristics of the airspace itself, and characteristics of the traffic per se, as they relate to complexity and workload (Arad, 1964; Mogford et al 1995; Majumdar & Ochieng, 2000). Grossberg (1989) made the distinction between static and dynamic aspects of the overall traffic pattern (cf. Rodgers et al., 1998). Static elements of the traffic include fixed sector characteristics such as sector size, placement of sector boundaries with respect to traffic flow, etc. Dynamic elements of the traffic pattern include the ephemeral aspects such as instantaneous traffic count, etc. A similar distinction was made by Chatterij and Sridhar (1998) who defined airspace complexity in terms of [1] structural elements and [2] flow complexity. Whereas the former relate to fixed geometric features of the airspace (e.g. number and intersections of airways), the latter refer to such as number and interactions of aircraft.

Arad (1964) distinguished three types of ATC task load in ATC: Background Load (the normal load associated with monitoring the screen, without any traffic present; Routine Load (resulting from the control of a standard aircraft); and Airspace Load (resulting from the separation of aircraft). Arad proposed the notion of Dynamic Elements of Load (DEL), as a means of creating a criterion against which ATC workload could be quantified. According to Arad, DEL consisted of the load represented by one standard, straight-and-level aircraft overflying the sector for one hour.

EUROCONTROL (2002b) has relied on a similar notion in developing the workload model for its Re-Organised ATC Mathematical Simulator (RAMS): ATC workload in this model derives from a combination of routine workload, conflict monitoring, and climb and descent monitoring.

Histon & Hansman (2002) claim that controllers rely on higher level organisations and conceptualisations of the traffic pattern in their perception of complexity. These three “structure-based abstractions” are: Standard flows, Traffic groupings; and Critical points.

Christien et al (2003) distinguished dynamic factors as being either flight related (e.g., number of aircraft, peak traffic hours, proximity to centre boundary) or interaction related. Interaction-related factors seem to capture some greater abstraction (Histon & Hansman, 2000) regarding the traffic flow, such as the flow entropy, or the time between conflict detection and resolution.

Chaboud et al (2000) distinguished between “real” and standard” conditions, with respect to workload and equipment. Whereas real workload results form the actual conditions (and limits the ability to compare across sites), “standard complexity” workload allows performance to be benchmarked using standard complexity parameters. So-called “standard equipment” workload evaluates complexity in terms of standard task duration, thus allowing cross-site comparisons independent of equipment differences. The authors argue that this characteristic makes it the most appropriate indicator for cross-site comparisons of complexity.

6. PAST RESEARCH

The preceding sections were meant to provide a brief theoretical background on the approaches that have been used to evaluate ATC complexity. In this course of this review, background on human factors aspects of cognitive complexity were highlighted, as well as some of the major themes, issues and debates within the body of literature. This section sets out to review the major empirical work that has been done, and to provide a list of candidate factors for evaluating ATC cognitive complexity.

It should be noted that literature sources differed in the depth of detail provided on indicators. This largely depended on the nature of the work (theoretical versus empirical). Where possible, details regarding definitions and measurement units were noted. Often (and especially in non-empirical work), however, indicators were mentioned only at a high level, without specification of either how they should be defined, nor how they should be measured (e.g. such indicators as "inadequate procedures" or route crossing or convergence points" (Christien et al., 2003) are mentioned as ATC complexity indicators only generally. In some cases, this was because a given article was theoretical in nature, or was a high level / summary review of work references elsewhere. For this reason, the accompanying list of Complexity Factors (Annex A) attempts to provide only primary references for each indicator.

6.1 Complexity Indicators

This section summarises selected research into the identification of ATC complexity indicators. Annex A to this report provides an overview of all identified complexity indicators.

<u>Davis (1963)</u> was perhaps the first to systematically examine the relationship between ATC complexity and controller workload. This study examined the influence on controller workload of the following factors

- Traffic density (defined as 50,65,80,100% of the actual studied sector)
- Complexity (defined as proportion of arrival and departure traffic to overflight traffic—30,50,70%)
- Number of airport terminals (1 versus 2)

The study found that workload (defined as total task time) responded to both traffic density and complexity (as defined in this study).

<u>Arad (1964)</u> examined the impact of various airspace factors on controller workload. Arad's work distinguished between background, routine and airspace load factors, and related these to conflict risk. Complexity factors included:

- Sector flow organisation
- Sector area
- Mean aircraft separation
- Climbing and descending aircraft
- Aircraft that have to be handed off vertically
- Aircraft that must be handed off to a terminal
- Pop up aircraft

- Frequency congestion
- Number of intersecting airways
- Number of military flights
- Mix of aircraft type
- Sector shape
- Presence of restricted airspace
- Proximity of sector boundary
- Time between conflict detection and resolution

A standardised airspace load factor was computed on the basis of data from 13 enroute centres. This study showed that routine controller load was affected by placement of sector boundaries in relation to main traffic flows. Arad noted that sector design could greatly influence controller workload. If sector boundaries are aligned with the main traffic flow, controllers enjoy more time and space. One criticism of Arad's work has been that the concept of flow organisation was left loosely defined (Rodgers et al., 1998).

Jolitz (1965) attempted to validate the workload formula of Arad (1964), but was unable to show that it could predict controllers' rated workload any better than could simple traffic count.

Schmidt (1976) developed a controller workload model based on the frequency of events that require a decision and action on the part of the controller. Schmidt's method relied on the execution time and frequency of observable tasks to calculate the so-called Control Difficulty Index:

$$CDI = \sum_{i} W_i E_i$$

Where: Wi = the event weighting based on task execution time
Ei= the expected number of events per hour

Potential factors from this study included
- potential conflicts between aircraft at route intersections
- potential overtaking manoeuvres on airways
- routine procedural events

The attractiveness of this approach lies in its implicit connection to the underlying decision making process. Unfortunately, it is not clear how to apply such a measure practically or cost-effectively. Whereas it is easy enough (albeit insufficient (cf. Pawlak et al 1996)) to count overt controller actions, it seems unlikely that one could ever feasibly tally the number of times a controller made a decision (for instance, how can one reliably determine in a "real world" setting when a controller has taken a decision not to take action?). Schmidt conducted field surveys to establish the weighting and frequency factors for various events. This led to the identification of the following as the most demanding events (in order of descending difficulty):
- Preventing a crossing conflict
- Preventing an overtake conflict
- Handoff
- Pointout (i.e. identifying a conflict or situation to another controller)
- Co-ordination with other controllers
- Pilot requests
- Traffic restructuring (i.e. rerouting)

Hurst and Rose (1978) explored the relationship between expert ratings of controller activity ("Pace ratings" (Kuhar, 1977)) and several factors, including

- hourly traffic
- peak traffic count
- communication time
- number of aircraft path changes

They found that traffic density accounted for 53% of the variance seen in activity and workload in enroute controllers.

Buckley et al. (1983) reported on two studies designed to establish the simulation and analysis feasibility of different techniques. In the first, he varied traffic density and sector geometry. Relevant factors included:

- number of aircraft
- fuel consumption
- time under control
- distance flown under control
- fuel consumption under control
- time within boundary
- path changes
- number of ground to air communications
- duration of ground to air communications

The second study collected a larger data set with a subset of the initial factor groupings. Results showed that sector geometry and traffic density interact for each of the workload measures. Factor analysis on the second set of data revealed four factors, which the researchers labelled *Confliction*, *Occupancy*, *Communication*, and *Delay*. In going from the first to second study, the researchers noted that the set of workload factors could be reduced with no appreciable impact on measurement accuracy, and with improved interpretability.

Stein (1985) conducted a simulation to explore the relationship between controller workload and

- total amount of traffic
- number of handoffs
- localised traffic density
- number of hand offs inbound
- number of hand offs outbound

Taskload was manipulated as the number of aircraft in the sector (low, medium, high) and localised traffic density (clustering). Regression analysis on ATWIT (subjective workload ratings) recorded once a minute showed that four factors (localised traffic density, number of handoffs outbound, total amount of traffic, and number of hand-offs inbound) together explained 67% of observed variance. This study showed the importance of localised traffic density on controller perceived workload: In his research, density was defined as the degree to which aircraft clustered within a small part of the airspace. This notion has intuitive appeal: it is not the overall airspace that matters, but the degree to which aircraft are squeezed together. As seems typical from the literature, there are various other ways to address the same notion of clustering.

Dynamic Density (approximately 1995 -). Some of the work into complexity was motivated by the concept of FreeFlight (RTCA, 1995) which has received increased attention over the last decade. Free Flight refers to the (partial or total) transfer of route selection and separation assurance authority from ground (ATC) to air (flightdeck). Closely coupled with the notion of free flight was the concept of Dynamic Density (Wyndemere, 1996; Laudeman et al 1998; Sridhar et al., 1998, Kopardekar, 2000; Kopardekar & Magyrits, 2002; Masalonis et al, 2003a; Masalonis et al., 2003b). Dynamic density was defined (Lauderman et al, 1996) as “a measure of control-related workload that is a function of the number of aircraft and the complexity of traffic patterns in a volume of airspace.”

In response to the RTCA’s Task Force 3 report regarding free flight (RTCA, 1995), a multi-organsation, multi-year research effort was initiated in the US to explore the development of a model that could incorporate traffic count and traffic complexity indicators in one. The aim was to benefit both the free flight initiative, and other ATM concepts such as dynamic resectorisation (Hadley et al., 2000; Wyndemere, 1996). This work was carried out over three phases, with partners including FAA, NASA, MITRE, Wyndemere (later Metron), and Titan Systems[4].

Using a number of variables to describe each traffic sector (eg. number of aircraft with 3-D Euclidean distance between 0-5 nautical miles excluding violations….), Dynamic Density was intended to permit the traffic-based workload of a controller to be predicted using objective information. It is speculated that this information can be used to predict highly complex situations in real time for up to 20 minutes before they happen, therefore allowing the additional demand to be met by preplanning and actions such as staff reassignment, alternate airspace configurations and modified traffic patterns.

The Dynamic Density notion was developed and validated operationally (Laudeman et al., 1998). A large data set was analysed using split half multiple regression (i.e., half of the data were used to set regression weights, and the other half were used to test these weights). The eight main factors contributing to Dynamic Density were (Laudeman et al., 1998):

- Heading change—number of aircraft making >15 degree heading change within 2 minute period
- Speed change—number of aircraft with an airspeed change of >10 kts (or .02 Mach) within a 2 minute period
- Altitude change-- number of aircraft making >750 ft altitude change within 2 minute period
- Minimum distance 0-5 miles-- number of aircraft with <5 nm. separation (in 3D space) to closest aircraft, within 2 minute period
- Minimum distance 5-10 miles—as above, for 5-10 nm. separation
- Conflict predicted 0-25 miles— number of aircraft predicted to be 1-25 nm (in 2D space) within next 2 minutes
- Conflict predicted 25-40 miles—as above, for 25-40 nm.
- Conflict predicted 40-70 miles—as above, for 40-70 nm

The results of the Dynamic Density composite measure showed that whereas traffic density by itself accounted for 33% of the explained variance in controller activity (r= 57), this was increased to 55% (r=.74) using the Dynamic Density equation. In 1998, the team presented the following regression weightings for each of the factors (higher weightings indicate higher predictive value):

[4] much of the work cited in this chapter (e.g., Wyndemere, 1996; Mogford et al., 1997) was conducted in support of the dynamic density research program.

- Heading Change (2.17)
- Conflict Predicted 25-40 (1.85)
- Conflict Predicted 40-70 (1.85)
- Minimum Distance 10 (1.18)
- Minimum Distance 5 (1.02)
- Altitude Change (.88)
- Speed Change (.15)
- Conflict Predicted 0-25 (.10)

Notice that as of 2000 (Sridhar, 2000), the Dynamic Density work had arrived at the following factor weightings (relative to a standard weighting of 1.00 for traffic density):

- Conflict (0,25) = 4.0
- Speed difference = 3.72
- Conflict (25,40) = 3.00
- Altitude change = 2.94
- Speed change = 2.45
- Heading change = 2.40
- Minimum distance = 2.45

Still later, a set of 33 complexity factors was agreed across the researchers. This list is not reproduced here, but can be found in Annex A.

A few shortcomings of the Dynamic Density work have been noted. First, factor weightings were applicable only in the sector in which they were collected and validated. The importance of Heading Change, for instance, was likely driven by the fact that the chosen sector had a large proportion of arrival traffic and vectoring (Laudeman et al., 1998). The researchers noted that more effort should be made to validate the results in other airspace. Second, the work still relied on observable behaviour as a criterion measure of controller workload, despite widely-acknowledged shortcomings of such an approach (cf. Pawlak et al., 1996). Third, Kopardekar and Magyrits (2002) suggested that the results of the Dynamic Density work could be extended and improved by further developing and testing the techniques using using non-linear techniques, including neural networks, genetic algorithms, and non-linear regression[5].

Notice that the work of the Dynamic density program is ongoing. Some of the studies reported elsewhere (e.g. the work of Chatterji & Sridhar, Wyndemere, and Kopardekar) were actually conducted under the umbrella of Dynamic Density. As of this writing (c.f. Masalonis, Callaham & Wanke, 2003), the program has yielded four separate complexity formulas, one from each of the main research teams.

Breitler, Lesko and Kirk (1996)— Cited in Kopardekar's (2000) review of the then-complete list of Dynamic Density factors, these researchers added the following relevant factors to the discussion:

- Number of sector sides;
- Number of main jet routes through a sector;
- Number of airports in the airspace;
- Numbers of Letters of Agreement (LOAs) for each sector;
- Number of entries (within last 15 minutes);
- Average altitude within the sector.

[5] based on personal communications with members of the dynamic density team, it seems that there has been interest in broader use of non-linear (e.g. neural network) analysis (cf. Chatterji & Sridhar, 2001).

Correlations with operational errors were quite low, the highest being subjectively rated "Ease of Transitioning" (r=.231, or less than 5% of explained variance). The only three significant correlations were found for ease of Transitioning, Number of fixes/airports; and Number of LOAs. In general, correlations coefficients were under 0.1 (i.e., less than .01 of variance explained).

Pawlak et al (1996), responding to the perceived over-reliance on behavioural indicators of ATC complexity, proposed a simple cognitive framework, involving the continuous cycle of formulating plans (evaluating the impact of a given control action), implementing the plans, monitoring the situation and evaluating the effectiveness of the plan (c.f. section 4.2). Using this framework as a basis, they developed an "approach for measuring and evaluating the perceived complexity of an air traffic situation, with an emphasis on the traffic characteristics that impact the cognitive activity of the controller". This involved interviewing a large number of controllers to sample their perceptions of complexity in a number of different situations. They identified the following factors as influencing perceived air traffic complexity:

- Special Use Airspace;
- Proximity of Potential Conflicts to Sector Boundary;
- Aircraft Density;
- Number of Facilities[6] ;
- Number of Aircraft Climbing or Descending;
- Number of Crossing Altitude Profiles;
- Variance in Aircraft Speed;
- Variance in Directions of Flight;
- Performance Mix of Traffic;
- Winds;
- Distribution of Closest Points of Approach;
- Angle of Convergence in Conflict Situation;
- Level of Knowledge of Intent of Aircraft;
- Separation Requirements;
- Co-ordination Required;
- Controller abilities;
- Equipment available.

Using the verbal protocol technique, controllers were asked to express their decision-making technique with respect to the traffic situation in a simulation environment. Following this, it was intended to produce a list of weightings for the perceived complexity produced by each factor and to validate this scale using simulations.

Chatterji & Sridhar (1997, cited in Kopardekar, 2000)— this (apparently-still) unpublished NASA Ames manuscript includes the following complexity factors:

- Maximum terrain elevation;
- Usual cloud ceiling;
- Volume of airspace available;
- Number of merging points;
- Number of neighbouring sectors that hand off traffic;
- Number of neighbouring sectors that accept handed off traffic;
- Number of sector operating procedures;
- Number of navaids available;

[6] This factor is not further defined, so it is not clear whether this refers to, for instance, number of adjacent centres.

- Aircraft mix;
- Sector boundary proximity;
- Reserved airspace proximity;
- Horizontal proximity;
- Vertical proximity;
- Time-to-go until conflict;
- Groundspeed variability;
- Shape of traffic geometry.

Of these, the first two seem less relevant to enroute control. It seems that (at least some of) these factors have by now made their way into the Dynamic Density index, as discussed shortly.

Pawlak & Brinton (1997) presented the results of this research in "free flight" simulations including controllers' rating of perceived complexity, subjective opinions of what additional technology or information the controllers would like to see to bring complexity to within "controllable limits" (e.g. timely intent information, accountability of the pilots in cases of separation violation, training etc.). However, whilst these reports present a possible framework, it does not fully present the relationship between the cognitive processes and how differing levels of complexity affect them, nor does it present the actual quantitative results of the weighting. For this reason, it is not yet possible to identify the (probably fuzzy) boundary between "controllable" and "uncontrollable".

Delahaye and Puechmorel 2000— responding to perceived weaknesses in the dynamic density and cognitive modelling approaches to complexity modelling, they search for intrinsic complexity factors, which derive directly from the location and speed of aircraft. These approaches fall into two classes: geometric and entropy-based. Based on their report[7], Kolmogorov Entropy for four types of convergences (again, lower entropy means higher predictability, and the lower bound for entropy is zero) were as follows:

- Parallel Flow 0
- Random Flow 8274
- Right Angle Crossing 64173
- Right Angle Crossing 487267

Notice that parallel flows, with an entropy of zero, are completely predictable. For the controller, there is no risk of conflict. What this author finds puzzling is that entropy should be so low for random flows—one would think that they represent by definition the greatest unpredictability.

EUROCONTROL (2000) reported on a comparison of EEC Bretigny and NATS UK approaches to evaluating traffic complexity. NATS relied on ATS output (i.e. service provided) and EEC used ATS workload as its criterion. WOODSTOCK provided the following complexity indicators

- Traffic counts;
- Traffic density;
- Number of sector entries per flight;

[7] There is a slight contradiction in the paper (between text and figure contents). As of this writing, the main author has not responded to a request for clarification.

- Number of conflicts;
- Measures of flight level changes;
- Average distance flown per aircraft;
- Traffic distribution by aircraft type.

Workload results (in terms of task frequency and duration) were assessed.

Workload ACC	Sector entries	Time per sector entry	Clearances	Time per clearance	Conflicts	Time per conflict
Real workload	Actual	Actual	Actual	Actual	Actual	Actual
Standard complexity workload	Standard	Actual	Standard	Actual	Standard	Actual
Standard equipment workload	Actual	Standard	Actual	Standard	Actual	Standard

The study concluded that "Standard Equipment Workload," which gives the workload assuming controllers had standard equipment and procedures (see the accompanying table) was the most appropriate way to assess complexity in European centres. The study also concluded that sectorisation heavily influences workload as computed by the EEC model.

Majumdar & Ochieng (2000) performed a series of multivariate statistical analyses (Multiple regression; Principal Components Analysis, and Factor analysis) on data from a total of 57 European sectors. Data were compared against European Airspace Model (EAM) workload. Among the 28 factors were:

- average instantaneous count;
- average navigational speed;
- bi-directional concentration;
- climb-cruise-descent profile;
- climb-cruise-flight profile;
- continuous descent profile;
- continuous climb profile;
- continuous cruise profile;
- cruise-descent flight profile;
- difference in upper and lower FLs;
- flights exiting in climb;
- flights entering from same ATC unit;
- flights entering from another ATC unit;
- flights entering in climb;
- flights entering in cruise;
- flights entering in descent;
- flights exiting to another ATC unit;
- flights exiting to same ATC unit;
- flights exiting in cruise;
- flights exiting in descent;
- flights in busiest 30 minutes;
- geographical concentration of flights;

- mean distance travelled;
- mean flight time;
- total climb flight time;
- total cruise flight time;
- total descent flight time;
- vertical concentration.

Results varied greatly across sectors. The study concluded that each sector has unique factors influencing its workload, and that new analysis methods might be warranted, including: weighted regression analysis, time series analysis, maximum likelihood analysis, spatially auto-correlated regression analysis, and non-linear regression.

Pfleiderer (2000) used Multidimensional Scaling (MDS)[8] to explore the notion of aircraft mix as a contributor to perceived workload. Aircraft mix is one of the often-cited (but often poorly-defined) complexity factors. In this study, controllers were asked to identify weight class, engine type and engine number, as well as provide estimates of cruise speed, climb rate, and descent rate, for a number of selected aircraft. The report concluded on the basis of MDS that controllers use a number of cues, obtained from a number of sources, to generate performance envelope stereotypes for aircraft. Results showed that these stereotypes are best captured through the use of:

- Aircraft engine type (i.e., piston, turboprop, jet) and
- Aircraft weight class (i.e., small, heavy, large)

Manning et al (2000; 2001) tried to identify objective (observable) measures of controller activity, as a proxy for controller workload. The researchers admit (Manning, 2000) the main limitation of such measures: although easy to systematically collect, they might not capture the richness of controller workload. Using a tool called POWER (Performance and Objective Workload Evaluation Research). Using two high altitude and two low altitude sectors, they compared POWER ratings with expert ratings of controller workload. Complexity factors were as follows:

- Number of controlled aircraft;
- Average control time;
- Heading, sped and altitude variation;
- Handoff count;
- Handoff time-to-accept;
- Alert count;
- Altitude changes;
- Data entry errors.

The researchers found that complexity was not related to any of the observable POWER measures. In particular, Manning et al. (2001) note that complexity, performance and workload were completely unrelated to: Time between hand-off initiation and acceptance; Number of data entry errors; Number of data entries; Number of route displays; Number of track reroutes; and number of strip requests. They suggested that certain POWER measures might have been related to sector static characteristics because of the structure of the chosen sectors. They further suggested that future work distinguish between static and dynamic sector characteristics.

[8] MDS can be thought of as a more general alternative to factor analysis

Schaefer et al (2001). Extending the work into Dynamic Density and the Tactical Load Smoother (TLS), this team reports on a study in which interviews with traffic managers were used to identify the following complexity factors in CANAC:

- Conflict position (i.e. is it close to sector boundaries?);
- Phase of flight;
- Height of conflict (a conflict in an upper sector is generally harder to solve due to the constrained airspace. The restricted number of available flight levels prevents vertical resolutions);
- Airspace structure (i.e. route definition and separation of flows);
- Weather;
- Traffic mix;
- Speed variation;
- Type of conflict (as it relates to time-to-go for resolution);
- Callsign density (i.e., callsign confusion potential: a large number of callsigns from the same carrier—e.g. KLM— creates potential misunderstanding and RT confusion);

This last factor, callsign density, appears a unique contribution of the team.

Kirwan, Scaife and Kennedy (2001) evaluated complexity in London area enroute and approach control. On the basis of group judgement exercises, they identified the top 12 complexity factors in UK airspace (for both enroute and approach) as follows:

- Volume/ flow/ growth rate of traffic - including the effect of 'bunching' of traffic at peak periods;
- Airspace design - including sector shape, number of levels in the sector, route structure and number of crossing points;
- Shared understanding - e.g. between adjacent sectors; between different ATM functions; etc.;
- Communications & co-ordinations - relating to time pressure on communications, both with radio-telephony (controller -pilot) and other calls on 'landlines';
- No soft option - lack of sufficient non-busy times to think and be proactive - related to so-called 'over-loads';
- Procedures - some procedures seen as overly complex or even clumsy, or with a high rate of change of sector design;
- Presentations to sector - handing over aircraft from one controller/sector to another (e.g. too early or too late, etc.);
- Human resources - relating to staffing issues;
- Non-standard flights - concerned with unusual traffic such as survey aircraft and training aircraft, for example;
- Aircraft performance - some areas have a large range of aircraft performance capabilities, from 'lows-and-slows' to fast-climbers;
- Military - largely relating to the unavailability of military danger areas for civil usage given the desire for increased civil capacity;
- Weather - e.g. turbulence, wind shear, thunderstorms, and weather interference for radio-telephony;

On the basis of this, the authors presented a complexity checklist to assist with airspace design. The 23 items from this checklist were as follows:

- have climb-throughs been minimised?
- Is there a maximum of one conflict point in the sector, being a 90 degree conflict point, and located away from the edge of the sector?
- Has number, location and orientation of holds been optimised?
- Has number of levels been optimised?

- Have intersecting corridors been reduced?
- Has traffic been standardised as far as possible?
- Is the sector long enough given the route profiles (e.g., long enough for climbs and descents)?
- Can a ‘dual carriageway’ concept be implemented where possible?
- Have opposite direction, head-on, and contra-flow traffic situations been avoided?
- Have speed limits and controls been used optimally?
- Has route complexity been minimised (e.g., not having routes at all points of the compass)?
- Have inbounds and outbounds been separated?
- Have conflicting departure patterns from two adjacent airports been avoided?
- Are associated procedures clear and straightforward?
- Are sector splits even?
- Have stepped SIDs been avoided?
- Have high performance SIDs been considered?
- Have funnelling and choke points been avoided/minimised?
- Have inbounds been based on reasonable and realistic aircraft performance profiles, rather than optimal ones?
- Have too many reporting points been avoided?
- Have environmental constraints been considered and their impact on complexity and workload minimised?
- Have knock-on effects onto adjacent sectors been calculated and assessed?
- Have plans adequately forecast the traffic increases, and will the design still work if these have been significantly under-estimated (e.g., by 20-40%)?

Chatterji and Sridhar (2001)— recognised the limitations of linear complexity formulations (i.e. single linear combinations of factors into a single regression formula). Their innovative approach therefore relied on the use of neural networks to non-linearly relate factors. The goal was to train the network using a sample of data, and validate performance of the network using another subset. Twelve hours of enroute data were analysed, and controller workload (rated on a 1-3 scale every 120 seconds) was compared to the influence of the following sixteen factors.

- Traffic density, proportional to the historical maximum for that airspace;
- Number of climbing aircraft, proportional to the historical maximum;
- Number of level aircraft, proportional to the historical maximum;
- Number of descending aircraft, proportional to the historical maximum;
- Average weighted horizontal distance between aircraft;
- Average weighted vertical distance between aircraft;
- Average minimum horizontal distance between aircraft;
- Average minimum vertical distance between aircraft;
- Minimum horizontal separation for an aircraft pair;
- Minimum horizontal separation for an aircraft pair;
- Number of aircraft within conflict timeframe;
- Average time-to-go to conflict;
- Smallest time-to-go to conflict;
- Groundspeed variation between aircraft;
- Groundspeed variation, proportionate to mean airspace groundspeed;
- The total conflict resolution difficulty based on time-to-go.

Inputs to the neural network were the sixteen complexity factors, and the output was the workload ratings. 80% of the entire data set was used to rate the network, the remaining 20% for testing the predictive value. To evaluate the impact of different factors (and combinations), 10 different sets of factors were used for training the network, from the simplest (traffic count only) to all sixteen. Validating, with a neural network, relies on feeding the trained network a new set of data, and to see whether the network can classify a new traffic pattern as low/medium/high (in this case) correctly (i.e., the same as a controller).

Results show that the predictive value of traffic count is useful mainly at low traffic levels. Their “simplest” trained network, relying only on traffic count, was able to correctly classify low workload patterns 99% of the time. For patterns rated medium and high, however, correct classification fell to only 54% and 11%, respectively. The authors speculated that this poor classification performance was a data artefact due to the small number of medium and high complexity training samples: Of the 2065 training samples, 1714 (83%) were low workload samples, 306 (15%) were medium, and only 45 (2%) were high. An alternative explanation of these results, however, centres on potential shortcomings in the traffic count measure itself. As other authors have speculated, the traffic count measure is a poor predictor of workload. It seems possible that predictive value of the measure would differ across the range of possible workload. For instance, whereas low workload might correlate well with traffic load, higher workload relates to more factors than simply number of aircraft on frequency.

Evidence favouring this second explanation was seen in the fact that classification performance improved significantly when the final six factors above (relating to conflict detection and resolution) were considered. The researchers concluded that the difficulty of detecting and resolving conflicts has a strong bearing on controllers' subjective workload. In summary, the neural network achieved its best classification performance when using the entire set of 16 factors.

The work of Chatterji and Sridhar (2001) seems notable for at least three reasons. First, it recognises the non-linear nature of the interaction between complexity factors, and so adopts a non-linear (neural network) approach to their analysis. Second, it explicitly recognises the limitations of using observable behaviour, rather than controller subjective response. Perhaps most importantly, however, it seems their work incorporates cognitive aspects into its choice of traffic geometry factors. For example, their use of (horizontal and vertical) minimum distance as complexity factors recognises that a single pair of proximate aircraft can force attention tunnelling, a demanding situation. Further, their work considered both the convergence angle and time-to-go (Chatterji & Sridhar, 1997) in defining complexity of pending conflict situations[9].

Histon et al (2001), Histon et al (2002) MIT researchers, working together with CENA/ENAC Toulouse, recognised that traditional geometric approaches to complexity have tended to overlook factors that contribute to complexity. Consistent with the work of Pawlak (Wyndemere, 1996; Pawlak et al., 1996; Pawlak & Brinton, 1997) they noted the importance of internal factors (e.g. controller strategy), as well as external factors (such as transient weather effects) on perceived complexity. Perhaps their most important contribution to the debate was the attention they paid to underlying structure, and the notion of structural abstractions, which they feel controllers use to understand and simplify a traffic pattern.

[9] The work of Chatterji & Sridhar (2001) might have overlooked the importance of controllers' subjective response to shallow convergence angles (c.f. section 10.7)

Based on live observations and focused interviews with controllers (terminal and en route) in US, Canada, and France, the team identified a series of complexity factors. Although there was no attempt to rank the factors, the team did note that the group of factors fell into three main categories, which they termed: *Airspace*, *Traffic*, and *Operational*. Of the many factors identified by the team, the following were those related to structural elements:

- Sector dimensions (shape, size, and "area of regard[10]");
- Spatial distribution of airways and navaids;
- Standard flows (number; Orientation relative to sector shape; Trajectory complexity; Interactions such as crossing points and merges);
- Co-ordination required with other controllers (Point outs; handoffs);
- Clustering of aircraft;
- Location of closest approach point in relation to sector border, merge point, etc;
- Sector transit time;
- Buffering capacity (with respect to traffic flow);
- Aircraft in holding pattern;
- Activation of Special Use Airspace;
- Noise abatement procedures.

Based on their analysis, the researchers created a generalised model of structure and complexity. This model noted the role of three simplifying abstractions (*standard flows, groupings and critical points*) that drive the controller's situation awareness (i.e., the ability to perceive, comprehend, and project future state of traffic).

The team's work into structural elements appears to be ongoing. Experimental validation of the Critical Points abstraction was carried out using a simplified ATC task (and university students) in which speed commands were used to merge traffic streams (Histon et al, 2002). As expected, proximity of critical points was associated with higher workload (using the Cooper Harper rating scale), number of separation violations, and average speed change per aircraft. Beyond validation work, the team also intends to incorporate the notion of structural abstractions into a complexity model. This effort too appears to be ongoing.

Athenes et al (2002)—As of this 2002 report, the team had not progressed to empirical work. Nonetheless, their approach seems valuable in that it explicitly responds to the following noted shortcomings of previous work, including

[1] An over-reliance on linear techniques, when the authors note the relationship between complexity and workload is a non-linear one;
[2] Disregard for auto-regulation of controller workload—that is, that changes in strategy (primarily related to time management) can allow controllers to modulate the impact of complexity drivers and, hence, perceived workload;
[3] Pre-occupation with problem solving, as opposed to the more relevant perceptual and decision making aspects of controlling, as sources of workload.

Similar to the earlier notion of Hancock, Chignell & Kerr (1988), they identify uncertainty and time pressure as two elements of a speed-accuracy trade-off controllers demonstrate in regulating workload, and which is captured in their proposed "Traffic Load Index" (TLI). In their view, TLI can be determined for any given aircraft as a function of time pressure (urgency) and uncertainty (which they

[10] Depending on the nature of traffic flows into and within the sector, the effective *area of regard* might extend well beyond the airspace boundaries (Histon et al., 2001) if, for instance, the controller must monitor traffic well up/down stream.

term "gravity," ranging from slight suspicion to evidence). Preliminary work suggests that TLI correlates better with NASA TLX (subjective workload ratings) and physiological workload indicators than does simple traffic count.

In summary, the work of Athenes et al (2002) is notable in that it is one of only a few (Histon et al, 2000 being another) that specifically addresses controllers' perceptual and decision making strategies, and how controllers themselves regulate the transformation of complexity and taskload into workload. Unfortunately for COCA, applicable results still seem some time off.

Kopardekar and Magyrits (2002)-- summarised the three phases of the Dynamic Density program. Phase I consisted of a Pilot Study to refine data collection procedures. Results of Phase I showed that controller and supervisor ratings were highly correlated (although supervisors tended to rate complexity higher), and that between-controller and between-supervisor ratings also correlated highly. In Phase II (the Data Collection Study), operational data were gathered from four en route centres (Atlanta, Cleveland, Denver, and Ft Worth). A total of 6480 complexity ratings were obtained. Phase III consisted of the Dynamic Density metric validation. This was done using regression analysis with the total Dynamic Density metric set. The focus was on the usefulness of both instantaneous and predicted (future) complexity.

Their overall conclusions were that the Dynamic Density metrics showed promise, and performed better than simple aircraft count (especially for predicted complexity). Two concrete suggestions were that [1] non-linear techniques be used with the data set, and [2] that fine tuning required more future experimental work to fine-tune weights, followed by [3] operational validation with a prototype metric.

Koros et al. (2003) recently reported on a field study of complexity in ATC towers. In the study, controllers from six control towers rated 29 complexity factors from the perspective of local and ground control. Factors were grouped into one of nine categories: Physical factors; Airspace characteristics; Weather; Ground operation; Equipment factors; Individual factors; ATC procedures; Distractions; and Training. Specific factors relevant to enroute operations are as follows:

- Traffic volume
- Aircraft performance differences
- Emergency operations
- Special flights (e.g., medical flights, helicopters, etc.)
- Overflights
- Unfamiliar pilots
- Pilots weak mastery of English
- Controller fatigue
- ATC procedures
- Equipment distractions (e.g. altitude alarms)
- Other distractions (not ATC related)
- On-the-job training

Results were found to be both site- and position-specific. Across all sites and across both positions, high traffic, frequency congestion, and runway/taxiway configuration were the leading complexity factors. Traffic density was found to influence workload (number of tasks, communication, co-ordination). Runway and taxiway configurations (some sites had more than 23 configurations possible) influence co-ordination and communication demands.

The Koros et al. (2003) study seems notable in at least three respects. First, it introduced several new factors to the literature. Second, it expanded the range of indicators to include training, pilot language facility, and distractions. Finally, whereas most of the literature concerns the enroute environment, this provides some candidate factors for evaluating ground and approach complexity.

Masalonis et al (2003)-- In two recent presentations (Masalonis, Callaham, Figueroa, & Wanke,2003; Masalonis, Callaham, & Wanke, 2003) researchers from MITRE CAASD review their work currently being carried out under the Dynamic Density program. They are particularly interested in usefulness of complexity measures for TFM (Tactical Flow Management) decision support, with a 120 minute time horizon. These researchers evaluated each of the four Dynamic Density equations in terms of:

- Ability to predict subjective complexity ratings, at up to 120 minutes
- Predictability
- Face validity (simply whether a test “looks valid” to examinees, and other untrained participants);
- Redundancy (via semantically driven factor analysis)

The researchers chose the following subset of 12 factors from the Wyndemere, FAA, and NASA set of Dynamic Density measures (the final eight were taken directly from the Wyndemere work):

- Number of aircraft in sector
- Sector volume (nm^3)
- Speed Change (number aircraft with airspeed change > 10 kts/.02Mach within 2 minutes)
- Aircraft count, normalised
- Aircraft density per sector, normalised
- Aircraft proximity pairs (within 8nm, and within 13 nm)
- Convergence angle (for aircraft pairs within 13 nm)
- Conflicts predicted (number of proximate pairs predicted to conflict)
- Conflict sector boundary proximity (number of conflict pairs close to sector boundary)
- Altitude change (number of aircraft with an vertical speed > 500 fpm)
- Aircraft Distribution (relative to sector structure)

Multiple regression was used to assess, for various samples of airspace, how well the 12 factors predicted controllers’ complexity ratings. Structured interviews were also used to cross-check the results of multiple regression analysis. Operational interviews with controllers from five sectors showed that the following factors had the highest factors weightings (in descending order) on controller rated complexity (on a 1-5 scale)[11]: Peak number of aircraft (4.63); Portion of sector unavailable due to weather (4.63); Weather at busy airports (4.63); Merging/crossing traffic (4.50); Influence of convective weather on adjacent sectors (4.38); Total number aircraft (4.25); Departure push near sector (4.25); and arrival push near sector (4.13). Eight additional factors showed smaller weightings.

In their results, Masalonis, Callaham & Wanke (2003) noted sector-specific differences in the predictive value of the factors, though these differences are small

[11] these factors were apparently chosen from a broader literature review, and so do not map exactly onto the Dynamic Density metrics (above)used for the quantitative analysis.

enough that a generalised set could be applicable to multiple airspace, at least in the near time (it is not clear what is meant by this “near term” distinction). In their 2003 Budapest presentation (Masalonis, Callahan & Wanke, 2003) they note that a centre-specific complexity model performed better than a general model for each of the four sectors evaluated (the difference was statistically significant $p<.01$ for three of the four). In summary, and although they seemed more optimistic than some researchers (e.g. Kopardekar: Manning) about the potential for a unified model, they do speculate that a multidimensional representation of workload might be more useful than a single equation combining all factors.

The work of Masalonis et al provides some possibly useful criteria for evaluating candidate factors. For instance, their concept of face validity is built on the rationale that:

- Adding another aircraft should never decrease complexity;
- That increasing speeds and reducing airspace should never decrease complexity;
- That moving an aircraft to increase its distance from all other traffic should never increase complexity (the author finds this assertion questionable);
- That the factor should be independent of orientation; and
- That a small change in airspace should never cause a large change in complexity.

Further, their criteria for evaluating redundancy assume that low inter-factor correlations (in statistical terms, that each factor contributes uniquely to the regression equation) is desirable.

6.2 Data Collection Methods

Pawlak et al. (1996) identified three distinct phases of constructing an ATC complexity model: Elicitation (initial identification of complexity indicators); Refinement of the indicators; and Validation of the model. This seems a useable framework to discuss the range of methods that have been used in collecting data on air traffic complexity. Notice that four general approaches have been used in studying ATC complexity: Experimental (generally simulation or laboratory) studies (Stein, 1985; Manning et al., 2000); Field research relying on Observational or Interview studies (Mogford et al., 1993; Histon et al, 2002), and Analytical studies (Soede et al., 1971). Some have used a combination of these approaches (Arad, 1964; Pawlak et al., 1996).

The following table highlights the main techniques that have been used to collect, refine and validate indicators of air traffic complexity (and the phase—Elicit, Refine or Validate—at which they seem useful). With slight variations, there seems to have been only a basic handful of data collection techniques used to study complexity.

Appendix B, which does not limit itself to previous ATC complexity research, provides a fuller list of potential data collection techniques.

Table 1. Data collection methods previously used to study ATC

Class	Technique	Sub-technique(s)	Phase E=Elicitation R=Review V=Validation	References	Notes
OBSERVATONAL					
	Field Observations		E, R, V	Averty et al. (2003)	
	Familiarisation		E	Histon et al, (2002)	
	Videotape analysis		E, R, V	Bailey & Willems (2002)	
	Over-the-shoulder assessments		E, R, V	Pawlak et al. (1996); Sollenberger et al (1997)	
	Behavioural analysis		E, R,V	Manning et al (2000;2001); Bailey & Willems (2002)	incl. Behavioural checklists (Manning et al, 2000)
		Activity analysis		Bailey & Willems (2002)	
		Communication analysis		Bailey & Willems (2002)	
		Voice command analysis		Davison (2002)	
INTERVIEW					
	Expert judgement sessions		E, R	Kirwan et al (2001)	
	Rating and ranking exercises		E, R	Rodgers et al (1998); Schaefer et al (1999);Kirwan et al (2001); Koros (2003)	incl. paired comparisons Delahaye & Puechmoreal (2000)
	Focus groups		E, R	Delahaye & Puechmoreal (2000); Histon et al (2002)	
	Critical decision making		E, R	Laudeman et al (1998); Meckiff et al (1998)	
	Verbal protocol analysis		E, R	Pawlak et al. (1996)	Incl. retrospective verbalisation
	Structured interview		E, R	Ahlstrom (2001)	
	Semi-structured field interview		E, R	d"Arcy & Della Rocco (2001)	
	Workload ratings		R, V		e.g. ISA, TLX, ATWIT
		Instantaneous			
		Post session			
	Questionnaires / Surveys		E, R	Laudeman et al (1998)	

Table 1. Data collection methods previously used to study ATC

Class	Technique	Sub-technique(s)	Phase E=Elicitation R=Review V=Validation	References	Notes
EXPERIMENTAL/ QUASI-EXPERIMENTAL					
	Static simulations		E, R	Boag (2002)	
	Small scale simulation			Pawlak et al. (1996)	
	Shadow mode trials		R, V		
	Physiological measurement		R, V	Averty et al (2003);Laudeman et al (1998); Chatterji & Sridhar (2001)	too intrusive for operational settings?
		Eye tracking measures		Källqvist (2002); Stein (1992)	
		Heart rate measures		Brookings et al. (1996)	
		Brain activity measures			
	Task performance		R, V		
		Primary task performance		Wierwille & Connor (1983)	
		Secondary task performance		Metzger & Parasuraman (1999)	
ANALYTIC					
	Task Analysis		E, R	Seamster et al (1993); Boehm-Davis et al	includes Cognitive Task Analysis
	Cognitive Modelling		V	Blom et al. (2001)	e.g. IPME, MIDAS, PUMA
	Monte Carlo simulation		V	Krozel & Peters (2000)	
	Analysis of operational data		R, V	Rodgers et al (1998); Gosling et al (2003)	e.g. incident review
	Analysis of modelled workload		R, V	Eurocontrol (1996)	e.g. EAM simulation model of workload

6.3 Workload Assessment Methods

It seems agreed from the literature that controller workload is an appropriate criterion measure for air traffic complexity. This section, therefore, reviews the techniques that have been used to assess workload, particularly in ATC settings.

Various measures have been used to assess ATC workload. These are generally categorised as *subjective*, *behavioural* or *physiological*[12]. Within each category, there are a number of specific indices available. A review of the various workload measurement techniques (and their relative trade-offs in terms of sensitivity, cost, intrusiveness, etc.) is beyond the scope of the present paper (several thorough reviews exist (cf Meshkati, Hancock & Rahimi, 1990; Kramer, 1991; Hilburn & Jorna, 2001). Following are some examples of workload measures that have been used in the past.

Subjective
NASA TLX (Brookings & Wilson, 1994)
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).

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)

Physiological
EEG, EMG and EOG (Costa, 1993)
Heart rate measures (Brookings & Wilson, 1994)
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 entropy (Hilburn, Jorna & Parasuraman, 1995)
Visual fixation frequency (Stein, 1992)

[12] A fourth method, workload modelling, is used during system development to predict workload imposed by future systems. The British PUMA system (Houselander & Owens, 1995) is one example. One noted strength of PUMA is its reliance on an underlying model of human attention (Multiple Resource Theory (Wickens, 1992)), although the same caveats about the use of behavioural data—PUMA considers task completion times-- still apply.

Workload measures in ATC: some empirical studies.

Overall, ATC workload evaluations have tended to rely on subjective measures. Although the use of subjective measures is attractive (they are inexpensive, easily collected, and perhaps offer better controller acceptance), they carry some important potential limitations, such as memory effects (Manning et al 2001), unwillingness to report (Hilburn & Jorna, 2001), and other biases.

Many studies in the past have attempted to relate overt behavioural measures (e.g., total radio communication time (Morrow, 1993)) directly to controller mental workload (Manning et al., 2001; Bruce et al., 1993). This use of behavioural measures is attractive at first glance, given that they can often be related directly to operational performance. A number of ATC-relevant tasks can be used as embedded indicators of ATC workload. For instance, the communication demands of a new ATC system (expressed as, say, total microphone key press time per hour (Hilburn & Nijhuis, 2000) appear to relate directly to both how hard the controller is working, and how efficient the system can be expected to perform. Schmidt (1976) developed an index of controller difficulty that relied on task execution time and frequency of decision or action forcing events. From this a list of seven tasks were identified as the most demanding (cf. section 6.1).

As has been noted with respect to behavioural measures of complexity, however (Pawlak et al., 1996), overt behaviour does not necessarily represent underlying controller effort (Newman et al., 1993). Much of the “workload” in ATC, after all, consists of non-observable mental activity. Controllers can, again, employ compensatory strategies— such as vectoring aircraft into a holding stack to “buy time”—that maintain the external appearance of steady activity (Sperandio, 1971; Koros et al., 2003).

7. LESSONS LEARNT

This section presents a summary of lessons learnt, regarding the identification of complexity factors, as well as data collection methods and workload assessment methods. Section 10 presents a summary of overall conclusions, and general lessons for COCA.

7.1 Workload as a Criterion Measure of Complexity

Ultimately, the COCA complexity model must be judged by its ability to predict some agreed criterion measure. The literature clearly suggests that controller workload is the best available criterion, despite the influence of other task and operator-related factors.

7.2 Subjective Versus Objective Measures of Workload

One area of debate within the workload assessment community is the relative value of objective (behavioural or physiological) measures and subjective techniques (such self-report measures such as ATWIT, ISA, and NASA TLX). There are several known limitations of subjective techniques, including: Potential memory effects if used post-session (Manning et al, 2001); Unwillingness to report damaging information (Kirwan et al., 2001); Context effects (Colle & Reid, 1998); Inaccessibility to skilled performance, and the sheer effort required to generate and collect self reports (Manning et al, 2001), all of which can bias the resulting data.

Despite these potential problems, ATC research has generally relied on subjective workload measures. This is due to their attractive cost, ease of use, and face validity. There are at least two additional considerations that recommend their use in the COCA project. First, COCA will likely follow a course of progressive refinement in developing its complexity model. The initial steps in this process will rely on small scale exploratory studies, in which physiological measures seem unfeasible. At least initially, COCA's emphasis should be on subjective workload measures.

A second important advantage for COCA is that such subjective measures might better tap into the subjective nature of workload itself. That is, an objective workload measurement indicator (and some reliable physiological ones exist) can never take the place of a controller's claim that workload has been reduced. Given the consensus view that workload is a subjective response to various inputs including, among others, traffic complexity, it is reasonable to investigate that response using subjective techniques and measures.

7.3 Direct versus Indirect Measures of Complexity

On a related point, there seems some disagreement in the literature on how reliably controllers can evaluate complexity. It seems widely agreed that cognitive complexity is a subjective phenomenon that is not directly observable. Nonetheless, there also seems consensus that proxy measures of complexity are valuable in determining what makes a controller's job difficult and by how much (Mogford et al., 1995), or in revealing structural abstractions controllers use to understand and simplify a given traffic pattern (Histon et al., 2001; Histon et al, 2002). If justifiable, this view provides the same sorts of advantages as subjective workload techniques— namely, they are attractive (especially during initial stages of a progressive refinement process), and also might tap into the subjective nature of complexity assessment. For both of these reasons, it seems logical to consider the use of direct observational / interview techniques, at least initially. Such techniques could always be followed by more “complex” (and expensive) indirect data collection methods later, when the model is (hopefully) more sophisticated.

7.4 Avoiding Self Report Bias

One potentially useful distinction to maintain is that between static (e.g. the airspace proper) and dynamic (e.g. transient traffic characteristics) that have been noted by several researchers (Rodgers et al., 1998; Grossberg, 1989). It is reasonable to assume, for instance, that controllers are in the best position to provide ratings concerning the static factors surrounding complexity. Such reports would not be time critical, and could presumably be collected after a session without risk of memory effects (at least for familiar airspace). Providing dynamic ratings, however, might prove more intrusive for the controller, and it is likely that an observer (be it a supervisor or another controller acting as over-the-shoulder observer) would be better positioned to provide these.

7.5 A Top-Down versus Bottom-Up Approach to Defining Complexity

For at least forty years, researchers have wrestled with the notion of ATC complexity. The influence of certain traffic characteristics (e.g. traffic count, or weather) on ATC complexity, is relatively easy to conceptualise. What has proven a more intractable problem, however, is how to capture and quantify the influence of underlying traffic structure on complexity. Terms such as “organisation flow” (Arad, 1964) or “structural elements” (Histon et al, 2001;Histon et al, 2002) have been used to describe the emergent properties of the traffic geometry per se. Although it is recognised as an important contributor to perceived complexity, the concept of organisation flow has been left loosely defined (Rodgers, 1998).

This search for these underlying structural or flow factors suggests that there are two fundamental approaches to defining ATC complexity. The first, which seems to have been more popular, is a bottom-up attempt to identify the constellation of relevant complexity factors (on the basis of say interviews or observation) from which a composite complexity formula can be constructed. Conversely, a top down approach would start from a criterion measure of controller response, and attempt to identify which complexity factors contributed to this response.

It would seem that a combination of the two approaches, might offer the most promise. The former could rely on eliciting all potential complexity factors, the second on controlled comparisons to evaluate how a criterion response (workload) varies with systematic manipulations of these factors (by the way, it is probably the greater experimental demands of the second approach that has made it less popular over the years). In short, if the likely range of potential complexity factors can be identified bottom up (and there is already a wealth of possible factors available from the literature), their effect can be assessed top-down by starting with an accepted criterion—namely, controller workload.

7.6 Progressive Refinement of the Complexity Model

Clearly no single indicator will capture the fullness of ATC complexity across all possible contexts. It is also argued, based on the literature, that no fixed set of indicators will apply to every airspace. Literature is clear on the potential for interactions, for instance on the basis of airspace (Pawlak et al., 1996); time of day (Kirwan 2001), equipment (Cheboud et al, 2001), etc. Refinement of the complexity model demands that it be applied to various types of airspace.

7.7 Interview Procedures

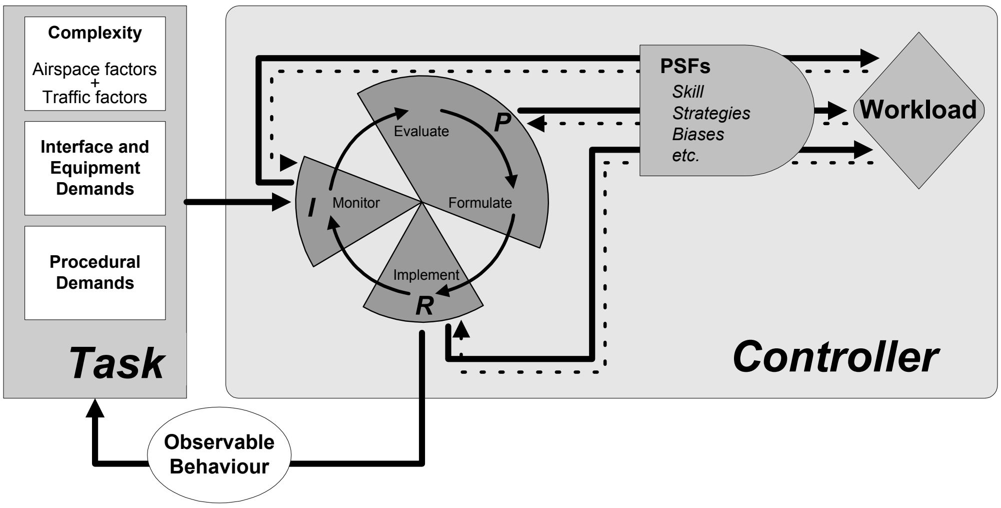
Studies have used interviews to elicit a variety of complexity-related information, such as the nature of the complexity factors, underlying decision making processes, and the relative importance of various types of information (Koros et al., 2003). Gromelski, Davidson & Stein (1992), in reviewing the techniques for field data collection in ATC, noted that less-structured conversational interviews can improve data collection by allowing the interviewer to

- Ask for examples to clarify a point;
- Explore the meanings of various phrases respondents use;
- Probe to ensure respondent understanding;
- Observe body language;
- Observe new topics raised by the respondent.

This view fits with the Focus Group approach that is often used in social science research. Focus groups are a form of loosely structured group discussion. The chief strength of the focus group procedure is that it exploits the social interaction nature of group interview, and uses this interaction (including, for example, interplay and modification of opinion) as an integral source of data. Focus groups are most useful in identifying the range of views, opinions or concerns present in a group. For that reason, their use in COCA is probably most appropriate during initial phase of factor identification, rather than in the later validation phase.

8. A PROPOSED MODEL OF ATC COMPLEXITY AND WORKLOAD

This section proposes a functional model of ATC complexity and workload that attempts to integrate the reviewed work on air traffic complexity and human information processing. This model distinguishes between taskload (or system) factors external to the controller, and the controller's internal processes (Hilburn & Jorna, 2001). Taskload consists of traffic complexity (which is driven by both airspace and traffic-related factors (Mogford et al, 1994; Mogford et al., 1995)) as well as by the demands of equipment, interface, and procedures (ibid.).



Proposed model of ATC complexity and workload.

Summed taskload is the input to the controller. Controller activity (the "pinwheel" of the controller portion of the diagram) consists of four elements: Monitoring, Evaluating, Formulating decisions, and Implementing decisions (Pawlak et al., 1996). In fact, these activities map well onto the traditional view of human information processing (Wickens, 1980), which distinguishes between Input (or *perception* or *stimulus*), Processing (which encompasses both the evaluation and formulation of a decision), and Response. These three are abbreviated *I*, *P* and *R* respectively in the diagram.

Each cycle of this I-P-R pinwheel has one of two outcomes-- either overt action (in which the controller acts on the taskload), or continuing the (non-observable) process of monitoring, evaluating, etc. "Implementation" in this sense can be deciding to postpone action (waiting to see what will happen). Again, observable activity does not capture all of a controller's processing.

In this model, controller workload is a response to this information-processing pinwheel—the demands of monitoring, evaluating, etc. However, the response is not a direct one, rather it is mediated by Performance Shaping Factors (PSFs) such as skill, fatigue, age, training, proneness to anxiety, etc. that can influence the resulting workload. Notice also that the influence of PSFs runs downstream (dashed arrows), through the allocation of attention to monitoring, evaluating/formulating, or implementing. For example, a controller might defer routine housekeeping tasks (e.g., not sharing turbulence reports with pilots), or might combine tasks in parallel (e.g., through joint RT clearances). In this model, such adaptive strategies do nothing to moderate the complexity of the underlying task. This is an important element of the model, since it encompasses not only systematic biases (e.g. as in decision making and perceptual biases that can colour monitoring and evaluation, as mentioned earlier) but also individual differences. PSFs are the noise in the complexity-taskload-workload transform. The end result of this entire process is controller workload, which is a function of both task demands and the controller's internal and subjective response to those demands.

How does this model advance the work of COCA? The following five immediate lessons come to mind regarding what we can hope to measure, and how we should go about eliciting complexity factors and refining a model of cognitive complexity:

- Observable behaviour is, again, a poor proxy for the workload of a controller, since many controller "activities" (e.g., evaluating, or deciding not to implement a response) are not observable;
- Complexity is not only the sum of the taskload faced by a controller (other aspects include interface and individual differences for instance, and these will always exert an influence (cf. Section 9.12)));
- Systematic and idiosyncratic biases will colour both the perception of workload, as well as the strategies brought to bear to maintain task performance;
- The above points not withstanding
- Although some studies have been confused on this point, individual operator factors (e.g., age, skill, etc.)are <u>not</u> complexity factors, and our search for complexity factors should overlook them;
- Workload is an internal and subjective response (albeit one that is at least partially accessible through physiological indicators).

It is believed that this functional model, while simple, fits both with what is known about human information processing and the literature on complexity and workload, and presents at least a passing level of face validity with respect to ATC operations. More importantly, it seems to afford a pragmatic way forward for the COCA project, and offer implications for what (and how) Complexity should be assessed in ATC.

9. GENERAL DISCUSSION

This chapter attempts to highlight some of the chief lessons that have been drawn from the preceding literature review. In most cases these lessons are based on the consensus view of the research community (as gleaned through the literature review). In others, they are built on an interpretation of past empirical work, through the lens of cognitive psychology. The following lessons are presented in no particular order.

9.1 No Complexity Indicator is Context-Free

There seems widespread agreement on the view that there is no single complexity indicator (or presumably composite of indicators) that applies equally well regardless of context. As Kirwan et al. (2001) noted, what works well in one setting might not work well in another site, or even at another time-of-day. The interactions between such factors can vary by site (Christien et al., 2003; Koros et al., 2003). Further, and despite of the admirable depth of their work, one report of Wyndemere's work into ATC complexity indicators (Pawlak et al., 1996) admits a large potential shortcoming: that their analysis was conducted on only one sector. The study goes on to concede that results might differ in another setting. Literature on NASA/FAA's dynamic density research (Laudeman, 1998) also acknowledged that different dynamic density measures performed better for different facilities. Christien et al (2003) also noted that the coefficients in their linear workload formula were "not the same for all sectors." Specifically, they note that, everything else equal, aircraft interactions will vary inversely with airspace volume.

9.2 Complexity Factors Do Not Always Interact in a Linear Way

On a related point, it seems agreed among several of the leading researchers in the field that one size can never fit all. What that means, in this case, is that no composite index composed of a linear combination of complexity indicators will adequately capture cognitive complexity of ATC in all contexts (cf. Athenes et al., 2002). A simple thought experiment is sufficient to illustrate the issue: Though traffic density (something as simple as aircraft per square kilometre[13]) is generally considered the best guess as to the taskload imposed upon a controller, it is easy to imagine an airspace in which traffic density per se has little or no bearing on the complexity of the controller's task. Indeed, complexity is low for a sector in which traffic flows are predictable (e.g., traffic is evenly spaced in-trail, and aircraft maintain following distances), somewhat regardless of traffic density.

Past attempts to assess ATC complexity have generally relied on linear combinations of individual complexity indicators (e.g., Pawlak et al., 1996). While the work of the Wyndemere group was commendable for its depth and approach (it recognised quite correctly, for instance, the need to move from physical to (non-observable) cognitive elements of ATC complexity), it ironically seems to have disregarded a critical aspect of human cognition. Namely, that what is complex in one context is not complex in another. As a result, attempts to define complexity as a linear composite of factors (e.g. "four parts aircraft density, to one part altitude transitions, to one part airspace volume, etc.") are inherently limited. One reference resulting from NASA/FAA's dynamic density research (a PowerPoint presentation on the Phase III results, provided 14 August, 2002) noted that the research partners,

[13] The researcher is still faced with the question of what units to use—aircraft per square kilometre in the entire airspace? In some defined subset? In some core area? (cf. Stein, 1985).

collectively, felt that the results could be improved through the use of different non-linear combinations of existing data.

9.3 Cognitive Complexity is More than Geometric Traffic Pattern

Some past attempts to assess air traffic complexity have relied on geometric relationships between aircraft (Histon, 2000), such as aircraft "clustering" behaviour (Cloerec et al, 1999). Various geometric approaches to airspace complexity have been proposed. In general, strictly geometrical approaches to complexity seem to have been beset with combinatorial problems in handling large numbers of aircraft (Durand & Granger, 2003). As Christien et al's (2003) presentation suggests, geometric measures of flow entropy do not always neatly capture complexity— in their study of hundreds of "elementary sectors" of European airspace, both high and low entropy sectors were associated with a large mix of climbing and descending traffic.

More critically (for the purposes of COCA), it seems that no geometric approach to date has fully captured the notion of complexity as it is perceived by the controller. Harwood et al. (1991) noted that controllers rely more on spatial and temporal patterns within traffic, than on the instantaneous positions of aircraft. And the work of Histon et al (2000;2001) suggests that abstracted underlying structures are derived that help controllers simplify and understand traffic patterns. It appears, though, that this work has a way to go.

9.4 The Need to Consider the Range of ATC Complexity

The rationale behind evaluating ATC complexity is that excessive complexity drives taskload, which indirectly drives workload, which raises the risk of overload, which ultimately sets an upper limit on sector capacity. Not surprisingly, therefore, the assessment of ATC complexity (and, by association, ATC workload) seems to have focused on the possible overload condition. However, operational data draw the underlying logic of this rationale into question: There is sufficient operational evidence (on the occurrence of ATC errors) as well as theoretical evidence (on human's poor monitoring performance and potentially-high workload under vigilance conditions (Parasuraman, 1987)) that under load might itself pose a serious threat to air safety (Jorna, 1993; Stager & Hameluck, 1990; Danaher, 1980). A review of ATC incidents in Canada (Stager, 1991), for example, showed that most occurred during low or moderate traffic load and normal traffic complexity. Similar data have emerged from studies of US ATC operational errors (Redding, 1992). In Europe, the 2000 mid-air collision over Germany occurred on a clear, quiet night. The suggestion has been made that controllers can adapt to heavy traffic peaks, but become error prone as traffic lightens (Fowler, 1980). Colle & Reid (1998) suggest that subjective workload ratings are increased, and controller judgement biased by low task difficulty. It seems therefore worthwhile to consider the range of possible conditions, including the potential impact of excessively low complexity on controller workload.

9.5 Decision Making Biases Contribute to the Perception of Complexity

What is it that makes a given traffic situation immediately recognisable to a controller as "complex?" Clearly, controllers bring more to the job than raw skills. They also bring a high level of domain expertise and in many cases years of experience. Part of what drives the perception of complexity on the basis of the task itself (i.e., the raw airspace and traffic characteristics (Majumdar & Ochieng, 2000)) is the expert recognition that controllers are able to make. Research on naturalistic decision making (Jensen, 1992; Klein, 1993) and decision making heuristics and related biases (Kahneman, Tversky & Slovic, 1973) indicate that there are systematic patterns in how experts perceive and evaluate data, and estimate future trends that can colour the perception of complexity.

9.6 Difficulty of Comparing Across Studies

The wide variety of measures that have been used to evaluate ATC complexity have complicated comparison between studies (Mogford et al., 1995). Despite the acknowledged consensus view that complexity causes workload, there seems confusion in the field about how to use the terms. Several studies used complexity factors such as keypad activity or pace ratings (Kuhar et al., 1976) or RT time and count (Buckley et al., 1983) that are better off seen as outputs measures (i.e., workload) rather than input factors (complexity). Finally, inconsistent factor definitions (and/differing amounts of detail provided) also complicated cross-study comparisons: Whereas one study would refer to "mix of aircraft type" (Grossberg, 1989), another would use the term "Traffic mix, slow versus fast aircraft" (Mogford et al., 1997). Further, citations to previous research (e.g.., Author B recounting Author A's metric list) in some cases either reworded, or omitted, certain factor definitions. Some interpretation was therefore required in compiling the metrics list that appears as Annex A. Despite these various problems-- the differences in definition, the confusion about certain terms, the details lost in the retelling-- there seems (as mentioned earlier) general agreement on a core set of perhaps 20-30 complexity metrics.

9.7 Complexity is Both Time and Space: Closure Angle and Time of Flight

It is known that shallow aircraft convergence angles are more demanding (Day, 1994; Wyndemere, 1996; Pekela & Hilburn 1998) in terms of conflict detection time. For instance, a closure angle of 90 degrees is generally less complex than a shallow angle of say 15 degrees.

However, the inverse relationship between the convergence angle and time-of-flight means that convergence angle alone is insufficient to describe complexity of a given aircraft convergence (Chatterji & Sridhar, 2001). For example, 90-degree convergences are geometrically easy to resolve than shallow convergences. However, the time available for resolution is smaller (cf. Chignell & Kerr, 1988). Similarly, head on situations seem easier to detect, but (because of high closure speed) are more difficult to resolve. Warren (1997) noted that whereas shallow convergence angles are harder to detect, but easier to resolve, head on conflicts are easy to detect but harder to resolve. Further, optimal resolution strategy differs with angle of convergence. Bilimoria et al (1996), for example, reported that speed change worked best (i.e. quickest) for closure angles of less than 19 degrees, and altitude resolutions work better for larger angles. Kopardekar and DiMeo (1997) specified a *Convergence Recognition Index* (CRI) that they claim (cf. Kopardekar, 2000) captures the degree of difficulty in recognising shallow convergence angles, and is based on how close the convergence angle is to 30 degrees.

The studies cited above were carried out in the context of free flight, and the search for optimised resolution manoeuvres. The focus was therefore on conflict resolution, as opposed to conflict detection difficulty (cf. Chatterji & Sridhar, 2001). In terms of controllers' perception of complexity, however, it is important to note that there seem to be systematic biases in terms of both convergence angle and time-of-flight. Whereas shallow angles are (to a point) more demanding in terms of conflict detection, it also seems controllers systematically tend to underestimate time-of-flight (Boudes & Cellier, 2000). On the basis of the above, it seems that both aspects of conflict situations—convergence angle and time-to-go-- should be considered in defining controller complexity.

9.8 Exporting US Results to Europe is Likely Just a Matter of Degree

Much of the available work was conducted in the US, much of this as part of the Dynamic Density program. One initial concern going into this literature review was that results from US studies might not be directly applicable to the European context. Although operational practices, traffic patterns and airspace constraints vary between US and Europe (Kirwan et al., 2001), little guidance was found on how to import US results. Some of the notable differences between US and European air traffic include the higher proportion of vertical transitions in the core area of Europe (and the resulting European distinction between upper and lower enroute airspace), and the greater occurrence of convective weather in the US (especially during the summer months, when storms can force large-scale, long distance reroutes).

In retrospect, this seems to be a minor concern. Although there are minor differences across studies in how complexity metrics are defined or expressed, on balance there seems a good deal of consensus on a core set of about 20-30 recommended metrics. Whereas the same basic set of complexity factors might apply to both the US and European context, factor weightings have to be adjusted across the two.

9.9 Known Aspects of Cognitive Performance

The literature on ATC complexity, and human factors, make it clear that the relationship between complexity and workload is an indirect one that is highly mediated by the influence of many individual characteristics. Whilst this poses obvious difficulties for ever fully capturing the notion of cognitive complexity mathematically, it would be overly pessimistic to conclude that the human factor must remain an unknown. In fact, the literature on human factors makes it clear that there are known aspects of human cognitive functioning that be incorporated into a predictive model of cognitive complexity.

As an example, take what is known about human short term memory and its limitations. It has been known for decades (Miller, 1956) that short term memory is limited to 7 ± 2 elements. In a task that requires repeating random number strings after a fixed delay, people almost universally show that they cannot retain more than 7 ± 2 elements in short term memory (this number can be as low as four under certain conditions (Cowan, 2001)). Notice, though, these elements need not be single digits, but can also be meaningful chunks (such as your own phone number[14]). The concept of chunking has proven useful in explaining how experts in chess (de Groot, 1965; Chase & Simon, 1973), contract bridge (Engle & Bukstal, 1978) and other domains can perform better than novices in tasks relying on short term memory.

Three general areas of human performance limitations can be identified. These are: Attention and Decision Making; Memory (especially short-term memory), and Perception. Empirical studies have revealed systematic aspects of human behaviour in each of these areas.

[14] Notice the difficulty you have in recognising your own telephone number read aloud if the speaker "chunks" the numbers in a way to which you are unaccustomed.

Example aspects of cognitive performance, which would be expected to have a bearing on cognitive complexity, include:

- Attention and memory:
  Short term memory (e.g., 7$\pm$2) limitations
  Selective attention limitations
  Sustained attention limitations
  Prospective memory (i.e., remembering to do something in the future) limitations.
  Input and response modality compatibility
- Perception
  Time estimation errors
  Perceptual errors in closure angle
  Data code (aural vs visual) and short term memory store compatibility problems
- Decision making
  Heuristic biases in pattern recognition (e.g. shallow planning)
  Memory availability problems

The proposed next phase of the COCA work intends to cross-check candidate complexity factors against a full list of potential cognitive performance problems. The aim would be to develop, on the basis of expert (human factors and ATC operations) expert judgments, a matrix of "factors-by-failures" that rates the degree to which each candidate complexity factor relates to specific potential cognitive failures. The aim would be to both [1] ensure that the complexity model adequately covers known aspects of cognitive performance, and [2] use the matrix as a "weeding" tool during later model refinement, to help reduce redundancy in the complexity model (i.e., to identify for removal potentially redundant complexity factors).

9.10 Potential Risks of Direct Complexity Estimates

The use of controller ratings has proven valuable in workload estimation (such as via the ISA or TLX instruments). Direct controller estimates of complexity, however, run the potential risk (at least in dynamic settings) of being heavily coloured by perceived workload. To date, there is no evidence that controllers can provide valid and reliable complexity ratings separate from rated workload.

There are ways of confronting this problem (through experimental design, analysis methods, and objective physiological-- as opposed to subjective-- workload measures) though these seem less feasible outside of controlled simulation environments. Although past studies have also used supervisor or over the-shoulder ratings of complexity, it is not clear that these are free from similar biases that can arise if one simply asks the controller directly (Thomas et al., 2002). In summary, and on the weight of the evidence, it seems that direct complexity rating/ranking methods should probably be limited to:

- Initial elicitation of complexity factors (cf. Pawlak et al., 1996);
- Subsequent controlled experimental work (i.e. simulation (Boag, 2002)); and
- Evaluation of static airspace factors, as opposed to dynamic traffic-related factors (cf. Mogford et al, 1995).

9.11 Non-Linear Approaches to Modelling Complexity

The vast majority of the literature on airspace complexity assumes a linear approach to capturing airspace complexity. For example, linear regression approaches (Laudeman, 1998; Masalonis et al., 2003) attempt to combine factors using fixed regression weights. The main shortcoming of this approach is clear: The single resulting regression equation tends to apply only to the particular context (sector, time of day, even weather), and does not generalise well to other contexts. What is it that makes a given traffic pattern so complex? In one case it might be the pattern of altitude transitions, in another it might be military activity, or weather, that restricts full use of airspace and thus limits controller options. A number of alternatives to linear regression have been proposed (e.g., Maximum Likelihood Analysis, Time-series analysis, Genetic Algorithms).

Non-linear approaches, on the other hand, start from the recognition that complexity factors combine in a non-linear way. Though the same constellation of factors might well apply across contexts, the relative importance of each differs by context. Several non-linear approaches have been used to model airspace complexity. Among them are dynamical systems modelling using a non-linear extension of Kolmogorov entropy (Delahaye et al., 2002).

The non-linear method that seems to have shown the most promise (or at least generated the most interest) is non-linear regression, typically by artificial neural networks (Chatterji and Sridhar (2001); Majumdar & Ochieng, 2001). In approaching non-linear regression, it might help to picture the resulting regression graphs of linear versus non-linear techniques. Whereas linear regression, as the name implies, results in a line that represents how factor weightings combine, non-linear regression results in a topology—picture a jagged volcano with a central peak. Cross sectional views of the radius (viewed in elevation) can be taken. An infinite number of such slices, or views can be taken—one for each given airspace context. The same factors (graphically, distance increments out from the volcano peak) apply across contexts, but the weightings (again graphically, the profile representing the relative height of the factor weightings) differs by context. In practice, a neural network is trained to the general topology, and applies to "learning" new profiles (i.e., contexts). Sridhar (2000) reviews work by the Dynamic Density team to investigate the use of neural network analysis of complexity factors. Training a neural net with samples of different complexity, they demonstrated 100% correct classification of the data.

This concept of non-linear regression has intuitive as well as theoretical appeal. It seems to fit well with what is known about naturalistic decision making (Klein, 1989; Klein, 1993), in particular that expertise in many fields is often more a process of pattern recognition than of action selection. In a variety of fields (whether it is fire control, trauma medicine, or air traffic control), once an expert recognises / diagnoses a situation (hopefully correctly) the course of action is clear to that expert. Athenes et al (2002) noted just this about air traffic control.

9.12 No Complexity Model is Perfect

It is important to note that, despite the great effort that has been devoted over the years to developing more predictive models of air traffic complexity, to date such models have generally managed to explain no more than 50% (r=.70) of the total variance in some criterion measure (Majumdar & Ochieng, 2000). The Dynamic Density index, for instance, while better than traffic density alone at predicting workload, still fails to account for a significant amount of peak workload (Kinnersly, 2000). In effect, "tuning" a linear model to certain influences (eg, flight count versus flight duration) can render a model insensitive to other influences (Mills, 1998).

Perhaps we can take some comfort in knowing that other domains, some much older than ATC, also continue to struggle with the modelling of complex phenomenon. Take, for example, the case of weather modelling. The computational dynamics of weather prediction are so complex that they can outstrip the capabilities of even the latest generation of supercomputers[15]. Despite all of the sophistication of today's weather models, and improvements over the decades in prediction accuracy (today's five day projections are now as accurate as three-day projections were only ten years ago (American Meteorological Society, 2000)), there are still shortcomings in weather modelling and its use.

Consider yourself an amateur meteorologist. If your single prediction indicator (a very simple model!) were that tomorrow's weather would be a repeat of today's, you would be correct a reasonable percentage of the time. In fact, such a simple model might be sufficient for your needs. If the cost of incorrectly forecasting sunshine is that you fail to carry an umbrella, and get wet on the 30-metre walk from your parking space to your office, the costs of being wrong do not seem very high. In that case, the utility of such a simple model seems adequate. What is surprising is that, with all of the advances in meteorological forecasting over the years, there has in some senses been very little practical improvement for the average person. Going back to the case of the hypothetical wet office worker, a sophisticated model that improves one-day forecasting from 80% to 85% accuracy is of questionable added value. As the old adage goes, "never let perfect be the enemy of good enough." Perhaps the same pragmatic approach should be adopted in approaching air traffic complexity—that any complexity model that performs better than traffic density alone represents an improvement.

[15] In fact, new non-linear techniques such as fuzzy logic have only very recently been incorporated into short term weather forecasting for aviation (cf. Hicks et al, 2003).

10. CONCLUSIONS

It is clear from the literature that a great deal of effort has been devoted over the years to determining what exactly drives ATC complexity. It is also clear (given that this research, after 40 years, is ongoing) that despite the variety of approaches applied, nobody has yet managed to solve the problem. As Mogford et al. (1994) put it "Complexity is complex." Past attempts to quantify ATC complexity have tended to rely on geometric properties of the traffic stream, or on observable controller behaviour. The main shortcoming of the latter is increasingly recognised—namely, that it does not fully consider the cognitive factors underlying complexity.This review set out to identify the range of ATC complexity factors that have been used, and to distil lessons about which factors might be most appropriate and useful for the COCA project. This effort seems to have been more successful on the first than the second score. That is, a number of complexity factors have clearly been used (see Annex A). However, beyond perhaps simple traffic density, the literature as a whole does not suggest that any specific factors are always valuable.

Rather, the literature underscores the point that there is no single factor, or fixed constellation of factors, that drives complexity in every context. What makes traffic "complex" in one situation (e.g., a large number of similar callsigns) might be different than what drives complexity in another context (e.g., proximity of active military areas, which limit control options). "Context" in this case can refer to separate airspaces, or even different times of day for the same airspace.

This effort started from the pragmatic realisation that a quantitative model will, by definition, always remain an abstraction. This is particularly true in the case of cognitive complexity modelling, which, due in large part to human quirks (or *individual differences*, in the language of human factors) will always have some statistical noise and non-linearity, and which will defy our efforts to relate complexity directly to workload. To address this shortcoming of most previous work into ATC complexity, there is increased interest (e.g., among the Dynamic Density team) in pursuing innovative means of factor weightings. In particular, there seems growing interest in pursuing such non-linear techniques as neural networks as a way to combine factors.

Although this effort set out to capture the cognitive elements underlying ATC complexity, the focus was on identifying <u>traffic</u> and <u>task related</u> factors that can be applied across individuals, and across different types of airspace. This is to be accomplished by an iterative refinement process incorporating subjective techniques (interviews, questionnaires, focus groups, expert paired comparisons, and verbal protocol exercises), together with objective (behavioural observation and experimental) data collection techniques to refine (and shorten) a list of factors that tap controllers' subjective perception of ATC complexity. What we ultimately aim for is a complexity model built upon a set of objective traffic and ATC task factors that

[1] Has been collected using a variety of techniques;
[2] Has been kept as broad as possible, so as not to a priori reduce the factor set;
[3] Has demonstrated validity / generalisability across various kinds of airspace;
[4] Appears, on the basis of literature review, factor elicitation and experimentation, to adequately address underlying cognitive issues that have been shown a relationship to cognitive complexity;
[5] Is refined through statistical development to balance the inherent trade off between prediction and interpretability (i.e., a large number of model factors might improve prediction, but at the cost of being unwieldy and un-interpretable); and

[6] Improves our prediction of controllers' subjective sense of complexity, beyond that currently offered by traffic density alone.

One consistent theme throughout this review has been the notion that the transform between complexity and workload is always context specific, and will always contain some statistical noise. Individual differences in decision making and control strategies, or slight differences in interfaces, mean that a generalised complexity index will never reach 100% predictive accuracy in all settings. This should be seen as inevitable, and not necessarily a sign of weakness in the approach. The goal of COCA should be to develop and refine a complexity index that is sensitive to complexity factors, robust enough to accommodate individual differences, and useful across the range of typical interfaces.

REFERENCES

Ahlstrom, U., Rubinstein, J., Siegel, S., Mogford, R. & Manning, C. (2001). Display concepts for en route air traffic control. DOT/FAA/CT-TN01/06. Atlantic City, New Jersey: FAA.
American Meteorological Society (2000). Policy Statement on Weather Analysis and Forecasting, Bulletin of the American Meteorological Society, 79, 2161-2163.
Arad, B.A. (1964). The controller load and sector design. Journal of Air Traffic Control, May, 12-31.
Arcy, J-F. d', & Della Rocco, P.S. (2001). Air traffic control specialist decision making and strategic planning—a field survey. Report DOT/FAA/CT-TN01/05. Atlantic City, New Jersey: FAA.
Athènes, S., Averty, P., Puechmorel, S., Delahaye, D. & Collet, C. (2002). Complexity and Controller Workload: Trying to Bridge the Gap. In Proceedings of the 2002 International Conference on Human-Computer Interaction in Aeronautics (HCI-Aero 2002). Cambridge, MA, USA: Massachusetts Institute of Technology.
Averty, P., Athenes, S., Collet, C. & Dittmar, A. (2002). Evaluating a new index of mental workload in real ATC situation using psychophysiological measures.
Baron, S. & Levison, W.H. (1977). "Display analysis using the optimal control model of the human operator," Human. Factors 19, 4 (Oct. 1977), 437-457
Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin, 1982, 1(2), 276-292.
Bell, D. (1982). Regret in decision making under uncertainty. Operations Research, 30, 961-981.
Bilimoria, K., Sridhar, B. & Chatterji, G.B. (1996). Effect of conflict resolution maneuvers and traffic density on free flight. AIAA Guidance, Navigation, and Control Conference, San Diego, June 1996.
Bisseret, A. (1970). Operational memory and structure of work. Bulletin de Psychologie, 24, 280-294.
Bisseret, A. (1971). Analysis of mental processes involved in air traffic control. Ergonomics, 14(5), 565-570.
Bisseret, A. (1981). Application of signal detection theory to decision making in supervisory control. Ergonomics, 24(2), 81-94.
Blom, H., Daams, J. & Nijhuis, H. (2000). Human cognition modeling in ATM safety assessment. Presented at the 3rd FAA/Eurocontrol R&D Seminar, Napoli, Italy.
Boag, C.C. (2002). Information processing and relational complexity in air traffic control. Unpublished Ph.D. Research Executive Summary.
Borg, C. G. (1978). Subjective aspects of physical and mental load. Ergonomics, 21, 215-220.
Boudes, N. & Cellier, J.M. (2000). Accuracy of estimations made by air traffic controllers. International Journal of Aviation Psychology. 10(2), 207-25.
Braby, C.D., Harris, D. & Muir, H.C. (1993). A psychophysiological approach to the assessment of work underload. Ergonomics, 36(9), 1035-1042.
Breitler, A.L., Lesko, M. & Kirk, M. (1996). Effects of sector complexity and controller experience on probability of operational errors in air route traffic. Report DTFA01-95-C-00002: FAA.
Brookings, J.B. & 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.
Brookings, J.B. & 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.
Bruce, D. S., Freeberg, N. E., & Rock, D. A. (1993). Influences of ATC system inputs and task parameters on controller workload pressure: The development of an explanatory model. Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting (pp. 108-112).
Buckley, E. P., DeBaryshe, B. D, Hitchner, N., & Kohn P. (1983). Methods and measurements in real time air traffic control system simulation (DOT/FAA/CT-TN83/26). Atlantic City NJ: DOT/FAA Technical Center.
Buckley, E. P., DeBaryshe, B. D, Hitchner, N., & Kohn P. (1983). Methods and measurements in real time air traffic control system simulation (DOT/FAA/CT-TN83/26). Atlantic City NJ: FAA.
Buckley, E., & Stein, E. (1992). Simulation research variable specifications (Unpublished manuscript).
Buckley, E.P. DeBarysche, B.D. Hitchner, N. & Kohn, P. (1983). Methods and Measurements in Real-Time Air Traffic Control Simulations. Report DOT/FAA/CT-83/26. Atlantic City, NJ, USA: Federal Aviation Administration.

Cardosi, K. & Yost, A. (2001). Controller and pilot error in airport operations: A review of previous research and analysis of safety data (DOT/FAA/AR-00-51). Washington, DC: FAA Office of Aviation Research.

Cardosi, K. M & 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.

Cardosi, K. M & 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.

Casso, N. & Kopardekar, P. (2001). Air Traffic Management System Development and Integration (ATMSDI) CTOD-2.3.2 Draft Guidelines Subtask 4: Human Factors Metrics Guidelines. Raytheon Corporation.

Chase, W.G. & Simon, H.A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.

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

Chatterji, G.B. (1999). Neural network based air traffic controller workload prediction. Proceedings of the AIAA Guidance, Navigation, and Control Conference, 1999.

Chaboud et al. (2000). Air Traffic Complexity: Potential Impacts on Workload and Cost. EEC Note 11/00. EUROCONTROL.

Chiles, W. D., & Alluisi E. A. (1979). On the specification of operator or occupational workload with performance-measurement methods. Human Factors, 21(5), 515-528.

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

Christoffersen, K., Hunter, C. N., & Vicente, K. J. (1994). Cognitive dipsticks: Knowledge elicitation techniques for cognitive engineering research (CEL 94-01). Toronto, Canada: University of Toronto, Cognitive Engineering Laboratory.

Cloerec, A., Zeghal, K. & Hoffman, E. (1999). Traffic complexity analysis to evaluate the potential for limited delegation of separation assurance to the cockpit. Presented at the cockpit IEEE/AIAA Digital Avionics Systems Conference, St Louis, Missouri, October 1999.

Colle, H.A. & Reid, G.B. (1998). Context effects in subjective mental workload ratings. Human Factors, 40 (4), 591-600.

Collins, W.E., Schroeder, D.J & Nye, L.G. (1991). Relationships of anxiety scores to screening and training status of air traffic controllers. Aviation, Space and Environmental Medicine, 62, 236-240.

Computer Resource Management, Inc. (1989). National Airspace System simulation support facility (NSSF) Data reduction for dependent parallel runways. (NSSF/89/002) Washington D.C.

Corker, K.M. (1999). Human performance simulations in the analysis of advanced air traffic management. In Proceedings of the 1999 Winter Simulation Conference.

Corker, K.M. & Smith, B. (1993). An architecture and model for engineering simulation. Paper presented at the AIAA Confrence on Computing in Aviation.

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

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24(1).*

Cox, M. (1992). Report on the cognitive aspects of the air traffic control task. Report IAM 706. Farnborough, Hampshire, UK: Royal Air Force Institute of Aviation Medicine.

Credeur, L., Capron, W. R., Lohr, G. W., Crawford, D. J., Tang, D. A., & Rodgers, W. G. (1993). A comparison of final approach spacing aids for terminal ATC automation. Air Traffic Control Quarterly, 1(2), 135-178.

Danaher, J.W. (1980). Human error in ATC system operations. Ergonomics, 22(5), 535-545.

Day, P.O.. (1994). Air Traffic Systems Strategies Studies (ATSRATS) Interaction Index, Roke Manor Research Ltd.

DeGroot, A (1965). Thought and Choice in Chess. The Hague, The Netherlands: Mouton

Deitch, E.L. (2001). Learning to Land: A Qualitative Evaluation of Pre-Flight and In-Flight Decision making Processes in Expert and Novice Aviators. Unpublished Ph.D. dissertation.

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

Delahaye, D., Farges, J., Schoenauer, M & Alliott, J. (1994). Genetic algorithms for air traffic assignment. Proceedings of the European Conference on Artificial Intelligence, ECAI.

Della Rocco, P. (1991). Fatigue and performance: Shiftwork in controllers of varying age (Dissertation Prospectus).
Durand, N. & Granger, G. (2003). A traffic complexity approach through cluster analysis. Presented at ATM 2003.Budapest, 23-27 June, 2003.
Endsley, M. & Rodgers, M. (1996). Attention distribution and situation awareness in air traffic control. In Proceedings of the 40th Annual Meeting of the Human Factors Society.
Endsley, M. R. (1994). Situation awareness in dynamic human decision making: Theory. In R. D. Gilson, D. J. Garland, & J. M. Koonce (Eds.), Situational Awareness in Complex Systems. Daytona Beach, FL: Embry Riddle Aeronautical University Press.
Engle, R.W. & Bukstal, L.H. (1978). Memory processes among bridge players of differing expertise. American Journal of Psychology, 81, 673-681.
EUROCONTROL (2002a). CARE/ASAS/Activity 2 Validation Framework WP3 Deliverable 5: Human Performance Metrics EUROCONTROL Reference: CARE/ASAS/NLR/02-034
EUROCONTROL (2002b). USA/ECAC complexity comparison. Draft report. January, 2002:
EUROCONTROL (2000). Investigating the air traffic complexity: Potential impacts on workload and costs. EEC Note 11/00.
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 (1996). Air traffic control model simulation studies- an introduction for potential users. EEC Note 2/AMS/1996.
FAA (2000). Airspace management Book—Metrics.
FAA (1984). Establishment and validation of en rouite sectors. DOT/FAA/Order 7210.46. Washington, D.C.: FAA.
FAA (1998). Position Classification Standard for Air Traffic Control: Series ATC-2152 terminal and en route. Washington, DC: FAA.
Fields, R.E., Wright, P.C., Marti, P. & Palmonari, M. (1998). Air traffic control as a distruibuted cognitive system. Proceedings of ECCE-9, the 9th European Conference on Cognitive Ergonomics, pp 85-90.
Fowler, F.D. (1980). Air traffic control problems: a pilot's view. Human Factors, 22, 645-653.
Fracker, M. L. & Davis, S. A. (1990). Measuring operator situation awareness and mental workload. Proceedings of the Fifth Mid-Central Ergonomics/Human Factors Conference (pp 23-25). Dayton, OH.
Fuld, R.B., Liu, Y., & Wickens, C.D. (1987). The impact of automation on error detection: Some results from a visual discrimination task. In Proceedings of the Human Factors Society, 31st Annual Meeting (pp. 156-160).
Galushka, J., Frederick, J., Mogford, R., & Krois, P. (1995). Plan view display baseline research report (DOT/FAA/CT-TN95/45). Atlantic City, NJ: FAA.
Gent, R. van, Hoekstra, J.M. & Ruigrok, R.C.J. (1997). Free flight with airborne separation assurance. In Proceedings of the Confederation of European Aerospace Societies (CEAS) 10th European Aerospace Conference, 20-21 October 1997, Amsterdam, The Netherlands.
Goh, J. & Wiegmann, D.A. (2002). Human factors analysis of accidents involving visual flight rules flight into adverse weather. Aviation, Space, and Environmental Medicine, 73, 817-822.
Granger, G., Durand, N. & Alliot, J.M. (2001). Optimal resolution of en route conflicts. Presented at the 4th USA/EUROPE ATM R&D Seminar. Santa Fe, New Mexico, USA.
Gromelske, S., Davidson, L. & Stein, E.S. (1992). Controller memory enhancement: Field facility concepts and techniques. DOT/FAA/CT-TN92/7. Atlantic City, New Jersey, USA: FAA.
Guttman, J.A., Stein, E., & Grimelski, S. (1995). The influence of generic airspace omn air traffic controller performance. DOT/FAA/CT-TN95/38. Atlantic City, New Jersey: FAA.
Hadley, J., Sollenberger, R., D'Arcy, J.F. & Bassett, P. (2000). Interfacility boundary adjustment. Report DOT/FAA/CT-TN00/06. Altantic City, NJ: FAA.
Hadley, G. A., Guttman, J. A., & Stringer, P. G. (1999). Air traffic control specialist performance measurement database (DOT/FAA/CT-TN99/17). Atlantic City, New Jersey: FAA
Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21(6)*, 803-831.
Hancock, P.A. & Caird, J.K. (1993). Experimental evaluation of a model of mental workload. Human Factors,3, 413-429.

Hancock, P.A., Chignell, M.H. & Kerr, G. (1988). Defining task complexity and task difficulty. P{resented at the XXIV International Congress of Psychology, Sydney, Australia.
Harris, R.L., Glover, B.L., & Spady, A.A. (1986). Analytic techniques of pilot scanning behavior and their application (NASA Technical Paper 2525).
Harris, R.L., Tole, J.R., Ephrath, J.R. & Stephens, A.T. (1982). How a new instrument affects a pilot's mental workload. Proceedings of the Human Factors Society 26th Annual Meeting, 1010-1013.
Hart, S. & Wickens, C.D. (1990).Workload assessment and prediction. In H.R. Booher (Ed.) MANPRINT: An emerging technology, advanced concepts for integrating people, machine, and organization. New York: Van Nostrand Reinhold.
Hart, S.G. & Stavelend, L.E. (1988). Development of the NASA-TLX (Task Load Index): results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.) Human mental workload. Holland: Elsevier.
Harwood, K., Roske-Hofstrand, R. & Murphy, E. (1991).Exploring conceptual structures in ATC. Proceedings of the Sixth International Symposium on Aviation Psychology, 29 April-2 May, pp. 466-473.
Hicks, T., Crawford, T. & Wilson, M. (2003). A Fuzzy Logic System for Automated Short Term Aviation Weather Forecasts. In Proceedings of the 3rd Conference on Artificial Intelligence Applications to the Environmental Science , Feb 2003.
Hilburn, B. & 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.
Hilburn, B. & Flynn, M. (2001). Air traffic controller and management attitudes toward automation: an empirical investigation. Presented at the 4th USA/EUROPE ATM R&D Seminar. Santa Fe, New Mexico, USA.
Hilburn, B., Bakker, M.W.P., Pekela, W.D. & Parasuraman, R. (1997). The Effect of Free Flight on Air Traffic Controller Mental Workload, Monitoring and System Performance. Proceedings of the Confederation of European Aerospace Societies (CEAS) 10th European Aerospace Conference, 20-21 October 1997, Amsterdam, The Netherlands.
Hilburn, B., Jorna, P.G.A.M. & 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.
Hooijer, J.S. & 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
Hopkin, V.D. (1979). Mental workload measurement in air traffic control. In N. Moray (Ed.) Mental workload: its measurement and theory. New York: Plenum Press.
Hopkin, V.D. (1994). Human performance implications of air traffic control automation. In M. Mouloua & R. Parasuraman (Eds.) Human Performance in Automated Systems: Current Research and Trends. Hillsdale, New Jersey, USA: Erlbaum.
Hunt, V.R. & Zellweger, A. (1987). Strategies for future air traffic control systems. Computer, February, 19-32.
Hurst, M.W. & Rose, R.M. (1978). Objective job difficulty, behavioral response, and sector characteristics in air route traffic control centres. Ergonomics, 21, 697-708.
Jex, H.R. (1988). Measuring mental workload: Problems, progress and promises. In P.A. Hancock and N. Meshkati (Eds.), Human Mental Workload. Amsterdam: North Holland.
Johannsen, G. & Rouse, W.B. (1979). Mathematical concepts or modelling human behavior in complex man-machine systems. Human Factors,21(6), 733-747.
Jorna, P.G.A.M. (1993). The human component of system validation. In JA Wise, VD Hopkin & P Stager (Eds.) Verification and validation of complex systems: human factors issues. Berlin: Springer-Verlag.
Kaempf, G.L. & Klein, G. (1994). Naturalistic decision making: THE Next Generation. IN N. Johnston, N. McDonald, & R. Fuller (Eds.), Aviation Psychology in Practice. Ashgate.
Kahneman, D. (1973). Attention and effort. Englewood Cliffs, New Jersey, USA: Prentice-Hall.

Kahneman, D., Slovic, P. & Tversky, A. (1982). Judgment under uncertainty: Heuristics and Biases. New York: Cambridge University Press.
Källqvist, C. (2002). Eye movements in visual scanning in simulated air traffic control: a field study. Unpublished Masters' Thesis, Lund University, Sweden.
Kalsbeek, J.W.H. (1976). Some aspects of stress measurement in air traffic control officers at Schiphol Airport. Symposium on stresses of air traffic control officers (University of Manchester, Dept. of Postgraduate Medical Studies), 39-42.
Kantowitz, B.H & Casper, P. (1988). Human workload in aviation. In E.L. Weiner & D.C. Nagel (Eds.), Human factors in aviation. pp. 157-187. San Diego: Academic Press.
Kinnersly, S.R. (2000). INTEGRA- Objective measures of ATM system safety: Literature survey. Eurocontrol report.
Kinney, G.C, Spahn, J. & Amato, R.A. (1977). The human element in air traffoic control: Observations and analyses of the performance of vcontrollers and supervisors in oprobviding ATC separation services. Report MTR-7655. MITRE Corporation.
Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an aid can and should go unused. Human Factors, 35, 221-242.
Kirwan, B., Scaife, R. & Kennedy, R. (2001). Investigating Complexity Factors in UK Air Traffic Management. Human factors and Aerospace Safety, 1(2), June 2001.
Kirwan, B. and Scaife, R. (2000) Factors Influencing Incident Rates – Results of LATCC Group Judgement Exercises. Technical Annex to R&D Report 0030. ATMDC, NATS, Bournemouth Airport.
Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. IEEE Transactions on Systems, Man, and Cybernetics, 19(3), 462-472.
Klein, G.A. (1989). Recognition-primed decisions. In W. Rouse (Ed.), Advances in Man-Machine Systems Research (pp. 47-92). Greenwich, CT: JAI Press.
Klein, G.A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G.A. Klein, J. Orasanu, R. Calderwood & C.E. Zsambok (Eds.), Decision making in action: Models and methods (pp. 138-147). New Jersey: Ablex Publishing.
Knecht, W., Smith, K. and Hancock, P. (1996). A dynamic conflict probe and index of collision risk. Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting.
Kopardekar, P & Magyrits (2002). Dynamic Density: Measuring and Predicting Sector Complexity. 21st Digital Avionics Systems Conference (DASC), Irvine, California.
Kopardekar, P. & DiMeo, K. (1997). Dynamic Density variables. PowerPoint briefing.
Kopardekar, P. (2000). Dynamic Density – A Review of Proposed Variables. FAA NAS Advanced Concepts Branch ACT-540: FAA.
Koros, A., Della Rocco, P.S., Panjwani, G., Ingurgio, V., & D'Arcy, J.F. (2003). Complexity in Air Traffic Control towers: A field study. Part 1: Complexity Factors. DOT/FAA/CT-TN03/14. Atlantic City, New Jersey: FAA.
Krebs, M.J., Wingert, J.W. & Cunningham, T. (1977). Exploration of an oculometer-based model of pilot workload (NASA Technical Report CR-145153). Minneapolis, Minnesota: Honeywell Systems & Research Center.
Krozel, J. (2000). Free Flight Research Issues and Literature Search. TR-00RT043-02. Los Gatos, California, USA: Seagull Technology Inc.
Lafon-Milon, M. T. (1981). Mental representation of vertical separation in the performance of air traffic control: Three dimensional representation of future states (Rapport INRIA nol. COR66). Le Chesnay, France: INRIA.
Laudeman, I., Shelden, S., Branstrom, R. & Brasil, C. (1998). Dynamic Density: An Air Traffic Management Metric. Report NAS/TM-1998-112226. Ames Research Center, Moffett Field, California: NASA.
Layton, C., Smith, P.J. & McCoy, C.E. (1994). Design of a cooperative problem-solving system for en-route flight planning: an empirical evaluation.
Leal de Matos P. (1998). The Development of Decision Support Models for European Air Traffic Flow Management, unpublished PhD Thesis, University of Warwick, UK.
Leighbody, G., Beck, J., & Amato, T. (1992). An operational evaluation of air traffic controller workload in a a simulated en route environment. 37th Annual Air Traffic Control Association Conference Proceedings (pp. 122-130). Arlington, Virginia: Air Traffic Control Association.

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

Lenorovitz, D.R. & Phillips, M.D. (1987). Human factors requirements engineering for air traffic control systems. In G. Salvendy (Ed.) Handbook of human factors. New York: Wiley.

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

Leroux, M. (1993). The PATS cooperative tools. In The proceedings of the scientific seminar on PHARE. Braunshweig, Germany: DLR Institute for Flight Guidance.

Leroux, M. (1993). The PATS cooperative tools. In The proceedings of the scientific seminar on PHARE. Braunshweig, Germany: DLR Institute for Flight Guidance.

Liu, Y., Fuld, R. & Wickens, C.D. (1993). Monitoring behavior in manual and automated systems. International Journal of Man-Machine Studies, 39, 1-15.

Majumdar, A. & Ochieng, W.Y. (2000). The factors affecting air traffic controller workload: a multivariate analysis based upon simulation modelling of controller workload. Centre for Transport Studies, Imperial College, London.

Majumdar, A. & Polak, J. (2001). Estimating the capacity of Europe's airspace using a simulation model of air traffic controller workload. Paaper No. 01-3250. Presented at the 80th Annual Meeting of the Transportation Research Board. Washington, DC: Transportation Research Board.

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

Masalonis, A., Callaham, M., Figueroa, Y. & Wanke, C. (2003). Indicators of Airspace Complexity for Traffic Flow Management Decision Support. Presented at the 12th International Symposium on Aviation Psychology, Dayton, Ohio.

Meckiff, C., Chone, R. & Nicolaon, J-P. (1998). The Tactical Load Smoother for multi-sector planning. Proceedings of the 2nd FAA/Eurocontrol ATM R&D Seminar. Orlando, Florida., Dec. 1998.

Mehadhebi, K. (1996). Comment representer la charge de travail du controleur. CENA.

Meister, D. (1986). Advances in human factors/ergonomics. New York: Elsevier.

Metzger, U. & Parasuraman, R. (1999). Free flight and the air traffic controller: Active control versus passive monitoring. Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting.

Michelson, A. (1993). The transfer of information in the dance language of honey bees. Journal of Comparative Physiology, 173, 135-141.

Mills, S.H. (1998). The combination of flight count and control time as a new metric of air traffic control complexity. Report DOT/FAA/AM-98/15. Oklahoma City: FAA.

Mogford, R. H., Guttman, J. A., Morrow, S. L., & 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. (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.

Mogford, R.H., Murphy, E.D., Roske-Hofstrand, R.J., Yastrop, G. & 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.

Moray, N., Lootsteen, P., & Pajak, J. (1986). Acquisition of process control skills. IEEE Transactions on Systems, Man, and Cybernetics, 16, 497-504.

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

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

Newman, M., Tattersall, A. & Warren, C. (1993). Modelling cognitive processes in air raffic control operators. In N. Johnson, R Fuller & N McDonald [Eds.] Aviation Psychology: Training and Selection. Aldershot, UK: Avebury Aviation, 118-123.

Niessen, C., Eyferth, K., & Bierwagen, T. (1999). Modelling cognitive processes of experienced air traffic controllers. Ergonomics, 42 (11), 1507-1520.

Nijhuis, H.B. (1994). A Study Into the Consequences of Data-Link for the Mental Workload of the Air Traffic Controller. PHARE/GHMI/1.1.1/NLR/001. Brussels, Belgium: Eurocontrol.
Nolan, M.S. (1990). Fundamentals of air traffic control. Belmont, California: Wadsworth.
O'Donnell, R. D. & Eggemeier, F. T. (1986). Workload assessment methodology. In K.R. Koff & J. Thomas (Eds.) Handbook of Perception and Human Performance, 2, Cognitive processes and performance, Wiley, New York.
Ozmore, R. E., Morrow, S. L. (1996). Evaluation of dual simultaneous instrument landing system approaches to runways spaced 3000 feet apart with one localizer offset using a precision runway monitor system, (DOT/FAA/CT-96/2). Atlantic City, NJ: DOT/FAA Technical Center.
Paivio, A. (1986). Mental representations: A dual coding approach. New York: Oxford University Press.
Parasuraman, R. (1987). Human computer monitoring. Human Factors, 29(6), 695-706.
Paul, L. (1990, October). Using simulation to evaluate the safety of proposed ATC operations and procedures (DOT/FAA/CT-TN90/22). Atlantic City, NJ: DOT/FAA Technical Center.
Pawlak, W.S., Brinton, C.S., Crouch, K. & Lancaster, K.M. (1996). A framework for the evaluation of air traffic control complexity. Presented at the AIAA national Conference.
Paylor, A.(1993). Air traffic control: today and tomorrow. Shepperton, Surrey, England: Ian Allen Ltd.
Perrow, C. (1999). Normal Accidents: Living with High Risk Technologies. Princeton, NJ, USA: Princeton Press.
Pfleiderer, E.M. (2000). Muldimensional scaling analysis of controllers' perceptions of aircraft performance characteristics. DOT/FAA/AM-00/24. Oklahoma City: FAA.
Phillipp, U., Reiche, D., & Kirchner, J.H. (1971). The use of subjective rating. Ergonomics, 14, 611-616.
Rasmussen, J. (1981). Models of mental strategies in process plant diagnosis. In J. Rasmussen & W. Rouse (Eds.), Human Detection and Diagnosis of System Failures. New York: Plenum, 1981.
Reason, J. (1988). Cognitive under-specification: its varieties and consequences. In B. Baars (Ed.) The psychology of error: a window on the mind. New York: Plenum Press.
Redding, R. E., Cannon J. R., Lierman, B. C., Ryder, J. M., Seamster, T. L., & Purcell, J. A. (1990). Cognitive task analysis of prioritization in air traffic control. Report to the Federal Aviation Administration. McLean, VA: Human Technology, Inc.
Redding, R.E. (1992). Analysis of operational errors and workload in air traffic control. In Proceedings of the Human Factors Society 36th Annual Meeting, 1321-1325.
Reiche, D., Kirchner, J.H. and Laurig, W. (1971). Evaluation of stress factors by analysis of radiotelecommunication in ATC. Ergonomics, 14, 603-609.
Remington, R.W., Johnston, J.C., Ruthruff, E. Gold, M. & Romera, M. (2000). Visual search in complex displays: Effects of traffic load, regularity, and conflict geometry on the detection of conflicts by air traffic controllers. Draft technical report. NASA Ames Research Center.
Rodgers, M.D., R.H. Mogford & L.S. Mogford (1995). Air traffic controller awareness of operational error development. In Proceedings of the 2nd Annual Situation Awareness Conference. Daytona Beach, Florida: Embry Riddle Aeronautical University.
Rodgers, M.D., Mogford, R.H., & Mogford, L.S. (1998). The relationship of sector characteristics to operational errors. DOT/FAA/AM-98/14). Washington, DC: FAA.
Rodgers, M.D. & Nye, L.G. (1993). Factors associated with the severity of operational errors at air route traffic control centers. In M.D. Rogers (Ed.) An Examination of the Operational Error Database for Air Route Traffic Control Centers. Report DOT/FAA/AM-93/22. FAA.
Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. Psychological Bulletin, 76, 45-48.
Roske-Hofstrand, R. J. (1989). Video in applied cognitive research for human-centered design. SIGCHI Bulletin, 21(2), 75-77.
RTCA (1995). Report of the RTCA Board of Directors Select Committee on Free Flight. Washington, DC: RTCA, Inc.
Sanderson, P. (1990). Verbal protocol analysis in three experimental domains using SHAPA. Proceedings of the Human Factors Society 34th Annual Meeting, (pp. 1280-1284). Santa Monica, CA: Human Factors Society.
Schaefer, D. (2001). Context sensitive speech recognition in the air traffic control simulation. EEC Note 02/2001. Eurocontrol.

Schaefer, D., Meckiff, C., Magill, A., Pirard, B. & Aligne, F. (2001). Air traffic complexity as a key concept for multi-sector planning. Presented at the Digital Avionics Systems Conference (DASC) 2001 meeting. Daytona Beach, Florida, USA.
Scerbo, M. W. (2001). Stress, workload and boredom in vigilance: a problem and an answer. In: Hancock, P. & Desmond, P. (Eds.) Stress, workload & Fatigue. Mahwah, New Jersey, USA: Erlbaum, 267 - 278.
Schaffer, M. (1991). The use of video for analysis of air traffic controller workload. Journal of Air Traffic Control, January - March, 35-37, 40.
Schmeidler, N & D'Avanzo, D.K. (1994). Development of staffing standards for air traffic control functions in tower cabs. Unpublished technical report.
Schmidt, D.K. (1978). A queuing analysis of air traffic controllers' workload. IEEE Transactions on Systems, Man and Cybernetics. SMCV-8(6),492-298.
Schneider, W. & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. Psychological Review, 84,1-66.
Seamster, T.L., Redding, R.E., Cannon, J.R., Ryder, J.M., & Purcell, J.A. (1993). Cognitive task analysis of expertise in air traffic control. International Journal of Aviation Psychology, 3 (4), 257-283.
Selcon, S. J., & Taylor, R. M. (1989). Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design. AGARD Conference Proceedings No. 478, Neuilly-sur-Seine, France.
Selcon, SJ (1990). Decision support in the cockpit. Probably a good thing? Proceedings of the Human Factors Society 34th Annual Meeting , pp 46-50, Santa Monica, California: Human Factors Society.
Sheridan, T.B. & Johannsen, G. (1976). Monitoring behavior and supervisory control. New York: Plenum.
Sheridan, T.B. (1987). Supervisory control. In G. Salvendy (Ed.) Handbook of human factors. New York: Wiley.
Simmons, B., Boan,L. & Massimini, P. (200).Simulation Analysis of Dual CRDA Arrival Streams to Runways 27 and 33L at Boston Logan International Airport . Report MTR 00W0000128. McLean, Virginia, USA: MITRE Corporation.
Sollenberger, R. L., Stein, E. S., & Gromelski, S. (1997). The development and evaluation of a behaviorally based rating form for the assessment of air traffic controller performance (DOT/FAA/CT-TN96-16). Atlantic City, NJ: DOT/FAA Technical Center.
Sperandio, J.C. (1971). Variations of operators' strategies and regulating effects on workload. Ergonomics, 14, 571-577.
Stager, P. & Hameluck, D. (1990). Ergonomics in air traffic control. Ergonomics, 33(4), 493-499.
Stager, P. (1991). The Canadian automated air traffic system (CAATS): an overview. In J. A. Wise et al. (Eds.), Automation and systems issues in air traffic control. Berlin: Springer-Verlag.
Stein, E.S. (1985) Air traffic controller workload: An examination of workload probe (Report FAA/CT-TN90/60). Atlantic City, New Jersey: FAA.
Stein, E.S. (1987). Where will all our air traffic controllers be in the year 2001? In Proceedings of the 4th Symposium on Aviation Psychology. pp. 195-201. Columbus, Ohio: Ohio Sate University.
Stein, E.S. (1988). Air traffic controller scanning and eye movements—A literature review. (DOT/FAA/CT-TN 84/24). Atlantic City Airport, New Jersey: Federal Aviation Administration Technical Center.
Stein, E.S. (1991). Air traffic controller memory: a field study (DOT/FAA/CT-TN90/60). Atlantic City, New Jersey: FAA Technical Center.
Stein, E.S. (1992). Air traffic control visual scanning (FAA Technical Report DOT/FAA/CT-TN92/16). Washington, DC: US Department of Transportation.
Tole, J.R., Stephens, A.T., Vivaudou, Eprath, A. & Young, L.R. (1983). Visual scanning behavior and pilot workload (NASA Contractor Report 3717). Hampton, Virginia: NASA Langley Research Center.
Thomas, R. P., Willem, B., Shanteau, J., Raacke, J., & Friel, B. M. (2001). CWS applied to controllers in a high fidelity simulation of ATC. Paper presented at the International Symposium on Aviation Psychology, Columbus, Ohio USA, March 2001.
Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185, 1124-1131.

Warm, J.S., Dember, W.N. & Hancock, P.A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.) Automation and Human Performance: Theory and Applications. Mahwah, NJ, USA: Lwrence Erlbaum.
Whitfield, D, Ball, RG, and Ord, G (1980). Some human factors aspects of computer-aiding concepts for air traffic controllers. Human Factors, 22(5), 569-580.
Whitfield, D., & Jackson, A. (1982). The air traffic controller's as an example of a mental model. In G. Johannsen & J.E. Rijnsdorp (Eds.), Analysis, design, and evaluation of man-machine systems: Proceedings of the IFAC/IFIP/IFORS Conference.
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.
Wickens, C.D. (1980). The structure of processing resources. In R. Nickerson and R. Pew (Eds.) Attention and performance VIII. Hillsdale, New Jersey: Erlbaum.
Wickens, C.D.(1992). Engineering Psychology and Human Performance. 2nd ed. New York: Harper Collins.
Wickens, C.D., Mavor, A.S. & McGee, J.P. (Eds.) (1997). Flight to the future: human factors in air traffic control. Washington, DC: National Academy of Sciences, National Research Council.
Woods, D. (1993). Process-tracing method for the study of cognition outside of the experimental psychology laboratory. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), Decision making in action: Models and methods. Norwood, NJ: Ablex.
Worm, A. (2001). On systems analysis and performance assessment in complex, high risk work environments. International Journal of Risk Assessment and management, 2(3/4).
Wyndemere (1996). An Evaluation of Air Traffic Control Complexity, Final report contract NAS2-14284. Boulder, Colorado USA: Wyndemere.
Young, F.W. & Hamer, R.M. (1987) Multidimensional Scaling: History, Theory and Applications Erlbaum, New York
Zijlstra, F.R.H & Doorn, L. van (1985). The construction of a scale to measure subjective effort. Technical Report, Delft University of Technology, Department of Philosophy and Social Sciences.

ANNEX A: COMPLEXITY FACTORS

The following table is an overview of the various complexity factors identified in this literature review. Notice that there were great differences across the literature in how much detail was provided for complexity factors. In some cases, factors were specified in great detail (including the measurement methods and units—such as “total number of flights effecting a heading change of more than 15 degrees, per hour”). In other cases, factors were mentioned without any clear indication of how to capture them operationally (e.g. “Military activity”). The occasional overlap between factors (below) reflects this inconsistency. Details are provided where available. Factors are categorised according to high level headings (e.g. *Airspace* factors or *Traffic Density* factors) to aid organisation of the table.

1. Aerodromes, number of airline hubs
2. Aerodromes, total number in airspace
3. Aircraft mix climbing and descending
4. Airspace, number of sector sides
5. Airspace, presence/proximity of restricted airspace
6. Airspace, proximity of sector boundary
7. Airspace, sector area
8. Airspace, sector boundary proximity
9. Airspace, sector shape
10. Airspace, total number of navaids
11. Conflicts, average flight path convergence angle
12. Conflicts, degree of flight path convergence
13. Conflicts, number of aircraft in conflict
14. Conflicts, number of along track
15. Conflicts, number of crossing
16. Conflicts, number of opposite heading
17. Conflicts, total time-to-go until conflict, across all aircraft
18. Convergence, presence of small angle convergence routes
19. Coordination, frequency of coordination with other controllers
20. Coordination, hand-off mean acceptance time
21. Coordination, hand-offs inbound, total number
22. Coordination, hand-offs outbound, total number
23. Coordination, number aircraft requiring hand-off to tower/approach
24. Coordination, number aircraft requiring vertical handoff
25. Coordination, number flights entering from another ATC unit
26. Coordination, number flights entering from same ATC unit
27. Coordination, number flights exiting to another ATC unit
28. Coordination, number flights exiting to same ATC unit
29. Coordination, number of communications with other sectors
30. Coordination, number of other ATC units acceptiing hand-offs
31. Coordination, number of other ATC units handing off aircraft
32. Coordination, total number LOAs
33. Coordination, total number of handofffs
34. Coordinations, total number required
35. Equipment status
36. Flight entries, number aircraft entering in climb
37. Flight entries, number aircraft entering in cruise
38. Flight entries, number aircraft entering in descent

39. Flight entries, number entering per unit time
40. Flight exits, number aircraft exiting in climb
41. Flight exits, number aircraft exiting in cruise
42. Flight exits, number aircraft exiting in descent
43. Flight Levels, average FL per aircraft
44. Flight Levels, difference between upper and lower
45. Flight Levels, number available within sector
46. Flight time, mean per aircraft
47. Flight time, total
48. Flight time, total time in climb
49. Flight time, total time in cruise
50. Flight time, total time in descent
51. Flight type, emergency / special flight operations, number
52. Flow organisation, altitude, number of altitudes used
53. Flow organisation, average flight speed
54. Flow organisation, complex routing required
55. Flow organisation, distribution of Closest Point of Approach
56. Flow organisation, flow entropy/structure
57. Flow organisation, geographical concentration of flights
58. Flow organisation, multiple crossing points
59. Flow organisation, number of altitude transitions
60. Flow organisation, number of current climbing aircraft proportional to historical maximum
61. Flow organisation, number of current descending aircraft proportional to historical maximum
62. Flow organisation, number of current level aircraft proportional to historical maximum
63. Flow organisation, number of intersecting airways
64. Flow organisation, number of path changes total
65. Flow organisation, routes through sector, total number
66. Flow organisation, vertical concentration
67. Other, controller experience
68. Other, level of aircraft intent knowledge
69. Other, pilot language difficulties
70. Other, radar coverage
71. Other, resolution degrees of freedom
72. Procedural requirements, number of required procedures
73. RT, average duration of Air-Ground communications
74. RT, callsign confusion potential
75. RT, frequency congestion
76. RT, frequency of hold messages sent to aircraft
77. RT, total number of Air-Ground communications
78. Separation standards (separation/spacing/standards)
79. Staffing
80. Time, total climb
81. Time, total cruise
82. Time, total descent
83. Traffic density, aircraft per unit volume
84. Traffic density, average instantaneous count
85. Traffic density, average sector flight time
86. Traffic density, localised traffic density / clustering
87. Traffic density, mean distance traveled
88. Traffic density, number flights during busiest 3 hours
89. Traffic density, number flights during busiest 30 minutes

90. Traffic density, number flights per hour
91. Traffic density, number of arrivals
92. Traffic density, number of current aircraft proportional to historical maximum
93. Traffic density, number of departures
94. Traffic density, total fuel burn per unit time
95. Traffic density, total number aircraft
96. Traffic distribution/dispersion
97. Traffic mix, aircraft type, jets vs props
98. Traffic mix, aircraft type, slow vs fast aircraft
99. Traffic mix, climbing vs descending
100. Traffic mix, military activity
101. Traffic mix, number of special flights (med, local traffic)
102. Traffic mix, proportion of arrivals, departures and overflights
103. Traffic mix, proportion of VFR to IFR pop up aircraft
104. Weather
105. Weather, at or below minimums (for aerodrome)
106. Weather, inclement (winds, convective activity)
107. Weather, proportion of airspace closed by weather
108. Weather, reduced visibility

ANNEX B: DATA COLLECTION METHODS

The following is a general overview of the types of methods that can be used to elicit, refine, and validate ATC complexity factors. This list expands that presented in table 1, section 6.2, and incorporates methods that have not necessarily been applied yet to the study of ATC workload and/or complexity. Four main categories of methods are distinguished: Observational, Interview-based, Experimental, and Analytic. Notice that many of the listed methods encompass a number of specific techniques ("Task Analysis," for instance, refers to a whole family of possible techniques and tools for decomposing task performance). The following also provides an assessment of the model development stage (Factor Elicitation, Refinement, or Validation abbreviated *E*,*R*, and *V* respectively) at which the method seems most useful to the COCA project. This assessment was based on knowledge of the types of data each method provides, but also on the costs and difficulties (e.g. intrusiveness, equipment costs, time, etc.) associated with each.

Method	Stage
OBSERVATONAL	
1. Field Observations	*E, R, V*
2. Familiarisation	*E*
3. Videotape analysis	*E, R, V*
4. Over-the-shoulder assessments	*E, R, V*
5. Behavioural analysis	*E, R,V*
Activity analysis	
Communication analysis	
Voice command analysis	
INTERVIEW	
6. Expert judgement sessions	*E, R*
Eg meta-evaluation	
7. Rating and ranking exercises	*E, R*
8. Focus groups	*E, R*
9. Cognitive walkthrough	*E, R*
10. Critical decision making	*E, R*
11. Verbal protocol analysis	*E, R*
12. Structured interview	*E, R*
13. Semi-structured field interview	*E, R*
14. Workload ratings	*R, V*
Instantaneous	
eg ISA	
Post session	
eg TLX, ATWIT	
15. Questionnaires / Surveys	*E, R*

Method	Stage
EXPERIMENTAL / QUASI-EXPERIMENTAL	
16. Static simulations	*E, R*
17. Small scale simulation	*E,R*
18. Shadow mode trials	*R, V*
19. Physiological measurement	*R, V*
Eye tracking measures	
Heart rate measures	
Brain activity measures	
20. Task performance	*R, V*
Primary task performance	
Secondary task performance	
ANALYTIC	
21. Task Analysis	*E, R*
22. Repertory Grid Analysis	*E,R*
23. Task modelling	*E,R*
eg MicroSAINT, HOS	
24. Cognitive Modelling	*V*
IPME, PUMA, MIDAS	
25. Monte Carlo simulation	*V*
26. Analysis of operational data	*R, V*
27. Analysis of modelled workload	*R, V*

ANNEX C: WORKLOAD INDICATORS

1. Air Traffic Workload Input Technique (ATWIT)
2. Auditory Choice Secondary Task Response Time
3. Bedford Workload Scale
4. Blink latency
5. Blink rate
6. Blink-saccade asynchrony
7. Blood Pressure
8. Brain Evoked Potentials (a.k.a. Evoked Response Potentials, or ERPs)
9. Card Sorting secondary task performance
10. Choice Reaction Time secondary task performance
11. Classification secondary task performance
12. Cooper Harper Rating Scale
13. Dichotic Listening performance
14. Electro-encephalographic (EEG) pattern
15. Event Related (Evoked Cortical) Potential
16. Galvanic Skin Response GSR
17. Head down time
18. Heart Rate
19. Heart Rate Variability
20. Instantaneous Self Assessment (ISA)
21. Lexical Decision secondary task performance
22. Magneto-encephalographic (MEG) activity
23. Mental Arithmetic performance
24. Modified Cooper-Harper Rating Scale
25. NASA Task Load Index (TLX)
26. Pilot Objective/Subjective Workload Assessment Technique (POSWAT)
27. PUMA
28. Pupil diameter
29. Radio Communications embedded task
30. Respiration rate
31. Scanning entropy (randomness)
32. Secondary task detection response time
33. Secondary task Interval production accuracy
34. Secondary Task signal detection hit rate
35. Secondary Task Time estimation accuracy
36. Serial Recall
37. Sternberg auditory memory search performance
38. Sternberg visual memory search performance
39. Subject Matter Ratings
40. Subjective Workload Assessment Technique (SWAT)
41. Taylor et al 7-point rating scale questionnaire
42. Time Estimation Secondary Task
43. Verbal Protocol Analysis
44. Visual Dwell Time
45. Visual Fixation Frequency
46. Visual Saccade Duration
47. Visual Saccade Rate