We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Air Traffic Complexity as a Source of Risk in ATM

Tomislav Radišić, Petar Andraši, Doris Novak, Biljana Juričić and Bruno Antulov-Fantulin

Abstract

In this chapter the connection between air traffic complexity and risks in air traffic management system will be explored. Air traffic complexity is often defined as difficulty of controlling a traffic situation, and it is therefore one of the drivers for air traffic controller's workload. With more workload, the probability of air traffic controller committing an error increases, so it is necessary to be able to assess and manage air traffic complexity. Here, we will give a brief overview of air traffic complexity assessment methods, and we will put the traffic complexity assessment problem into a broader context of decision complexity. Human reliability assessment methods relevant to air traffic management will be presented and used to assess the risk of loss of separation in traffic situations with different levels of complexity. To determine the validity of the human reliability assessment method, an analysis of conflict risk will be made based on the real-time human-in-the-loop (HITL) simulations.

Keywords: air traffic complexity, risk, human reliability assessment, air traffic control, simulation

1. Introduction

Humans are at the core of every complex system in the world, and that is true for the air traffic management (ATM) system as well. While extremely resourceful and capable of dealing with unexpected circumstances, humans are also prone to errors. Although significant technological, organizational, operational, and other advances have been made in recent decades, catastrophic accidents driven by human errors are still a regular, albeit increasingly rare, occurrence. Recently, the realization that complete elimination of all human errors will probably never be achievable has took hold [1]. As with any system that requires high degree of safety, ATM system solves this issue by employing multiple levels of risk and safety management, each providing a layer of the safety net. Nevertheless, methods for reducing the human error are still widely used and being researched. These methods, which consider the effect of human error on risk and reliability, are generally classified under the name of human reliability analysis (HRA). This chapter explores the applicability of HRA as part of the overall risk assessment with a focus on air traffic complexity issues.

There are many motivations for performing a risk or reliability analysis. In most cases it is to reduce the potential for system failure caused by humans. In this case the risk analysis can be used either in the design process or during the operation.

Sometimes it is needed to change or restructure the organizational design in a manner which ensures at least the same level of safety as before. In other cases, risk analysis can be performed as a part of licensing arrangements where the operator is tasked with assuring that a system meets a safety target. Or it can be used during the decision-making process where an operator chooses one of the possible systems to procure. In many of these cases, HRA will be undertaken as part of the more comprehensive risk assessment process.

Air traffic controllers (ATCOs) are at the core of the ATM. They are the central node where most important safety-related tactical decisions are made. Their job is to gather information and process them with the goal of reaching solutions which ensure safe and cost-efficient air traffic. One of their main tasks is prioritization of actions because human mental capacity is limited and it has been shown that ATCOs frequently deal with information overload [2]. Previous research showed that overload usually causes performance decline [3]. Air traffic complexity is one of the main factors driving the increase in the ATCO workload, so it is a reasonable assumption that increased complexity will result in increased errors due to decay in ATCO performance. Therefore, it is important to be able to assess air traffic complexity as a possible source factor of risk.

In this chapter, the connection between air traffic complexity, controller workload, HRA, and risk assessment will be made. For that purpose, in Section 2, a brief overview of complexity will be made, starting with definition and ending with assessment methods. In Section 3, a broader area of decision complexity and inherent difficulty of making the correct decision in a complex system will be presented. In Section 4, a very brief overview of HRA methods relevant to ATM system will be presented, as well as an HRA method developed specifically for use in ATM. In the latter parts of Section 4, an example of how to include the air traffic complexity into the HRA will be shown, and, in comparison, risk analysis based on real-time human-in-the-loop (HITL) simulations will be presented.

2. Air traffic complexity

2.1 Definition and purpose

The Random House dictionary defines complexity as "the state or quality of being complex; intricacy", and complex as "composed of many interconnected parts; compound; composite", "characterized by a very complicated or involved arrangement of parts, units", *and* "so complicated or intricate as to be hard to understand or deal with" [4]. While this example uses *complicated* to define *complex*, some other sources argue that there is a major difference between the two. Collins English Dictionary states that [5]:

> *Complex is properly used to say only that something consists of several parts. It should not be used to say that, because something consists of many parts, it is difficult to understand or analyze.*

On the other hand, Cilliers, in his seminal book on the topic, claims exactly the opposite [6]:

> *If a system—despite the fact that it may consist of a huge number of components—can be given a complete description in terms of its individual constituents, such a system is merely complicated. [...] In a complex system, on the other hand, the interaction among constituents of the system, and the interaction between the*

> *system and its environment, are of such a nature that the system as a whole cannot be fully understood simply by analysing its components.*

One example of such thinking is presented by Snowden in [7]. He claims that the aircraft can be considered complicated due to many parts. Once disassembled and analyzed, the function of all parts and their relationships can be determined. Human organizations and systems are, on the other hand, complex. They are made up of many interacting agents, with agent being any component of the system with identity. Agents can have multiple identities based on the context, i.e., a person can assume group identity or switch between formal and informal identities based on the environment. As these identities change, the components of the system change, the rules an agent follows change, and interactions between the components change. This makes it impossible to distinguish between the cause and effect because they are intertwined [8].

In the context of air traffic control, complexity was rarely clearly defined, perhaps due to assumed common knowledge. One notable exception is Meckiff (et al.) who stated that the air traffic complexity can be most easily defined as *difficulty of monitoring and managing a specific air traffic situation* [9]. It is intuitively clear that it is easier for the air traffic controller to monitor the airspace sector in which aircraft trajectories do not intersect and there are no level changes than the sector in which there are a lot of merging traffic flows and aircraft often change levels. As such, air traffic complexity could also be defined as a number of potential aircraft-aircraft and aircraft-environment interactions during a given time frame. Not all of these interactions require the same level of attention, urgency, or, ultimately, controller workload to resolve.

Complexity is *not* the same as traffic density. Obviously, the number of aircraft in a sector (also known as density, traffic load, or traffic count) directly influences the air traffic complexity. This number, however, is not the only indicator of the level of complexity, especially if one wishes to compare different sectors of airspace [10–12]. Two traffic situations can have equal density but vastly different complexity. Due to two different types of interactions, some researchers have chosen to make a distinction between **airspace complexity** (also static, structural) and **air traffic complexity** (also dynamic, flow complexity [13]) which is influenced by the airspace complexity. This distinction will be used in this chapter as well. Unless explicitly stated, complexity will from now on refer exclusively to air traffic complexity.

Complexity is *not* a synonym for workload, although it has been proven multiple times that the increase in complexity results in increase in workload which in turn limits the airspace sector capacity [14, 15]. Mogford et al. [11] reviewed numerous research articles in search of complexity and workload relationship. They concluded that the complexity is actually a source factor for controller workload (**Figure 1**). However, complexity and workload are not directly linked. Their relationship is mediated by several other factors, such as equipment quality, individual differences, and controller cognitive strategies [11].

Controller cognitive strategies can be improved through training and experience that is readily seen when comparing experienced and inexperienced controllers. However, if one takes into consideration an average controller with average training, only two avenues to reduced controller workload remain—increasing equipment quality and decreasing complexity.

2.2 Previous research on air traffic complexity

Complexity was a common research topic since the early days of modern ATC operations. First papers that mention complexity were written in the early
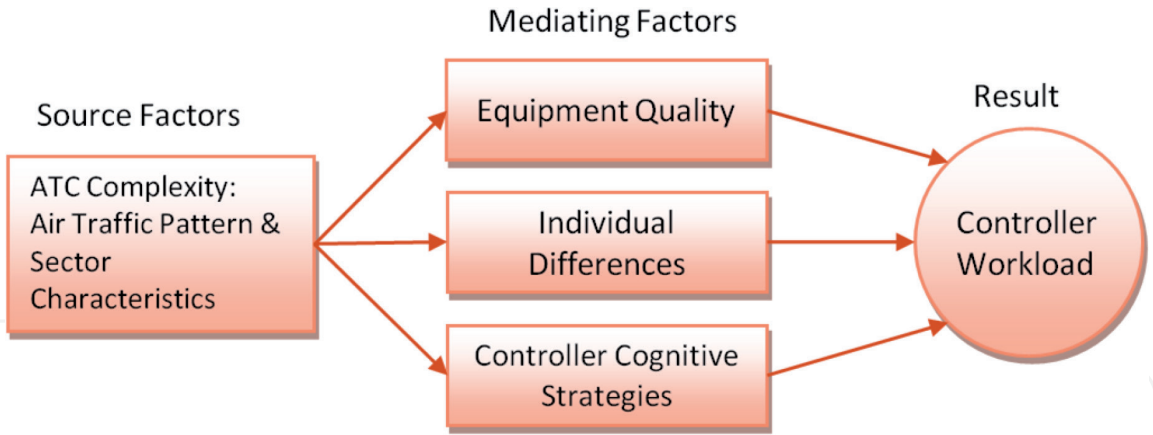
Figure 1.
The relationship between air traffic complexity and workload.

1960s [16]. Since then, dozens of papers and reports were written on the topic of complexity—excellent reviews of those papers were written by Mogford [11] and Hilburn [17]. Instead of writing a completely new literature review, this chapter will present important research paths, ideas, methods, and facts, which are relevant to the present research.

It needs to be noted that most of the early research was conducted in order to better define factors that affect workload. Today, most of those factors, with present understanding and definitions, would probably be called complexity factors. Some studies were nonempirical and lack exact definitions and measurement methods for complexity indicators. Those studies were excluded from this short review to give more room to those studies with experimentally validated complexity factors.

Schmidt [18] approached the problem of modelling controller workload from the angle of observable controller actions. He created the control difficulty index, which can be calculated as a weighted sum of the expected frequency of occurrence of events that affect controller workload. Each event is given a different weight according to the time needed to execute a particular task. Though the author conducted extensive surveys to determine appropriate weights and frequencies for various events, this approach can only handle observable controller actions, which makes it very limiting.

Hurst and Rose [19], while not the first to realize the importance of traffic density, were first to measure the correlation of expert workload ratings with traffic density. They concluded that only 53% of the variance in reported workload ratings can be explained by density.

Stein [20] used Air Traffic Workload Input Technique (ATWIT), in which controllers report workload levels during simulation, to determine which of the workload factors influenced workload the most. Regression analysis proved that out of the five starting factors, four factors (localized traffic density, number of handoffs outbound, total amount of traffic, number of handoffs inbound) could explain 67% of variance in ATWIT scores. This study showed the importance of localized traffic density which is a measure of traffic clustering. Technique similar to ATWIT will be used throughout the next three decades, including a modified ATWIT scores that will be used in this research.

Laudeman et al. [21] expanded on the notion of the traffic density by introducing dynamic density which they defined as a combination of "both traffic density (a count of aircraft in a volume of airspace) and traffic complexity (a measure of the complexity of the air traffic in a volume of airspace)." Authors used informal interviews with controllers to obtain a list of eight complexity factors to be used in

dynamic density equation. The only criterion was that the factors could be calculated from the radar tracks or their extrapolations. The intention was to obtain an objective measure of controller workload based on the actual traffic. Their results showed that the dynamic density was able to account for 55% of controller activity variation. Three other teams [13, 22, 23] working under the Dynamic Density program developed additional 35 complexity indicators (factors), which were later successfully validated as a group by Kopardekar et al. [24]. Unfortunately, it was later shown that the complexity indicator weights were not universal to all airspace sectors, i.e., they had to be adjusted on a sector-by-sector basis [25]. This shortcoming, while making the dynamic density technique difficult to implement for operational purposes, has no influence if one wishes to compare two concepts of operations under similar conditions (similar sector configuration). Furthermore, same authors [24] suggested that, due to possibly nonlinear interactions between complexity factors, the dynamic density performance could be improved by using nonlinear techniques such as nonlinear regression, genetic algorithms, and neural networks.

Almost the same group of authors will use multiple linear regression method 5 years later to determine which subset of complexity indicators will correlate well with the controller's subjective complexity ratings [26]. After extensive simulator validation, results of this study showed that there are 17 complexity indicators that are statistically significant. Top five complexity indicators were sector count, sector volume, number of aircraft under 8 NM from each other, convergence angle, and standard deviation of ground speed/mean ground speed. Similar work was done by Masalonis et al. [27] who selected a subset of 12 indicators and Klein et al. [28] who selected a subset of only 7 complexity indicators, though with less extensive experimental validation.

In a similar vein, Bloem et al. [29] tried to determine which of the complexity indicators had the greatest predictive power in terms of future complexity. The authors concluded that there is a significant difference in predictive power of different complexity indicators. To complicate the matter further, they concluded that the subset of the complexity indicators that had the best predictive power changed depending on the prediction horizon.

To calculate potential impact of air traffic complexity on workload and costs, in 2000 the EUROCONTROL has given the same set of traffic data to UK National Air Traffic Services (NATS) and the EUROCONTROL Experimental Centre (EEC) with a task of independently devising a method of measuring the level of service [30]. While NATS has estimated ATS output (the service provided), the EEC has estimated the ATS workload needed to deliver the service. Both "were found to produce reasonably consistent results," with an additional note that further analysis should be done before the final parameters for determining ATS provider costs are established. By 2006 EUROCONTROL's Performance Review Commission finalized the complexity indicators to be used for ANSP benchmarking [31]. For this method the European airspace is divided into 20 NM X 20 NM X 3000 ft. cells, and for each cell the duration of potential interactions is calculated. Aircraft are "interacting" if they are in the same cell at the same time. The ratio of the hours of interactions and flight hours is the so-called adjusted density. In addition, the "structural index" is calculated as a sum of potential vertical, horizontal, and speed interactions. The final complexity score is calculated as a product of adjusted density and structural index. All in all, only four complexity indicators are used for this analysis, and no validation of any sort was presented in the report. It was noted, however, that shifting the starting position of the grid by 7 NM caused the ANSP ranking to change dramatically (up to 16 places in an extreme case). Nonetheless, this method is still used for ANSP benchmarking.

First to consider measuring complexity during TBO were Prevot and Lee [32]. They coined the term trajectory-based complexity (TBX) which is a measure of complexity in TBO. The basis of the TBX calculation is a set of nominal conditions—nominal sector size, nominal number of transitioning aircraft, and a nominal equipage mix. Any difference to nominal operations causes a modification to the TBX value. Authors do not explain the method to determine the nominal conditions except that they can "be defined through knowledge elicitation sessions on a sector by sector basis or based upon more generic attributes." The TBX value is then a number of aircraft that would produce the same workload under the nominal conditions as do aircraft under real conditions (e.g., the TBX of 20 means that the workload is equal to the aircraft count of 20 under nominal conditions even though there are actually only 16 aircraft in the sector). The advantage of this method is that it gives a single complexity value that can be easily related to aircraft count and is thus very user-friendly and self-explanatory (unlike many other complexity metrics). However, this study included only six complexity indicators with weights that were determined in an ad hoc manner and hardly any validation with actual subjective complexity. Only one of those complexity indicators was indirectly related to TBO (number of aircraft with data-link). Many human-in-the-loop simulation runs were performed in which the controllers had to give workload scores which were then compared with TBX value and simple aircraft count. While the authors claim that the subjective workload score correlated better with the TBX value, there was no objective correlation assessment presented. Finally, the authors have not compared the effect of fraction of TBO aircraft on air traffic complexity.

In a subsequent paper by same authors, the relationship between workload and data-link equipage levels was explored [33]. It was concluded that the workload ratings correlated much better with the TBX score than with the aircraft count for varying data-link equipage levels.

Prandini et al. have developed a new method of mapping complexity based exclusively on traffic density [34]. This method is applicable only to the future concept of aircraft self-separation and does not take into account the human factors at all.

Gianazza [35–37] proposed a method for prediction of air traffic complexity using tree search methods and neural networks. This method is based on the assumption that the air traffic complexity in historic flight data increased prior to the splitting of the collapsed sector into two smaller ones and decreased prior to collapsing the sectors into a larger one. The neural network was trained using this historical data, and then it could predict future increase in air traffic complexity. Tree search method was then used to determine the airspace configuration which yields lowest workload and complexity for the given air traffic pattern.

Lee et al. [38] have proposed that airspace complexity can be described in terms of how the airspace (together with the traffic inside it and the traffic control method) responds to disturbances. The effect of disturbances on control activity needed to accommodate that disturbance is what defines complexity in their opinion. The more control activity needed, the more complex the airspace is. They propose a tool, airspace complexity map, which should help to plan the airspace configuration and the future development of ATM.

In Radišić et al. [39], authors used domain-expert assessment to test the effect of the trajectory-based operations (TBO) on air traffic complexity. ATCOs were recruited to perform human-in-the-loop (HITL) simulations during which they were asked to provide real-time assessment of air traffic complexity. Linear regression model was used to select, among 20 most used complexity indicators, those indicators which correlated best with subjective complexity scores. Six indicators were used to generate a predictive linear model that performed well in conventional

operations but less so under TBO. Therefore, the authors defined and experimentally validated two novel TBO-specific complexity indicators. A second correlation model combining these two novel indicators with four already in use generated much better predictions of complexity than the first model. Nonetheless, the best correlation that was achieved was R = 0.83 (R^2-adjusted = 0.691). In subsequent work, the authors attempted to achieve better prediction by using artificial neural networks; however, similar results were obtained. This indicates that there is some variation in subjective complexity scores provided by ATCOs that cannot be explained by traffic properties. Indeed, it might be the case that ATCOs introduce a degree of noise into the complexity scores due to difficulty of maintaining the consistent scoring criteria [40].

Wang et al. [41] in their work used network approach to calculate air traffic complexity based on historical radar data. Their assumption is that air traffic situation is essentially a time-evolving complex system. In that system aircraft are key waypoints; route segments are nodes; aircraft-aircraft, aircraft-keypoint, and aircraft-segment complexity relationships are edges; and the intensities of various complexity relationships are weights. The system was built using a dynamic weighted network model.

Xue et al. [42] in their work analyzed three complexity indicators for simulated UAS traffic: number of potential conflicts, scenario complexity metric, and number of flights. Scenario complexity metric is based on cost of pairwise conflict which is defined as deviation from the original path. To perform analysis on around 1000 scenarios at different density levels, authors had to develop a UAS simulator. Analysis was done using Pearson and ACE statistics methods.

Future concept of operations will involve usage of far wider range of air traffic controller tools; therefore, it is expected that new complexity indicators related to interaction of controllers and equipment will have to be developed. Furthermore, novel complexity assessment methods are needed due to limits of current techniques.

2.3 Complexity estimation methods

In this section, several air traffic complexity estimation methods will be examined in greater detail. All complexity estimation methods are based on the traffic data which describes a traffic situation. Since the complexity is a psychological construct, the most relevant estimator of complexity in a given traffic situation is the air traffic controller. The air traffic controller can look at the traffic data and decide whether a traffic situation is complex or not. All other methods are just attempts at approximating the level of complexity as estimated by the controller. The main problem with expert-based estimation is the inconsistency between controllers, where one controller gives a different complexity estimate than the other. Therefore, most other methods seek ways to make the complexity estimate without human input. Ideally, those other methods would be validated by comparing them to the expert, i.e., controller's estimate; however, this is not always the case.

Three main methods of control-based (i.e., based on ATCOs' experience of complexity as a driver for workload and, subsequently, limiting factor of airspace capacity) air traffic complexity estimation will be presented here:

- Expert-based air traffic complexity estimation—where an expert, in most cases an air traffic controller, gives their estimate of the complexity

- Indicator-based air traffic complexity estimation—where the values of complexity indicators, derived from traffic data, are used to determine the level of complexity

- Interaction-based air traffic complexity estimation—where the complexity is estimated on the basis of the number of aircraft interactions in a given airspace cell (this method could be broadly defined as a very narrow indicator-based complexity estimation method due to a very low number of indicators)

- Others—methods based on other principles, such as counting the number of clearances [43], evaluating proximity based on probabilistic occupancy of airspace [44], measuring sensitivity to initial conditions of the underlying dynamic system called Lyapunov exponents (i.e., assessing predictability of traffic) [45], and many others

3. ATC operations and the decision domains

The decision-making process needs to be adapted to the context in which the operations take place. It is often seen that one kind of decision-making, completely adapted to its environment and therefore useful, cannot be easily transferred to another environment. This is often the case with accomplished engineers being notably less successful after moving into the managerial role.

Classification of such environments and appropriate decision-making modes is sometimes attempted with the goal of making rules about the best ways to manage each context. However, this is not an exact science because there are multiple factors that can change the decision-making context depending on who the person making the decision is, or how experienced they are. Nevertheless, there is still utility in being aware of the environment in terms of decision contexts and learning how to detect when the environment shifts from one domain to another.

One such classification attempt is the *Cynefin* framework [7]. It was developed in the early 2000s as a tool for decision-making, and it proposes five decision domains:

- Simple (also, obvious)—In this domain the situation is well known and stable. The cause–effect relations are established and rarely change. Following procedures and best practices is the best course of action to ensure efficient realization of goals. Decision-making process is usually made of the sense-categorize-respond steps. A major issue in this domain is the overreliance on patterns and routine behavior which stifles innovation and precludes any change. This has caused many issues in the past when organizations were not willing to adapt to changes or innovate, but on the other hand, this has also created many opportunities for disruption by newcomers.

- Knowable (also, complicated)—This domain includes environments in which not everything is known but everything can be understood with enough time and effort. In *knowable* domain the experts can work rationally towards solutions by sensing the environment, analyzing the data, and applying the best practices. In contrast with *simple* domain, where the main part of the task is applying the best practices, in *knowable* domain most of the effort is spent analyzing the situation.

- Complex—In this domain are environments or systems which cannot be analyzed by breaking them down into smaller pieces, analyzing them individually, and creating the big picture based on the analysis of individual components. The very act of interacting with the system introduces changes which cannot always be predicted. The main mode of management of complex systems is through observation of patterns, finding ways to sustain those patterns we

desire, and disrupting those we do not. One particular phenomenon that arises in complex systems is the so-called retrospective coherence. The state of the system seems logical and coherent once it is retrospectively analyzed; however, current state of the system could hardly be anticipated in advance because there are many other equally plausible system states.

- Chaotic—Chaotic systems cannot be analyzed for cause and effect relationships. Patterns are not visible, and if one waits for patterns to emerge, the damage could become disastrous. It is in these conditions that the system is most difficult to manage but also most capable of change, for better or for worse.

Air traffic control is all about making decisions, so it is not a novel idea to apply the Cynefin framework to the ATC operations even though Cynefin was originally proposed for business-related decision-making [46]. Air traffic control is a complex system with numerous human and machine agents, organized in deep layers of components glued by multiple communication modes and protocols. This is even more apparent in air traffic management systems. Although someone might look at the routine ATC operations and consider them simple, or even mechanistic, such thinking is a sure way towards probably costly failure. In our opinion, ATC operations can be assigned to all domains depending on the traffic situation or changes in states of the system:

- During the nominal low-traffic ATC operations, the traffic situation is easy enough in terms of workload to be considered as belonging to the *simple* domain. The ATCO needs to sense the traffic situation or a particular part of it, usually by looking at the radar screen and talking to the pilots. Then they need to categorize the task that needs to be performed in order to ensure safe and efficient traffic. The task can be categorized as any of the numerous routine ATC tasks, e.g., conflict resolution, clearing or initiating climbs or descents, managing exit flight level constraints, etc. Then the ATCO acts by issuing a command or a clearance. This process occurs many times an hour, and some parts of it are trained to such a degree that the ATCO is often not even conscious of them.

- In nominal high traffic ATC operations, the number of interactions rises and so does the difficulty of maintaining safe and efficient air traffic. The situation needs to be sensed and then analyzed for all the tasks that need to be performed. Tasks are often prioritized based on the urgency and difficulty. A lot more time is spent on this analysis than in low-traffic situation. The ATCO then solves the issues by applying solutions that are considered to be best practice. There are multiple ways of solving an issue, and all of them are correct if the safety is maintained and flight efficiency is not unreasonably reduced. Unless there is some source of major uncertainty present, such as adverse weather conditions, this type of operations is best described as belonging to the *knowable* domain.

- In off-nominal operations of any traffic level or nominal operations with a major source of uncertainty, such as adverse weather, the decision context often enters the *complex* domain. The traffic situation evolves into unpredictable directions which can be completely explained only post hoc. Systemic complexity management measures, such as regulations, are undertaken to ensure safety because continuing with business as usual could lead, with unacceptable probability, to incidents or accidents. Nonetheless, these measures

are sometimes not enough or are compounded with additional issues which altogether cause the loss of situational awareness for the ATCO or the pilots. Incidents lurk in these conditions.

- Operations in the *chaotic* domain should never happen in ATC. The whole system is designed to prevent such occurrences. However, history has shown us that there are sequences of events that can throw the whole system into a disarray and shift the decision context very quickly from the *simple* into the *chaotic* domain. One example of such a sequence is Croatia Control's area control center (ACC) outage of 2014 when flooding due to unprecedented rainfall combined with human error and organizational deficiencies caused the complete loss of power to all ATC systems for 2 h [47]. When radar screens went blank, quick-thinking ATCOs used their personal mobile phones to contact ACCs of neighboring countries to warn them of potential conflicts, thus preventing midair collisions. This incident clearly illustrates how quickly a situation can go from bad (*complex* domain, operations in adverse weather) to worse (*chaotic* domain, complete loss of power).

It should be noted here that air traffic complexity should not be confused with *complex* domain in the Cynefin framework. Air traffic complexity is present in all decision domains, usually being lower in the *simple* domain and higher at the other end of the spectrum in the *chaotic* domain.

The main purpose of this classification of decision contexts in ATC is to help make ATCOs and supervisors aware of the different environments that are possible behind the seemingly unchanging radar screen. Another purpose, which will be discussed in the next section of this chapter, is to lay down the framework for assessing risks associated with air traffic complexity.

4. Assessing risks associated with air traffic complexity

Complexity in ATM is often split into two parts: airspace complexity (static complexity) and air traffic complexity (dynamic complexity). It is generally agreed that both dynamic and static components of complexity can affect controller workload and influence the probability of occurrence of an ATC (i.e., controller) error. Dynamic complexity relates to the factors describing air traffic complexity, i.e., it can include factors such as traffic volume, climbing/descending traffic, mix of aircraft type, military area activity, and types of aircraft intersection. Static factors, on the other hand, encompass factors related to the airspace, such as airspace structure, proximity of reporting points to sector boundaries, and standing agreements between ANSPs.

In a human factors study, areas rated as some of the biggest contributors to risk in ATM are workload, human error, allocation of function, and situational awareness [48]. As mentioned previously, air traffic complexity is a measure of difficulty of controlling the air traffic in a given sector; therefore, it is a direct contributor to workload. In a sense, ATCO's job is to make correct decisions, whereas air traffic complexity is a factor that makes the search for the right decision more difficult. Therefore, increased complexity can directly increase the probability of a wrong decision being made because the size of the search space increases faster than the set of correct solutions. Here lies the main connection between air traffic complexity and risk. Probability of human error (i.e., human error risk) increases with increased complexity. Thus, it is reasonable to assess the complexity-related risks from the human reliability assessment point of view.

EUROCONTROL investigated the possible relationship between ATM system complexity and safety. They tried to develop a complexity hazard and operability (HAZOP) technique with the main objective being to trial this approach and evaluate its utility for safety assessment and obtain feedback on its acceptability with operations personnel [49]. The attempt at developing complexity HAZOP was unsuccessful due to difficulty of adjusting the HAZOP technique to the complexity issues. Therefore, in this section only HRA methods will be presented.

4.1 Human reliability assessment

This section will provide a brief overview of the HRA methods and their development over the years; however, for a more thorough review of HRA methods, one can find more information in [46, 50]. Only those methods that are in some way relevant to HRA in aviation will be considered.

HRA can be defined as "any method by which human reliability is estimated" [51], and it is generally presented as having three main parts: (1) identifying possible human errors and contributors, (2) modelling human error, and (3) quantifying human error probabilities. These methods were first developed in nuclear power safety systems.

In the early models of HRA, human was often considered as just another part of the system. For example, in [52], a *technique for human error-rate prediction* (THERP) was developed based on the techniques used in nuclear power plant risk management, i.e., a straightforward event tree analysis was performed. Each human action (e.g., reading a display, operating a lever) was given a human error probability (HEP) as a probability with a value from 0 (least probable) to 1 (most probable). Sample of values for different errors can be seen in **Table 1**. The values assigned to each error type came from authors' experience and from earlier studies performed in the defense sector.

THERP also specified performance shaping factors (PSF) which were used to modify the nominal HEPs based on the context of the action (e.g., time pressure, human-machine interface, etc.). A list of possible PSFs for one error is given in **Table 2**. One can notice that there are no error multipliers associated with each PSF. It is the duty of the assessor to define the maximum affect that each PSF could have on HEPs. Criticism of THERP was mostly that it was too difficult to apply because of quite detailed decomposition of tasks that it relied on a database of HEPs which was never really validated and that it took very broad and casual definitions of human performance factors.

Error	HEP
Failure to perform rule-based actions correctly when written procedures are available and used (with recovery)	0.025
Inadvertent activation of a control; select wrong control on a panel from an array of similar-appearing controls identified by labels only	0.003
Omitting a step or important instruction from a formal or ad hoc procedure	0.003
Omitting an item of instruction when use of written procedures is specified (<10 items)	0.001
Checking the status of equipment if that status affects one's safety when performing his tasks	0.001
Turn rotary control in the wrong direction when there is no violation of populational stereotypes	0.0005
Errors of commission in check-reading analog meters with easily seen limit marks	0.001

Table 1.
Examples of HEPs given in [52].

1	Stress level of the operator	
2	Rate at which the operator must process signals	
3	Frequency with which a particular display is scanned	
4	Whether a written checklist is used to direct the operator to specific displays	
5	Relationship of the displays to annunciators or other attention-getting devices	
6	Extent to which the information needed for operator decisions and actions is displayed directly	
7	Human factors engineering related to the design and arrangement of the displays	

Table 2.
Examples of PSFs for errors related to reading unannunciated displays [52].

Another version of this type of model was done in human error assessment and reduction technique (HEART) [53]. The database of HEPs was much smaller and more generic, so it was more flexible and easier to apply than THERP. Instead of highly detailed errors, the focus is on a handful of generic task types for which probabilities of failure are given (**Table 3**). This simplification has made the HEART technique much more accepted outside the nuclear power industry for which the THERP was designed.

The author has identified the human factors he found relevant by searching the human factors literature and assigned relative weights to them, identified impacts of errors, and suggested a set of human error data which should enable higher reliability of the system. Instead of calling them PSFs, the author called them error-producing condition (EPC) and provided the multipliers for each. Multipliers are used to increase the nominal human unreliability in cases where there are circumstances that increase the probability of human error. Some of the EPCs are shown in **Table 4**.

More generic error types have led to confusion when trying to apply it to a specific industrial application. This problem was addressed by developing specialized versions of HEART for specific industries. One such derivative will be discussed shortly.

These models are characterized by defining two broad categories of errors: errors of omission (when human operator fails to make an action) and errors of commission (when operator makes a wrong action). These simplifications were later put, at least partially, into the context of actual human behavior which knows many other ways of committing an error. For example, [54, 55] included contextual effect such as stress, organizational culture, and tiredness into the model, whereas [56, 57] also included the possible variation in operator's responses and recovery actions undertaken once the errors have been noticed. By taking into account the context of human behavior, these techniques have made a qualitative step forward in comparison to the THERP and HEART, so they are generally called second-generation HRA techniques. This did not, however, improve their adoption in the industry because simpler and more flexible techniques, such as HEART, are more usable and sustainable. For this reason, the first HRA technique developed specifically for ATM was based on HEART technique. It was developed in 2008 and named Controller Action Reliability Assessment (CARA) [58].

4.2 Human reliability assessment in ATC

Compared to HEART, CARA's generic task types were developed to better suit the needs of HRA in ATM (**Table 5**). To make sure that the task types are in line with the commonly used models of ATCO tasks, the basis for task development was

Generic task	Proposed nominal human unreliability	5th–95th percentile bounds
Totally unfamiliar, performed at speed with no real idea of likely consequences	0.55	0.35–0.97
Shift or restore system to a new or original state on a single attempt without supervision or procedures	0.26	0.14–0.42
Complex task requiring high level of comprehension and skill	0.16	0.12–0.28
Fairly simple task performed rapidly or given scant attention	0.09	0.06–0.13
Routine, highly practiced, rapid task involving relatively low level of skill	0.02	0.007–0.045
Completely familiar, well-designed, highly practiced, routine task occurring several times per hour, performed to highest possible standards by highly motivated, highly trained, and experienced person, totally aware of implications of failure, with time to correct potential error but without the benefit of significant job aids	0.0004	0.00008–0.009
Respond correctly to system command even when there is an augmented or automated supervisory system providing accurate interpretation of system state	0.00002	0.000006–0.0009

Table 3.
Generic tasks and proposed human unreliability in HEART technique [53].

Error-producing condition	Maximum predicted increase in unreliability when going from good conditions to bad
Unfamiliarity with a situation which is potentially important but which only occurs infrequently or which is novel	×17
A shortage of time available for error detection and correction	×11
A low signal-to-noise ratio	×10
A means of suppressing or overriding information or features which is too easily accessible	×9
No means of conveying spatial and functional information to operators in a form which they can readily assimilate	×8
A mismatch between an operator's model of the world and that imagined by a designer	×8

Table 4.
EPCs and their multipliers as proposed in the HEART technique [53].

found in EUROCONTROL's studies. Literature and ergonomics database reviews were undertaken to find the data which supports new values of HEPs for each generic task type. Where more than one error probability for a given task was found in the literature or the databases, geometric mean was used to establish a single value. Furthermore, uncertainty bounds of each HEP were determined using the single sample t-test [59].

EPCs used in CARA were, like general task types, developed by adjusting EPCs from HEART and other techniques (most notably SPAR-H [60] and CREAM [61]). To ensure that the CARA EPCs closely follow the well-established contextual structure used in ATC, they were modelled to fit the Human Error in ATM (HERA) [62] classification structure. For initial consideration, CARA EPCs' maximum affect

Task context	Generic task type	HEP	Uncertainty bounds
A. Offline tasks	A. Offline tasks	0.03	—
B. Checking	B1. Active search of radar or FPS, assuming some confusable information on display	0.005	0.002–0.02
	B2. Respond to visual change in display (e.g., aircraft highlighted changes to low-lighted)	0.13	0.05–0.3
	B3. Respond to unique and trusted audible and visual indication	0.0004	—
C. Monitoring for conflicts or unanticipated changes	C1. Identify routine conflict	0.01	Holding value'
	C2. Identify unanticipated change in radar display (e.g., change in digital flight level due to aircraft deviation or corruption of datablock)	0.3	0.2–0.5
D. Solving conflicts	D1. Solve conflict which includes some complexity. Note for very simple conflict resolution consider use of GTT F	0.01	Holding value'
	D2. Complex and time pressured conflict solution (do not use time pressure EPC)	0.19	0.09–0.39
E. Plan aircraft in/out of sector	E. Plan aircraft in/out of sector	0.01	Holding value'
F. Manage routine traffic	F. Routine element of sector management (e.g., rule-based selection of routine plan for an aircraft or omission of clearance)	0.003	Holding value'
G. Issuing instructions	G1. Verbal slips	0.002	0.001–0.003
	G2. Physical slips (two simple choices)	0.002	0.0008–0.004

Holding values are to be updated once more data is available.

Table 5.
Generic task types used in CARA technique [59].

values were taken from HEART, SPAR-H, and CREAM by selecting the most similar EPCs and then picking the one with the highest value (**Table 6**). It is expected that with further refinement of underlying data, the maximum affect values will be adjusted to better suit the actual values in ATC.

For the first time here, one can see that the traffic complexity was taken into account (EPC 17) with maximum affect of 10. CARA User's Manual provides additional information about this EPC, adding three anchor points for this EPC [63]:

- Higher than normal traffic levels with some non-routine conflicts to solve (EPC multiplier 0.1)

- Higher than normal traffic levels with some non-routine conflicts requiring constrained solutions; possibility of secondary conflicts (conflict resolution can lead to a second conflict) (EPC multiplier 0.5)

- High traffic levels with unusual patterns of traffic requiring problem solving and a number of future conflicts requiring resolution (EPC multiplier 1.0)

EPC multipliers are used to scale the EPC affect from its maximum value to the actual value for the situation that is being assessed, thus getting the actual

HERA element	CARA EPCs	Maximum affect
Documentation/ procedures	1. Shortfalls in the quality of information conveyed by procedures	5
Training and experience	2. Unfamiliarity and adequacy of training/experience	20
	3. On-the-job training	8
Workplace design/HMI	4. A need to unlearn a technique and apply one which requires the application of an opposing philosophy—stereotype violation	24
	5. Time pressure due to inadequate time to complete the task	11
	6. Cognitive overload, particularly one caused by simultaneous presentation of non-redundant information	6
	7. Poor, ambiguous, or ill-matched system feedback—general adequacy of the human-machine interface	5
	8. Trust in system	—
	9. Little or no independent checking	3
	10. Unreliable instrumentation	1.6
Environment	11. Environment—controller workplace noise/lighting issues, cockpit smoke	8
Personal factor issues	12. High emotional stress and effects of ill health	5
	13. Low vigilance	3
Team factor issues	14. Difficulties caused by team coordination problems or friction between team members	10
	15. Difficulties caused by poor shift hand-over practices	10
Pilot-controller communication	16. Communications quality	—
Traffic and airspace issues	17. Traffic complexity	10
	18. Unavailable equipment/degraded mode—weather issues	—
Weather	19. Weather	—
Non-HERA: organizational culture	20. Low workforce morale or adverse organizational environment	2
Non-HERA: cognitive style	21. Shift from anticipatory to reactive mode	10
	22. Risk taking	4

Table 6.
CARA EPCs and values of their maximum affect [59].

effect (AE). As is the case with many HRA techniques, some expert opinion is needed here to determine where the assessed scenario falls on the scale of 0.1–1.0. An example of human error risk calculation is given in the next section.

4.3 Using CARA to assess the effect of complexity on ATCO error risk

To better show how CARA is used to assess the effect of complexity on ATCO error risk, a simple example will be used. In this example, we suppose that the ATCO is working on an en route sector with moderately high air traffic complexity. Weather is calm and there are no failures in any of the air or ground equipment. In these conditions, we might want to assess the probability that the ATCO will not notice a conflict.

To do this, we select a generic task type (GTT) that best suits our situation. Here, it is *C1. Identify routine conflict* with HEP of 0.01. Appropriate EPC to select in this case is the EPC *17*: *traffic complexity* with maximum affect of 10. Also, we use our expertise to determine that the current traffic situation is moderately complex, so we use EPC multiplier to determine the assessed effect (AE) equal to 0.4. Calculating the probability (P) of ATCO's failure to detect the conflict is then calculated using Eqs. 1–3.

$$P = GTT \times ((EPC - 1) \times AE + 1) \tag{1}$$

$$P = 0.01 \times ((10 - 1) \times 0.4 + 1) \tag{2}$$

$$P = 0.046 \tag{3}$$

The result shows that the probability of ATCO failing to notice a conflict in a moderately complex situation is 0.046 or 4.6%. The −1 and +1 in Eq. 1 are added to ensure that the resulting EPC is more than 1 without needlessly increasing the EPC (e.g., if only the final +1 was added). Conversely, the probability of ATCO identifying a conflict is equal to 95.4%. These probabilities are valid for a situation with only one ATCO; however, en route ATC operations are usually performed with two ATCOs handling a sector (planning and executive ATCOs). The probability that both ATCOs will fail to notice the conflict is equal to 0.046 x 0.046 = 0.0021 which is to say that approximately 1 in 500 conflicts in moderately complex traffic situations will not be identified (step 1 in **Figure 2**). Fortunately, ATC tools, such as short-term conflict alert (STCA), will sound the alarm in that case, and the ATCO will have the opportunity for a timely recovery.

This calculation showed how to use CARA to determine probability of a single event. Events can be chained into probability trees to calculate the probability of a sequence of events. Building on the previous example, we can calculate the probabilities of further events after the conflict was identified or after a conflict was missed. First possibility, and a more probable one, is that the conflict was identified. Next step for ATCOs is to solve it. Let us assume that this task can be assigned to the *D1. Solve conflict which includes some complexity* GTT which is assigned HEP of 0.01. Using a GTT with the same HEP as in previous example, in combination with same EPC for traffic complexity, will yield the same error probability of 0.046 (step 2 in **Figure 2**). If ATCO notices that the conflict is not solved, they will make another attempt to solve it (step 3 in **Figure 2**). This can be considered a recovery action for the previous error (not solving the conflict). It is up to the assessor to analyze the traffic situation and operational procedures to determine how many attempts an ATCO could have before the STCA alarm rings. Modelling of additional tools, such as separation tool which helps ATCO to determine whether the conflict resolution action was successful or not, can assist the assessor in determining the most accurate sequence of events.

If the conflict was missed or the ATCO could not solve it in time, STCA will sound the alarm. This usually occurs 2 min before the loss of separation. ATCOs' response to the STCA can be modelled using the *B3. Respond to unique and trusted audible and visual indication* GTT which is assigned HEP of 0.0004. Due to short time until loss of separation, it is reasonable to use EPC number *5: time pressure due to inadequate time to complete the task* which is assigned maximum affect value of 11. Since this GTT only relates to noticing and responding to the STCA, the
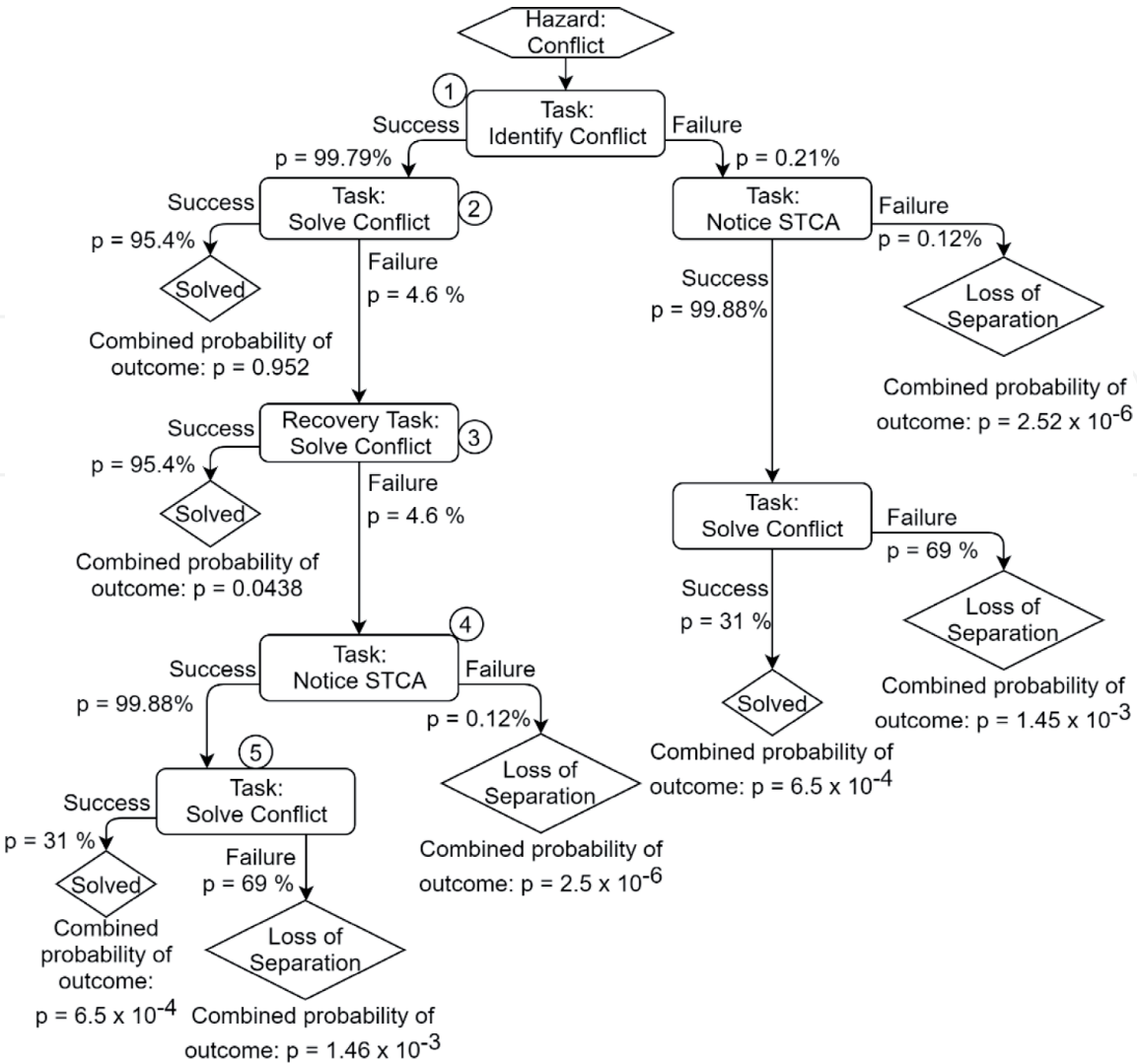
Figure 2.
Probability tree for conflict resolution in moderately complex traffic according to CARA.

actual effect of this EPC will be on the lower side, so the multiplier is set to 0.2. Calculation of the error probability is then made with Eqs. 4–6.

$$P = GTT \times ((EPC - 1) \times AE + 1) \tag{4}$$

$$P = 0.0004 \times ((11 - 1) \times 0.2 + 1) \tag{5}$$

$$P = 0.0012 \tag{6}$$

This calculation shows that the probability of not noticing the STCA alarm will be 0.12% (step 4 in **Figure 2**). Once the ATCO notices the STCA, they will make another effort to solve the conflict. This time, the appropriate GTT is *D2: complex and time pressured conflict solution* which is assigned HEP value of 0.19 with confidence interval between 0.09 and 0.39. The assessor should use expert guidance to determine which value should actually be used; in this example, 0.15 will be used. In addition, assessor could add two EPCs, one for time pressure ((*5*) *time pressure due to inadequate time to complete the task*) and one for complexity ((*17*) *traffic complexity*); however, CARA User Manual states that the EPC 5 should not be

combined with GTT D2 and neither should EPC 5 and 17 be used together [63]. This prevents overly pessimistic results. Therefore, only EPC 17 will be included in the assessment. Like in previous steps of this example, we will use 0.4 as EPC multiplier to determine the assessed effect. The calculation is given by Eqs. 7–9.

$$P = GTT \times ((EPC - 1) \times AE + 1) \tag{7}$$

$$P = 0.15 \times ((10 - 1) \times 0.4 + 1) \tag{8}$$

$$P = 0.69 \tag{9}$$

This calculation shows that, in complex traffic situation, the probability of a conflict not being solved under time pressure (STCA alarm) will be 69% (step 5 in **Figure 2**). In comparison, if the traffic is not complex, the probability of failure will be only 15%. Obviously, assessor should adjust the values of GTTs and EPCs to better suit the situation being assessed, so these probabilities are in no way final.

Finally, the probability of each outcome can be calculated by multiplying the probabilities of each event that led to that outcome. For example, if one wishes to calculate the probability that the conflict will be solved only after two failed attempts and an STCA alarm, step 5 in **Figure 2**, they should multiply probabilities of all events leading to that outcome as seen in Eqs. 10–12.

$$P_{outcome} = P_{event1} \times P_{event2} \times ... \times P_{eventN} \tag{10}$$

$$P_{outcome} = 0.9979 \times 0.046 \times 0.046 \times 0.9988 \times 0.31 \tag{11}$$

$$P_{outcome} = 6.5 \times 10^{-4} \tag{12}$$

The last step in this process is to sum up all the probabilities of a favorable outcome (conflict solved) versus all the probabilities of an unfavorable outcome (loss of separation). In this example, the probability of the favorable outcome is 99.71% versus the probability of an unfavorable outcome which is 0.29%.

To better appreciate the effect of traffic complexity on the risk of human error, comparison with the traffic situation which is not complex can be made by excluding the traffic complexity EPC from the calculation. This calculation is omitted here for brevity, but the same method without the traffic complexity EPCs yields probability of a loss of separation below 3.5×10^{-5} per conflict (approximately 1 in 28,600 conflicts). That is two orders of magnitude less probable than in the case with moderate complexity (0.29% or 1 in 345). On the other hand, if the traffic is highly complex, the assessor might use higher EPC multiplier for complexity, all the way up to 1. In that case, the probability of an unfavorable outcome, i.e., loss of separation, is 2% (1 in 50) which is 7 times more probable than in the example above (**Table 7**).

4.4 Using simulations to assess the effect of traffic complexity on risk

In addition to CARA, another method for assessing risks related to air traffic complexity is by conducting simulations. Simulation is a core method for ATM research and training, with different purposes requiring different levels of fidelity

	Low complexity	Moderate complexity	High complexity
p(solved)	0.999965	0.9971	0.98
p(loss of separation)	0.000035	0.0029	0.02

Table 7.
Comparison of probabilities to solve the conflict in traffic situations with different levels of complexity.

and simulation scope. Fidelity refers to the level of similarity between the simulated environment and the actual operations. Simulation scope can be broadly divided into strategic and tactical simulations. Strategic simulation tools (e.g., EUROCONTROL's NEST) are used to analyze current and forecast future ATM situation on a global level. On the other hand, tactical simulation tools are used to accurately simulate ATC operations on a sector level (e.g., ATCoach by UFA or Micronav's BEST Radar Simulator) [64]. For studies involving human factors, tactical real-time human-in-the-loop simulations provide the most reliable results.

Most representative results are produced when the simulator satisfies these requirements:

- Realistic working environment

- Accurate and versatile aircraft models

- Representative ATC tool operation

- Built-in stochasticity

- Human voice communication

- Research-level data logging

- Suitable meteorological model

- Suitable system and sub-system failure modelling

We used HITL simulations to assess the effect of trajectory-based operations (TBO) on air traffic complexity; for more information about that study, see [39]. Here we will provide a brief description of the methodology used and additional analysis of human errors made during that experiment. This will enable comparison of the simulation with the results obtained from CARA.

4.4.1 Example of an HITL simulation methodology

Simulation scenarios were developed based on the actual flight data. To measure complexity in conventional and trajectory-based operations, each simulation scenario had to be developed in three versions: conventional operations, 30% aircraft flying TBO, and 70% aircraft flying TBO.

Ten suitably experienced air traffic controllers were recruited to perform simulations. They all held professional air traffic controller licenses and had operational experience in Zagreb CTA Upper North sector (where the simulated traffic situations would take place). Before the actual experiment began, each controller received training in order to get accustomed with the simulator interface and operational procedures (though they were designed to closely resemble

their actual working environment). The training consisted of an introductory lecture, pre-simulator briefing, simulator runs, and post-simulator briefing. One pseudo-pilot was used for all simulation runs. The controller could communicate with the pseudo-pilot only via voice communication (through headset) or data-link.

Each controller performed three scenarios for each of the three types of the operations, each corresponding to different traffic loads—low, medium, and high (9 runs in total). *Low* scenarios were modelled to represent off-peak traffic, *medium* scenarios to represent peak traffic, and *high* scenarios to represent future peak traffic loads with 15% higher peak traffic. To prevent order of simulation scenarios affecting results, each controller was randomly assigned order in which he or she will perform different versions of the scenario (conventional, 30% TBO, 70% TBO). The order in which scenarios with different traffic loads (low, medium, high) were performed was, however, fixed and known to ATCOs. This enabled controllers to assess complexity more consistently.

During each simulation run, a subjective complexity measurement (SCM) tool opened every 2 min, accompanied by nonintrusive aural notification. The tool consisted of seven buttons (1–7), and the controller had to click on the one which was closest to the perceived level of air traffic complexity. The controller's complexity assessment was time-stamped and stored.

In addition to the subjective complexity measurement scores, objective complexity indicators were also calculated in real time, time-stamped, and stored. For the purpose of calculating new complexity indicators post-simulation, all aircraft states were stored for each time step of the simulation (1 s). Aircraft state included all data that pertained to the specific flight at that point in time (e.g., position, velocity, heading, mass, pitch, bank, throttle, drag, climb mode, acceleration mode, assigned flight level/speed/heading, route, etc.).

All other available information was also stored. Human-machine interactions were recorded in-application, while an additional application was used to record radar screen and voice communication.

4.4.2 Simulation results and comparison with CARA

Overall, 88 simulator runs were performed, each lasting for approximately 50 min. Though it is very difficult to ascertain the number of potential and actual conflicts, the frequency of STCA alarms and loss of separation occurrences can be used to assess the risk that air traffic complexity introduces. Before going into further details, it must be noted that the probabilities presented herein are accurate only for this particular set of scenarios in this particular airspace controlled by these particular ATCOs, even if the sample size issues are disregarded. These probabilities should not be used for making real-life operational decisions and are presented here as an example of the human reliability analysis that can be produced from real-time HITL simulations.

In **Figure 3**, all 88 simulation runs are plotted, showing scenario complexity and number of STCA alarms for each. Blue dots represent simulation runs which had only STCAs, whereas red dots show those runs in which loss of separation also occurred. ATCOs were not allowed to give additional complexity scores once the loss of separation occurred, thus preventing that event from influencing their opinion. Separation minima were 5 NM horizontally and 1000 ft. vertically. Complexity scores were calculated as an average of the ATCO's subjective complexity scores made during the peak 20 min of the simulation run [39]. Correlation coefficient between these two variables, complexity and number of STCAs, is 0.71, which indicates a somewhat strong correlation.

Figure 3.
Number of STCA alarms vs. scenario complexity (red dots represent loss of separation).

First thing to notice is that most of the simulator runs, 58 out of 88, finished with zero STCAs. Of the remaining 30, only 5 were in *medium* traffic load scenarios, i.e., scenarios with traffic loads equal to current peak traffic. The remaining 25 were all in *high* traffic load scenarios which were designed with 15% higher peak traffic loads.

Next thing to notice is that, even though the complexity scores are highly subjective, it is very rare to have scenarios with complexity higher than 4 and no STCAs (only 4 out of 33 or 12%). This indicates that the ATCOs are bunching most of the scenarios into the lower half of the scale, perhaps underestimating the actual difficulty of managing the traffic situations.

In terms of HRA, it is interesting to calculate the probability that the STCAs will be resolved before the loss of separation occurs. Overall probability of human error in this case is only 0.155 (11 out of 71) compared to the figure calculated by CARA in the example presented in the previous section, which was 0.69. Surprisingly, this probability will not change much even if the scenarios were filtered by complexity. For example, for scenarios with complexity above 5, the probability of an STCA turning into a loss of separation is 0.175 (10 out of 57). For scenarios with complexity above 6, the probability is only slightly higher at 0.189 (7 out of 37). Here, ATCOs obviously show significant compensatory effects which should be included into CARA or modelled more precisely by assessors using the existing GTTs and EPCs.

On the other hand, the probability that the simulation run will contain at least one loss of separation rises sharply with complexity. For the lower half of the complexity scale, this probability is zero. If we consider all scenarios with complexity score equal to or above 4, the probability of loss of separation is 0.33 (11 out of 33). For scenarios with the score equal to or above 5, the probability is 0.5 (10 out of 20), and for scenarios with the complexity score above 6, the probability is 0.538 (7 of 13). This shows that even though the probability of an STCA turning into loss of separation is lower than expected by CARA, the number of conflicts rises to the level at which the loss of separation becomes extremely probable.

As for the *Cynefin* framework, it could be applied here only in broad brushes. One could argue that the first quarter of the complexity scale in these simulations

maps to the *simple* domain because there are no STCAs. Second quarter, with only a couple of STCAs which were quickly resolved, perhaps maps to the *complicated* domain. The third quarter could be mapped to the *complex* domain because there are many STCAs, but only two were not resolved in time. Finally, the last quarter of the scale arguably maps to the *chaotic* domain due to high probability of loss of separation which indicates that the ATCOs had lost the immediate control of the situation. Notwithstanding the *Cynefin* framework, it is clear that the ATM system should be designed to keep the complexity in the lower half of the scale and serious efforts are needed to achieve this in the face of the rising traffic demand.

5. Conclusion

In this chapter we have shown how the air traffic complexity, through increasing the difficulty of finding the correct solution to the traffic conflict, influences human error probability and, consequently, risk in ATM as well. CARA HRA technique was used to show an example of calculation that can be used to assess the probability of a loss of separation in traffic situations with low, moderate, and high complexity.

Like other HRA techniques, CARA also relies on an expert assessor who must be able to correctly model the ATC operations by choosing the appropriate GTTs and EPCs. This process is very sensitive to small changes in the initial conditions because adding or omitting a single probability calculation often results in an order of magnitude different final probabilities. This problem is further exacerbated by uncertainty in modelling the ATC operations. For example, it is nearly impossible to determine beforehand how many opportunities to resolve a conflict will an ATCO have before a loss of separation occurs. In the example shown in Section 4.3, we used two attempts before an STCA sounded the alarm and one attempt afterwards. If any of those attempts were omitted, the probability of a loss of separation would have increased by a significant amount (up to 10 times). Furthermore, different ATCOs will use different strategies to solve a conflict, especially if the conflict solution implies secondary potential conflicts, which makes modelling of ATC operations in CARA even more difficult. This is not to say that CARA should not be used for HRA or as a part of broader risk assessment. It just means that CARA should be used with caution and that the results should be considered more as an indication of a risk instead of as an exact quantification of risk.

To better illustrate the accuracy of CARA and to show an additional method for risk assessment, we have presented a brief analysis of a simulation-based risk modelling. During the HITL simulations, which included complexity assessment, STCA alarms and loss of separation occurrences were identified and recorded. Expectedly, it was shown that the number of STCAs quite strongly correlates with the perceived level of air traffic complexity. More interesting was the fact that the probability of STCA turning into loss of separation was much smaller than the one predicted by CARA. Also, it almost did not change with the increase of complexity which indicates presence of strong compensatory effects.

On the other hand, the human error probability for a conflict, defined as a probability of a failure to solve the conflict resulting in a loss of separation, increases with the increase in complexity. Of all 88 simulation runs, zero losses of separation occurred in scenarios with complexity below 4 (55 simulation runs). However, for simulation scenarios with score above 6, loss of separation occurred in 54% of simulation runs. This increase can somewhat be explained by higher traffic loads, leading to more conflicts which then led to more occurrences of loss of separation. The truth is, however, that the increase in traffic was not such that the number of

conflicts should rise to the levels achieved in the simulations. Simulation scenarios with high traffic load had only 15% more flights than scenarios with medium traffic load. It is the complexity of the traffic situation that precluded the ATCOs from being aware of all possible interactions and from solving the conflicts before it was too late. Though the sample size in the simulation study was quite small, it is clear that the model developed by the assessor in the CARA technique should be adjusted to reduce the probability of failing to solve the STCA.

In conclusion, both CARA and simulator study have a place in risk analysis in ATM. Best results are achieved when the simulations are performed to gather the probabilities of human error in a specific environment and when CARA is used to integrate the individual probabilities into a big picture assessment of ATM risks. The simulation study showed that the air traffic complexity is not only a large source of uncertainty but that it correlates nonlinearly with probability of loss of separation. This makes it difficult to model in common HRA techniques, with results having large error margins, but the greatest error would be to not model it at all.

Author details

Tomislav Radišić*, Petar Andraši, Doris Novak, Biljana Juričić
and Bruno Antulov-Fantulin
Faculty of Transport and Traffic Sciences, University of Zagreb, Zagreb, Croatia

*Address all correspondence to: tradisic@fpz.hr

IntechOpen

References

[1] Kontogiannis T, Malakis S. A proactive approach to human error detection and identification in aviation and air traffic control. Safety Science. 2009;**47**:693-706. DOI: 10.1016/j. ssci.2008.09.007

[2] Roske-Hofstrand RJ, Murphy RD. Human information processing in air traffic control. 1998

[3] Thackray RI, Mark Touchstone R. Effects of high visual taskload on the behaviours involved in complex monitoring. Ergonomics. 1989;**32**:27-38

[4] Random House, Inc. Random House Dictionary; 2014

[5] Collins English Dictionary— Complete & Unabridged 10th ed. HarperCollins Publishers; 2014

[6] Cilliers P. Complexity and Postmodernism. London: Routledge; 1998. E-book edition ur

[7] Snowden D. Complex acts of knowing: Paradox and descriptive self-awareness. Journal of Knowledge Management. 2002;**6**:100-111

[8] Juarrero A. Dynamics in Action. Cambridge, MA: MIT Press; 1999

[9] Meckiff C, Chone R, Nicolaon J-P. The tactical load smoother for multi-sector planning. In: Proceedings of the 2nd USA/Europe Air Traffic Management Research and Development Seminar. 1998

[10] Kirwan B, Scaife R, Kennedy R. Investigating complexity factors in UK air traffic management. Human Factors and Aerospace Safety. 2001;**1**:125-144

[11] Mogford RH, Guttman J, Morrow S, Kopardekar P. The Complexity Construct in Air Traffic Control: A Review and Synthesis of the Literature. McKee City, NJ: CTA Inc.; 1995

[12] Athènes S, Averty P, Puechmorel S, Delahaye D, Collet C. Complexity and controller workload: trying to bridge the gap. In: 2002 International Conference on Human-Computer Interaction in Aeronautics. Cambridge, MA, USA: Massachusetts Institute of Technology; 2002

[13] Chatterji G, Sridhar B. Measures for air traffic controller workload prediction. In: 1st AIAA, Aircraft, Technology Integration, and Operations Forum. 2001. p. 5242

[14] Christien R, Benkouar A, Chaboud T, Loubieres P. Air traffic complexity indicators & ATC sectors classification. In: The 21st Digital Avionics Systems Conference, 2002. Proceedings, vol. 1, IEEE. 2002. pp. 2D3-2D3

[15] Majumdar A, Ochieng W. Factors affecting air traffic controller workload: Multivariate analysis based on simulation modeling of controller workload. Transportation Research Record: Journal of the Transportation Research Board. 2002:58-69

[16] Davis CG, Danaher JW, Fischl MA. The Influence of Selected Sector Characteristics upon ARTCC controller activities. Arlington: The Matrix Corporation; 1963

[17] Hilburn B. Cognitive complexity in air traffic control: A literature review. EEC Note. 2004;**4**:27-50

[18] Schmidt DK. On modeling atc work load and sector capacity. Journal of Aircraft. July 1976;**13**:531-537

[19] Hurst MW, Rose RM. Objective job difficulty, behavioural response, and sector characteristics in air route traffic control centres∗. Ergonomics. 1978;**21**:697-708

[20] Stein E. Air Traffic Controller Workload: An Examination of Workload Probe (Report FAA/CTTN90/60). Atlantic City, New Jersey: FAA; 1985

[21] Laudeman IV, Shelden S, Branstrom R, Brasil C. Dynamic density: An air traffic management metric; 1998

[22] An Evaluation of Air Traffic Control Complexity. Boulder, CO, USA: Wyndemere; 1996

[23] Kopardekar P. Dynamic density: A review of proposed variables. In: FAA WJHTC Internal Document Overall Conclusions and Recommendations. Federal Aviation Administration; 2000

[24] Kopardekar P, Magyarits S. Dynamic density: Measuring and predicting sector complexity [ATC]. In: The 21st Digital Avionics Systems Conference, 2002. Proceedings, vol. 1, IEEE. 2002. pp. 2C4-2C4

[25] Kopardekar P, Magyarits S. Measurement and prediction of dynamic density. In: Proceedings of the 5th USA/Europe Air Traffic Management R & D Seminar. Vol. 139. 2003

[26] Kopardekar PH, Schwartz A, MagyaritsS,RhodesJ.Airspacecomplexity measurement: An air traffic control simulation analysis. International Journal of Industrial Engineering: Theory, Applications and Practice. 2009;**16**:61-70

[27] Masalonis A, Callaham M, Wanke C. Dynamic Density and Complexity Metrics for Real-Time Traffic Flow Management. Budapest, Hungary: An; 2003

[28] Klein A, Rodgers M, Leiden K. Simplified Dynamic Density: A Metric for Dynamic Airspace Configuration and NextGen Analysis. Orlando, USA: IEEE; 2009

[29] Bloem M, Brinton C, Hinkey J, Leiden K, Sheth K. A robust approach for predicting dynamic density. In: 9th AIAA Aircraft Technology, Integration, and Operations (ATIO) Conference. 2009

[30] Chaboud T. Investigating the Air Traffic Complexity: Potential Impacts on Workload and Costs. Bruxelles, Belgium: EUROCONTROL; 2000

[31] A.C.E. Working Group. Complexity Metrics for ANSP Benchmarking Analysis, Bruxelles. Belgium: EUROCONTROL; 2006

[32] Prevot T, Lee P. Trajectory-Based Complexity (TBX): A Modified Aircraft Count to Predict Sector Complexity during Trajectory-Based Operations. Seattle, USA: An; 2011

[33] Lee P, Prevot T. Prediction of Traffic Complexity and Controller Workload in Mixed Equipage NextGen Environments. Boston, USA: An; 2012

[34] Prandini M, Putta V, Hu J. Air traffic complexity in future air traffic management systems. Journal of Aerospace Operations. 2012;**3**:281-299

[35] Gianazza D, Guittet K. Selection and evaluation of air traffic complexity metrics. In: 2006 IEEE/AIAA 25th Digital Avionics Systems Conference; IEEE. 2006. pp. 1-12

[36] Gianazza D. Smoothed traffic complexity metrics for airspace configuration schedules. In: ICRAT 2008, International Conference on Research in Air Transportation. 2008

[37] Gianazza D. Forecasting workload and airspace configuration with neural networks and tree search methods. Artificial Intelligence. 2010;**174**:530-549

[38] Lee K, Feron E, Pritchett A. Describing airspace complexity: Airspace response to disturbances.

Journal of Guidance, Control, and Dynamics. 2009;**32**:210-222

[39] Radišić T, Novak D, Juričić B. Reduction of air traffic complexity using trajectory-based operations and validation of novel complexity indicators. IEEE Transactions on Intelligent Transportation Systems. 2017;**18**:3038-3048

[40] Andraši P, Radišić T, Novak D, Juričić B. Subjective air traffic complexity estimation using artificial neural networks. Promet-Traffic & Transportation. 2019;**31**:377-386

[41] Wang H, Song Z, Wen R. Modeling air traffic situation complexity with a dynamic weighted network approach. Journal of Advanced Transportation. 2018;**2018**:3-10

[42] Xue M, Do M. Scenario complexity for unmanned aircraft system traffic. In: AIAA Aviation 2019 Forum. 2019. p. 3513

[43] Tobaruela G, Schuster W, Majumdar A, Ochieng WY, Martinez L, Hendrickx P. A method to estimate air traffic controller mental workload based on traffic clearances. Journal of Air Transport Management. 2014;**39**:59-71

[44] Prandini M, Putta V, Hu J. A probabilistic measure of air traffic complexity in 3-D airspace. International Journal of Adaptive Control and Signal Processing. 2010;**24**:813-829. DOI: 10.1002/acs.1192

[45] Idris H, Delahaye D, Wing D. Distributed trajectory flexibility preservation for traffic complexity mitigation. In: ATM 2009, 8th USA/Europe Air Traffic Management Research and Development Seminar. 2009

[46] French S, Bedford T, Pollard SJT, Soane E. Human reliability analysis: A critique and review for managers.

Safety Science. 2011;**49**:753-763. DOI: 10.1016/j.ssci.2011.02.008

[47] Završno izvješće o specifičnom ATM događaju—Gubitak napajanja nove zgrade HKZP-a (FInal Report on Specific ATM Event—Loss of Power at the SIte of the New CCL Building). Agencija za istraživanje nesreća u zračnom, pomorskom i željezničkom prometu; 2015

[48] Krois P, Rehmann J. Assessing human factors risks in air traffic management research. 2005 International Symposium on Aviation Psychology. 2005:419

[49] Airspace Complexity & Safety Analysis Hazard & Operability (HAZOP). 2005. p. 47. DOI: 10.1163/1570-6664_iyb_SIM_org_39214

[50] Kirwan B. A Guide To Practical Human Reliability Assessment. CRC Press; 2017

[51] Swain AD. Human reliability analysis: Need, status, trends and limitations. Reliability Engineering & System Safety. 1990;**29**:301-313

[52] Swain AD, Guttmann HE. Handbook of Human-Reliability Analysis with Emphasis on Nuclear Power Plant Applications. Final Report. Sandia National Labs; 1983

[53] Williams J. HEART—A proposed method for achieving high reliability in process operation by means of human factors engineering technology. In: Proceedings of a Symposium on the Achievement of Reliability in Operating Plant. Birmingham: Safety and Reliability Society; 1985

[54] Hollnagel E. Human reliability analysis. Context and Control. Vol. 1. 1993. p. 336

[55] Barriere M, Bley D, Cooper S, Forester J, Kolaczkowski A, Luckas W, et al. Technical basis and implementation

guidelines for a technique for human event analysis (ATHEANA). NUREG-1624, Rev2000. p. 1

[56] Boring R. Dynamic Human Reliability Analysis: Benefits and Challenges of Simulating Human Performance. Idaho National Laboratory (INL); 2007

[57] Mosleh A, Chang Y. Model-based human reliability analysis: Prospects and requirements. Reliability Engineering & System Safety. 2004;**83**:241-253

[58] Kirwan B, Gibson H. CARA: A human reliability assessment tool for air traffic safety management—Technical basis and preliminary architecture. In: The Safety of Systems. Springer; 2007. pp. 197-214

[59] Gibson WH, Kirwan B. Application of the CARA HRA Tool to Air Traffic Management Safety Cases. EEC; 2008

[60] Gertman D, Blackman H, Marble J, Byers J, Smith C, et al. The SPAR-H Human Reliability Analysis Method. US Nuclear Regulatory Commission; 2005. p. 230

[61] Hollnagel E. Cognitive Reliability and Error Analysis Method (CREAM). Elsevier; 1998

[62] Isaac A, Shorrock ST, Kirwan B. Human error in European air traffic management: The HERA project. Reliability Engineering & System Safety. 2002;**75**:257-272. DOI: 10.1016/ S0951-8320(01)00099-0

[63] Kilner A, Gibson H. Controller Action Reliability Assessment (CARA) User Manual. EUROCONTROL; 2009

[64] Radišić T, Novak D, Juričić B. Development and validation of an ATC research simulator. In: INAIR 2015 International Conference on Air Transport. 2016