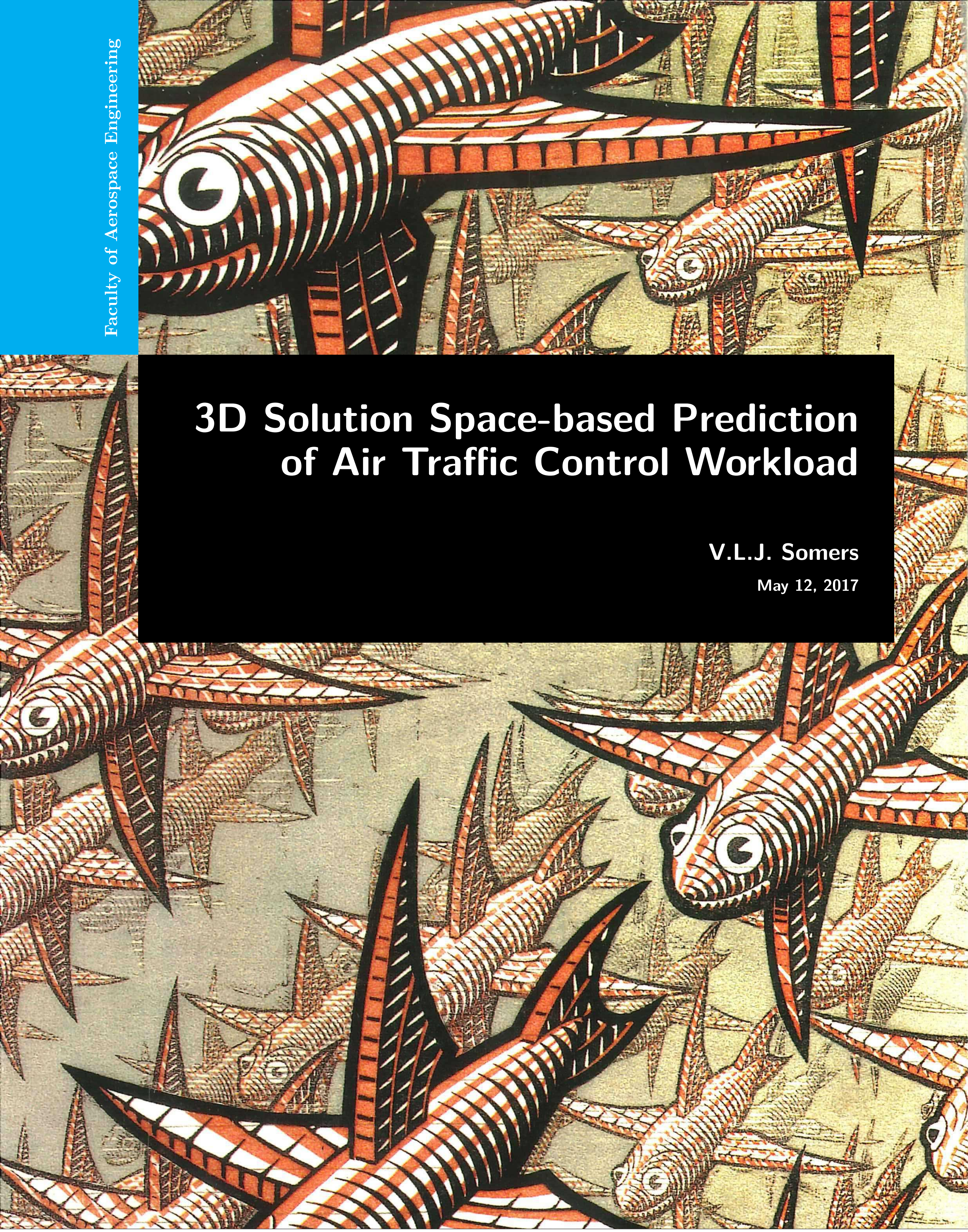
Faculty of Aerospace Engineering



# 3D Solution Space-based Prediction of Air Traffic Control Workload

V.L.J. Somers

May 12, 2017

Delft University of Technology

Challenge the future

# 3D Solution Space-based Prediction of Air Traffic Control Workload

Master of Science Thesis

For obtaining the degree of Master of Science in Aerospace Engineering at Delft University of Technology

V.L.J. Somers

May 12, 2017

Faculty of Aerospace Engineering · Delft University of Technology

Delft University of Technology

Delft University Of Technology
Department Of
Control and Simulation

The undersigned hereby certify that they have read and recommend to the Faculty of Aerospace Engineering for acceptance a thesis entitled **"3D Solution Space-based Prediction of Air Traffic Control Workload"** by **V.L.J. Somers** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: May 12, 2017

Readers:

dr.ir. C. Borst

prof.dr.ir. M. Mulder

dr.ir. M. M. van Paassen

dr.ir. M. Voskuijl

# Acknowledgments

The past months have been an interesting and incredible journey, over mountains and through deep valleys, with victories and drawbacks, but with a never ending sunny landscape in front of me. Looking over my shoulder, back to the road of graduation, I look back at this beautiful slanted landscape full of opportunities and challenges I had the privilege to tackle. With a smile on my face, I can think back to the many memories made. It has been a weird time, the end of an era, the crowning glory to life as a student.

However, I could not have done it without the support of family, friends and especially, my supervisors. With their never ending feedback and criticism, they might sometimes drove me crazy, but more often, they drove my master thesis to greater excellence.

Therefore, first of all I would like to acknowledge my supervisors, because I am very grateful for all their advise, knowledge and guidance. Clark, who was always available for more questions and brainstorm sessions. Max, who guided me onto the path of human-machine interaction in the first place and with a quicker email reply than my inbox could handle. Even more, such sharp spot-on answers and feedback to my struggles, I don't think I ever experienced more efficient and satisfying meetings. And finally, René, for his out of nowhere inexplicable good ideas and suggestions.

Next, this master thesis could not have been established without the participation and knowledge of a number of other people. First, I would like to thank everyone that participated in the experiment for their time, dedication, enthusiasm and stories. Second, Gustavo for taking time to explain his work and pre-test my experiment. And finally, Eus for showing me the ACC center at Schiphol, taking the time for showing me around, explaining everything and answering all my questions to give me a better understanding and feeling of the work of an air traffic controller.

On a different level, I think my graduation and studies would have been a lot harder without the mental support and love of friends and family. I specifically would like to thank my parents for their worries, for helping out were needed, for keeping believing and always motivating and inspiring me.

And last, but definitely not least, I would like to thank the people of room SIM 0.08 for making graduation at C&S feel like being part of one big very weird family and being at the faculty feel like coming home.

# List of Figures

# List of Tables

# Acronyms

| | |
|---|---|
| **AAA** | Amsterdam Advanced Air Traffic Control |
| **ACC** | Area Control Center |
| **APP** | Approach Center |
| **ATC** | Air Traffic Control |
| **DCT** | Direct |
| **EFL** | Executive Flight Level |
| **EXQ** | Execute |
| **FBZ** | Forbidden Beam Zone |
| **FL** | Flight Level |
| **HDG** | Heading |
| **ISA** | Instantaneous Self Assessment |
| **PZ** | Protected Zone |
| **PZN** | Protected Zone |
| **RSME** | Rating Scale Mental Effort |
| **SPD** | Speed |
| **SSD** | Solution Space Diagram |
| **SVE** | Speed Vector |
| **TOC** | Transfer of Control |
| **TWR** | Tower |
| **VO** | Velocity Obstacle |

# List of Symbols

## Greek Symbols

| | | |
|---|---|---|
| $\alpha$ | Cutting plane angle | [rad] |
| $\beta$ | Step angle | [rad] |
| $\gamma$ | Flight path angle | [rad] |
| $\mu$ | Average | |
| $\phi$ | Viewing angle | [rad] |
| $\sigma$ | Standard deviation | |
| $\Theta$ | Tangent point of the envelope equation | [rad] |
| $\theta$ | Angular coordinate along the circular velocity vector | [rad] |
| $\theta_1$ | Minimum $v_x$ angular coordinate of the upper arc of the vertical cross section | [rad] |
| $\theta_2$ | Maximum $v_x$ angular coordinate of the upper arc of the vertical cross section | [rad] |
| $\theta_3$ | Minimum $v_x$ angular coordinate of the lower arc of the vertical cross section | [rad] |
| $\theta_4$ | Maximum $v_x$ angular coordinate of the lower arc of the vertical cross section | [rad] |

## Roman Symbols

| | | |
|---|---|---|
| $A$ | 0-1 matrix | |
| $AC$ | Aircraft | |
| $g$ | Grid size | [m/s] |
| $i$ | Controlled aircraft index | |
| $N$ | Number of aircraft visible in the controlled air space | [-] |
| $R$ | Radius of the protected zone | [m] |

| | | |
|---|---|---|
| $r$ | Radius of the circular velocity vector | [m] |
| $\mathbf{r}$ | Position vector | [m] |
| $t_0$ | Start time | [s] |
| $t_c$ | Time to collision | [s] |
| $\mathbf{v}_B$ | Velocity vector of the observed aircraft | [m/s] |
| $\mathbf{v}_c$ | Collision course velocity vector | [m/s] |
| $VOL$ | Volume | [m/s] |
| $v_x$ | X-component of the velocity vector | [m/s] |
| $v_y$ | Y-component of the velocity vector | [m/s] |
| $v_z$ | Z-component of the velocity vector | [m/s] |
| $x_m$ | $v_x$ coordinate of the grid middle point | [m/s] |
| $Z$ | Vertical velocity | [m/s] |
| $z_1$ | Lower $v_z$ coordinate of the cylindrical protected zone | [m/s] |
| $z_2$ | Upper $v_z$ coordinate of the cylindrical protected zone | [m/s] |

## Subscripts

| | |
|---|---|
| $con$ | Controlled aircraft |
| $max$ | Maximum value |
| $mean$ | Mean value |
| $min$ | Minimum value |
| $obs$ | Observed aircraft |
| $sum$ | Summation |
| $x$ | X-component |
| $y$ | Y-component |
| $z$ | Z-component |

# Contents

# Chapter 1

# Introduction

## Background

An air traffic controller once said he experienced controlling the air traffic as being in the middle of a fish bowl with the aircraft as fish swimming from and towards him. Another described his job as a constant effort to try to read the newspaper.

These phrases indicate the multiple aspects and dimensions of the air traffic controller job. Still much is unknown about this fascinating occupation and its belonging workload. However, it is still the main limiting factor for traffic growth (Abdul Rahman, 2014). To accommodate for its high workload, many investments are already made into the development of new Air Traffic Control (ATC) procedures and new sector designs. However, what their exact influence will be on the workload and demand on ATC remains to be seen. This makes research into ATC workload predictions only more interesting and relevant (Boag, Neal, Loft, & Halford, 2006).

There are many more reasons why ATC workload is crucial. Another reason is that high workload can lead to defunction, less efficiency and a higher air safety risk resulting in unnecessary accidents, as mentioned by Boag, Neal, Loft and Halford (2006). This probability that errors occur largely increases when a controller works within the margins of his or her capacity. Although it has to be noted, this holds for both ends, low workload can also increase the error probability due to boredom and inefficiency. Moreover, the longer the controller works at high workload, the larger the chance of stress reactions, which decreases efficiency and makes the controller more error prone.

Finally, when the workload can be objectively estimated beforehand, less large scale, expensive tests will be needed. This not only has an economical benefit, but in addition implies that new sector designs and task assignments can be better optimized. Hence, this can more successfully lead to workload reduction and lift the growth barrier that was mentioned earlier (Gaillard, 1993).

## Problem Statement

Metrics for assessing air traffic control workload are highly in demand. In the last decades already many metrics have been developed to deal with this problem. However, most are either too case specific or too subjective to be able to give reliable objective predictions of future air traffic situations and belonging workload (d'Engelbronner, 2010).

To start closing this gap and find a solution to estimate workload objectively, the solution space metric (Hermes et al., 2009) has been developed. The solution space combines sector geometric and aircraft kinematic aspects to define the subset of all possible velocity and heading vectors while satisfying the safety, i.e., separation, constraints. The main idea behind the solution space metric is that ATC workload is inversely related to this set consisting of all available solutions for the controlled aircraft, the 'solution space'. The smaller the solution space, the less maneuver space for the aircraft and the more complicated it will be to resolve potential conflicts and the higher the ATC workload.

So far this method has shown promising results in 2D test cases (Abdul Rahman, Borst, Mulder, & Paassen, 2015; Mercado Velasco, Mulder, & van Paassen, 2010). However, the 2D solution space still misses one important aspect of reality: the third dimension, i.e., altitude. Although the first research into the 3D case has been done (Zhou, 2011), a fully developed model is still lacking. Hence, the development of the 3D solution space might be the workload metric ATC has been looking for. Moreover, all previous validation experiments have been conducted using low fidelity simulators. Therefore, the project goal can be defined as the development and evaluation of a 3D solution space-based metric for ATC workload, within a simulation of higher fidelity.

The main aspects of the problem statement and this thesis, the multidimensional aspect of the air traffic controller job combined with the underlying mathematical structure of the solution space and the update to 3D, all come together in the print by Escher called 'depth', portrayed on the cover of this thesis.

## Research Questions

For the development and evaluation of the 3D solution space metric, three main research questions have been defined, guiding the different parts of the research to be carried out:

- How to construct a well-defined and feasible 3D solution space metric for ATC workload? *(Part IIB - Chapter 4)*
  - How is ATC workload and complexity defined? *(Part IIA - Chapter 2)*
  - What are other/current ATC workload metrics and what are they based on? *(Part IIA - Chapter 2)*
  - How does the solution space method work? *(Part IIA - Chapter 3)*
  - What other 2D and 3D geometric metrics exist for areas outside ATC and what can be learned from them? *(Part IIA - Chapter 3)*

- How do we test the performance of the constructed 3D solution space metric? *(Part I and Part IIB - Chapter 5)*
  - Which scenarios can be used for verification and validation? *(Part IIB - Chapter 5)*
  - Which other metrics can it be compared to? *(Part IIB - Chapter 5)*
  - How to set up a relevant human-in-the-loop experiment for validation? *(Part IIB - Chapter 5)*
- What can we conclude on the constructed 3D solution space metric in its functionality to predict ATC workload based on the analyses of the human-in-the-loop experiment? *(Part I)*
  - How does the metric perform in the human-in-the-loop experiment? *(Part I)*
  - How does it perform compared to other metrics? *(Part I)*

## Research Structure

The report starts with a presentation of the main work as IEEE paper. This part will focus mainly on the third and most important research question about the performance of the 3D solution space metric in its functionality to predict ATC workload. The answers to the two subquestions will be discussed in the results section of the paper. Next, part II will feature the preliminary research, giving more background to the problem and the first set-up of the research, covering the first two research questions in more detail. The result of the first research question, the design of the 3D solution space metric, can be found in Part IIB - Chapter 4. The subquestions, leading to the final design, are discussed in part IIA, were the first two are covered in Chapter 2 and the final two in Chapter 3. The second research question and sub questions are more elaborated on in Part IIB - Chapter 5 with the more detailed experiment design being presented in the experiment section of the IEEE paper. Finally, the report will end with multiple appendices, giving more background to the experiment and featuring more results than could be discussed in the IEEE paper.

# Part I

# IEEE paper

# Design and Evaluation of a 3D Solution Space Metric for Air Traffic Control Workload

V.L.J. Somers, C. Borst, M. Mulder and M.M. van Paassen
Control & Simulation, Department Control & Operations
Delft University of Technology, Delft, The Netherlands
Email: vera.somers@gmail.com, {c.borst, m.mulder, m.m.vanpaassen}@tudelft.nl

***Abstract*—Air Traffic Control (ATC) workload is the main limiting factor for air traffic growth. Although predicting workload is important, a proper objective ATC workload metric is still not available. Previous research has shown that using the solution space diagram as a workload prediction metric is promising. This metric is based on the concept that ATC workload is inversely related to the size of the set of all available solutions. The current solution space metric however, does not incorporate altitude. In this paper a 3D solution space metric is described and evaluated. An experiment has been conducted to test the relation of the 3D solution space metric with workload and compare it to a conventional workload metric, the aircraft count, and a quasi-3D metric: the 2D layered solution space. Participants were required to separate air traffic on predefined flight routes, while rating their subjective workload at regular time intervals with an Instantaneous Self Assessment-based method. Weak correlations with workload were found for all tested metrics and no significant differences were found between them. Excluding aircraft that had been given a transfer of control did not yield a significant improvement. Although no significant differences were found, the 2D layered metric showed more promising results than the 3D solution space-based metric, indicating that air traffic controllers might think in 2D layers over fixed altitude ranges rather than considering the complete 3D physical solution space.**

## I. Introduction

Air Traffic Control (ATC) will remain to be one of the main limiting factors for air traffic growth, due to the manual aspect of the job [1]. To accommodate the high workload demanded of the air traffic controllers, many investments are already made into the development of new ATC procedures and new sector designs. However, what their influence will be on workload remains to be seen [2]. This makes research into ATC workload predictions only more interesting and relevant. Even more, when the workload can be objectively estimated beforehand, less large scale, expensive tests will be needed for new route and sector designs. This not only has an economical benefit, but would also mean that new sector designs and task assignments can be better optimized [3].

Research by Hilburn and Jorna [4] has shown that workload can be separated in two main parts, the taskload, i.e., the part of the work demand imposed on the controller purely by the task, and the individual operator factors such as expertise, training, strategy and resource management. To predict the average experienced workload, objective metrics for workload assessment focus on the taskload [5]–[8]. The most simplified metric for controller workload is the aircraft count, or, the Static Density [6]. Although this metric has been proven to quickly give acceptable results, its main limitation is that it cannot take into account other important air traffic parameters such as the interaction between aircraft and flight characteristics [9], [10]. Moreover, research by Lee [11] has shown that the aircraft count has a non-linear relationship with workload. More complex metrics that aim to take the dynamic interaction between aircraft into account include the Dynamic Density [12] and the Traffic Load Index [5]. The downside of these metrics is that they have to be tuned from subjective ratings for specific situations in a specific sector to give reliable results [5], [6]. Therefore these metrics are likely to be too case specific or too dependent on subjectivity to be able to give reliable objective predictions of future air traffic situations and associated workload [13].

To start closing this gap and find a solution to estimate workload objectively, the solution space metric [14] has been developed. This metric combines sector geometric and aircraft kinematic aspects and uses the intersection of the velocity obstacle and the aircraft flight envelope to define the so called solution space. The main idea behind this concept is that ATC workload is inversely related to the size of the set of all available solutions for the controlled aircraft. The larger the intersection of the velocity obstacle is with the aircraft flight envelope, or the more velocity obstacles of observed aircraft the flight envelope intersects, the smaller the solution space will be for the controlled aircraft. Hence, the more complicated the situation will be to resolve potential conflicts and the higher the ATC workload.

So far, this method has shown promising results in 2D test cases [1], [15]. However, the 2D solution space still misses one important aspect of reality: the third dimension, i.e., altitude. Although the first attempts for the 3D case have been done, a fully developed model is still lacking. Moreover, all previous validation experiments have been conducted using low fidelity simulators. Therefore, the goal of this research is the development and evaluation of a 3D solution space-based metric for ATC workload, within a simulation of higher fidelity.

First of all, the existing 2D analytical model of the solution space, as described by Mercado Velasco et al. [16], was extended to 3D. Next, to validate the 3D solution space-based metric and to test its performance compared to existing metrics, an experiment was performed with a 3D ATC simulator based on the Amsterdam Advanced Air Traffic Control (AAA) system, excluding wind and radio communication. In the experiment, participants were asked to indicate their experienced subjective workload at regular time intervals while controlling the air traffic. Afterwards these ratings can then be compared with calculated taskload values from different metrics: the newly developed 3D solution space metric, the aircraft count and a 2D layered solution space metric, which considers the 2D solution space on different flight levels. Using results of the analyses, it can be determined whether the developed 3D solution space-based metric indeed provides a good objective ATC workload estimate.

Furthermore, the new higher fidelity set-up gives the opportunity to test the contribution to the experienced workload of aircraft that have been given a transfer of control. Finally, it can be investigated if air traffic controllers think in layers or in actual 3D space by comparing the quasi-3D layered solution space with the new developed 3D metric.

This paper starts with a more elaborated description of the 2D solution space in Section II to give a background on the proposed method. Section III explains the proposed 3D solution space metric and the mathematical concepts. The different metrics that will be compared and tested for their relation with workload are explained in Section IV. Next, the experiment set-up and choices made to validate the new metric and to test the hypotheses are elaborated in Section V, with the workload results, correlation analysis and the metrics ability to predict workload being presented in respectively Section VI, VII and VIII. A further in-depth discussion of the results and the observations is given in Section IX. Finally, Section X presents the final conclusions and recommendations.

## II. The Solution Space

The solution space, as first defined and developed by Van Dam, Mulder and Van Paassen [17] as a separation assistance tool for pilots, finds its origin in the robotics field in the velocity obstacle method, described by Fiorini and Shiller [18]. The velocity obstacle method defines a collision cone for object avoidance by taking tangent lines from the controlled object's position to the observed object in the velocity field. By translating this cone with the relative velocity vector an area is created in which no velocity vector for the controlled object can be selected, otherwise a collision will occur. The strength of this method is that by using relative velocity to translate the relative collision cone to an absolute collision cone, it can see a dynamic situation as static and only needs the current position and velocity of the objects controlled and observed, which makes it a first order method [19]. Although the method was originally developed for one-on-one avoidance, it can easily be extended to multiple objects, which results in multiple, possibly overlapping, velocity obstacles [18].

Van Dam, Mulder and Van Paassen [17] applied the velocity obstacle method to aircraft navigation by considering the minimum spatial separation required to maintain the safety goal for aircraft. The protected zone (PZ), i.e., the area that other aircraft are not allowed to intrude to avoid loss of separation [20], is taken as the object to avoid. The PZ is defined in the 2D horizontal plane as as a circle with a radius of 5 NM centered around the aircraft based on minimal separation requirements established by [21]. Van Dam, Mulder and Van Paassen [17] called the resulting collision cone the Forbidden Beam Zone (FBZ).

Using the FBZ the solution space can be defined. Since the controlled aircraft is also bound to a minimum and maximum velocity vector, the so called flight or performance envelope can be superimposed on the FBZ to obtain the solution space: the subset of all possible velocity and heading vectors while satisfying the safety, i.e., the separation, constraints, see Figure 1. Hermes et al. [14] also studied the solution space, but then calculated it by taking the different parts of the aircraft trajectory over time separately. This quickly gets complex and time consuming when longer and more difficult trajectories are observed. Therefore, Mercado Velasco et al. [16] simplified this approach to make it more applicable. It should be noted that the solution space was originally more introduced from a pilot's perspective, whereas Mercado Velasco, Mulder and Van Paassen [15] also first links this solution space and hence, the geometry and complexity of the airspace, to air traffic control workload. By redefining the FBZ, Mercado Velasco et al. [16] opened the way for fairly easy calculations when considering aircraft intent, acceleration, heading and a new way to round off for finite time consideration.

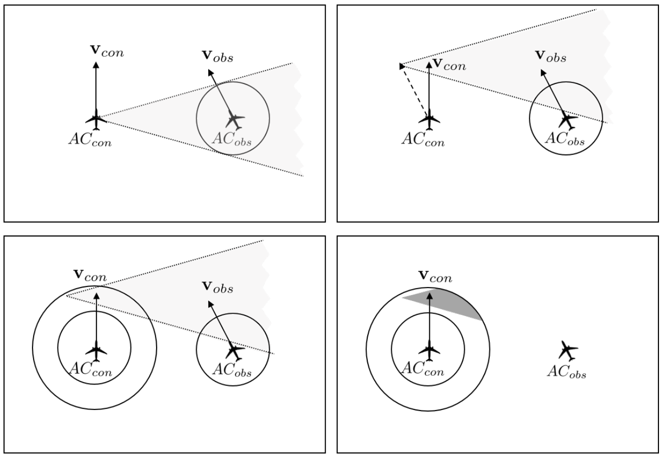



Fig. 1: Construction of the 2D solution space, where $AC_{con}$ and $AC_{obs}$ represent, respectively, the controlled and observed aircraft and $\mathbf{v}_{con}$ and $\mathbf{v}_{obs}$ are the controlled and observed aircraft's velocities, adapted from Abdul Rahman [22]

The solution space as first defined by Mercado Velasco, Mulder and Van Paassen [15] and later improved in [16], can be constructed as follows: Instead of considering how to avoid collision, the trajectory the controlled aircraft should fly to obtain collision with the observed aircraft is taken. This results in the collision velocity vector $\mathbf{v}_c$. By varying the time to when collision would occur, $t_c$, the bisector as visualized in Figure 2 is obtained. As discussed not only the aircraft is defined as object to avoid, but also intruding the protected zone will lead to loss of separation. Therefore, around this set of velocity vectors a set of projected circles has to be taken. All these circles combined form the forbidden beam zone, see Figure 2.

This family of circles can be described by the Parametric Equation 1:

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \mathbf{v}_c(t_c) + r(t_c) \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}$$
$$\forall \quad \theta \in [-\pi, \pi]; \quad t_c \in [t_0, \infty\rangle \tag{1}$$

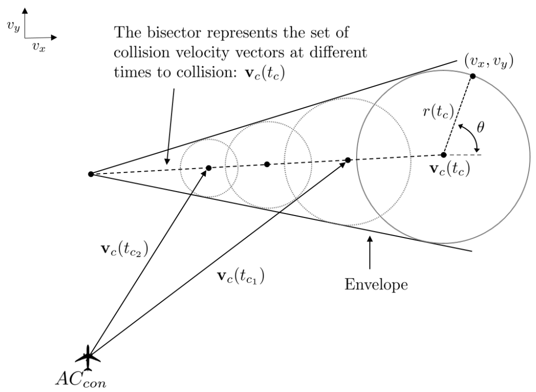



Fig. 2: Family of circles and its envelope, adapted from Mercado Velasco et al. [16]

where the velocity coordinates of the family of circles are described by $(v_x, v_y)$ and $r(t_c)$ is the radius of the corresponding circle at time to collision $t_c$. All points within the envelope of this circular velocity set will result in loss of separation. Therefore, the intersection of the envelope equation and the performance envelope of the controlled aircraft will result in the solution space. As can be seen it is purely based on geometric and aircraft kinematic properties [14], [23].

The solution space diagram has originally been developed for collision avoidance and pilot visual aid. Via Hermes et al. [14], d'Engelbronner et al. [23], Mercado Velasco, Mulder and Van Paassen [15] and Abdul Rahman et al. [24] the idea has formed to use it as an ATC workload metric. However, the solution space diagram (SSD) itself only contains visual information and cannot be applied directly as a metric [23]. For workload prediction, the size of the intersection area of the FBZ and the flight envelope has to be considered. Research by Abdul Rahman [22], indicates that ATC considers six attributes: flight level, flight path, longitudinal separation, relative velocity, direction of flight after reporting points and lateral separation. The 2D solution space metric as described, takes into account five of these parameters. Transforming it to 3D will include all six.

In 3D the protected zone is defined as a cylinder to include an additional vertical separation constraint of 1000ft [21]. A first approach to translate the 2D solution space to 3D by using a cylindrical protected zone has been performed by Zhou [25]. There, tangent lines to the top and bottom of the cylinder formed two collision cones. These cones are then connected with a tangential block to construct the 3D FBZ. To calculate the overlapping volume with the 3D performance envelope, a numerical approximation is done by using a voxelisation method [26]. The limitations of this first research are that it not only uses a rough numerical approach, but only the 3D solution space for aircraft interaction one on one is considered. Moreover, the method has only been tested with a small experiment with 4 participants (of which 1 had to be discarded) with different levels of expertise. The obtained results were promising; however, due to the small sample size and limitations, no conclusions could be drawn about the 3D solution space as a predictor for workload.

## III. Towards a 3D Solution Space Metric

For the design of the new 3D solution space-based metric, the 2D solution space metric as analytically defined by Mercado Velasco et al. [16] and described in the previous section will be updated to a feasible 3D version.

### A. A circle in 3D

The first transition to 3D can be made by putting the family of circles as described in Equation 1, in a 3D space. This results in Equation 2 and can be visualized as a 3D collision cone, illustrated in Figure 3.

$$\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} = \mathbf{v}_c(t_c) + r(t_c) \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{bmatrix} \quad \forall \quad \theta \in [-\pi, \pi]; \quad t_c \in [t_0, \infty\rangle \tag{2}$$

In Equation 2 and Figure 3 the collision course velocity vector $\mathbf{v}_c(t_c)$ is defined as in Equation 3. Here $\mathbf{r}_{obs}(t_c)$ is the position of the observed aircraft at time to collision $t_c$, which is defined as Equation 4. $\mathbf{r}_{con}(t_0)$ is the position of the controlled aircraft at time $t_0$.

$$\mathbf{v}_c(t_c) = \frac{\mathbf{r}_{obs}(t_c) - \mathbf{r}_{con}(t_0)}{t_c - t_0} \tag{3}$$

$$\mathbf{r}_{obs}(t_c) = \mathbf{r}_{obs}(t_0) + \mathbf{v}_B \cdot t_c \tag{4}$$

Here $\mathbf{v}_B$ is the velocity vector of the observed aircraft. In Equation 3 $r(t_c)$ is defined as Equation 5, where R is the radius of the protected zone. Time, velocities and distances are all taken in SI units.

$$r(t_c) = \frac{R}{t_c - t_0} \tag{5}$$

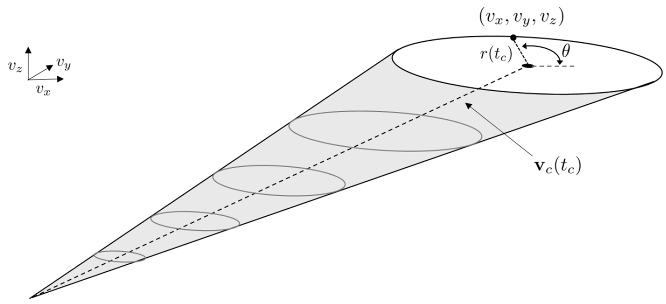



Fig. 3: 3D collision cone emerging from the envelope of a family of circles in 3D

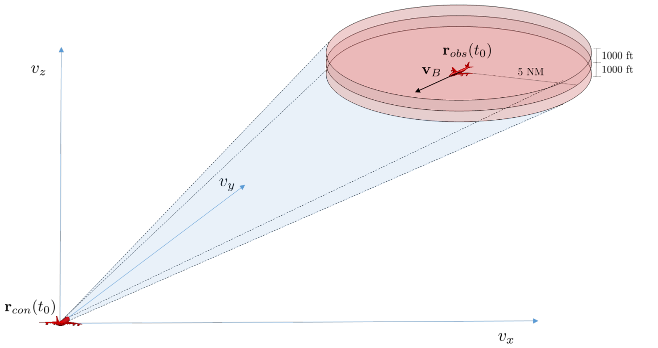



Fig. 4: Observed aircraft with cylindrical protected zone in relation to the controlled aircraft

## B. *3D FBZ Cylindrical Protected Zone*

For an aircraft in 3D space, the protected zone is not simply a circle in a 3D environment. The separation requirements as mentioned before consist of 5 NM horizontal separation and 1000 ft vertical separation, illustrated in Figure 4, which results in a cylinder around the aircraft with a radius of 9260 meter and a height of 609.6 meter in SI units.

Assume that this cylinder is built up from infinite many circles ranging from the lower to the upper height as described by the vertical separation requirements. This implies that the 3D FBZ can be seen as infinite many of the collision cones as described before, resulting from these circles in 3D. To understand how this works and to get a parametric equation for the complete 3D figure, first consider the horizontal cross section, at for example $v_z = 10\,[m/s]$, of multiple of these collision cones, see Figure 5.

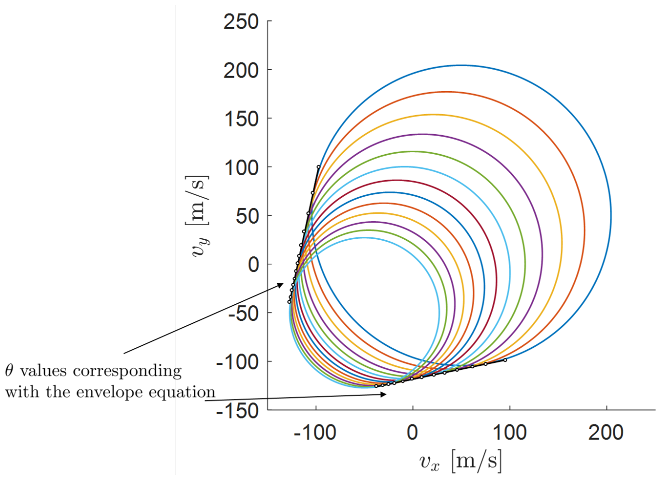



Fig. 5: Horizontal cross section of multiple collision cones at $v_z = 10\,[m/s]$

Now at a certain vertical velocity $v_z = Z$ for each family holds:

$$v_{c,_z} = Z, \tag{6}$$

which gives, using Equation 3:

$$\frac{r_{obs,_z}(t_c) - r_{con,_z}(t_0)}{t_c - t_0} = Z, \tag{7}$$

where $r_{obs,_z}(t_c)$ defines which circle, with accompanying circle family as defined by Equation 2, of the original cylinder is considered.

Now assume for simplicity that the altitude of the controlled aircraft at the start time, as well as the start time itself are taken as zero, i.e., $r_{con,_z}(t_0) = 0$ and $t_0 = 0$. These simplifications are not necessary for actual calculations (as performed in the validation experiment), however, they are done here to give a clearer view of how the analytical set up of the metric works. Applying these assumptions and combining Equations 4 and 7, results in Equation 8:

$$r_{obs,_z}(t_0) + v_{B,_z} \cdot t_c = Z \cdot t_c \tag{8}$$

Again for clarity the assumption is made that $v_{B,_z}$ is not dependent on $t_c$, i.e., no acceleration is taken into account. This is not done in actual calculations, but makes the derivation of $t_c$ easier for visualization. This results in Equation 9 and finally, Equation 10. Note that the time to collision $t_c$ is a function of $r_{obs,_z}(t_0)$, the vertical position of the observed aircraft at time $t_0$.

$$r_{obs,_z}(t_0) = (Z - v_{B,_z}) \cdot t_c \tag{9}$$

$$t_c = \frac{r_{obs,_z}(t_0)}{Z - v_{B,_z}} \tag{10}$$

Hence, at a certain vertical speed $Z$ of the controlled aircraft, for every circle that can be traced back to a circle of the original cylinder by $r_{obs,_z}(t_0)$ there is an accompanying $t_c$ value. Hence, for a certain $Z$ a new 2D family of circles emerges, see Equation 11. Because $t_c$ is a function of $r_{obs,_z}(t_0)$, $r_{obs,_z}(t_0)$ is taken as the new variable. Here $z_1$ and $z_2$ are respectively the lower and upper height of the Protected Zone (PZ), where $z_2 > z_1$.

$$\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} = \mathbf{v}_c(r_{obs,_z}) + r(r_{obs,_z}) \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{bmatrix}$$
$$\forall \quad \theta \in [-\pi, \pi]; \quad r_{obs,_z}(t_0) \in [z_1, z_2] \tag{11}$$

The envelope equation, as illustrated in Figure 5, for this family of circles can be found by slightly adapting the envelope equations as used by Mercado Velasco et al. [16]. This results in Equation 12:

$$\begin{vmatrix} \frac{\partial v_x}{\partial t_c} & \frac{\partial v_x}{\partial \theta} \\ \frac{\partial v_y}{\partial t_c} & \frac{\partial v_y}{\partial \theta} \end{vmatrix} = 0 \tag{12}$$

where:

$$\dot{v}_{c_x} = \frac{\partial v_{c_x}}{\partial t_c} \quad \dot{v}_{c_y} = \frac{\partial v_{c_y}}{\partial t_c} \quad \dot{r} = \frac{\mathrm{d}r}{\mathrm{d}t_c}$$

Finally, the values of $\theta$, corresponding with the envelope as demonstrated in Figure 2, are calculated with Equations 13 and 14.

$$\theta = 2\arctan(\Theta) \tag{13}$$

$$\Theta = \frac{-\dot{v}_{c_y} \pm \sqrt{|\dot{\mathbf{v}}_c|^2 - \dot{r}^2}}{-\dot{v}_{c_x} + \dot{r}} \tag{14}$$

Now the envelope equation at $v_z = Z$ is known, $Z$ can be varied to obtain the complete 3D figure. This is done by varying the time to collision, which is related to the value of $Z$ of one horizontal family of circles.

### C. 3D performance envelope

The solution space approach uses the intersection of the performance envelope and the FBZ for workload calculations. Now the FBZ has been expressed in a parametric form, the performance envelope should also be transformed to 3D.

The performance envelope is determined by multiple factors, namely the stall speed, idle thrust, maximum thrust, fastest climb, steepest climb and maximum speed, see Figure 6. Instead of rotating this side view around 360 degrees, both for simplicity and to make it more realistic, a different approach is taken. Air traffic controllers will not use the full performance envelope to give commands to aircraft, only a part of it is used. For example, an air traffic controller will not actually consider 'idle thrust' as a reasonable command when avoiding conflicts and controlling the air space. Therefore, a rectangle is taken within this performance envelope using the minimum and maximum Indicated Airspeed for the particular aircraft. The maximum rate of descent and rate of climb is dependent on the altitude and therefore, the performance envelope will not only be different per aircraft, but also change with the flight level. An example of the simplified performance envelope is displayed in Figure 6.

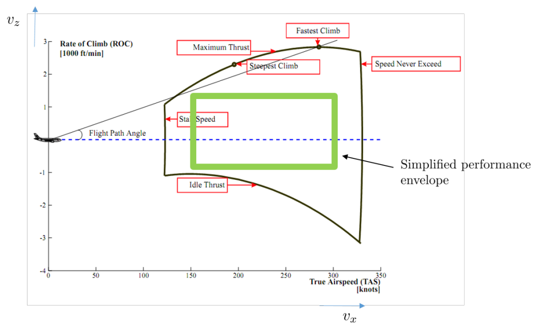



Fig. 6: The simplified performance envelope indicated with a green rectangle on the original performance envelope of the Cessna Citation I in a trimmed flight condition at 16405 feet altitude, adapted from Heylen et al. [27]

### D. Intersection FBZ and performance envelope

The 3D aircraft flight envelope is constructed by partly rotating a rectangle around the $v_z$-axis. Another view to this is imagining the flight envelope as a 'book' consisting of a finite number of pages, see Figure 7. Each of these pages represents a vertical cutting plane, as also described in [28], illustrated in Figure 8.

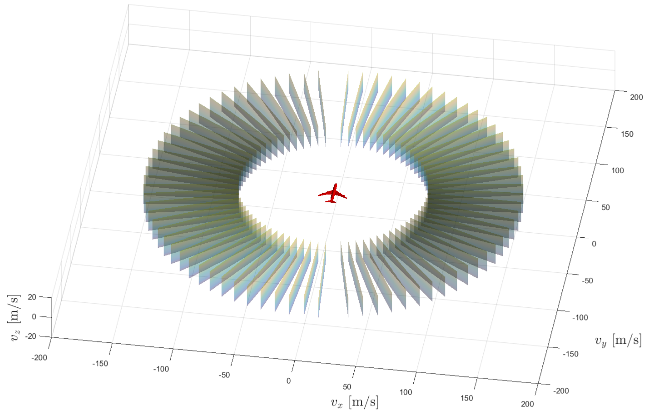



Fig. 7: Vertical cutting planes of the performance envelope forming a 'book'

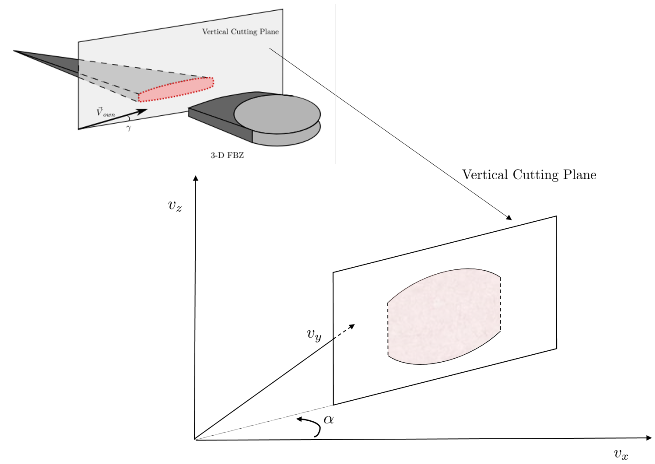



Fig. 8: The vertical cutting plane and the resulting intersection at angle $\alpha$, adapted from Brantegem [29]

The volume, resulting from the intersection of the 3D FBZ and the rotated aircraft performance envelope, can be constructed by obtaining the vertical intersections of all 'book pages' (the vertical cutting planes with the 3D FBZ) and rotating these intersection areas around the $v_z$-axis over the step angle '$\beta$' taken between the vertical cutting planes, as illustrated in Figure 9.

### E. Vertical cross section

The first step, as illustrated in Figure 8, is to obtain the cross section of the 3D FBZ with each 'page'. This is determined by combining the vertical cutting plane equations

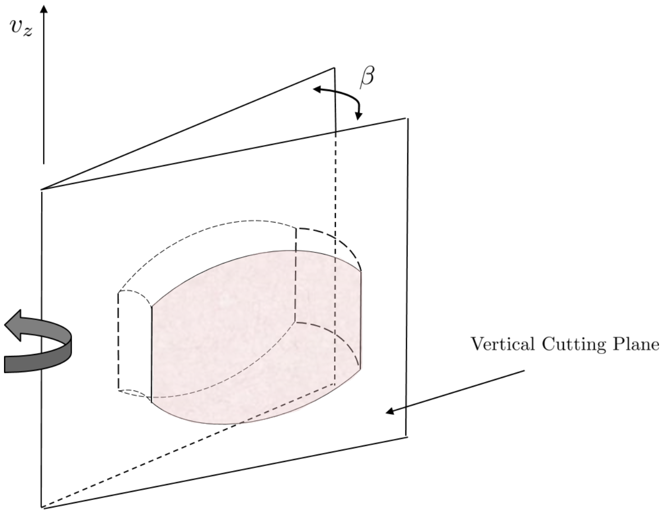



Fig. 9: Rotating the vertical cross section over step angle $\beta$

with the 3D FBZ equations. The resulting shape consists of two arc equations, which are a function of $\theta$, and two vertical connecting lines. Note that the two arc equations can be traced back to the top and bottom curved surfaces of the 3D FBZ.

The resulting vertical cross section, although a 2D shape, is still given in 3D coordinates. To express it in 2D coordinates, it can be rotated around the $v_z$-axis over angle $\alpha$, aligning it with the $v_x v_z$-plane, such that all $v_y$ coordinates are 0 and only $v_x$ and $v_z$ remain. This is illustrated in Figure 10.

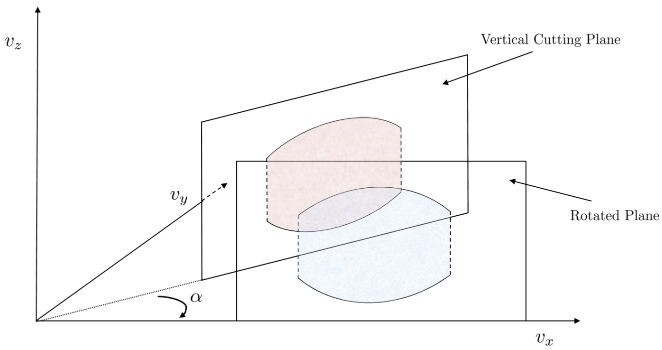



Fig. 10: Vertical cross section and its rotated image on the $v_x v_z$-plane

To determine the volume in between two 'pages', the vertical cross section is rotated over the angle '$\beta$', as demonstrated in Figure 9. This can be analytically calculated using shell integration for parametric equations. To determine the complete volume of the intersection of the 3D FBZ and the aircraft performance envelope, these calculations have to be performed for every page in the 'book'; the 3D aircraft performance envelope, where also has to be taken into account if the intersection lies within the boundaries defined by the performance envelope.

## F. Grid approach

The previously described analytical method works when there is interaction between only two aircraft. However, when multiple aircraft are considered, there may exist multiple vertical cross sections on each vertical cutting plane. When these vertical cross sections overlap, using the previously described method on the individual aircraft would result in taking the overlapping area twice.

An option could be to determine the intersection points and subtract the volume created by the intersection area by using revolution integrals again. When visualized it is straightforward where the overlapping area is. However, on a large scale with many possible overlapping areas and points, doing this analytically is not computation time effective.

Therefore, it is proposed to switch from a complete analytical approach to a combined grid approach, where a grid, with grid size 'g', is placed over the vertical cutting planes. The grid is depicted as a zero matrix where grid points that lie inside the intersection will be switched to a 1. This is visualized in Figure 11. For each grid point, it is determined whether it is inside or outside the cross section. The criterion for inclusion is illustrated in Figure 12.

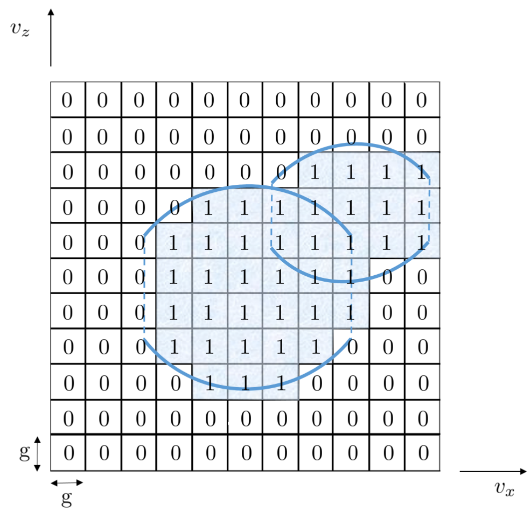



Fig. 11: Using a grid with gridsize 'g' to approach the total solution space on a single vertical cutting plane

The equation for the rotational volume of a single grid point is defined by:

$$VOL = x_m \cdot \beta \cdot g^2, \tag{15}$$

where $x_m$ is the $v_x$ coordinate of the grid middle point. The total volume can be determined by using the 0-1 matrix $A$ as illustrated in Figure 11 and results in:

$$VOL = A \cdot x_m \cdot \beta \cdot g^2 \tag{16}$$

It should be noted that for time efficiency, first the boundary points of the 2 arc equations are determined, after which the closest inside grid $v_x$ coordinates are found. An additional

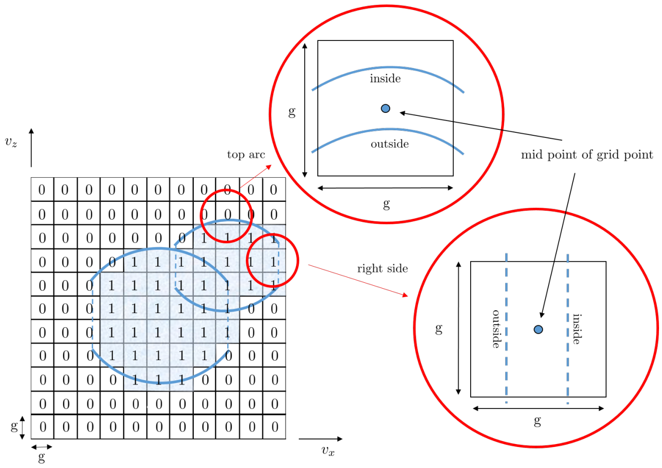



Fig. 12: Determining if a grid point is in or outside the cross section, where the mid point is taken as the reference

advantage of this is that it can be checked beforehand whether the $v_x$ coordinates lie completely, partly or not in the range of the performance envelope.

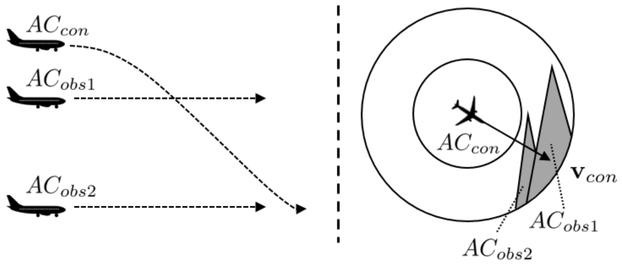
For the $v_x$ coordinates within the determined boundaries, the accompanying $v_z$ boundaries are determined. Afterwards, all points between the determined boundaries will be switched to a 1. Finally, with Equation 16, the total volume can be calculated, which will be used for the workload percentage calculations.

## IV. Workload Metrics

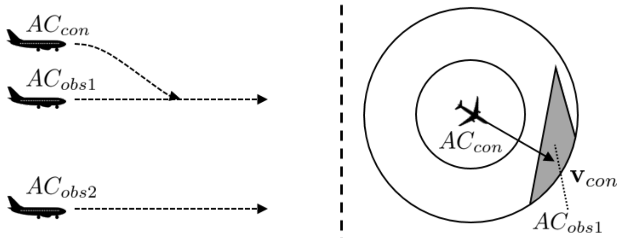
To compare the correlation with workload, two other metrics will be tested besides the 3D solution space metric. First of all, the aircraft count will be used, because it is a well tested and proved metric. Secondly, the 2D layered solution space metric will be used.

### A. 2D layered solution space

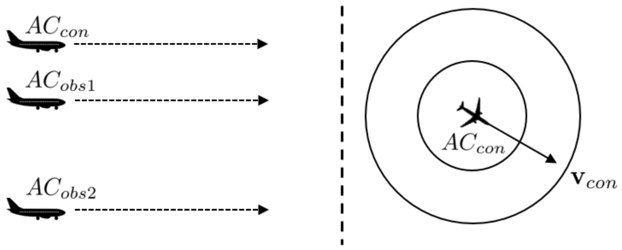
The 2D layered solution space is a quasi 3D metric and uses the 2D solution space projected at different flight levels. As illustrated in Figure 13, it considers the flight levels that will be crossed and takes the corresponding 2D solution spaces and projects them on a 2D plane. This results in the difference compared to the 3D solution space metric, that this metric for example will indicate zero workload if all aircraft are on different flight levels and are not climbing or descending, see Figure13c. The 3D solution space metric on the other hand, will still indicate workload in this situation. Although the 2D layered solution space metric is more conservative, by only taking the 2D SSD from certain flight levels instead of considering the actual 3D space, it is probable that air traffic controllers consider a 2D problem and layers, i.e. flight levels, due to current ATC practice. Especially since current radar screens present the controllers with the 3D problem on a 2D screen. Therefore, this metric might be more compatible than a complete 3D solution space metric. Even more, the 2D layered solution space metric is also quicker and more straightforward to apply than the hybrid 3D solution space-based metric. Therefore, if it gives better or similar results than the 3D metric, it not only gives an insight in how air traffic controllers see the sector, it might also be considered a more applicable metric.




(a) Controlled aircraft descending over multiple flight levels




(b) Controlled aircraft descending to lower flight level




(c) All aircraft at different flight levels with no descends or climbs

Fig. 13: 2DSSD layered approach, adapted from Lodder [30]

### B. Three different solution space metrics

To test the solution space-based metrics to a greater extent, three different metrics, based on research by Zhou [25], will be tested, namely:

- $SSD_{mean} = \frac{\sum_{i=1}^{N} SSD_i}{N}$
- $SSD_{max} = max(SSD_i)$
- $SSD_{sum} = \sum_{i=1}^{N} SSD_i$

where $N$ is the number of aircraft visible in the controlled air space of the simulator and $i$ the aircraft that is seen as the controlled aircraft for which the complete solution space is calculated. The first two metrics will give an idea if controllers consider more the average workload induced by all the aircraft or if their workload is more an effect of the most difficult task at that moment. Finally, the sum is taken as a metric that takes the amount of traffic into account as well and hence, can be seen as a combination of the solution space metric with the static density.

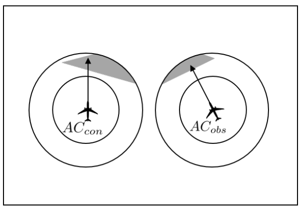
### C. Viewing angles

Research by d'Engelbronner et al. [23] has shown that the full 360 degrees aircraft performance envelope had higher correlations with workload than considering only that part that is within pre-determined viewing angles. Although for 3D this might be different, it is assumed this result can be extended to 3D and the full redefined performance envelope will be considered for workload calculations.

### D. Transfer of Control

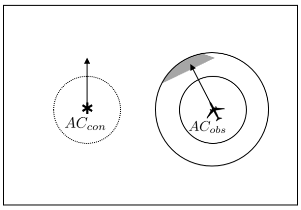
Before aircraft leave an air sector they get a transfer of control to the adjacent sector. After TOC is given, the aircraft can no longer be controlled. Up to now no research has been done if leaving these aircraft out of the solution space calculations, has any effect on the predicting effect of the metric. Therefore, for all metrics both the case in which all aircraft are considered are taken into account and the situation that aircraft that have been given a transfer of control are no longer taken as possible controlled aircraft. They will remain being included as observed aircraft in the solution space calculations of other aircraft and hence, their presence will still have an effect and to a lesser extent be taken into account. This is visualized for the solution space metrics in Figure 14. For the aircraft count, aircraft that have been given a transfer of control will not be counted.

$SSD_{mean} = 0.105$
$SSD_{max} = 0.14$
$SSD_{sum} = 0.21$




(a) Including all aircraft

$SSD_{mean} = 0.07$
$SSD_{max} = 0.07$
$SSD_{sum} = 0.07$




(b) Excluding TOC aircraft

Fig. 14: Influence of excluding aircraft that have been given a transfer of control as controlled aircraft to the SSD metrics

## V. Experiment Design

To investigate the correlation of the 3D solution space metric as described in the Section III with workload, a human-in-the-loop experiment was performed. Using a 3D ATC simulator, participants had to control and separate the air traffic, while at regular time intervals rating their experienced subjective workload using the Instantaneous Self Assessment (ISA) scale [31]. The experiment also allowed the investigation how the 3D solution space metric performs compared to the aircraft count and quasi 3D layered solution space and if the expertise level of the participants influences the correlation.

### A. Experiment set-up

The experiment was performed with a simplified, high fidelity, three-dimensional simulator, based on the AAA system used in the Netherlands. Figure 15 displays the ATC simulator's main screen, showing one aircraft as an example and the pop-up workload bar on the left. The traffic could be controlled by using a separate command window, demonstrated in Figure 16, which could be operated by either using the touchscreen function or by using the mouse. To control the traffic, the participant had to select the aircraft by clicking on it with the mouse, afterwards commands could be given with the command window.

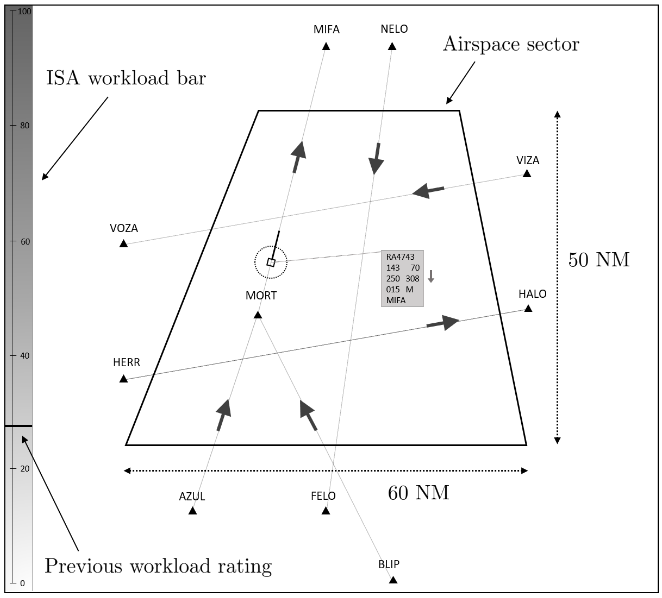



Fig. 15: The three-dimensional simulator screen with the different way points, routes, the ISA workload bar on the left and one aircraft visible in the middle

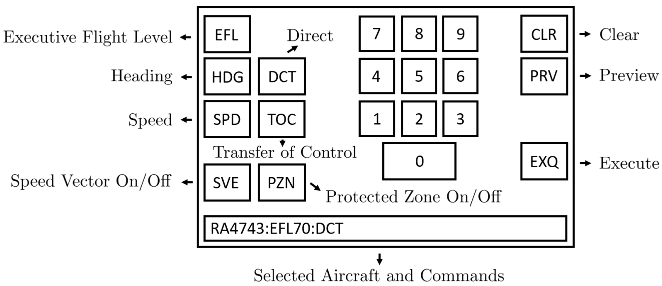



Fig. 16: The command window participants had to use to control the aircraft

For simplification, to minimize training effects and to be able to purely test the part of the workload caused by the traffic in the sector, the simulator had different simplifications compared to the actual AAA system. Aircraft types were limited to three categories: light, medium and heavy. The type was displayed in the aircraft label. Violations were made visible by changing the color to orange for a caution that loss of separation will occur within 120 seconds, and red for a warning for actual loss of separation within 60 seconds. Commands given by the touchscreen are directly followed by the aircraft, without delay or radio contact necessary. In this way the actual influence imposed by the traffic could be tested and the performance of the metric for the prediction of the task load can be assessed. Finally, participants had the option to turn the speed vector and a separation circle of 5NM on/off to help them separate the traffic. For simplification it was assumed that all different aircraft types had the same 5 NM protected zone.

## B. Participants and Instructions

In total ten participants (all male) participated in the experiment, divided into two expertise level groups. Of the ten participants, four were retired Air Traffic Controllers and six had either completed a multiple day extensive ATC-course and/or had worked as a researcher in the ATC field. The professional controller group had experience ranging from 33 - 35 years of experience ($\mu$ = 34.5 yrs, $\sigma$ = 0.9 yrs). Of this group, two were Area Control Center controllers and two Tower Approach Control controllers.

Participants were instructed to separate the air traffic and hand them over at the adjacent sector at predefined flight levels. Traffic could be separated by giving speed, heading and/or altitude commands and had to be given a transfer of control (TOC) before leaving the sector. After a transfer of control is given the aircraft can no longer be controlled. The specific instructions were the following:

- Inbound traffic coming from AZUL and BLIP and going to the northern waypoint MIFA, has to be merged and leave the sector between FL 70 - 100.
- Outbound traffic from NELO to FELO has to leave the sector at FL 200.
- Over flights towards HALO have to be handed over at FL210.
- Over flights towards VOZA leave the sector at the same flight level as they enter (FL140).
- Aircraft have to be given a transfer of control before they leave the sector.
- When aircraft are given a transfer of control they have to be separated (at least 1000 ft vertically and 5NM horizontally) from each other and should not be involved in any conflicts.

## C. Scenarios and variables

The experimental run consisted of four scenarios of 20 minutes each. Each scenario was different to be able to test the limitations and characteristics of the workload metrics. In general two scenarios had low traffic and two high. The detailed differences are summarized in Table I.

TABLE I: Scenario characteristics

| | Scenario A | Scenario B | Scenario C | Scenario D |
|---|---|---|---|---|
| Traffic | High | Low | High | Low |
| Traffic mix | yes | partly | yes | yes |
| Merges | +/- | - | + | +/- |
| Overtakes | +/- | - | + | +/- |
| Crossings | +/- | +/- | - | - |
| Deviating aircraft | - | - | - | + |

The different scenarios (A-D) were presented in a different pre-defined order as shown in Table II, to eliminate possible training effects. Here P1-P4 indicate the professional group and C1-C6 the course/research experienced participants.

TABLE II: Scenario order

| | Participant | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Experiment Run | P1 | P2 | P3 | P4 | C1 | C2 | C3 | C4 | C5 | C6 |
| 1 | A | B | C | D | A | C | D | B | A | D |
| 2 | B | A | D | C | C | A | B | D | D | A |
| 3 | C | D | A | B | D | B | A | C | B | C |
| 4 | D | C | B | A | B | D | C | A | C | B |

During the experiment the expertise (retired and extended ATC course/research) was a between-subject variable. The different scenarios are a within-subject variable. The independent variables are the different scenarios and metrics and the dependent variable the corresponding correlations with the ISA ratings.

## D. Workload estimation

As discussed by d'Engelbronner et al. [23], controlling air traffic is a dynamic job. With the situation changing over time, the workload will most likely change with it and will, therefore, be a time dependent variable. As a result, the workload should be measured at several points in time.

The subjective workload, as experienced by the participants, will be measured every 30 seconds by using the Instantaneous Self-Assessment workload bar. Although there are many techniques to measure perceived workload during a human-in-the-loop experiment, ISA is one of the simplest tools to quickly get acceptable workload ratings [8]. The ISA method is specifically developed by the UK Civil Aviation Authority to get immediate subjective workload ratings during air traffic control tasks. The original ISA scale requires ratings from 1 to 5 [31], [32]. However, to get more precise feedback a rating-scale from 0 to 100 will be used as displayed in Figure 15.

The ISA method can gather time-based workload ratings, which suits a dynamic experiment with changing task demands. Therefore, it is more accurate than post-task methods. Furthermore, it is very practical and has shown to be consistent with other subjective workload measures [31], [32]. A disadvantage is that it can interfere with the primary task performance. However, research by Farmer and Brownson [32] has shown that it still has the lowest intrusiveness compared to other dynamic metrics. The main drawback and risk is however, that the ISA workload method is very subjective and therefore, only reliable if the participants rate their perceived workload accurately. Because the experiment and validation of the metric succeeds or fails with the workload ratings given, the importance of the workload ratings is stressed to the participants. To help participants rate their workload more accurately a horizontal bar indicates the previously given workload rating, see Figure 15. Also sound will be audible as the workload bar pops up to remind participants to give their workload rating.

The metric workload values of the different metrics that will be assessed (the aircraft count, 2D layered solution space and the new developed 3D solution space-based metric) will be calculated every 3 seconds as all sector and traffic information will be logged every 3 seconds. Because the ISA ratings are only logged every 30 seconds, the mean and maximum of the metrics over that time can be used to be compared to the given ISA ratings. This is illustrated in Figure 17.

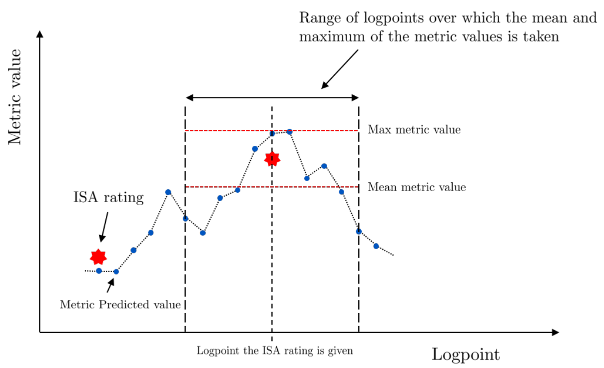



Fig. 17: The mean and maximum value of the metric data can be taken to compare to the ISA ratings

Finally, to get more feedback on the subjective workload experienced overall, also post-task subjective workload is asked from the participants. Firstly, by means of a horizontal slider, see Figure 18, ranging from very low to very high that appears at the end of each scenario. Secondly, by requiring the participants to rate the overall scenario workload on the Rating Scale Mental Effort (RSME). This scale indicates with sentences in the native language which numbers (ranging from 0 to 150) represent what effort level. Both these measurements can be used to test the reliability of the given ISA ratings. The difference in scale however, makes it interesting to use both post-task measurements.

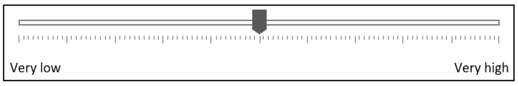



Fig. 18: The horizontal slider workload bar used for post-task workload measurement

### E. Experiment procedure

Prior to the experiment all participants were asked to fill out a short pre-questionnaire to determine their ATC experience and signed an informed consent form. Next, they were briefed with an instruction manual and additional verbal explanation on the experiment goal and set-up.

Following the briefing, the training period started to get the participants familiar with the rules of the sector and the command window. The training consisted of 9 scenarios, however as participants quickly picked up the commands, training scenario 2 was skipped. After the training phase, the four experiment scenarios, all lasting 20 minutes, were presented in random order. For training and experimental runs the same sector was used to eliminate training effects. For realism, the sector, visualized in Figure 15, and scenarios were based on the Amsterdam Sector South.

The experiment ended with a post-questionnaire to gather information on the simulator realism, scenarios and given workload ratings.

### F. Simulation characteristics

The simulations were run at twice the real time speed to be able to acquire more data in the given experiment time. A possible negative side effect of this could be that aircraft performances can be experienced as unrealistic. As mentioned previously, three different types of aircraft are used: light, medium and heavy, with performance envelopes as given in Table III.

The aircraft label of the traffic, as illustrated in Figure 15 indicates not only the type of aircraft, but also the aircraft ID, current flight level, executive flight level, indicated airspeed, true airspeed, heading and destination waypoint. Since no wind was taken into account, the true airspeed is equal to the ground speed.

### G. Hypotheses

The goal of the experiment is to determine if the developed 3D solution space metric could be a good objective ATC workload metric and is better in predicting ATC workload than existing metrics. Prior to the experiment the following main hypotheses were established.

- **H1A.** If the 3D solution space-based metric is related to ATC workload, the 3D solution space metric will show significant correlation with subjective ATC workload indications.
- **H1B.** If the 3D solution space-based metric is a better predictor for ATC workload than existing metrics (e.g., the aircraft count and 2D layered SSD), the 3D solution space metric will significantly show better correlation with the subjective workload ratings than the aircraft count and 2D layered SSD.
- **H2.** If professional air traffic controllers think in flight levels and project the 3D space on a 2D plane, the 2D layered solution space metric will significantly show better correlation with the subjective workload ratings than the 3D solution space-based metric, for the professional group.
- **H3.** If aircraft that have been given a transfer of control do not impose significant workload on the controller, leaving out aircraft that have been given a transfer of control as possible controllable aircraft in the workload metric will lead to significantly higher correlations. In previous experiments all aircraft in the sector were always taken into account. However, it is hypothesized that controllers do not take these aircraft into account and thus leaving out the aircraft that have been given a transfer of control as controlled aircraft, will lead to a better metric and better correlations.

## VI. Workload Results

The validity of the experiment results depends on the reliability of the subjective workload ratings. Therefore, first the workload ratings are analyzed. On an additional note, every

TABLE III: Performance Envelopes

| | IAS | | Rate of Climb | | | Rate of Descend | | |
|---|---|---|---|---|---|---|---|---|
| Type | min | max | FL 0-100 | FL 100-200 | FL 200-300 | FL 0-100 | FL 100-200 | FL 200-300 |
| Light | 180 kts | 250 kts | 2370 fpm | 2375 fpm | 1455 fpm | 2400 fpm | 3600 fpm | 3285 fpm |
| Medium | 200 kts | 290 kts | 2490 fpm | 2265 fpm | 1570 fpm | 1985 fpm | 2700 fpm | 3285 fpm |
| Heavy | 230 kts | 350 kts | 3160 fpm | 2770 fpm | 2040 fpm | 1345 fpm | 1755 fpm | 2205 fpm |

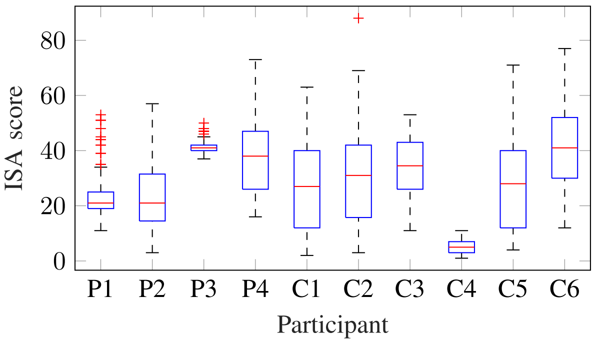



Fig. 19: ISA score distribution per participant

scenario started out with an empty sector and aircraft starting to fly in. To eliminate the effects of this start up phase, the first 5 minutes of the experiment data were removed. The remaining 15 minutes of data were used for the analyses.

### A. *Outlier analysis*

Multiple outliers were found for participants P1 and P3, as shown in the boxplot in Figure 19. In both cases these outliers are consistent with consequently higher ratings at that time, hence these outliers fit in the behavioral pattern of the corresponding participant. For participant C2 one outlier was detected. At this moment in time, loss of separation occurred, so it seems plausible the workload was indeed very high.

Because all outliers can be explained, none are rejected and all obtained ISA workload data were used for further analysis.

### B. *Post-task workload measurements*

In order to validate the consistency of the given ISA scores, the horizontal slider (Figure 18) that appeared at the end of every scenario and the RSME scores were compared to the ISA scores given in the corresponding scenario. To determine what these post-task measurements exactly represent, they were tested against the last ISA score given, the mean and maximum of the ISA scores given in the last minute of the scenario, the mean and maximum of the last 2.5 minutes of ISA scores, the mean and maximum of the last 5 minutes and the mean and maximum of the complete scenario.

A Shapiro-Wilk test showed that not all parameters had a normal distribution. Therefore, all correlations were calculated using the non-parametric Kendall's tau. The correlations and their significance can be found in Table IV. It should be noted that correlations calculated using Kendall's tau are different compared to correlations obtained with the parametric Pearson's method and can not be compared. It can be seen that the horizontal slider had highest correlations with the

TABLE IV: Correlations RSME and horizontal workload bar with ISA scores

| | RSME | | Horizontal Slider | |
|---|---|---|---|---|
| | R | p | R | p |
| Last value | 0.4146 | $< 0.01$ | 0.3071 | $< 0.01$ |
| Last 1 minute mean | 0.3834 | $< 0.01$ | 0.3064 | $< 0.01$ |
| Last 1 minute max | 0.4052 | $< 0.01$ | 0.3239 | $< 0.01$ |
| Last 2.5 minutes mean | 0.4864 | $< 0.01$ | 0.3255 | $< 0.01$ |
| Last 2.5 minutes max | 0.5049 | $< 0.01$ | 0.4347 | $< 0.01$ |
| Last 5 minutes mean | 0.5356 | $< 0.01$ | 0.3308 | $< 0.01$ |
| Last 5 minutes max | 0.5339 | $< 0.01$ | 0.4363 | $< 0.01$ |
| Overall scenario mean | **0.5788** | $< \mathbf{0.01}$ | 0.3719 | $< 0.01$ |
| Overall scenario max | 0.5362 | $< 0.01$ | **0.4793** | $< \mathbf{0.01}$ |

maximum ISA scores, especially the overall maximum ISA score (R=0.4793, p<0.001), whereas for the RSME score the highest correlation was found with the overall average ISA score (R=0.5788, p<0.001). Because the different workload parameters have good correlations, the ISA data are accepted for further analyses.

### C. *Questionnaire*

In the post-questionnaire, participants were asked to rate how confident they were about their own workload ratings and the difficulty level of the scenarios. Most participants indicated they felt confident about their workload ratings, with eight indicating they felt confident or very confident about it. This underlines the acceptance of the ISA scores for further metric analysis. Concerning the difficulty level of the scenarios, with 1 representing way too easy and 5 way too hard, five participants rated the experiment scenarios with a 3 and three with a 2.

### D. *Number of Commands*

To analyze the relation between the number of commands given by the participants and their workload ratings, the number of commands and the instructions given per command were compared to the horizontal slider score, the RSME, the average ISA score and the maximum ISA score. From Table V it can be concluded that the number of commands had a significant correlation with the maximum ISA scores, whereas the instructions given per command indicated no significant correlation. The correlations that were found, were negative correlations, indicating that more instructions per command corresponds with lower workload. However, these correlations were small and non-significant.

Next, the number of commands and instructions per command were analyzed on scenario and expertise level. Figures 20 and 21 show the actual datapoints plotted next to the boxplots. With a Kruskal Wallis test it was determined if there was a significant difference between the two different expertise groups. For the number of commands no significant difference was found as was to be expected from Figure 20. It was also

TABLE V: Correlations number of commands with workload

| | Number of Commands | | Instructions per Command | |
|---|---|---|---|---|
| | R | p | R | p |
| Horizontal slider | 0.3497 | $< 0.05$ | -0.0450 | 0.6988 |
| RSME | 0.2277 | $< 0.05$ | -0.1895 | 0.0939 |
| ISA mean | 0.3577 | $< 0.05$ | -0.1854 | 0.0989 |
| ISA max | **0.4915** | $<$ **0.05** | -0.0791 | 0.4893 |

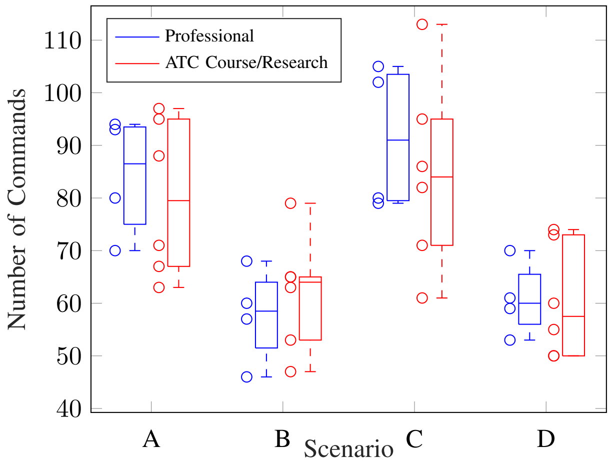



Fig. 20: Number of commands per expertise level per scenario

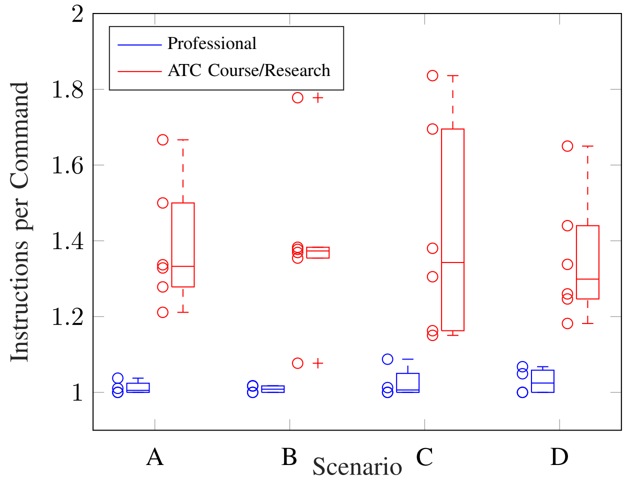



Fig. 21: Instructions given per command per expertise level per scenario

analyzed if there existed a relation between the number of commands given per scenario per expertise group and the given ISA z-scores, visualized in Figure 24. However, no relation could be found, indicating that if a certain expertise group indicated higher workload than the other group, not necessarily a higher number of commands was given compared to the other group for that specific scenario.

The instructions given per command had a significant difference for the Kruskal Wallis test ($p<0.05$). As can be seen from Figure 21, the professional group gave less instructions per command.

The command variables were also compared on a scenario level. Because the number of commands had a normal distribution per scenario, a sphericity of 0.470 and homogeneity of variances, an ANOVA was performed. The results indicated a significant difference between the scenarios ($F(3,24)=59.532$, $p<0.05$). With a post-hoc Bonferroni test significant differences were found for scenario A and B, A and D, B and C, C and D. Only for scenarios A and C ($p=0.255$) and B and D ($p=1$) a significant difference could not be found.

For the instructions per command the non-parametric Friedman test was done to test the difference between scenarios, since this parameter was not normally distributed for each scenario. No significant differences were found between the scenarios ($\chi^2(3)=0.626$, $p=0.890$).

The specific commands given for the different expertise groups per scenario were examined, see Figures 22 and 23. It can be seen that the professional group gave less direct to waypoint (DCT) commands and slightly more speed (SPD) commands. This is corresponding with observations made during the experiment. The professional group tends to make the job easy for themselves, something they also indicated themselves during the experiment. Hence, if the aircraft heads roughly in the right direction, they do not deem it necessary to give a DCT command. The ATC course/research group on the other hand still focuses to a great extent on trying to be the perfect controller and spends time (if available) on handing over aircraft as perfect as possible. Both groups however, were still above the minimum threshold of commands that was needed to successfully complete the scenarios.

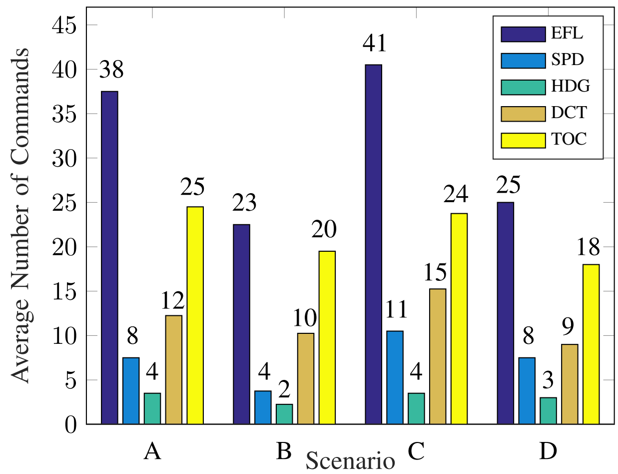



Fig. 22: Instructions per scenario for the professional group

## E. Group differences

As shown above, for the number of commands no significant differences were found between the groups, whereas for the instructions given per command a difference was found. Using a Kruskal Wallis test on the ISA data no significant difference could be determined between the expertise groups. To compare the ISA scores of the different participants, the z-scores are taken. The z-scores per scenario per group are illustrated in Figure 24.

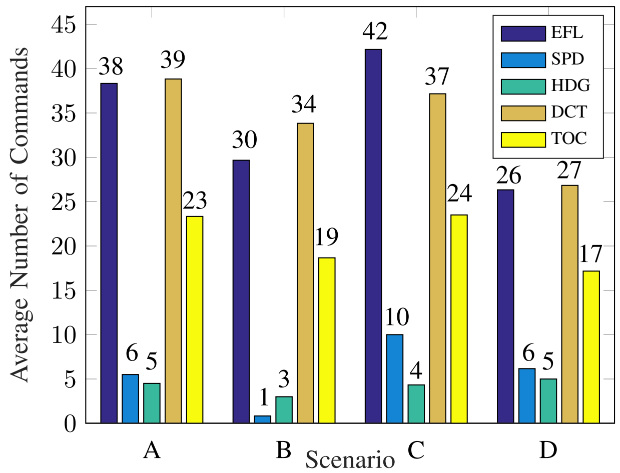



Fig. 23: Instructions per scenario for the ATC course/research group

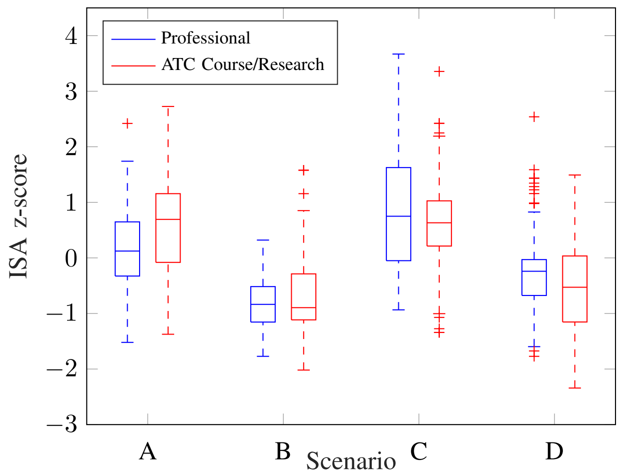



Fig. 24: ISA z-score per expertise group per scenario

An interesting difference that was noted during the experiment, is that loss of separation only occurred in the ATC course/research group, in total four times.

### F. Scenarios

The scenarios were designed to feature different characteristics, see Table I. To test if these characteristics also have an influence on the experienced workload, the scenarios are compared to each other for number of commands, instructions given per command and ISA z-score. For the Number of Commands significant differences were found between most scenarios, except A and C, and B and D. For instructions given per command no significant differences were found. For the ISA z-score, as illustrated in Figure 24, first a Shapiro-Wilk test has been used to validate that the z-score distributions per scenario had a normal distribution. Therefore, next an ANOVA was performed with a post-hoc Bonferroni test. The ANOVA indicated a significant difference ($F(3,24)=24.456$, $p<0.05$). With the conservative Bonferroni test, it could be determined that scenario A was significantly different compared to B and D, and scenario B was significantly different compared to C. Similar to the number of commands, for A and C ($p=1$), B and D ($p=0.349$), similar distributions could not be excluded.

For the different metrics the scenario differences were also analyzed. For normally distributed metrics an ANOVA with post-hoc Bonferroni and for non-normal distributed metrics a Friedman test with post-hoc Wilcoxon test was done. Significant differences ($p<0.05$) were found between all scenarios, except scenarios A and C. It can be concluded that scenario differences are distinct enough to analyze the workload and metrics on a scenario level.

All these analyses combined it can be concluded that scenarios A and C are more difficult than scenarios B and D and hence, scenarios with higher traffic and more merges and overtakes lead to higher workload. Crossings and deviating aircraft did not have a significant influence in the difference in experienced difficulty level of the scenarios.

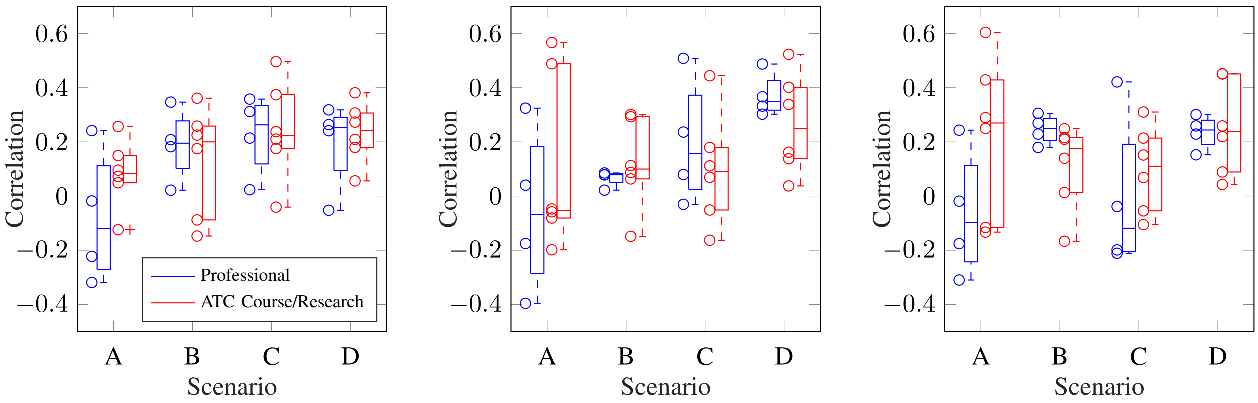
## VII. Correlation Analysis

Different correlation analyses have been performed. First of all, the correlation of the different metrics with the corresponding ISA z-scores is determined. As explained previously, the correlation between the ISA ratings with both the average values and maximum values of the metrics, as visualized in Figure 17, is taken. Next, the cross-correlation is considered between the different solution space metrics to understand their differences and corresponding controller strategies. Finally, the influence of excluding transfer of control aircraft on workload correlation is investigated.

### A. Three Metrics

In the previous section, scenario differences have been established and also expertise difference can not be completely excluded. Therefore, correlations will be calculated at an individual participant level for each metric with the corresponding ISA z-score. To determine the significant difference between metrics, it will first be determined if the correlations have a normal-distribution for each metric and scenario with the Shapiro-Wilk test due to the limited sample size. Next, a Kruskal Wallis test will be used to determine if there is a difference between expertise level. If no difference is found and the data meet the assumptions for a repeated measures ANOVA, this will be performed with if necessary a post-hoc Bonferroni test to check for significant differences between scenarios. Otherwise a non-parametric Friedman test is performed with if necessary a post-hoc Wilcoxon test.

In total 1440 correlations were calculated. In Figure 25 the correlations for the maximum values, as explained in Figure 17, of the aircraft count, 2D layered $SSD_{sum}$ and the 3D $SSD_{sum}$ are illustrated. It can be seen that all metrics have a weak correlation with the ISA z-score. With the previously described method no significant differences were found between the metrics, except for scenario C where a significant difference was found between the aircraft count and the SSD metrics. An example of the time trace and corresponding scatter plots of a high correlation ($R=0.6038$) with the ISA score is demonstrated in Figure 26 and a weak negative correlation ($R=-0.13302$) in Figure 27. Figure 26b




(a) Aircraft Count (b) 2D Layered $SSD_{sum}$ (c) 3D $SSD_{sum}$

Fig. 25: Correlations for taking the maximum values compared to the ISA z-score for the different metrics

indicates that the high correlation is quite reliable and there clearly exists a correlation, because the graph has an oval shape with a couple of outliers. The weak negative correlation scatter plot, see Figure 27b, on the other hand shows that here the correlation is less reliable. The high scattering of the data indicates no clear existence of any correlation between the metric and the ISA z-score.

To get a final idea about the correlations of the different metrics with the ISA z-scores, also the overall correlation was calculated, neglecting individual and scenario differences. The results can be found in Table VI. It can be seen that taking the sum of the SSD gives better results. It can also be noted that the 3D SSD metric has the lowest correlations, although no significant differences were determined on an individual level. Taking the mean or the maximum value compared to the ISA z-score, as visualized in Figure 17, does not seem to have an influence.

TABLE VI: Correlations metrics with ISA z-score

| | Mean | | Max | |
|---|---|---|---|---|
| | R | p | R | p |
| Aircraft Count | **0.3852** | $<$ **0.05** | **0.3987** | $<$ **0.05** |
| 2D Layered $SSD_{mean}$ | 0.3091 | $<$ 0.05 | 0.3069 | $<$ 0.05 |
| 2D Layered $SSD_{max}$ | 0.3246 | $<$ 0.05 | 0.3172 | $<$ 0.05 |
| 2D Layered $SSD_{sum}$ | **0.3888** | $<$ **0.05** | **0.3907** | $<$ **0.05** |
| 3D $SSD_{mean}$ | 0.3105 | $<$ 0.05 | 0.2954 | $<$ 0.05 |
| 3D $SSD_{max}$ | 0.2788 | $<$ 0.05 | 0.2362 | $<$ 0.05 |
| 3D $SSD_{sum}$ | **0.3778** | $<$ **0.05** | **0.3714** | $<$ **0.05** |

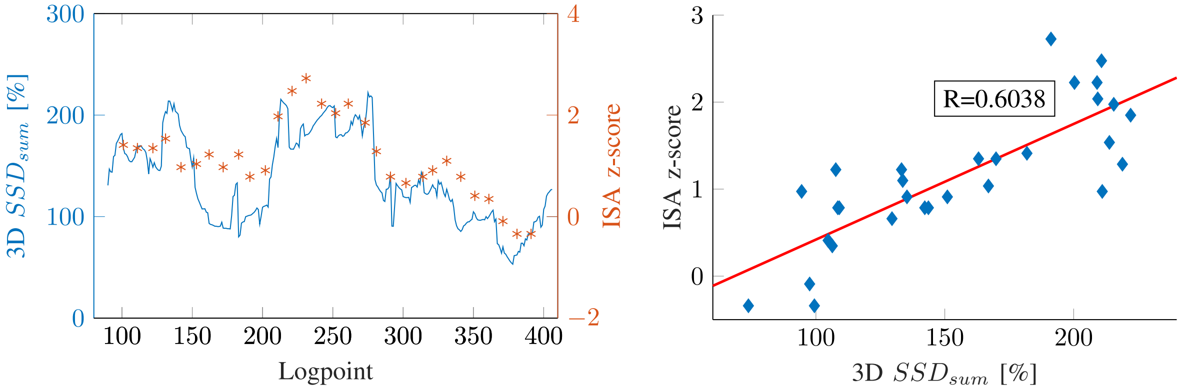
### B. *Cross Correlations*

For a better understanding of the differences between the 2D layered SSD and 3D SSD, the cross correlations were analyzed. Both metrics look at the solution space, however the 2D SSD is more conservative. The 2D metric only looks at the relevant flight levels on a 2D level, whereas the 3D SSD takes all space into account. For the $SSD_{mean}$ cross correlations between the 2D layered and 3D metric were found from 0.3116 to 0.3860, for the $SSD_{max}$ weaker correlations were found from $-0.0459$ to 0.1866 and the $SSD_{sum}$ had the highest correlations from 0.3871 to 0.4468. An example of the $SSD_{mean}$ cross correlation is illustrated in Figure 28a. The peak at logpoint 124 for the 2D layered SSD is caused by aircraft flying in a trail at the same flight level. This causes relatively high SSD scores on that flight level, which is translated to the total 2D SSD score, whereas the 3D SSD takes into account the free airspace above and beneath. The peak and dip for respectively the 3D SSD and 2D layered SSD at logpoint 247 are caused by climbing/descending aircraft with no overlapping flight levels, resulting in no extra solution space for the 2D layered, whereas the 3D SSD still takes these aircraft into account.

The $SSD_{max}$ has no to weak cross correlation for the two metrics. An example is given in Figure 28b. The peaks at logpoints 150 and 367 are caused by the situation as visualized in Figure 29, where an aircraft is cleared to climb over another aircraft with overlapping protected zones. Because they are not at the same flight level, this is not a loss of separation. However, the 2D layered SSD considers aircraft at the same flight level if they are cleared to climb, hence considers it a loss of separation, giving a 100% workload score. This is a control approach a professional controller would never execute. This phenomenon therefore, only occurred in the ATC course/research group. The peak for the 3D SSD at 306 is caused by the exact opposite effect. Aircraft separated by 1000 ft, the minimum vertical separation requirement, that are not climbing or descending do not give any SSD area for the 2D SSD, whereas they do increase the SSD percentage for the 3D SSD metric.

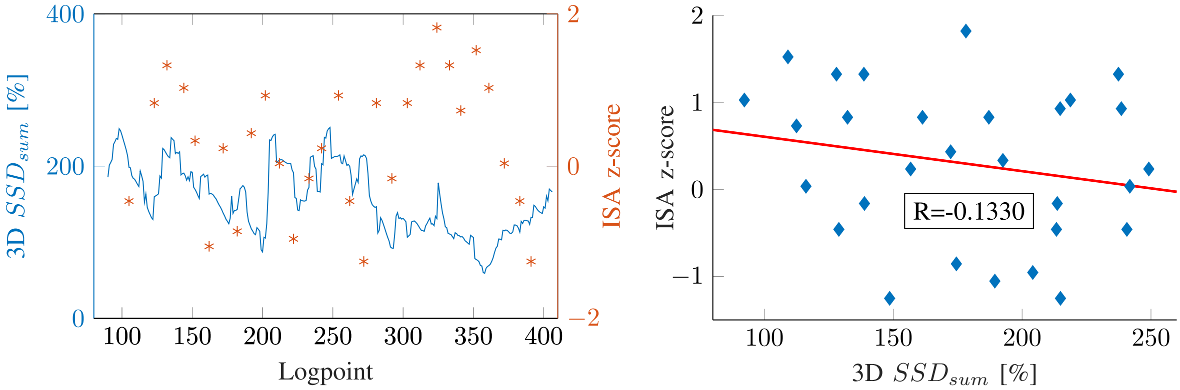
### C. *TOC influence*

The influence of leaving out aircraft that have been given a transfer of control, as visualized in Figure 14, is analyzed by means of the same correlation procedures as described in the previous sections. The resulting correlations for taking the maximum values compared to the ISA z-score for the different TOC excluded metrics are shown in Figure 30. Comparing these figures to Figure 25 no significant differences can be seen. This is confirmed by the individual correlation analysis results, where, as expected, no significant differences were found between the metrics. Although on an individual level,




(a) Time Trace (b) Scatter Plot

Fig. 26: Correlation 3D $SSD_{sum}$ with ISA z-scores for participant C5, scenario A




(a) Time Trace (b) Scatter Plot

Fig. 27: Correlation 3D $SSD_{sum}$ with ISA z-scores for participant C3, scenario A

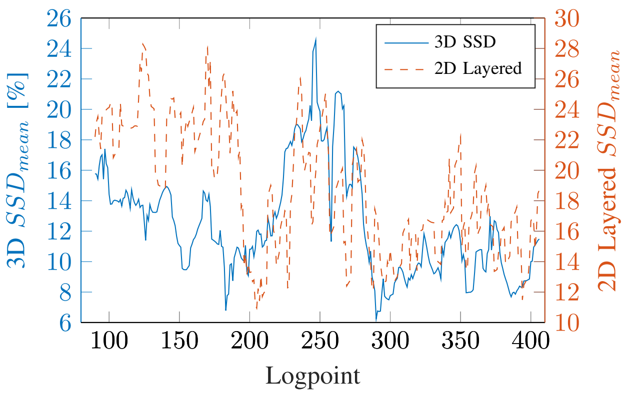
sometimes larger differences were found, in general no specific higher or lower correlations were found when excluding the TOC aircraft.

Another comparison was done, by again neglecting individual differences and taking the correlations for all data combined. The resulting correlations can be found in Table VII. The aircraft count shows a minor improvement, whereas the SSD metrics indicate somewhat lower values compared to Table VI. The same trends as before can be noticed. The $SSD_{sum}$ metrics give better correlations. The aircraft count seems to perform better compared to the SSD metrics, where the 2D layered SSD is again slightly better than the 3D SSD. However, with the statistical analysis at an individual level no significant differences were found, except for scenario C where the aircraft count (both including and excluding transfer of control) performed significantly better than the other metrics.

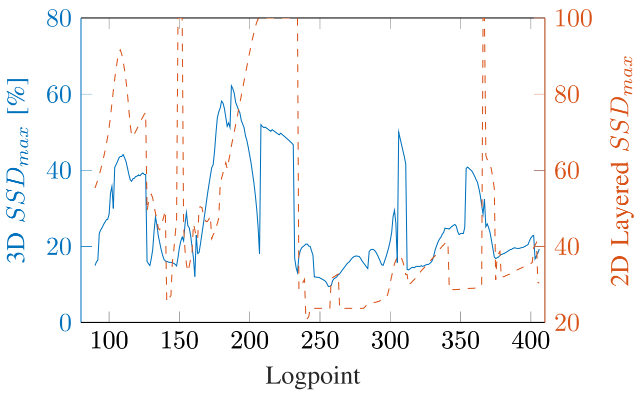
The cross correlations between the metrics including and excluding TOC aircraft were also calculated for each scenario. For the aircraft count a cross correlation from -0.0237 for scenario C to 0.5396 for scenario B was found. This underlines the fact that the aircraft count excluding TOC aircraft performs best for high workload scenario C, whereas for low workload scenario B, both aircraft count metrics showed weak correlations. The 2D layered metrics had cross correlations ranging from 0.3980 to 0.6042, indicating that no major improvements were accomplished by excluding TOC aircraft. The same holds for the 3D SSD metric, where cross correlations from 0.4369 to 0.7648 were found.

TABLE VII: Correlations metrics excluding TOC aircraft with ISA z-score

| | Mean | | Max | |
|---|---|---|---|---|
| | R | p | R | p |
| Aircraft Count | **0.3852** | **$<$ 0.05** | **0.4055** | **$<$ 0.05** |
| 2D Layered $SSD_{mean}$ | 0.2649 | $<$ 0.05 | 0.2490 | $<$ 0.05 |
| 2D Layered $SSD_{max}$ | 0.3211 | $<$ 0.05 | 0.3095 | $<$ 0.05 |
| 2D Layered $SSD_{sum}$ | **0.3874** | **$<$ 0.05** | **0.3841** | **$<$ 0.05** |
| 3D $SSD_{mean}$ | 0.2397 | $<$ 0.05 | 0.2260 | $<$ 0.05 |
| 3D $SSD_{max}$ | 0.2634 | $<$ 0.05 | 0.2346 | $<$ 0.05 |
| 3D $SSD_{sum}$ | **0.3340** | **$<$ 0.05** | **0.3287** | **$<$ 0.05** |




(a) $SSD_{mean}$ for participant C1, scenario A




(b) $SSD_{max}$ for participant C1, scenario D

Fig. 28: Cross Correlations between the 2D Layered SSD and 3D SSD

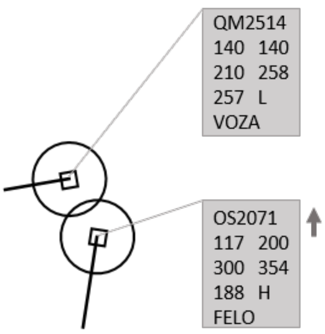



Fig. 29: Aircraft at different flight levels with overlapping protected zones as seen from the 2D layered SSD approach

To further understand what the influence of excluding transfer of control aircraft does for the performance of the solution space metrics, and to exclude the factor that this approach should have also been applied to observed aircraft, it was also investigated whether excluding TOC aircraft not only as controlled aircraft, but also as observed aircraft gave better correlations. This resulted in similar correlations as the original tested metrics.

## VIII. Metric Predictability

For further analysis of the different metrics' performance, their ability to predict workload is investigated. A linear relation can not be assumed between the metrics and most likely does not exist, therefore, a random forest regressor [33] is used. Random forest regressors do not have a distributional assumption, contrary to linear regression. Even more, random forest regressors are in general the most accurate learning algorithm available. Another advantage is that they can take many variables as input values for the algorithm to obtain more accurate predictions. This gives the option to not only use the average or maximum metric values over a period of time as was done before, but all of these variables can be taken into account at the same time.

For each participant three scenarios with known metric values and ISA z-scores were taken to predict the ISA z-scores for the fourth scenario. The following values were taken as an input to the random forest regressor: the exact metric value at the logpoint the ISA rating was given, the five metric values before and after the logpoint the ISA rating was given and the mean and maximum of all previously described metric values. To get a more precise prediction this was repeated 1000 times, where the average predicted value was taken as the predicted value. 1000 times was established as the number where the average predicted value stabilizes. It should be noted that due to the small amount of data, the predicted model will not be as accurate as it could have been when larger amounts of data would have been available. However, it is another indication of how the metric performs besides the correlation analysis. The mean absolute error of the predicted values and the actual given ISA z-scores can be found in Figure 31. It can be noticed that the metrics do not perform significantly different.

To be able to interpret these mean absolute errors, the time traces for participant C5 of the predicted and actual ISA z-score for the 3D $SSD_{sum}$ are given as an example in Figure 32.

It can be seen that the metrics give high mean absolute errors for the low workload scenario B. Although they predict the trend of the metric quite well, they are off by 1 standard deviation in z-score. Scenario A and D are predicted, given the small amount of data, quite well.

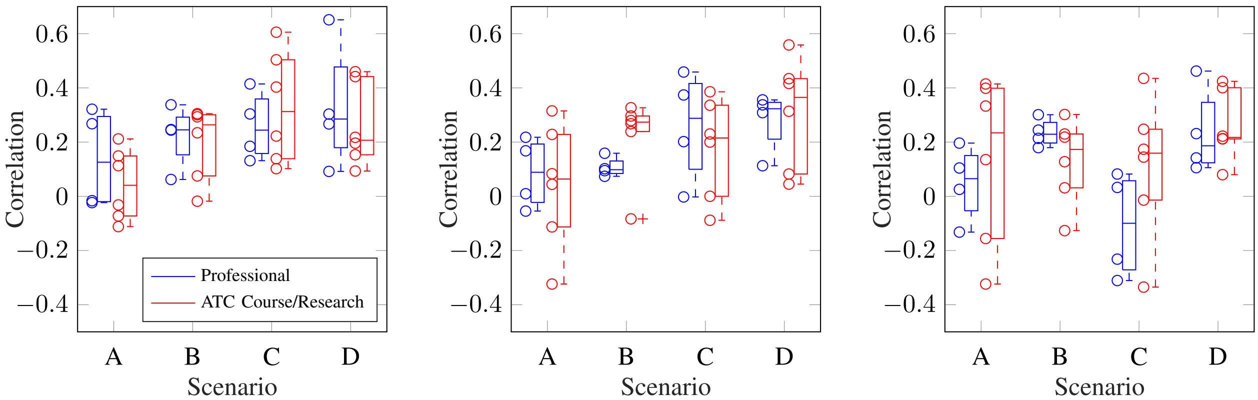
## IX. Discussion

### A. Post-task workload

The post-task workload ratings given after each scenario on the horizontal slider had the highest correlation with the overall maximum ISA z-scores. The RSME scores on the other hand showed higher correlations with the average ISA z-scores. A possible explanation for this is that the horizontal slider popped up right after each scenario and required instant feedback, when the highest workload peak is still in the most present short-term memory. Whereas for the RSME, participants had more time to consider the whole scenario, giving them the opportunity to rate it more compared to the average workload.

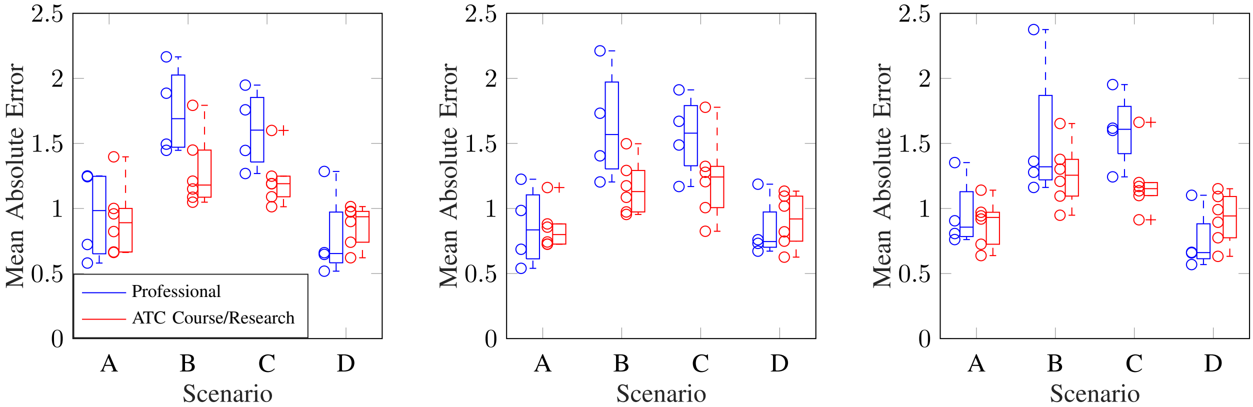
### B. Number of Commands

For the number of commands the highest correlation was found with the maximum ISA z-score. Weaker correlations




(a) Aircraft Count (b) 2D Layered $SSD_{sum}$ (c) 3D $SSD_{sum}$

Fig. 30: Correlations for taking the maximum values compared to the ISA z-score for the different metrics excluding TOC aircraft




(a) Aircraft Count (b) 2D Layered $SSD_{sum}$ (c) 3D $SSD_{sum}$

Fig. 31: Mean absolute errors for the predictions made with the different metrics per scenario per group

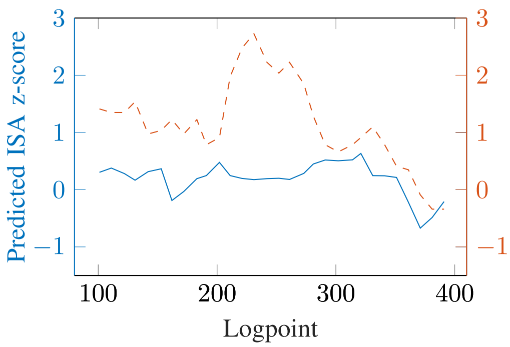
were found with the other post-task workload measurements. Although it can not be used as a predictive metric, it did show to be a good indication of workload. This is contradicting research by d'Engelbronner et al. [23], who found very weak relations for number of commands with the subjective workload ratings. This might have to do with the different set-up of experiments. D'Engelbronner et al. [23] used a 2D click and drag set-up, whereas the described 3D experiment uses an external command window with more command options such as 'transfer of control'. Hence, more interaction with the aircraft is required, making it more realistic and resulting in higher correlations with the subjective workload.

For the instructions given per command no significant correlations were found with the workload. Also no distinct differences were found between the number of given instructions per command and the different scenarios. For the given instructions per command, it can be concluded that they are not an indication for workload. However, this parameter did indicate a difference between the different groups of expertise.

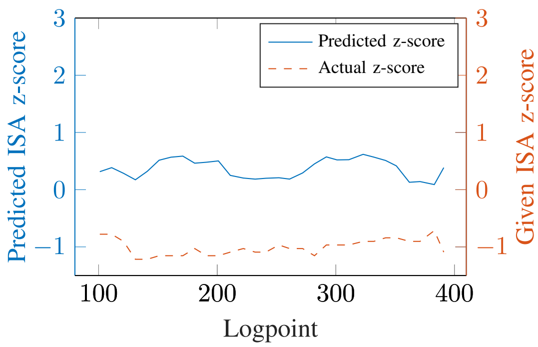
With respect to the specific instructions given, the professional group gave less DCT commands. This is corresponding with a difference in strategy and training. Professional air traffic controllers aim to make their workload as low as possible. For example, they do not want to give themselves more workload by giving a DCT command if the aircraft already travels in the right direction. The ATC course/research group on the other hand aimed to be certain that every aircraft is handled as perfect as possible. This resulted in ATC course/research participants even giving multiple DCT commands to the same aircraft.

## C. Group differences

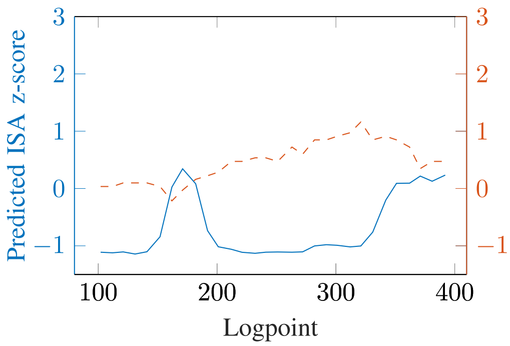
By analyzing the ISA z-scores, the different parameters and metrics, no distinct differences were found between the different expertise groups. The only parameter indicating a difference was the given instructions per command. The professional group had a preference for giving one instruction per command, whereas the ATC course group gave multiple. This indicates a difference in strategy and mindset and probably has



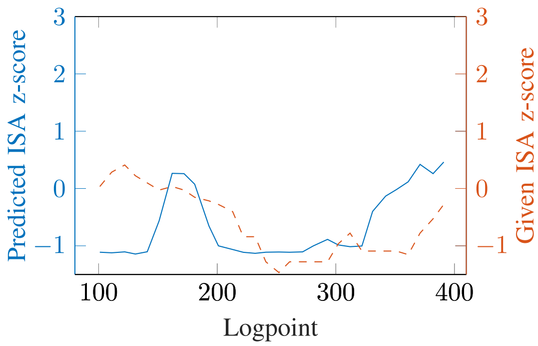

(a) Scenario A, Mean Absolute Error = 0.9759




(b) Scenario B, Mean Absolute Error = 1.3782




(c) Scenario C, Mean Absolute Error = 1.1984




(d) Scenario D, Mean Absolute Error = 0.6316

Fig. 32: The predicted ISA z-score compared to the actual give ISA z-score for C5 using the 3D $SSD_{sum}$ metric

to do with their training. Less commands to one aircraft at the same time, decreases the chances of mistakes in real life with radio contact. Their indication during the experiment of overall low workload for the scenarios, confirmed they had time for a strategy of giving one command per aircraft. The fact that they are more experienced was also notable when looking at loss of separation, which did not occur in the professional group, but did occur four times in the ATC course group.

### *D. Scenario differences*

From Figure 24 and the number of commands it became clear that scenarios A and C had a significant higher workload than scenarios B and D. The post-hoc comparison of A and C and B and D, showed that scenario A was not significantly different than C and that B was not significantly different compared to D. This indicates that the traffic intensity had a high influence on the perceived workload, since both A and C had a higher traffic intensity.

According to the different metrics, scenarios A and C were also significantly more difficult than the other scenarios. However, the metrics indicated scenario D as being significantly more difficult than scenario B, whereas the ISA scores showed no distinct difference between the two scenarios. This could be caused by the fact that scenario D included deviating aircraft, which, however, due to the low traffic intensity had no influence on the experienced subjective workload. An observation during the experiment was that deviating aircraft mainly influenced professionals, who think in patterns, whereas ATC course/research participants did not always seem to notice the change in behavior of the aircraft as they handled every aircraft individually.

These results indicate that the metrics can be used for scenario evaluation. However, also the more straightforward RSME and horizontal slider can be a first indication if no dynamic results are required.

### *E. Correlation analysis*

Compared to previous research by Abdul Rahman et al. [1] and d'Engelbronner et al. [23], overall lower correlations were found. This may be caused by multiple differences compared to previous experiments. First of all, an external command window was used instead of direct manipulation. This brings the experiment closer to reality, but makes it harder to compare to previous research. Secondly, changing the experiment set-up to a 3D scenario changes the problem of controlling and separating the air traffic. Although this on the one hand makes the experiment and results more realistic, it also gives the controller more maneuver space, making the task easier when no outside influences such as radio, delay or no flight zones are added. However, these were not added to be able to purely measure the task imposed workload and to get a first idea about the performance of the 3D SSD. This might indicate that the actual controller workload is highly influenced by these external factors. Finally, it should be noted that in general situations with lower workload are harder to rate, because it is more difficult to distinguish differences for participants. The professional control group clearly had the control strategy to keep their workload low, resulting in the difficulty that their workload differences are harder to indicate and lower correlations results are more likely to be found.

No significant differences were found between the different metrics for scenarios A, B and D, indicating that all metrics perform at the same level. For scenario C the aircraft count performed significantly better than the SSD metrics. Therefore hypothesis 1B, that the 3D SSD is better than current metrics, is rejected. Hypothesis 1A, about the performance of the 3D SSD, can due to the weak correlations in combination with no significant differences between the metrics neither be retained nor rejected and no final conclusion can be made about the current developed 3D solution space metric.

Although no statistical differences were found, it appeared that the $SSD_{sum}$ metric had slightly higher correlations with the subjective workload compared to the other SSD metrics. This is probably due to the fact that taking the sum, also takes into account the number of aircraft. This could also be seen from the fact that scenarios A and C with the highest traffic density had the highest workload ratings. This is even more emphasized by the aircraft count performing significantly better than the other metrics for scenario C. The fact that traffic density has an impact on and high correlation with the workload might had to do with the set-up of the experiment. Similar to real life, every aircraft had to be given at least one command, namely a transfer of control, therefore all traffic had to be monitored, resulting in more traffic implying more workload.

Another possible explanation of the weak correlations, is the fact that a small number of participants was used, making it difficult to filter out individual differences. As can be seen from Figure 19, the participants rated quite differently. Although taking the z-score filters out these individual differences, for more clarity and more well-founded conclusions a higher sample size is needed.

The intent for comparing the 2D layered SSD and the 3D SSD with cross correlations was to identify if controllers think in layers/flight levels or consider the complete 3D space. Although no significant differences were found, the 2D layered SSD performed slightly better. However, during the experiment it was noticed that ATC course/research participants sometimes cleared aircraft to a flight level, while the protected zones were still overlapping. This is something an actual controller would never do and gives a 100% workload peak for the 2D layered SSD. Therefore, with only 4 retired controllers participating in the experiment, the 2D layered SSD perhaps did not significantly perform better as might have been the case for a complete controller participant experiment. Hypothesis 2, that the 2D layered SSD metric would perform better for professional controllers, is, therefore, neither retained nor rejected.

The 3D SSD metric on the other hand, although it does not perform better than the aircraft count or 2D layered SSD, might still be the preferable metric when looking at new sector design and traffic routes, since it contains more information about the complete 3D space.

### *F. TOC influence*

Excluding aircraft that have been given a TOC for the metrics was expected to give better results. However, no significant differences were found. For the aircraft count, excluding these aircraft did seem to have a minor positive effect. For the SSD metrics however, this trend was not seen. This might be an indication that aircraft that have been given a transfer of control are still monitored and do have an influence on the workload. Hypothesis 3, that excluding transfer of control aircraft as controlled aircraft improves the predictability for workload is, therefore, rejected.

### *G. Metric predictability*

From analyzing the metric predictability for workload, it became clear that the metrics also performed at the same level prediction wise. The metrics gave the worst workload predictions for scenarios B and C. For scenario B, deemed the easiest scenario, over-estimations for the workload were indeed expected and can be seen in Figure 32b. For scenario C, deemed the most difficult, this result was more surprising. However, the same principle might have happened: predicting it from one high workload and two low workload scenarios, gave bad predictions. In general, scenario C gave weak workload correlations with the SSD metrics.

### *H. Future work*

The performed 3D experiment encompassed a higher fidelity than previous experiments, which could have influenced the correlations that were found. Also the small participant sample size, of which only four were retired air traffic controllers, could have been an important factor. Therefore, more research with a higher sample size and more experienced controllers is recommended. Furthermore, to add more reality and difficulty to the experiment, no fly zones, flight level limitations and delays can be introduced.

Moreover, it is recommended to focus more on research into an updated 2D layered SSD metric as this metric is more aligned with current ATC practise and therefore, expected to be more promising. Especially, because no linear relations were found between the exact 3D solution space and workload, a 2D method will offer easier and quicker calculations, making it easier to implement for real-time simulation in the future. It might also be an idea to update the 2D layered SSD by instead of taking the 2D SSD at multiple flight levels, taking horizontal cross sections of the actual 3D solution space metric.

A different application and opportunity for the developed 3D model can be the evaluation and designing of new traffic routes and air sectors, because the 3D SSD gives more insight into the available 3D space and possible conflicts than the aircraft count or 2D layered SSD.

## X. CONCLUSION

In this paper the development and evaluation of a 3D solution space-based metric for air traffic control workload has been presented. It can be concluded that the 2D solution space was successfully updated to a 3D model. A human-in-the-loop experiment, however, did not give conclusive answers as to the correlation of the newly developed metric with workload.

For the 3D solution space as well as the aircraft count and a quasi-3D layered solution space metric, weak correlations were found with workload. However, the 2D layered solution space is deemed more promising due to its alignment with current ATC practice and easier implementation compared to the complicated exact 3D solution space. No final conclusions can be made, but it is recommended to shift research from the exact 3D solution space volume estimation more towards the modeling of a 2D problem in a 3D space, hence, from a 3D solution space metric towards a more elaborated quasi-3D layered metric.

## References

[1] S. M. B. Abdul Rahman, C. Borst, M. Mulder, and M. M. van Paassen, "Sector complexity measures: A comparison," *Jurnal Teknologi*, vol. 76, no. 11, pp. 131–139, 2015.

[2] C. Boag, A. Neal, S. Loft, and G. S. Halford, "An analysis of relational complexity in an air traffic control conflict detection task." *Ergonomics*, vol. 49, pp. 1508–1526, 2006.

[3] A. W. K. Gaillard, "Comparing the concepts of mental load and stress," *Ergonomics*, vol. 36, no. 9, pp. 991–1005, 1993.

[4] B. Hilburn and P. G. A. M. Jorna, "Workload and air traffic control," in *Stress, Workload and Fatigue*, P. A. Hancock and P. A. Desmond, Eds. Lawrence Erlbaum Associates, 2001.

[5] S. Athènes, P. Averty, S. Puechmorel, D. Delahaye, and C. Collet, "Atc complexity and controller workload: Trying to bridge the gap," in *Proceedings of the International Conference on HCI in Aeronautics*. AAAI, 2002, pp. 56–60.

[6] B. Hilburn, "Cognitive Complexity in Air Traffic Control a Literature Review," *EEC note*, no. 04, 2004.

[7] J. Crutchfield and C. Rosenberg, "Predicting subjective workload ratings: A comparison and synthesis of operational and theoretical models," FAA Civil Aerospace Medical Institute and The Boeing Company, Tech. Rep. DOT/FAA/AM-07/6, March 2007.

[8] S. M. B. Abdul Rahman, M. M. van Paassen, and M. Mulder, "Using the solution space diagram in measuring the effect of sector complexity during merging scenarios," pp. 1–25, 2011.

[9] J. Djokic, B. Lorenz, and H. Fricke, "Air traffic control complexity as workload driver," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 6, pp. 930–936, 2010.

[10] R. H. Mogford, J. A. Guttman, S. L. Morrow, and P. Kopardekar, "The Complexity Construct in Air Traffic Control: A Review and Synthesis of the Literature," FAA, Tech. Rep. DOT/FAA/CT-TN9S/22, July 1995.

[11] P. U. Lee, "A Non-Linear Relationship between Controller Workload and Traffic Count," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49, no. 12, 2005, pp. 1129–1133.

[12] P. Kopardekar and S. Magyarits, "Dynamic density: Measuring and predicting sector complexity," in *Proceedings of the 21st Digital Avionics Systems Conference*, vol. 1. Inst. of Electrical and Electronics Engineers, October 2002, pp. 2C4–1 – 2C4–9.

[13] S. Loft, P. M. Sanderson, A. Neal, and M. Mooij, "Modeling and Predicting Mental Workload in En Route Air Traffic Control : Critical Review and Broader Implications," *Human Factors*, vol. 49, no. 3, pp. 376–399, 2007.

[14] P. Hermes, M. Mulder, M. M. van Paassen, J. H. L. Boering, and H. Huisman, "Solution-Space-Based Complexity Analysis of the Difficulty of Aircraft Merging Tasks," *Journal of Aircraft*, vol. 46, no. 6, pp. 1995–2015, 2009.

[15] G. A. Mercado Velasco, M. Mulder, and M. M. van Paassen, "Analysis of Air Traffic Controller Workload Reduction Based on the Solution Space for the Merging Task," in *AIAA Guidance, Navigation, and Control Conference*, 2010, pp. 1–18.

[16] G. A. Mercado Velasco, C. Borst, J. Ellerbroek, M. M. van Paassen, and M. Mulder, "The Use of Intent Information in Conflict Detection and Resolution Models Based on Dynamic Velocity Obstacles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2297–2302, 2015.

[17] S. B. J. van Dam, M. Mulder, and M. M. van Paassen, "Ecological interface design of a tactical airborne separation assistance tool," *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, vol. 38, no. 6, pp. 1221–1233, 2008.

[18] P. Fiorini and Z. Shiller, "Motion Planning in Dynamic Environments using Velocity Obstacles," *I. J. Robotic Res*, vol. 17, pp. 760–772, 1998.

[19] ——, "Motion planning in dynamic environments using the relative velocity paradigm," *1993 IEEE International Conference on Robotics and Automation*, vol. 1, pp. 560–566, 1993.

[20] RTCA, "Airborne conflict management: Application description v2.5," FAA, Tech. Rep. RTCA SC- 186, March 2002.

[21] ICAO, "Procedures for air navigation services," Tech. Rep. Doc 4444 ATM/501 (PANS-ATM), 15th ed., 2007, ch. 5.

[22] S. M. B. Abdul Rahman, "Solution Space-based Approach to Assess Sector Complexity in Air Traffic Control," Ph.D. dissertation, Delft University of Technology, 2014.

[23] J. d'Engelbronner, M. Mulder, M. M. van Paassen, S. de Stigter, and H. Huisman, "The Use of the Dynamic Solution Space to Assess Air Traffic Controller Workload," in *AIAA Guidance, Naviagtion and Control Conference*, 2010, pp. 1–21.

[24] S. M. B. Abdul Rahman, C. Borst, M. Mulder, and M. M. van Paassen, "Measuring Sector Complexity : Solution Space-Based Method," *Advances in Air Navigation Services*, pp. 11–34, 2010.

[25] W. Zhou, "The 3d solution space," Unpublished Master's Thesis, Delft University of Technology, 2011.

[26] M. W. Jones, "The Production of Volume Data from Triangular Meshes Using Voxelisation," *Computer Graphics Forum*, vol. 15, no. 5, pp. 311–318, 1996.

[27] F. Heylen, S. van Dam, M. Mulder, and M. M. van Paassen, "Design and Evaluation of a Co-planar Separation Assistance Display," in *AIAA Guidance, Navigation and Control Conference and Exhibit*, 2008, pp. 1–6.

[28] J. Ellerbroek, K. C. R. Brantegem, M. M. van Paassen, and M. Mulder, "Design of a coplanar airborne separation display," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 3, pp. 277–289, 2013.

[29] K. Brantegem, "Ecological 2-D Coplanar Airborne Separation Assurance System," Unpublished Master's Thesis, Delft University of Technology, 2011.

[30] J. Lodder, J. Comans, M. M. Van Paassen, and M. Mulder, "Altitude-extended solution space diagram for air traffic controllers," in *International Symposium on Aviation Psychology*, 2011, pp. 475–480.

[31] A. J. Tattersall and P. S. Foord, "An experimental evaluation of instantaneous self-assessment as a measure of workload," *Ergonomics*, vol. 39, no. 5, pp. 740–748, 1996.

[32] E. Farmer and A. Brownson, "Review of Workload Measurement, Analysis and Interpretation Methods," Eurocontrol, Tech. Rep. CARE-Integra-TRS-130-02-WP2-1-0, March 2003.

[33] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

# Part II

# Preliminary Report

# Part II-A

# Air Traffic Control Workload Metrics

***A literature review***

# Chapter 2

# Workload and Complexity in ATC

To be able to develop a good, feasible, objective 3D Solution Space-based metric for ATC workload, first a proper literature review about the research carried out up to now, is needed. Therefore, in this chapter is looked into what ATC workload consists of and what metrics have been developed so far and their advantages and disadvantages.

## ATC Workload

Since ATC workload is one of the limiting growth factors of world air traffic and is additionally crucial in many other elements as discussed before, it can be stated that it is worth looking into. The first step is to look at what workload is on itself and what it entails. Especially the difference between taskload and workload is of great importance. (Hilburn & Jorna, 2001) Taskload is the complexity of the task of ATC or as seen in relation to workload: it is the level of work demand imposed on the 'average' controller. Workload on the other hand is the level of work demand of the individual controller and is, hence, influenced by individual aspects such as expertise, training, strategy and resource management. (Athènes, Averty, Puechmorel, Delahaye, & Collet, 2002; Hilburn, 2004; Crutchfield, 2005; Abdul Rahman, Paassen, & Mulder, 2011) The relation between taskload and workload can be seen in Figure 2-1. Metrics for workload are actually always trying to predict the (average) taskload, such that it can be used as a prediction for the (individual) workload.

Another term often mentioned in literature concerning ATC workload is the so called mental workload. Mental workload is defined as "the term used to describe the mental cost of accomplishing task requirements" (Boag et al., 2006, p. 1508). Hence, it is more or less similar as the general term workload discussed before, since it is the load experienced by the individual controller and influenced by subjective elements. Therefore, the (mental) workload is impossible to asses objectively. (Hermes et al., 2009) However, it is possible to design an objective metric to predict the average experienced workload (taskload), which on itself is more useful, since it is the objective, subject-independent side of workload. (d'Engelbronner, 2010)

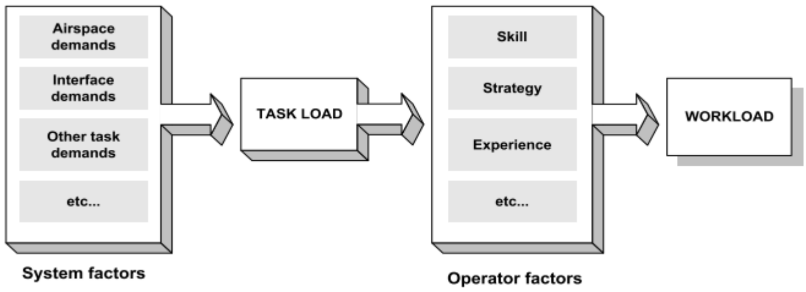



**Figure 2-1:** Taskload and Workload relation (Hilburn & Jorna, 2001)

## ATC Complexity

When discussing ATC workload, ATC complexity is often mentioned as well. This is due to the relation between complexity and workload as can be seen in Figure 2-2 and Figure 2-3. Many studies go even further into stating that ATC complexity, seen as characteristic variables to define air traffic situations, is the main contributor to ATC workload. (Athènes et al., 2002; Mogford, R. H. ; Guttman, J. A. ; Morrow, S. L. ; Kopardekar, 1995)

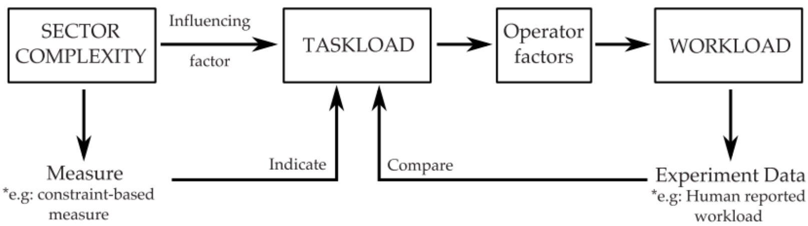



**Figure 2-2:** Relationship between sector complexity, taskload and workload (Abdul Rahman, 2014)

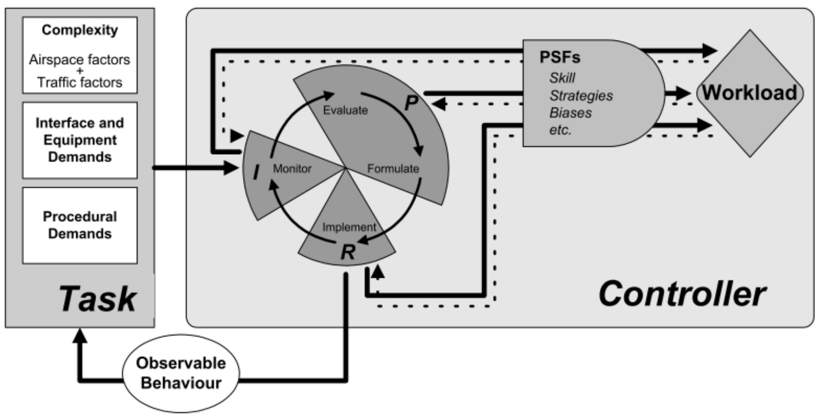



**Figure 2-3:** Process from task complexity to controller workload (Hilburn, 2004)

Another interesting aspect can be seen in Figure 2-3 when looking at the controller segment of the diagram. It consists of four parts: Monitoring, Evaluating, Formulating and Implementing. (Hilburn, 2004) The interesting aspect is that only the last part, actually executing the decisions process done earlier, is objectively observable. (Djokic, Lorenz, & Fricke, 2010) And as can be seen complexity plays an important part here. Moreover, many different complexity variables that show correlation with ATC workload have been defined by different studies. (Abdul Rahman, 2014; Hilburn, 2004) Therefore, complexity variables are worth looking into, since they can provide a first estimation for workload metrics.

## Existing ATC Workload Metrics

It was found that at the moment the most common methods for assessing taskload (or average workload) are aircraft count, dynamic density, traffic load index and the new upcoming Solution Space Diagram (SSD) method. Therefore, the pros and cons of these methods will be discussed.

### Aircraft Count

This method, as the name already gives away, uses the number of aircraft to be controlled as a metric for workload. The idea behind it is that the more aircraft are interacting in the area to be controlled, the higher the workload. Though this metric is very easy to use and apply and gives quickly basic good results, the main limitation of this metric is, it is does not take into account other important air traffic parameters such as the interaction between aircraft and flight characteristics. (Djokic et al., 2010) For example different situations with the same number of aircraft, can give very different workloads depending on if they interact in more or less complex ways (Mogford, R. H. ; Guttman, J. A. ; Morrow, S. L. ; Kopardekar, 1995), see Figure 2-4. Moreover, studies have demonstrated there is a non-linear relation between aircraft count and workload. (d'Engelbronner, 2010; Hermes et al., 2009)

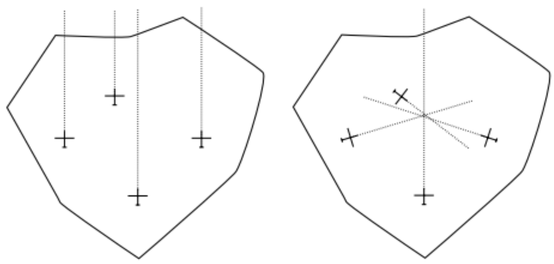


**Figure 2-4:** Different situations with the same no. of aircraft can result in different workload (Hermes et al., 2009)

### Dynamic Density

To deal with the limitations of the aircraft count and take into account more complexity parameters that influence workload and especially the dynamic contributions to the complexity, the dynamic density was developed. The dynamic density is defined as "the collective effort of all factors, or variables, that contribute to sector-level air traffic control complexity or difficulty at any point in time" (Kopardekar & Magyarits, 2002, p. 2C4-1). Hence, it can be seen as sort of a cost function including different complexity variables each given their own weights. These weights are determined and calibrated by subjective ratings of workload in a specific sector and/or situation. Therefore, they, on one hand, are way more dedicated to predicting certain workload, however, on the other hand it makes the dynamic density very case specific and not extrapolatable to other situations and sectors. Furthermore, they are still based on observable behavior and subjective ratings. (Athènes et al., 2002; Hilburn, 2004) And have been lacking to be able to accurately enough predict ATC workload for new scenarios. (Loft, Sanderson, Neal, & Mooij, 2007) This because the parameter weights are only applicable for the sector and scenario they are calibrated for. (Abdul Rahman, 2014; Hilburn, 2004)

### Traffic Load Index

In order to improve the dynamic density and in an effort to try to combine geometric and dynamic aspects, the traffic load index was invented. This method uses weights on aircraft to estimate the workload for the controller. Each aircraft starts out with a load of 1. If an aircraft is involved in a possible conflict or other workload intense scenario the weight is increased up to a maximum of 3.5. The ratings for each aircraft are adjusted with each radio update and the summed total rating of all aircraft in the sector give a traffic load index which indicates the ATC workload in that sector at that time. (Athènes et al., 2002)

Related to the traffic load index, is the theory of relational complexity. This theory states that task demand and cognitive capacity can be linked. Boag et al. (2006) used this theory to consider pair-wise relations among aircraft. Similar to the traffic load index, each of these interacting pair of aircraft, gets a relational complexity or load assigned. Where the relational complexity equaled the number of variables the controller has to process. All interacting pair relational complexities are summed, such that a total load on a scale from 0 to 12 was obtained.

The advantage of the traffic load index and similar theories is that it takes not only the aircraft count, but in addition uses the dynamic behavior and effects of time pressure and uncertainty. (Hermes et al., 2009) However, it still has as a downside that it relies on subjective ratings and researchers to establish and calibrate the appropriate loadings for the aircraft in the sector. And hence, is extremely case specific, which can be an advantage, but for general objective workload estimation also a huge disadvantage. (d'Engelbronner, 2010)

### Solution Space

A final more upcoming method is the Solution Space. The idea behind this metric is that ATC workload is inversely related to the space consisting of all available solutions for the controlled aircraft, the 'solution space'. Hence, the smaller the solution space, the less maneuvering space the controller has to resolve conflict, and the higher the ATC workload. This method is based on the so called forbidden beam zones as researched by Van Dam (2008) and uses the relative velocity plane. (d'Engelbronner, 2010) This way it provides an objective new method and takes into account many of the complexity variables as mentioned before. First tests also show high correlations between the solution space and tested subjective workload. (d'Engelbronner, 2010; Hermes et al., 2009) Currently the main limitation of the solution space diagram is, it is only completely developed in 2D and hence, doesn't take into account the complexity of vertical separation.

In comparison with the other metrics mentioned before, the solution space performs the same or better. For even better understanding of the possibilities and differences, see Table 2-1. In this table the criteria of an objective complexity measure are set out and compared for the different metrics discussed.

**Table 2-1:** Comparison of different ATC metrics, adapted from (Abdul Rahman, 2014)

| Metric | Independent of Controller | Independent of Sector Lay-Out | Captures traffic dynamic | Captures future condition | Output |
|---|---|---|---|---|---|
| Aircraft Count | No | No | No | No | Scalar |
| Dynamic Density | No | No | Yes | Yes | Scalar |
| Traffic Load Index | No | No | Yes | Yes | Scalar |
| Solution Space | Yes | Yes | Yes | Yes | Diagram/Scalar |

As can be seen and established from the comparison the solution space gives promising results and has as main advantage the objectivity and general application compared to the other methods. Therefore, next is looked closer into the solution space and the options to extend it to 3D.

# Chapter 3

# Solution Space

## 2D Solution Space

Since the solution space gives the most promising results so far, a 3D metric will be designed using that method to evaluate ATC workload. However, to be able to apply this method, one first needs to properly understand how this method works and what the underlying thoughts are. Therefore, in this section the development and idea of the solution space will be discussed. Next, other 2D Velocity Obstacle (VO) methods will be looked into to see if something can be learned from them.

### Velocity Obstacle

The first step towards a new way of approaching collision avoidance was done by Tychonievich (1989) in his research. To be able to look at a problem of two moving objects in 2D in a stationary sense, the collision cone obtained by tangent lines towards the observed object in the horizontal plane, is translated with the help of the relative velocity vector, see Figure 3-1. This is the basis of the developed velocity obstacle method by Fiorini and Shiller (1998) and collision cones as described by Chakravarthy and Ghose (1998). By using relative velocity the collision cone can be translated such that if the current object controlled has a velocity vector that falls into this cone, collision will occur and, hence, a velocity vector can be selected that avoids this set, the 'velocity obstacle', and collision avoidance can be obtained. The downside of these methods is they are only developed for rectangular motion and in the horizontal plane only; hence, only valid for 2D path planning problems. (Abdul Rahman, 2014; Chakravarthy & Ghose, 2012)

The strength of the velocity obstacle method is that by using relative velocity to translate the relative collision cone to an absolute collision cone, it can see a dynamic situation as static and only needs the current position and velocity of the objects controlled and observed, which makes it a first order method (Fiorini & Shiller, 1993) Though the method was originally developed for one on one avoidance, it can easily be extended to multiple objects,

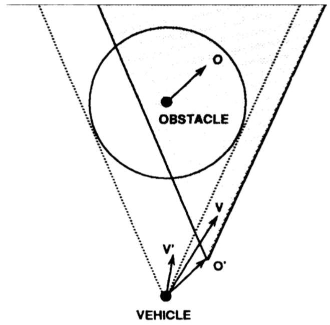



**Figure 3-1:** Moving obstacle avoidance cone (Tychonievich et al., 1989)

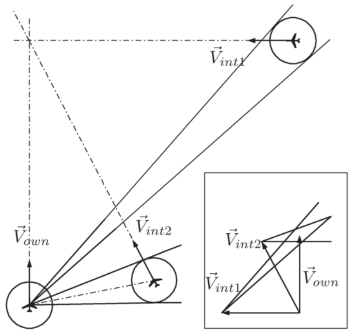



**Figure 3-2:** Two forbidden beam zones due to conflict with two intruders (van Dam et al., 2008)

which results in multiple velocity obstacles. (Fiorini & Shiller, 1998)

To bring this method closer to current aircraft navigation, the so called Forbidden Beam Zone (FBZ) was developed by Van Dam (2008). This is related to the velocity obstacle, but then by looking at the minimum spatial separation to maintain the safety goal for aircraft. This is based on the definition of the so called Protected Zone (PZ) around aircraft, in the 2D horizontal plane a circle with a radius of 5 NM centered around the aircraft in which no other objects are allowed, otherwise âĂIJloss of separationâĂİ does occur. (RTCA, 2002) Hence, looking back at the collision trajectories, this protected zone is the object with which collision should be avoided. Thus the separation constraints enforced by the PZ result, seen in a relative velocity field, in the forbidden beam zone. Which can be created by tangent lines resulting from the own position and tangent to the PZ, translated with the relative velocity vector, see Figure 3-2. Interesting characteristics of the FBZ are for example that it expands or opens up the closer the aircraft get; hence, the bigger the FBZ gets.

## Construction of the 2D Solution Space

Using the FBZ the solution space can be defined. Since the controlled aircraft is in addition bound with a minimum and maximum velocity vector, the so called flight or performance envelope is superimposed on the FBZ to obtain the solution space: the subset of all possible velocity and heading vectors while satisfying the safety and separation constraints, see Figure 3-3. Hermes (2009) also looked at the solution space, but then calculated by looking separately at the different parts of the aircraft trajectory over time. Which gets complex and time consuming quickly while longer and more difficult trajectories are observed. Therefore, Mercado Velasco (2015) simplified this approach to make it more applicable. The solution space was originally also more introduced from a pilot's perspective. Whereas Mercado Velasco (2010) also first links this solution space and hence, the geometry and complexity of the airspace, to air traffic control workload. By redefining the FBZ, Mercado Velasco (2015) opened the way for fairly easy calculations when considering aircraft intent, acceleration, heading and a new way to round off for finite time consideration.

The solution space as defined by Mercado Velasco first in (2010) and later improved in (2015), can be constructed as follows. Instead of looking at how to avoid collision, is looked at the

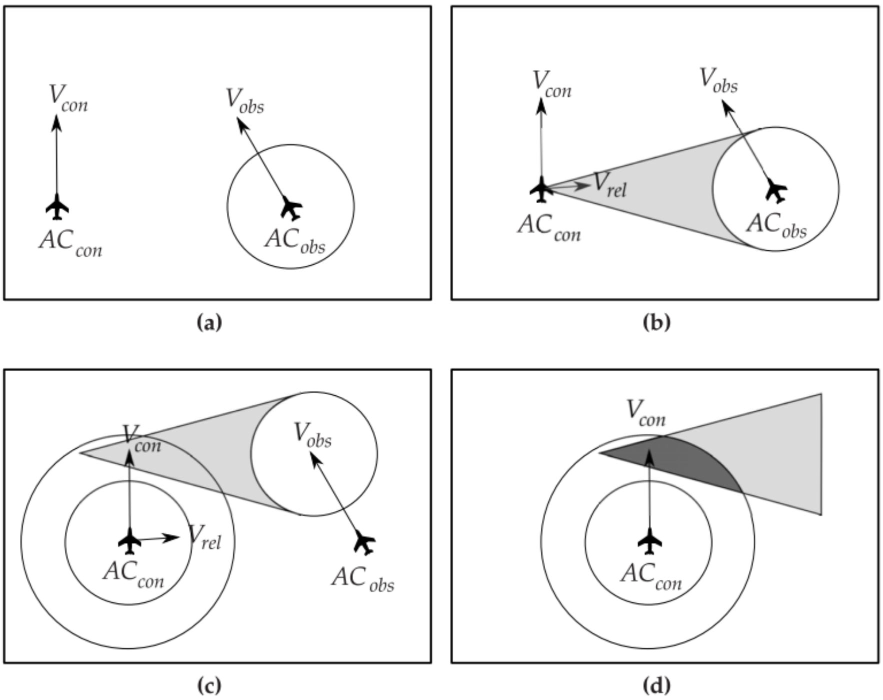



**Figure 3-3:** Construction of the 2D solution space (Abdul Rahman, 2014)

trajectory the controlled aircraft should fly to obtain collision with the observed aircraft, see Figure 3-4, and the belonging collision velocity vector, see Figure 3-5. However, as can be seen, not only the aircraft is defined as collision, in addition intruding the protected zone will lead to loss of separation. Therefore, around this set of velocity vectors a set of projected circles has to be taken. All these circles combined form the forbidden beam zone, see Figure 3-6, and this family of circles can be described by the Parametric Equation 3-1. Hence, the intersection of the envelope equation of this circular velocity set and the performance envelope of the controlled aircraft will result in the solution space. As can be seen it is purely based on sector geometric and aircraft kinematic properties. (Abdul Rahman, 2014; Hermes et al., 2009; d'Engelbronner, 2010)

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \mathbf{v}_c(t_c) + r(t_c) \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}$$
$$\forall \quad \theta \in [-\pi, \pi]; \quad t_c \in [t_0, \infty\rangle \tag{3-1}$$

Though the solution space was originally more developed for collision avoidance and pilot visual aid, via Hermes (2009), d'Engelbronner (2010), Mercado Velasco (2010) and Abdul Rahman (2010) the idea has formed to use it as an ATC workload metric. However, the solution space diagram itself only contains visual information and cannot be applied directly as a metric. (d'Engelbronner, 2010) Therefore, for workload should specifically looked into the calculating of the intersection area of the FBZ and the flight envelope. According to Abdul Rahman (2014), ATC considers six attributes: flight level, flight path, longitudinal separation,

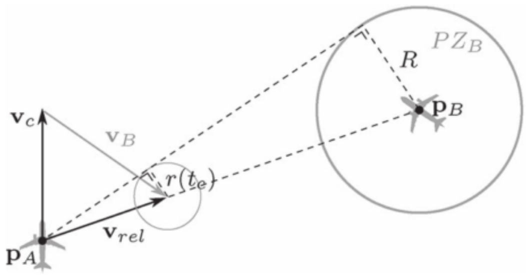



**Figure 3-4:** Obtaining a circular velocity set (Mercado Velasco et al., 2015)

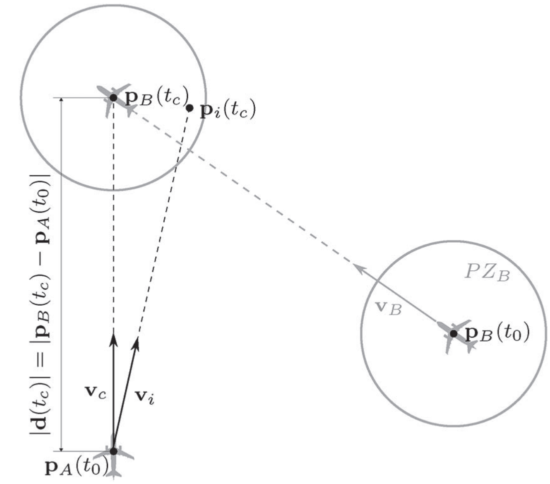



**Figure 3-5:** Construction of the collision velocity vector(Mercado Velasco et al., 2015)

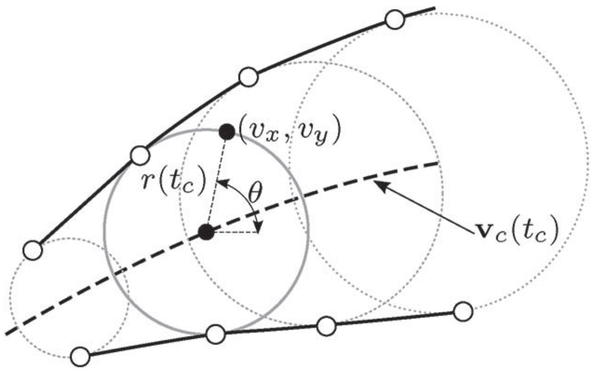



**Figure 3-6:** Family of circles and envelope equation (Mercado Velasco et al., 2015)

relative velocity, direction of flight after reporting points and lateral separation. The 2D solution space metric as described, takes into account five of these conditions, transforming it to 3D will include all six.

## Other 2DVO applications

Though using the method of Mercado Velasco (2015) so far looks promising to extend to 3D, it is also wise to look at other applications of the 2D velocity obstacle or similar methods, to see if they indicate any ideas for extension to 3D. One such a similar geometric method is defined by Bilimoria (2000) and consists of a closed-form analytical solution for the optimal heading changes, velocity changes and combinations for conflict resolution in the horizontal plane. This is done by a geometric interpretation of the aircraft trajectory by looking at minimizing trajectory deviations and separation distances, which also results in two tangent lines from the protected zone of the observed aircraft to the controlled aircraft. Another similar method was developed by Damas and Santos-Victor (2009), which uses the so called 'forbidden velocity map'. Similar to the velocity obstacle here every observed object also results in a cone shaped forbidden velocity region. What makes this method different from the original velocity obstacle is that when applied to avoidance, the velocity obstacle method always looks at the area outside of the collision cone as the possible solution set, whereas Damas focuses on the forbidden velocity set by using velocity polar coordinates; hence, the inside area of the collision cone. By doing this, his collision cone or forbidden velocity map can easily be extended for speed uncertainty.

Additions and improvement to the standard VO method were also done. So developed Chakravarthy and Ghose (1998) a collision cone approach that is valid for multiple shaped objects by using kinematic equations and relative velocity components. Shiller (2007) on the other hand looked at non-linear trajectories and created a non-linear velocity obstacle, which is also still first order and results in a warped cone. This method, by using a translation, only needs the current velocity and path curvature of the observed object and hence, can be used for observed objects with unknown trajectories.

Another improvement of the velocity obstacle method for robot applications was done by Snape (2011) with the hybrid reciprocal velocity obstacle. This entails the assumption that other observed objects or robots are cooperating and also actively trying to avoid collision. Therefore, not as drastic trajectory deviations are needed as with the normal VO by incorporating the cooperation of the observed object. However, should be noted, though this is a nice feature for robots of the same type, in non-cooperative situations or non-communicative situations this method can never be applied. Jenie (2013) also builds on this idea with the selective velocity obstacle method. It is called selective, since it takes into account certain pre-programmed or pre-settled flight rules, which changes the optional avoidance manoeuvres outside of the VO.

Finally, Lodder (2011), by using the 2D solution space at different flight levels, sets a first step towards the 3D version. Here aircraft flying at certain predefined flight levels above and under the controlled aircraft are taken into account by taking their 2D projected solution space at the flight level the aircraft is flying or intending to fly to. In addition airspeed and response time effects were taken into account when calculating the solution space per flight level. However, it should be noted that in practice controllers block intermediate flight levels if they give an altitude clearing. Hence, they do not actually take into account the exact begin and end time of the aircraft at a certain flight level. Therefore, these effects can actually be neglected.

# 3D Solution Space

Now the 2D Solution Space is properly understood, the transition to 3D can be made. First, 3D velocity obstacles for spherical protected zones will be looked upon, since this is the easiest transition from 2D to 3D. Unfortunately the PZ is not of a sphere shape, as discussed next, but a cylindrical shape. Therefore, next the existing cylindrical models are discussed including their current limitations. Finally, since now is looked into the 3D solution space, that is the 3D volume left over when intersecting the flight envelope and the 3D collision cone, intersections of the 3D collision cone are discussed.

### Collision Cones for Spherical objects

The first and most straightforward transition from a 2D to a 3D velocity obstacle is done by taking, instead of a circle in a horizontal plane, a spherical protected zone. This was already done relatively early by Fiorini and Shiller (1993), see Figure 3-7. Goss and Subbarao (2004) also look at analytical solutions for collision avoidance in a 3D environment by using a spherical protected zone and a mixed geometric and collision cone method. In the same way as the 2D velocity obstacle described by Bilimoria (2000) and Chakravarthy and Ghose (2011) this 3D approach uses the current position coordinates and instantaneous velocities vectors as input in an Earth-fixed axes system. Next, it looks at the geometric heading and deviation angles of the vectors in combination with the relative velocity and the line-of-sight vector to describe the collision cone. The optimal collision avoidance strategy is then obtained by minimizing the separation distance: the tangential solutions of the sphere. The result, however, is similar to the one obtained by Fiorini as seen in Figure 3-7.

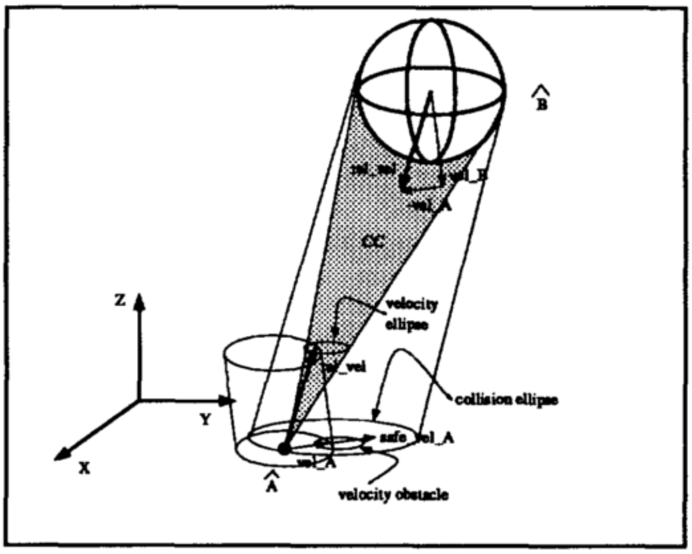



**Figure 3-7:** 3D velocity obstacle with spherical protected zone (Fiorini & Shiller, 1993)

Carbone (2006) uses a similar geometric collision cone based approach. The difference is, however, in the avoidance strategy, which in this case considers possible separate lateral, longitudinal and speed changes. Luongo (2009) improves this, by finding simultaneous change solutions: optimal solutions for simultaneous change of all control variables. However, where Carbone does especially take the aircraft dynamic limitations into account for the analytical solution, which is also necessary for air traffic control, Luongo uses several simplifying assumptions such as neglecting the flight envelope and dynamic constraints. Furthermore, no changes in speed vector after $t_0$ are allowed and the speed vector has to be a step input. All limitations which make it less interesting for an ATC case.

Finally, Jenie (2016) defines in addition to the velocity obstacle, so called 'diverging zones', the 'reachable velocity' and 'avoidance planes'. Though this was relatively straightforward in 2D as described in (Jenie et al., 2013). For the 3D case this is more of a struggle and this mainly ends up being resolved by defining solutions on previously mentioned avoidance planes, which are cutting planes of the collision cone. Therefore, Jenie looks further into these conic sections and the parameters that describe and influence them, which can be an interesting starting point for defining the volume of the solution space as will be discussed later in the conic intersection section. The main interesting idea that can be taken from it is that the velocity obstacle cone is defined in parametric equations in Euclidean space for easy obtaining of the conic section of the velocity obstacle and avoidance planes.

### Flight envelope and protected zone

Taking a spherical protected zone results in a relatively easy and straightforward transition to 3D, however, a spherical protected zone is not the best approach in every case. The literature discussing spherical protected zones is mainly designed for UAVs, where the flight path and velocity is 360 degrees flexible. However, in airplane and hence, air traffic control cases, the PZ can not simply be taken as a sphere and as Carbone (2006) mentioned, an aircraft is bound by certain lateral and vertical maneuvering limits, resulting in a different flight envelope and protected zone.

This protected zone, or the area that other aircraft are not allowed to intrude to avoid loss of separation, is based on the minimum separation distances. The common separation minimums for aircraft are defined as 5 NM in the horizontal plane (the circle as used in 2D) and a 1000 ft vertical separation, which results in a cylinder in the 3D space. (Ellerbroek, Brantegem, van Paassen, & Mulder, 2013; Brantegem, 2011)

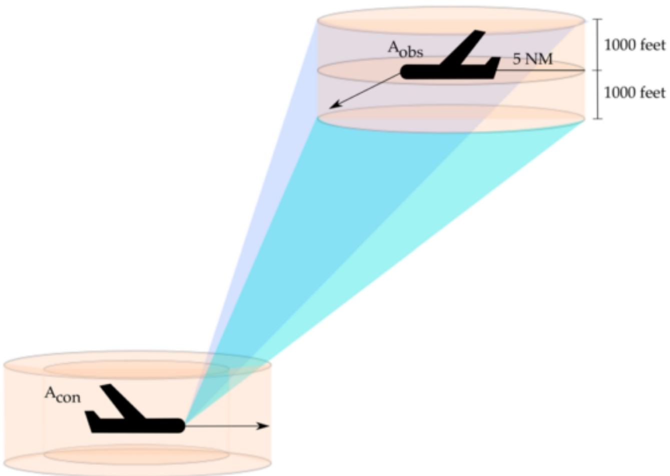



**Figure 3-8:** 3D cylindrical protected zone, 3D flight envelope and resulting 3D forbidden beam zones (Abdul Rahman, 2014)

The flight or performance envelope of the controlled aircraft in 3D on the other hand is obtained by first looking at the flight envelope in the horizontal and vertical plane, bounded by minimum/maximum speed vector and climb/descent limitations respectively. (Ellerbroek et al., 2013) To transform this to 3D, Zhou (2011), simply rotates the performance envelope 360 degrees around the vertical axis, combined with the PZ this results in Figure 3-8.

However, this is only valid when taking the assumption that the controlled aircraft is able to spontaneously change its speed vector over 360 degrees.

### Collision Cones for Cylindrical objects

A step closer to the 3D solution space design can be done by looking into the first approximations and geometric optimizations found using cylindrical instead of spherical protected zones. Luongo (2010) for example made an analytical algorithm for the collision avoidance using a cylindrical safety bubble. The same way as done for the spherical version (Luongo et al., 2009), an avoidance region is created with the help of relative speed vectors and a simultaneous change of all control variables, namely velocity, slope and track angles, to minimize trajectory deviation, while preventing loss of separation. This results in four solutions, one tangential line for each surface of the cylinder. The downside, however, is though already an improvement by using a cylinder, still the same assumptions are taken regarding the velocity and not taking into account the flight envelope and aircraft dynamics.

Chakravarthy and Ghose (2011) go even further, by not only looking at cylindrical surfaces, but generalizing the problem for quadric surfaces. For example ellipsoids, hyperboloids and cylinders belong to this group. Though exact analytical solutions are obtained for differential surfaces such as ellipsoids, the more complicated non-convex shaped surfaces (such as cylinders) are approached by using spatial cones and planar sections of the 3D collision cone and union of these collision cones to approximate the total collision cone. A slight expansion of this was done in (Chakravarthy & Ghose, 2012); however, still only general collision and avoidance conditions were obtained with the construction of planar sections by looking at planes consisting the relative velocity vector and the controlled and observed object. However, these methods are only valid for constant velocities and offer quite a long-winding method instead of one that can easily be translated for collision avoidance to a metric for ATC workload.

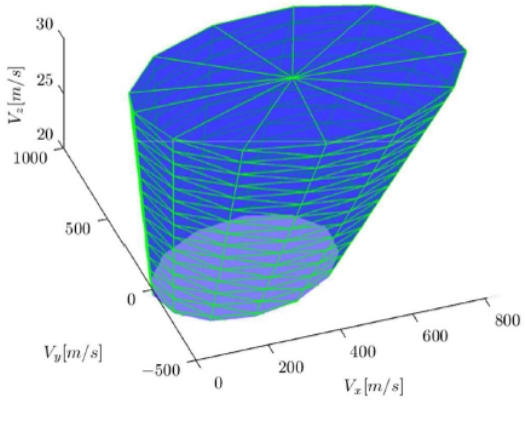



**Figure 3-9:** Triangulated surface simplified 3DSSD (Zhou, 2011)

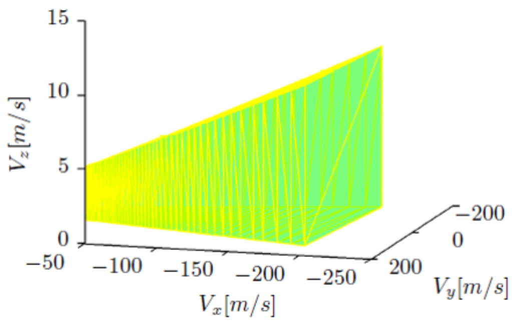



**Figure 3-10:** Triangulated surface of the tangential planes (Zhou, 2011)

Finally, the closest to an optimal analytical solution using the solution space and elliptical protected zones, is the method by Zhou (2011). This is done by looking at the top and the bottom circle of the cylinder and constructing overlapping collision cones from there by putting a tangential block between them. By looking at the intersection of these cones and the 360 degrees rotated flight envelope an approximation for the solution space is obtained, see Figure 3-8.

However, though a beginning towards an analytical form is made, the final method consist of numerical approximations of the volume of the 3D solution space. This is done by creating meshed surfaces of the different constructed 3D Figures (3D flight envelope, collision cone bottom circle, collision cone top circle and 3D tangential block), see Figures 3-9 and 3-10 for examples. Finally, a voxelisation method, illustrated in Figure 3-11, is used to determine if 3D cubes are insides or outside the meshed surface. After which all 3D cubes that lie in all the different 3D meshed figures are counted. As can be noted, many assumptions of the volume that forms the 3D solution space (such as neglecting some realistic aircraft dynamic limitations) were done instead of an easy to use analytical closed form. This can partly be attributed to the fact that later improvements of the 2D analytical mathematical form as described by (Mercado Velasco et al., 2015) were done after Zhou's work.

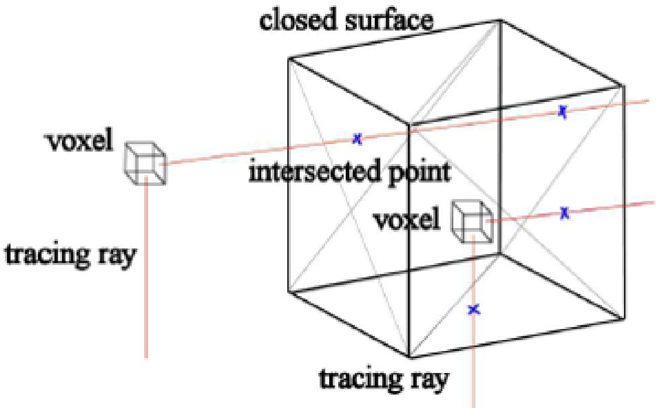



**Figure 3-11:** Ray intersection method (Zhou, 2011)

## Coplanar and Conic Intersection

As mentioned Chakravarthy and Ghose (2012) already looked at conic intersections to obtain approximations for collision cones for quadric surfaces. Therefore, it might be interesting to look at other literature that looks into coplanar usage and conic intersections. So uses Ellerbroek (2011) conic intersections of the 3D forbidden beam zone for a coplanar display that gives the pilot insight into future loss of separation, see Figure 3-12 for an example of a vertical crossection and visualization of the 3D forbidden beam zone, which has sort of a flattened cone shape. The 3D forbidden beam zone is fairly easily obtained by defining it as "the volume that contains all vectors, originating from the own aircraft, that point in the direction of an intruder's protected zone" (Brantegem, 2011, p. 25). Even more, the mathematical calculations behind these cutting planes are straightforward and can be found in (Brantegem, 2011), where is looked at aligning the vectors of the cutting planes with the mathematical form of the 3D forbidden beam zone.

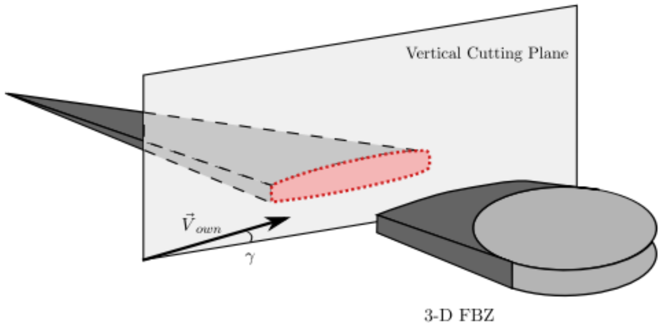



**Figure 3-12:** Vertical crossection of the 3D forbidden beam zone (Brantegem, 2011)

# Part II-B

# Air Traffic Control Workload Metrics

***Towards a 3D solution space-based metric***

Chapter 4

# 3D Solution Space-based Metric

Having established that the solution space is promising, however, an analytical form for 3D still lacks, the design for a new metric can be done. Starting out with the analytical parametric equations from 2D, a transition is made to a 3D metric. Afterwards, the new metric is more critically looked upon with a sensitivity analysis and finally, preliminary conclusions will be done.

## Circle in 3D

The first transition to 3D can be made, by putting the family of circles as described by Mercado Velasco (2015), see Equation 3-1, in a 3D space. This results in Equation 4-1 and visualized results in a 3D collision cone, see Figure 4-1. Here, for a reference, already the aircraft 3D performance envelope is plotted to give an idea of the part of the collision cone that is within the optional speed ranges of the controlled aircraft. This will be elaborated on in the next sections.

$$\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} = \mathbf{v}_c(t_c) + r(t_c) \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{bmatrix}$$
$$\forall \quad \theta \in [-\pi, \pi]; \quad t_c \in [t_0, \infty\rangle \tag{4-1}$$

In Equation 4-1 the collision course velocity vector $\mathbf{v}_c(t_c)$ is defined as in Equation 4-2. Where $\mathbf{r}_{obs}(t_c)$ is the position of the observed aircraft at time to collision $t_c$, which is defined as Equation 4-3. $\mathbf{r}_{con}(t_0)$ is the position of the controlled aircraft at time $t_0$.

$$\mathbf{v}_c(t_c) = \frac{\mathbf{r}_{obs}(t_c) - \mathbf{r}_{con}(t_0)}{t_c - t_0} \tag{4-2}$$

$$\mathbf{r}_{obs}(t_c) = \mathbf{r}_{obs}(t_0) + \mathbf{v}_B \cdot t_c \tag{4-3}$$

Here $\mathbf{v}_B$ is the velocity vector of the observed aircraft. Finally, $r(t_c)$ from Equation 4-2 is defined as Equation 4-4, where R is the radius of the protected zone. Time, velocities and distances are all taken in SI units.

$$r(t_c) = \frac{R}{t_c - t_0} \tag{4-4}$$

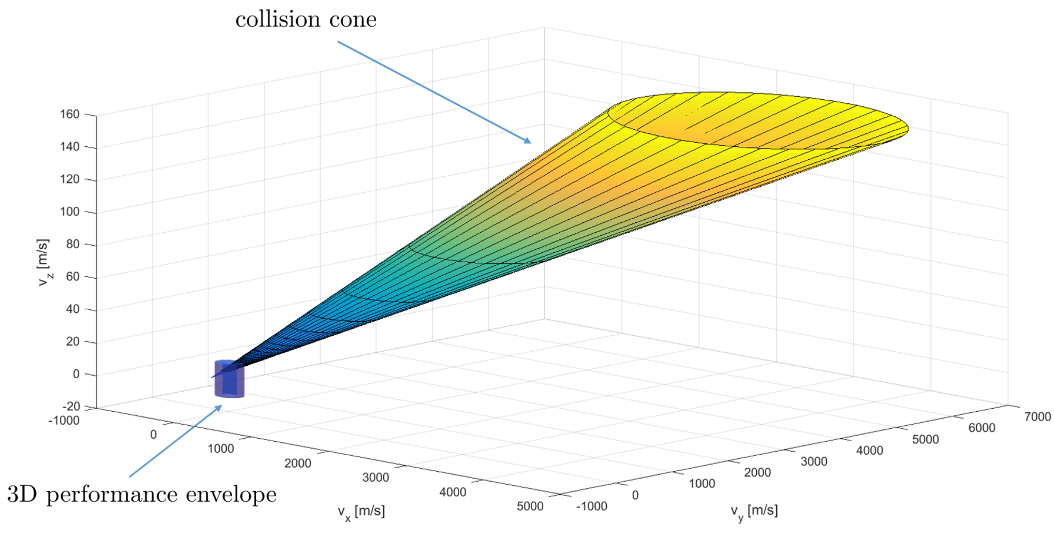



**Figure 4-1:** Collision cone due to a circle in 3D. The aircraft 3D performance envelope is illustrated at (0,0,0).

## 3D FBZ Cylindrical Protected Zone

However, for an aircraft in 3D space the protected zone is not simply a circle in a 3D environment. The separation requirements as mentioned before consist of 5 NM horizontal separation and 1000 ft vertical separation, illustrated in Figure 4-2, which results in a cylinder around the aircraft with a radius of 9260 meter and a height of 609.6 meter in SI units.

Therefore, say this cylinder is built up from infinite many circles ranging from the lower to the upper height as described by the vertical separation requirements. This implies that the 3D FBZ can be seen as built up from infinite many of the collision cones as described before, resulting from these circles in 3D. See Figure 4-3 for 5 of these collision cones, resulting from circles that are part of the protected zone cylinder.

To understand how this works and to get a parametric equation for the complete 3D figure, first, is looked at the horizontal cross section at $v_z = 10\,[m/s]$ of multiple of these collision cones, see Figure 4-4.

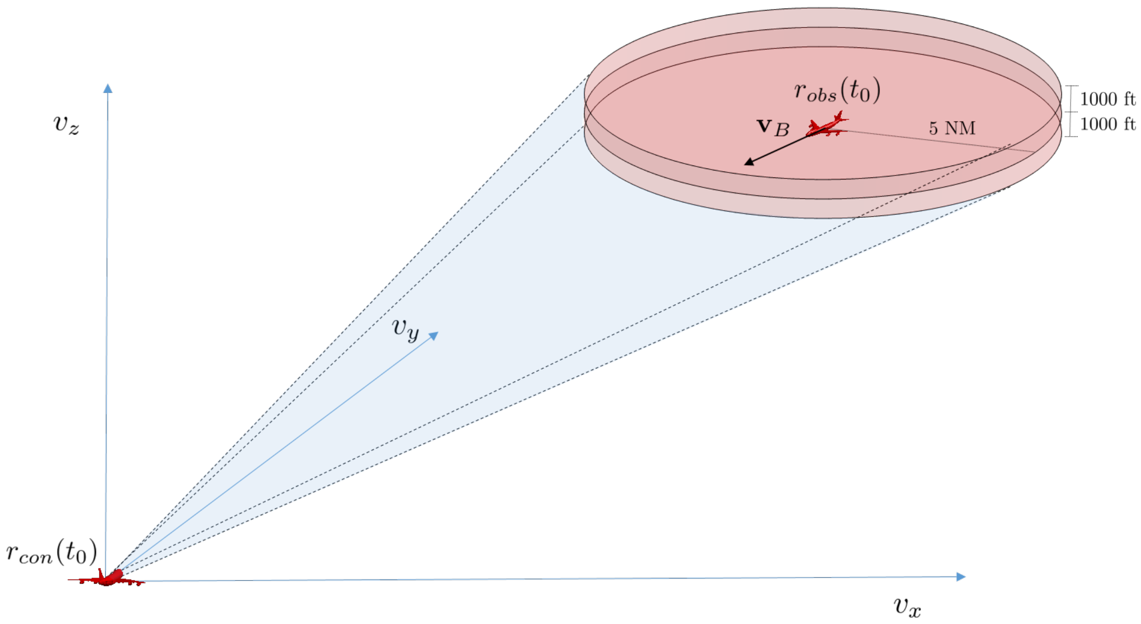



**Figure 4-2:** Observed aircraft with cylindrical protected zone in relation to the controlled aircraft

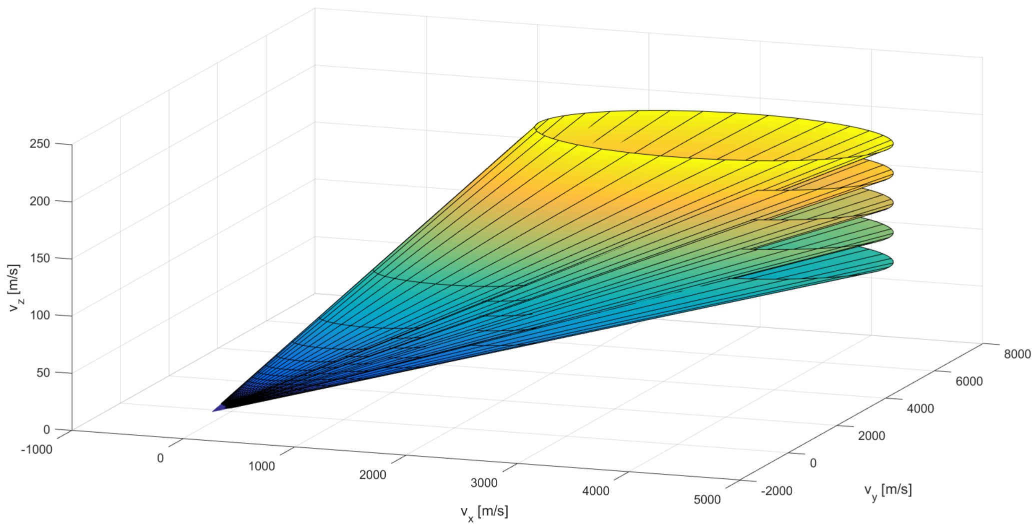



**Figure 4-3:** 5 collision cones as part of the protected zone cylinder

Now at a certain vertical velocity $v_z = Z$ holds for each family:

$$v_{c,z} = Z, \tag{4-5}$$

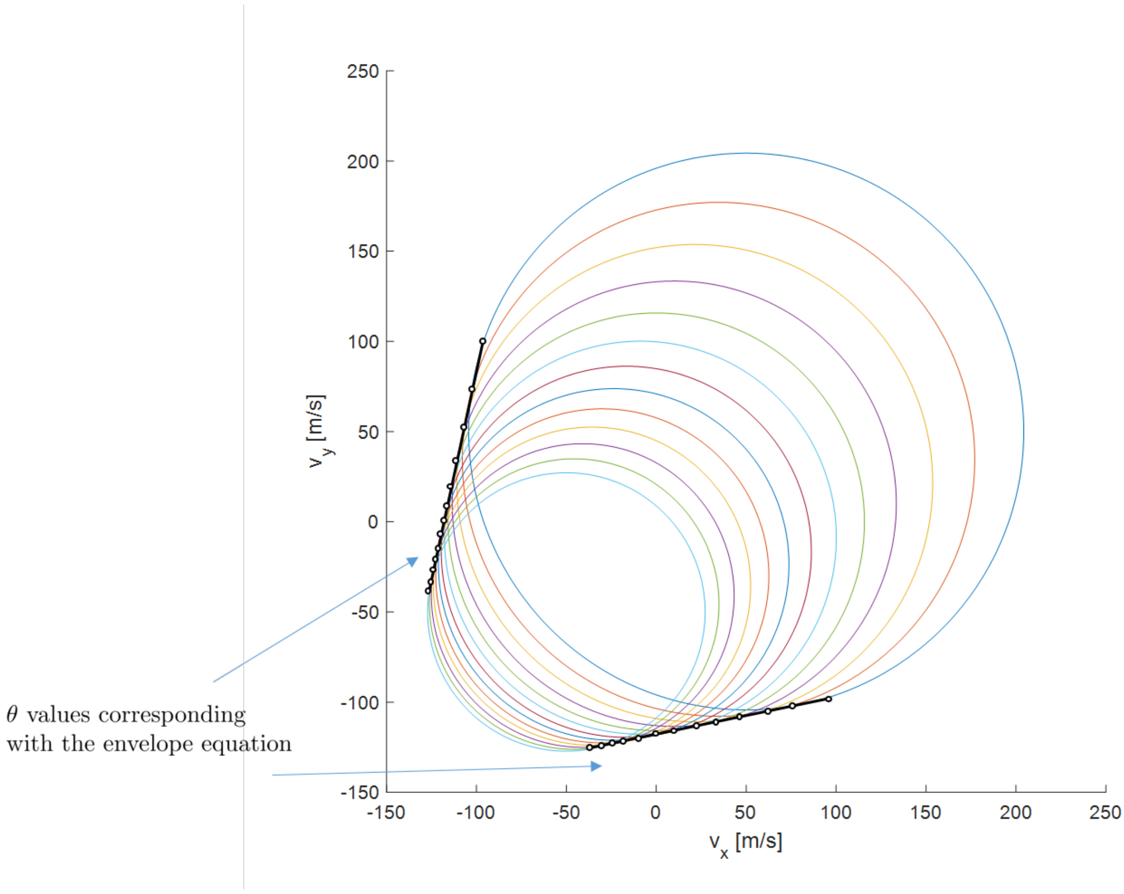



**Figure 4-4:** Horizontal cross section of multiple collision cones at $v_z = 10\,[m/s]$

which gives, using Equation 4-2:

$$\frac{r_{obs,z}(t_c) - r_{con,z}(t_0)}{t_c - t_0} = Z, \tag{4-6}$$

where $r_{obs,z}(t_c)$ defines which circle, with accompanying circle family as defined by Equation 4-1, of the original cylinder is looked at.

Now for simplicity the assumption is taken that $r_{con,z}(t_0) = 0$ and $t_0 = 0$. These simplifications are not necessary for actual calculations, however, they are done here to give a clearer view of how the analytical set up of the metric works. Applying these assumptions and combining Equations 4-3 and 4-6, results in Equation 4-7:

$$r_{obs,z}(t_0) + v_{B,z} \cdot t_c = Z \cdot t_c \tag{4-7}$$

Again for clarity the assumption is made that $v_{B,z}$ is not dependent on $t_c$. This is not done in actual calculations, but makes the derivation of $t_c$ easier by hand for visualization. This results in Equation 4-8 and finally, Equation 4-9. Note that the time to collision $t_c$ is a function of $r_{obs,z}(t_0)$, the vertical position of the observed aircraft at time $t_0$.

$$r_{obs,z}(t_0) = (Z - v_{B,z}) \cdot t_c \tag{4-8}$$

$$t_c = \frac{r_{obs,z}(t_0)}{Z - v_{B,z}} \tag{4-9}$$

Hence, at a certain vertical speed $Z$ of the controlled aircraft, for every circle that can be traced back to a circle of the original cylinder by $r_{obs,z}(t_0)$ there is an accompanying $t_c$ value. Hence, for a certain $Z$ a new 2D family of circles emerges, see Equation 4-10. Because $t_c$ is a function of $r_{obs,z}(t_0)$, $r_{obs,z}(t_0)$ is taken as the new variable. Here $z_1$ and $z_2$ are respectively the lower and upper height of the PZ, where $z_2 > z_1$.

$$\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} = \mathbf{v}_c(r_{obs,z}) + r(r_{obs,z}) \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{bmatrix}$$
$$\forall \quad \theta \in [-\pi, \pi]; \quad r_{obs,z}(t_0) \in [z_1, z_2] \tag{4-10}$$

The envelope equation, as illustrated in Figure 4-4, for this family of circles can be found by slightly adapting the envelope equations as used by Mercado Velasco (2015), see Equation 4-11.

$$\begin{vmatrix} \frac{\partial v_x}{\partial t_c} & \frac{\partial v_x}{\partial \theta} \\ \frac{\partial v_y}{\partial t_c} & \frac{\partial v_y}{\partial \theta} \end{vmatrix} = 0 \tag{4-11}$$

Where holds,

$$\dot{v}_{c_x} = \frac{\partial v_{c_x}}{\partial t_c} \quad \dot{v}_{c_y} = \frac{\partial v_{c_y}}{\partial t_c} \quad \dot{r} = \frac{\mathrm{d}r}{\mathrm{d}t_c}$$

Finally, the values of $\theta$ corresponding with the envelope equation as demonstrated in Figure 3-6 are calculated with Equations 4-12 and 4-13.

$$\theta = 2\arctan(\Theta) \tag{4-12}$$

$$\Theta = \frac{-\dot{v}_{c_y} \pm \sqrt{|\dot{\mathbf{v}}_c|^2 - \dot{r}^2}}{-\dot{v}_{c_x} + \dot{r}} \tag{4-13}$$

Now the envelope equation at $v_z = Z$ is known, $Z$ can be varied to obtain the complete 3D figure. This is done by varying the time to collision, which is related to the value of $Z$ of one horizontal family of circles. The resulting figure can be found in Figure 4-5. It is demonstrated how the 3D FBZ is constructed by taking tangent lines towards the translated PZ cylinder.

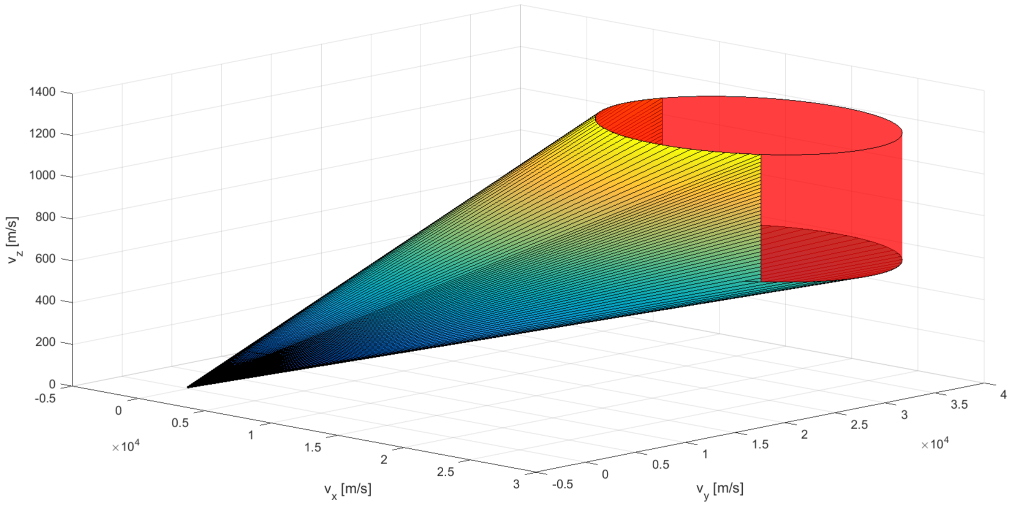



**Figure 4-5:** 3D Forbidden Beam Zone

## 3D performance envelope

The solution space approach uses the intersection of the flight performance envelope and the forbidden beam zone for workload calculations. Now the forbidden beam zone has been expressed as an easy to use parametric form, the performance envelope should also be transformed to 3D.

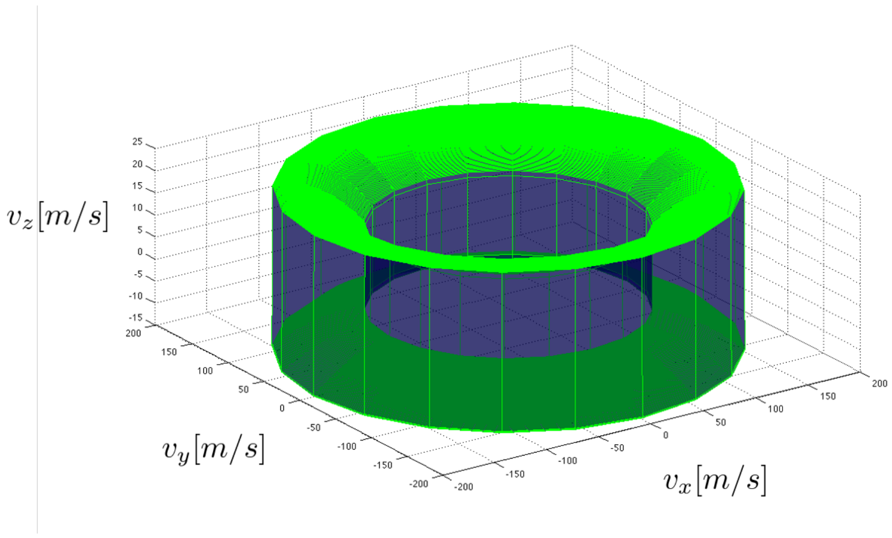
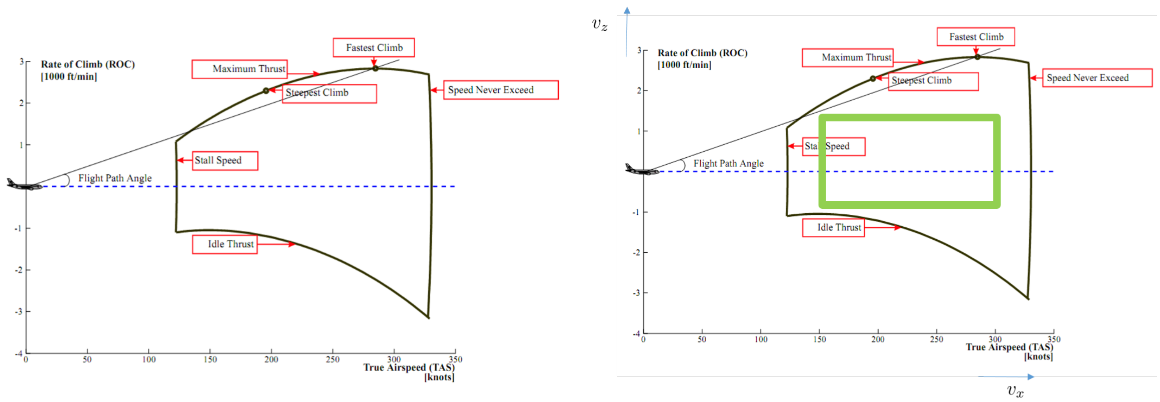



**Figure 4-6:** 3D performance envelope (Zhou, 2011)

Zhou (2011) had done this by rotating the performance envelope as illustrated in Figure 4-6 around 360 degrees. For both simplicity and to be more realistic, the approach here will be slightly different. Since air traffic controllers will not use the full performance envelope to give commands to aircraft, only a part of it is used. For example, an air traffic controller will not really consider 'idle thrust' as a reasonable command when avoiding conflicts and controlling the air space. Therefore, a rectangle is taken within this performance envelope using the minimum and maximum True Airspeed as indicated (in the simulator) for the particular aircraft. The maximum rate of descent and rate of climb is dependent on the altitude and therefore, the performance envelope will not only be different per aircraft, but also change with the flight level. Furthermore, if the controlled aircraft is already climbing or descending only the top or bottom part of the performance envelope will be taken, since it is unlikely the aircraft will instantly change from climb to descent and vice versa. See Figure 4-7 for an example of the simplified performance envelope.




**(a)** Example of an performance envelope

**(b)** Adaption of the performance envelope

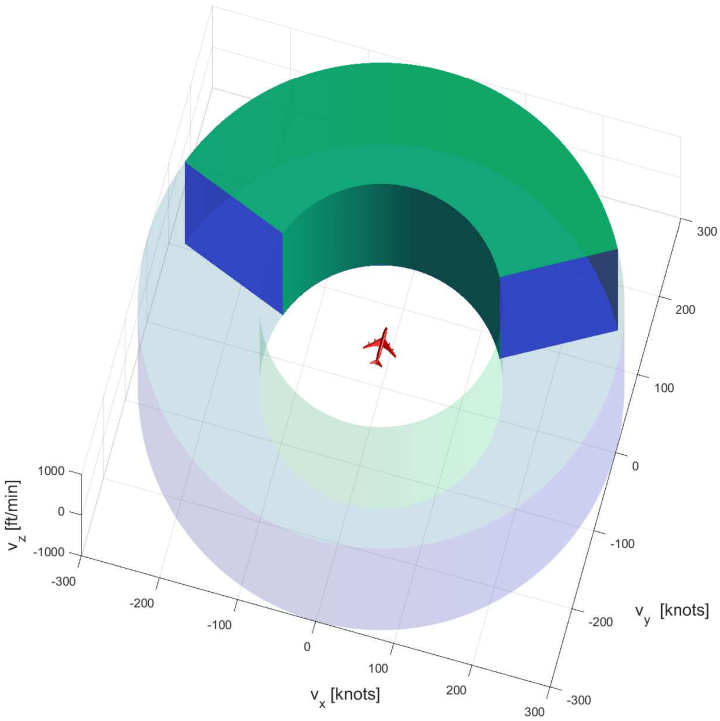
**Figure 4-7:** Performance envelope of the Cessna Citation I in a trimmed flight condition at 16405 feet altitude, adapted from (Heylen et al., 2008)

Furthermore, since it is unlikely an aircraft will be sent 180 degrees in the other direction, specific 'viewing angles', $\phi$, will be used. See for the resulting 3D flight envelope Figure 4-8a. The reference frame is taken such that the controlled aircraft is facing with its nose towards the positive $v_y$ axis, as illustrated in Figure 4-2. Hence, the viewing angles, which are related to the heading of the aircraft, are taken a certain angle $\pm\phi$ from the $v_y$ axis, displayed in Figure 4-8b. Finally, note that to be able to integrate this with the 3D FBZ, the aircraft performance envelope will have to be converted to SI units.

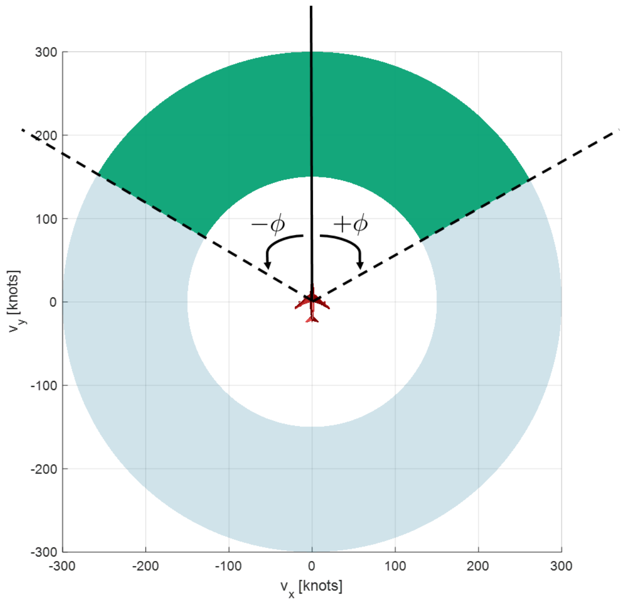
## Intersection FBZ and performance envelope

The 3D aircraft flight envelope is constructed by partly rotating a rectangle around the $v_z$-axis. Another approach to this is viewing the flight envelope as a 'book' consisting of a finite number of pages, see Figure 4-9. Each of these pages represents a vertical cutting plane, as previously described by Brantegem (2011), illustrated in Figure 4-10.

The volume, resulting from the intersection of the 3D FBZ and the rotated aircraft performance envelope, can be constructed by obtaining the vertical intersections of all 'book pages' (the vertical cutting planes with the 3D FBZ) and rotating these intersection areas around the $v_z$-axis over the step angle '$\beta$' taken between the vertical cutting planes, as illustrated in Figure 4-11.

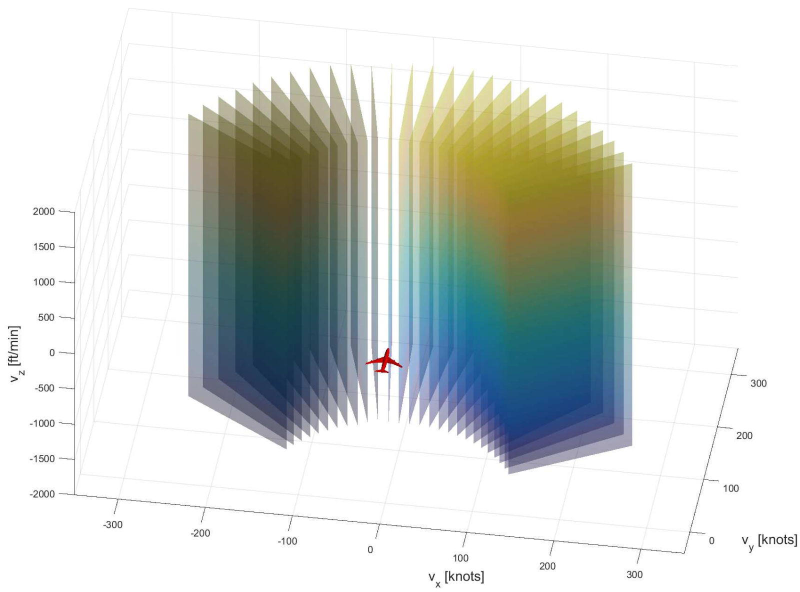



**(a)** 3D aircraft flight performance envelope

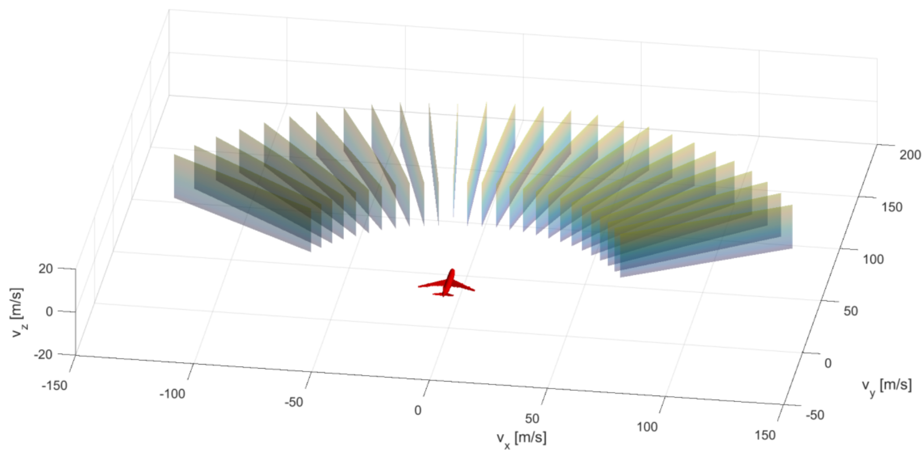



**(b)** Viewing angles $\phi$

**Figure 4-8:** Aircraft performance envelope using specific viewing angles




**(a)** Original




**(b)** Converted to SI units

**Figure 4-9:** Vertical cutting planes of the performance envelope forming a 'book'

## Vertical cross section

The first step, as illustrated in Figure 4-10, is to obtain the cross section of the 3D FBZ with each 'page'. This is determined by combining the vertical cutting plane equations with the 3D FBZ equations. The resulting shape consists of two arc equations, which are a function of $\theta$, and two vertical connecting lines. Note that the two arc equations can be traced back to the top and bottom curved surfaces of the 3D FBZ.

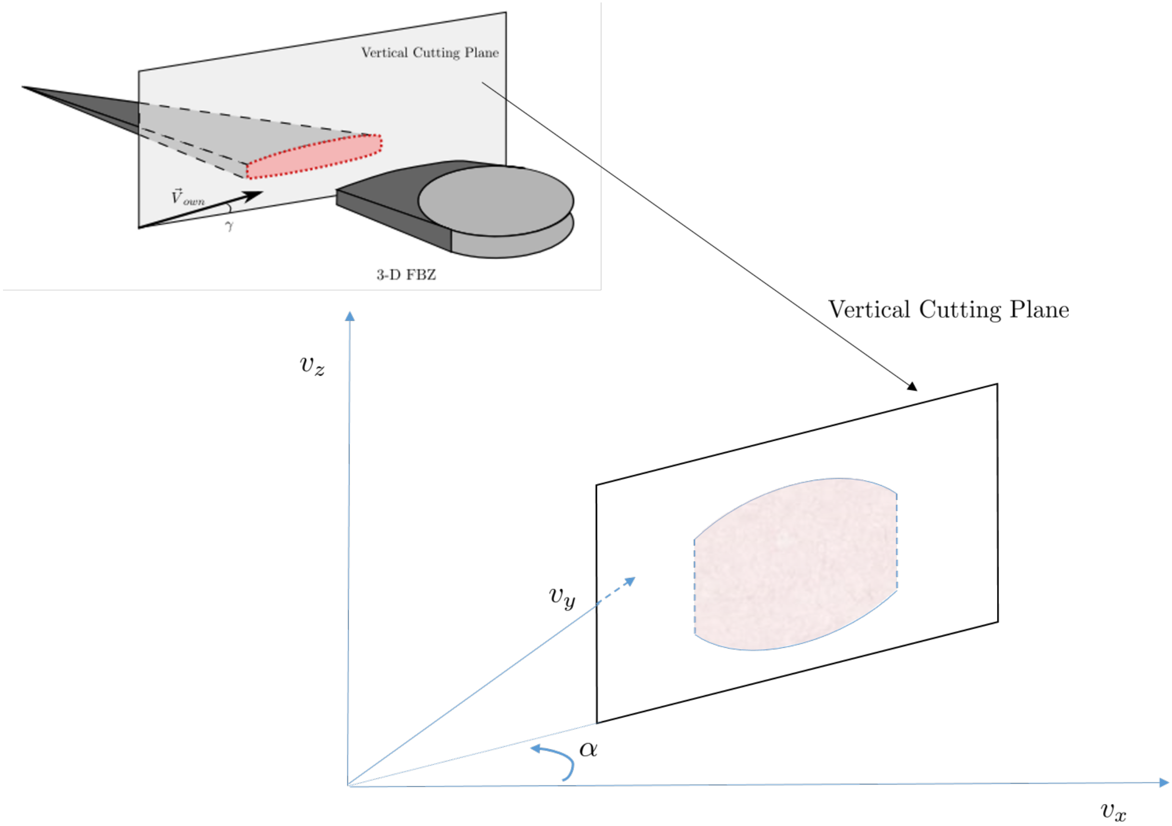



**Figure 4-10:** The vertical cutting plane and the resulting intersection

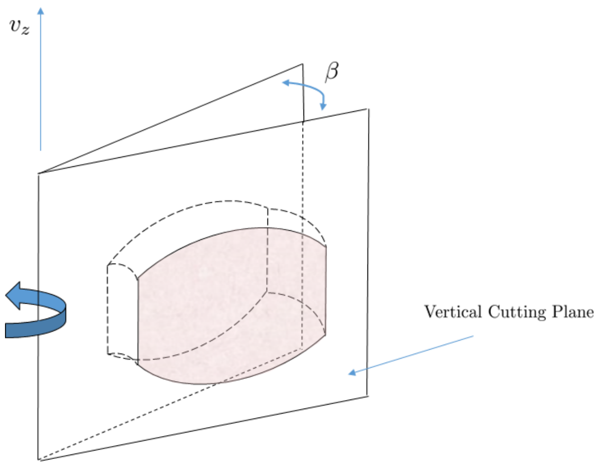



**Figure 4-11:** Rotating the vertical cross section over step angle $\beta$

The resulting vertical cross section, though a 2D shape, is still given in 3D coordinates. To express it in 2D coordinates, it can be rotated around the $v_z$-axis over angle $\alpha$, aligning it with the $v_x v_z$-plane, then all $v_y$ coordinates are 0 and only $v_x$ and $v_z$ remain. This is illustrated in Figure 4-12.

To determine the volume in between two 'pages', the vertical cross section is rotated over

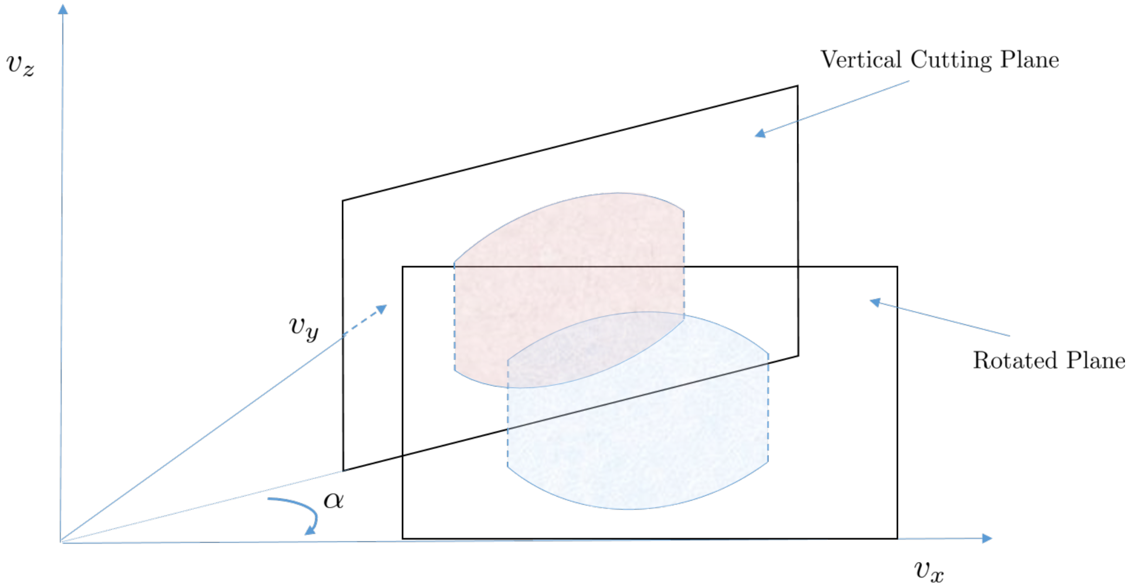



**Figure 4-12:** Vertical cross section and it's rotated image on the $v_y$-axis

the angle '$\beta$', as demonstrated in Figure 4-11. This can be analytically calculated using shell integration for parametric equations, see Equation 4-14 and Figure 4-13. Here it is assumed that $\theta_1 < \theta_2$ and $\theta_3 < \theta_4$. If this is not the case the limits will be opposite.

$$VOL = \underbrace{\beta \int_{\theta_1}^{\theta_2} v_x \cdot v_z \cdot \frac{\mathrm{d}v_x}{\mathrm{d}\theta}\ \mathrm{d}\theta}_{\text{upper}} - \underbrace{\beta \int_{\theta_3}^{\theta_4} v_x \cdot v_z \cdot \frac{\mathrm{d}v_x}{\mathrm{d}\theta}\ \mathrm{d}\theta}_{\text{lower}} \tag{4-14}$$

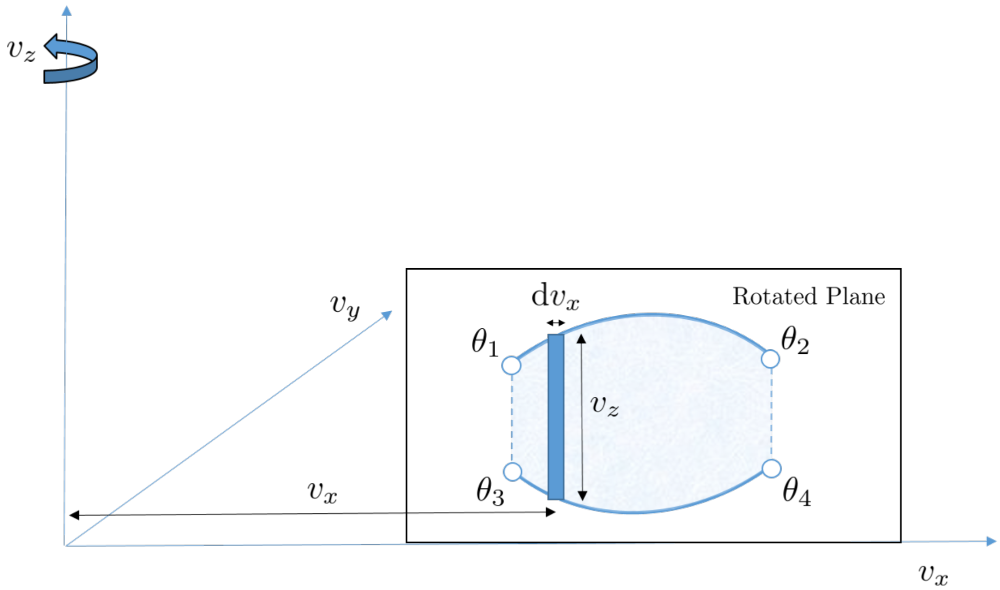
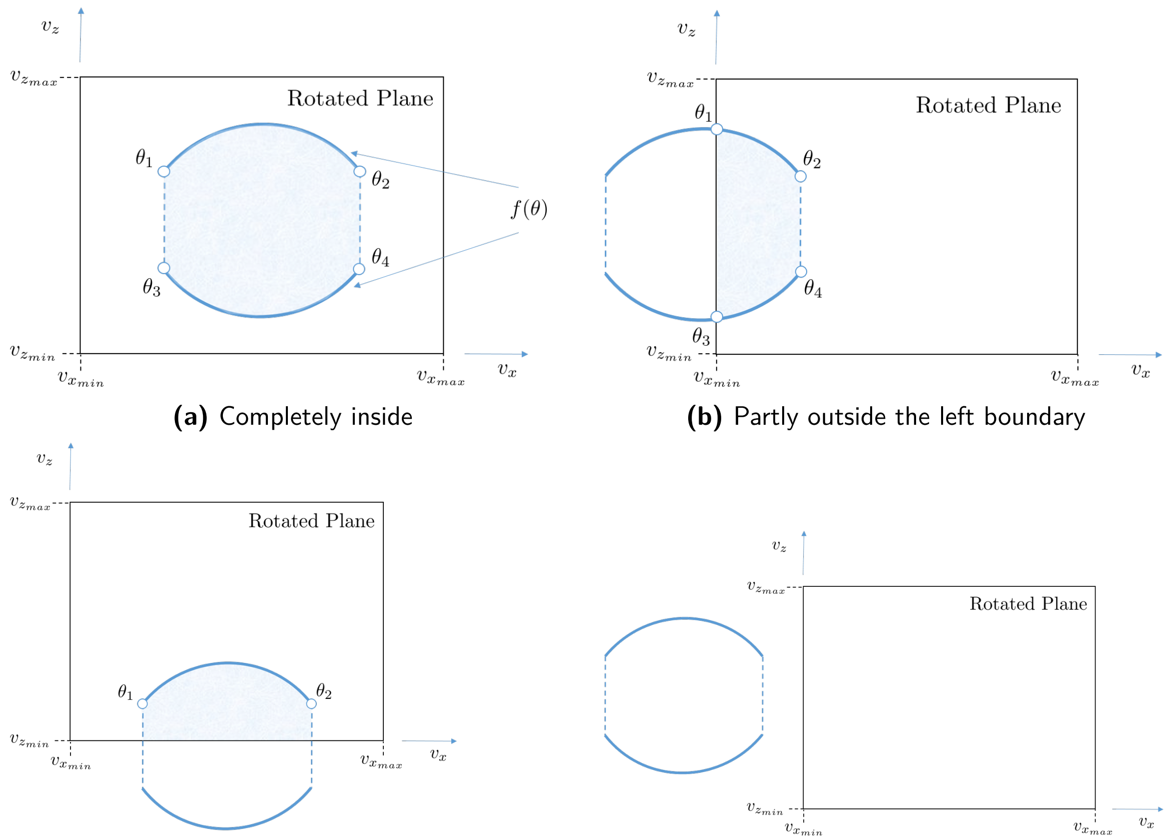



**Figure 4-13:** Determining the volume between pages with shell integration




**(a)** Completely inside

**(b)** Partly outside the left boundary

**(c)** Partly outside the lower boundary

**(d)** Completely outside the boundaries

**Figure 4-14:** The vertical cross section of the 3D FBZ inside and outside of the performance envelope boundaries

Figure 4-14 demonstrates how the limits are taken when the vertical cross section of the 3D FBZ is on different positions in comparison to the boundaries defined by the performance envelope of the aircraft. If the cross section is only partly within these boundaries, different limits or functions will have to be taken. If, as illustrated in Figure 4-14b, the cross section crosses the left or right boundary, the crossing points will be set as the new limits. Second, if as Figure 4-14c the cross section crosses the upper or lower boundary, this boundary line at height $v_z = v_{z_{min}}$ or $v_z = v_{z_{max}}$ will replace the corresponding upper or lower function in Equation 4-14. Finally, if the cross section completely falls outside of the 'page', Figure 4-14d, there is no volume to be calculated.

The above described method is for a single 'page'. To determine the complete volume of the intersection of the 3FBZ and aircraft performance envelope, these calculations have to be done for every page in the 'book'; the 3D aircraft performance envelope.

## Grid approach

The previously described analytical method works when there is interaction between only two aircraft. However, when multiple aircraft are considered, there may exist multiple vertical cross sections on each vertical cutting plane. When these vertical cross sections overlap, using Equation 4-14 on the individual aircraft would result in taking the overlapping area twice.

An option could be to determine the intersection points and sub tract the volume created by the intersection area by using revolution integrals again. However, while visualized it is straightforward where the overlapping area is, on a large scale with many possible overlapping areas and points, doing this analytically is not computation time effective.

Therefore, it is proposed to switch from a complete analytical approach to a combined grid approach, where a grid, with grid size ‘g’, is placed over the vertical cutting planes. The grid is depicted as a zero matrix where grid points that lie inside the intersection will be switched to a 1. This is visualized in Figure 4-15. For each grid point, it is determined whether it is inside or outside the cross section. The criterion for inclusion is illustrated in Figure 4-16.

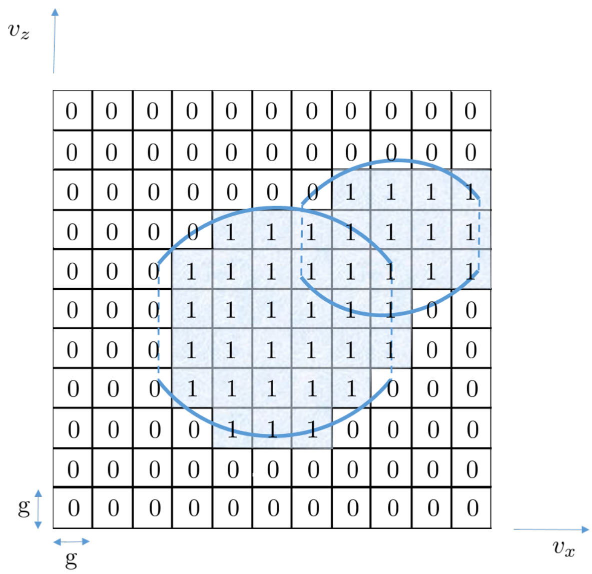



**Figure 4-15:** Using a grid with gridsize ‘g’ to approach the total solution space on a single vertical cutting plane

The equation for the rotational volume of a single grid point is derived in Appendix A and is defined by:

$$VOL = x_m \cdot \beta \cdot g^2, \tag{4-15}$$

where $x_m$ is the $v_x$ coordinate of the grid middle point. The total volume can be determined by using the 0-1 matrix $A$ as illustrated in Figure 4-15 and results in:

$$VOL = A \cdot x_m \cdot \beta \cdot g^2 \tag{4-16}$$

It should be noted that for time efficiency, first the boundary points of the 2 arc equations are determined, after which the closest inside grid $v_x$ coordinates are found. An additional

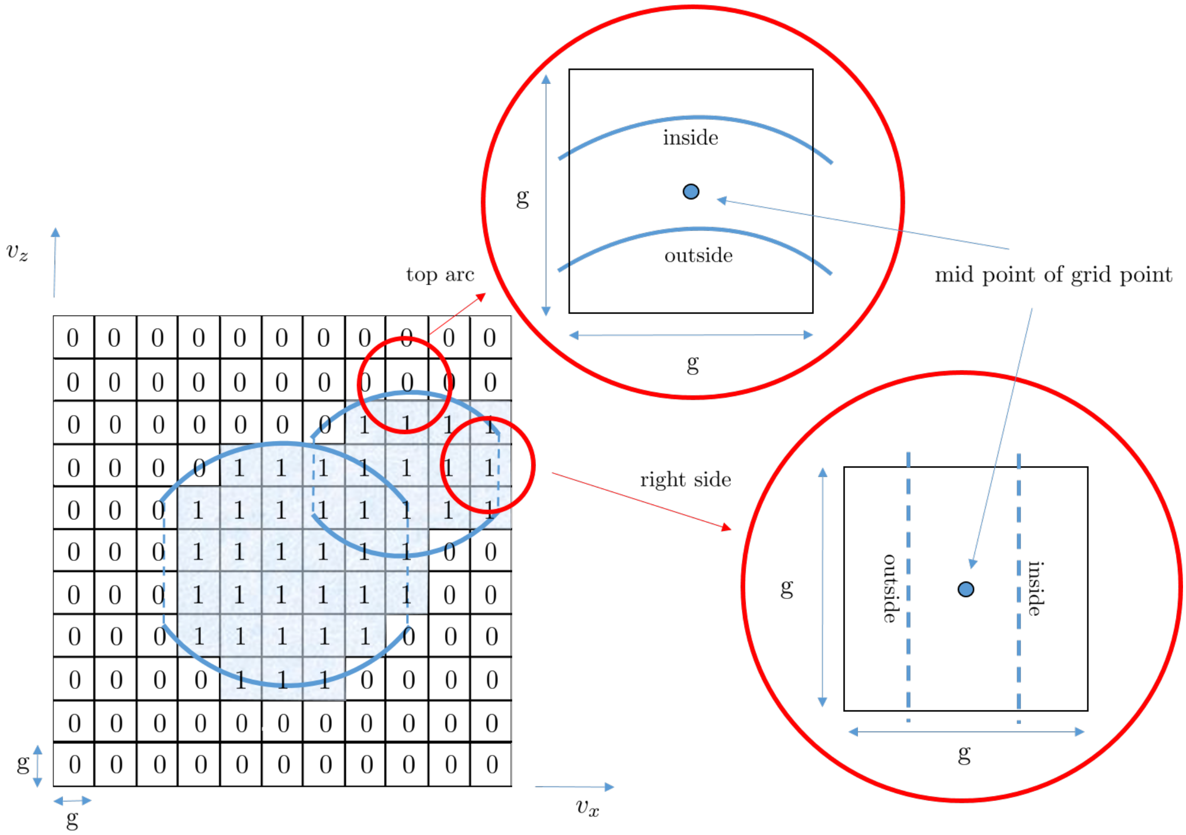



**Figure 4-16:** Determining if a grid point is in or outside the cross section

advantage of this, it can be checked beforehand if the $v_x$ coordinates lie completely, partly or not in the range of the performance envelope. These boundary points are the same as visualized as $\theta_1, \theta_2, \theta_3$ and $\theta_4$ in Figure 4-14.

Due to the parametric and geometric nature of the equations, going over all the grid $v_x$ coordinates in the determined boundaries to find the accompanying $v_z$ coordinates of the upper and lower arc, is not computation efficient. Therefore, an approach is taken to do the calculation the other way around.

First, for both the upper and lower arc many data points are generated in between the boundaries $(\theta_1, \theta_2, \theta_3, \theta_4)$ as displayed in Figure 4-17 to numerically approximate the arc equations. Next, for each grid point inside these boundaries is determined which generated point on the arc line has the closest $v_x$ coordinate, where is corrected for the curvature. For these $v_x$ coordinates, the accompanying $v_z$ coordinate is known as well. Note that this adds the requirement that sufficient points have to be generated to find satisfying values on the curve.

Afterwards, all the points between the determined $v_z$ boundaries will be switched to a 1. Finally, with Equation 4-16, the total volume can be calculated, which will be used for the workload percentage calculations.

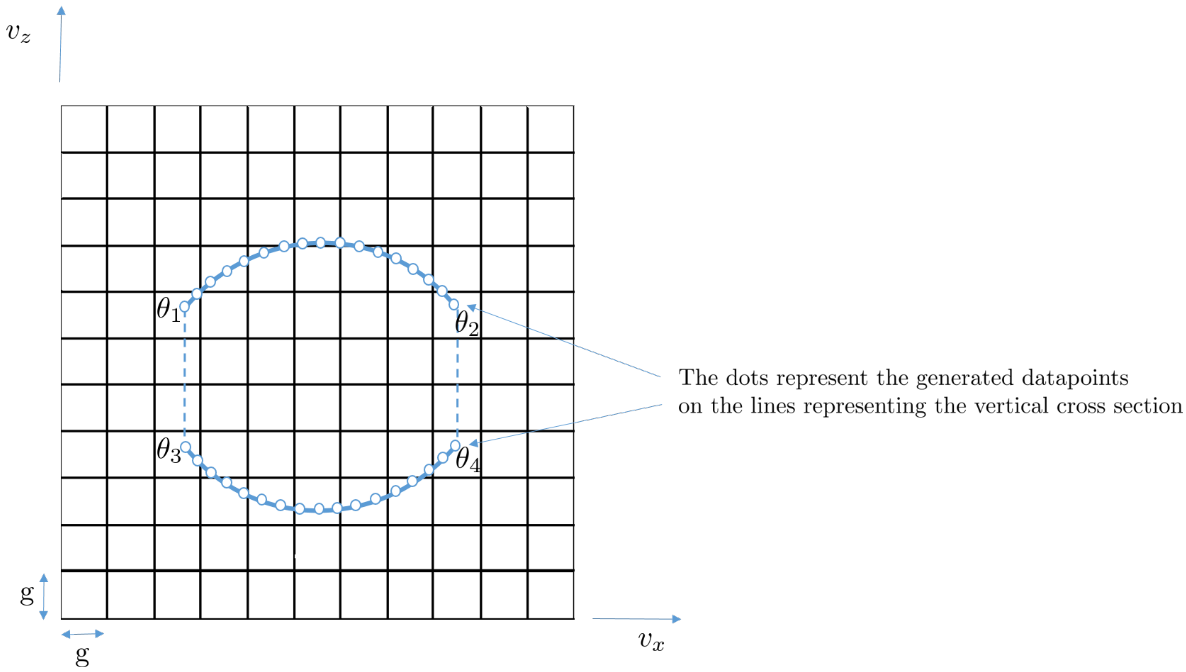



**Figure 4-17:** A grid with grid size 'g' is lain over the vertical cutting plane and vertical cross section

## Sensitivity Analysis

To verify the programmed metric works and to test its sensitivity, test were done where a 3D object with a known volume, a cone, was put between the viewing angles '$\phi$' to let the metric calculate the volume with different parameters, see Figure 4-18. These calculated volumes were then compared to the actual volume, to determine how sensitive the metric is to these parameters. As a reference should be noted that the radius and the height of the cylinder was taken as 3.21 [m/s]. Also for the performance envelope, $v_{x_{min}} = 0$ [m/s] and $v_{x_{max}} = 15$ [m/s], the same values were taken for the vertical speed $v_z$. The minimum and maximum speeds were taken such that the complete cone would fall within it and hence, are not based on any actual values.

The different parameters that will be tested are the grid size, step angle '$\beta$' and the number of points that are generated to numerically approximate the arc equations as explained before. When the approximated volume is calculated it is expressed as a percentage compared to the actual volume of the cone. In the final experiment also other noise will be in place, therefore, parameters that give approximated volumes up to 1 percent deviation of the actual volume are considered acceptable, hence, in the range of 99 % - 101% of the actual volume.

### Grid size

The first parameter that was tested, was the grid size. The page step angle was set to $\pi/180$ [rad] (or 1°) and the number of generated data points to 500000. By taking sufficient data

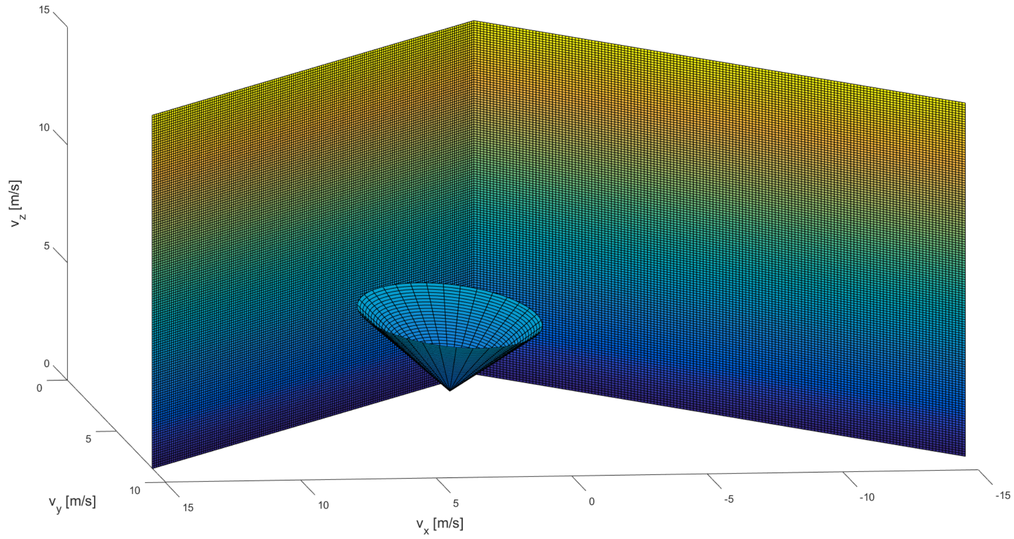



**Figure 4-18:** Cone within the viewing angles for sensitivity analysis

points to approach the arc lines, it is prevented that an inaccurate approach to the real lines, will give the wrong impression that a smaller grid size can not improve the metric.

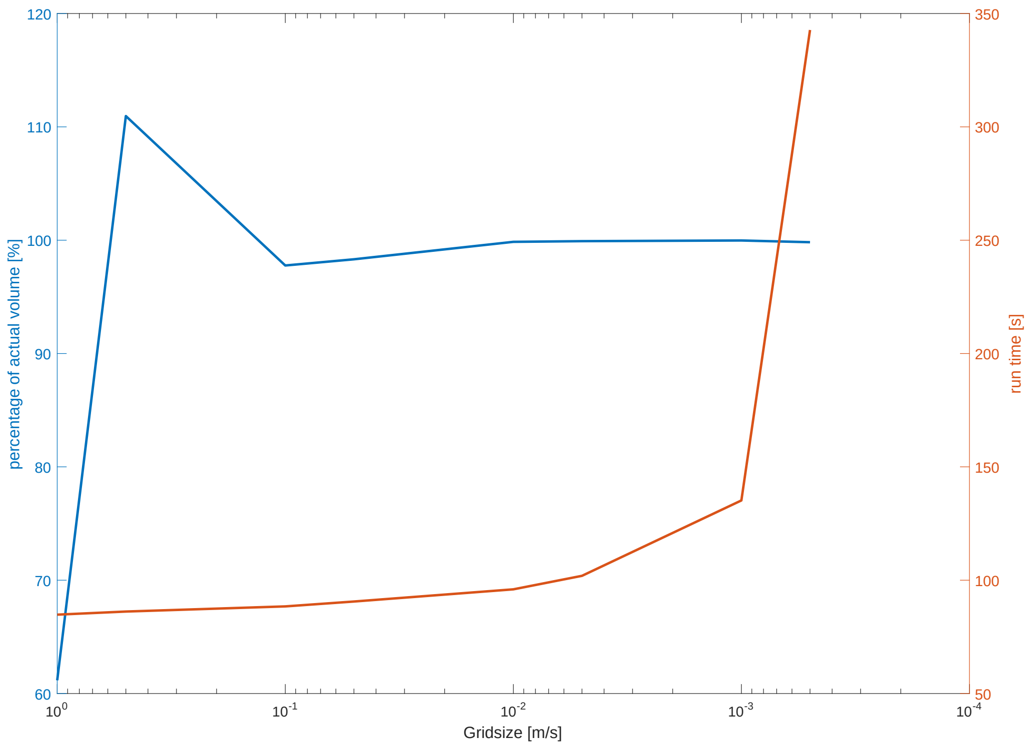



**Figure 4-19:** Sensitivity Analysis - Grid size (Step Angle = $\pi/180$, generated points = 500000)

In Figure 4-19 can be seen that from 0.1 the metric starts to get close to 100% of the actual volume, where from 0.01 the improvements in volume are minimal, but the running time starts to drastically increase. Therefore, in this case a grid size of 0.01 would be ideal. For the other sensitivity analysis a grid size of 0.05 is taken though, to see variations and influences of the other parameters better. However, this implies that for the other parameters never a 100 % accurate volume can be found, since the grid size will stay the limiting factor. Finally, for the actual metric, is worked on a different scale, scaling the grid size as well. However, the 3D forbidden beam zones will vary in size, the further away the aircraft, the smaller the forbidden beam zone, the smaller the grid size needed to get the most accurate volume. However, it should be noted, that if an aircraft is that far out, the smaller the influence on the actual solution space and the smaller the variations in volume matter.

## Step angle

The sensitivity analysis for the page step angle '$\beta$' is illustrated in Figure 4-20. The page step angle is defined as the angle between two pages of the performance envelope book. As a reference point should be noted that $\pi/180$ rad $\approx 0.0175$ rad $\approx 1$ °. The first interesting aspect from the graph is that the percentage stays around 98%, hence, the step angle '$\beta$' does not influence the approximated volume much as long as it is not taken too low. However, the running time increases to high values with a smaller $\beta$ starting from $\pi/360$ rad $\approx 0.0087$ rad. Furthermore, the estimated volume does not keep on increasing with a smaller page step angle. Therefore, it can be concluded that for the final metric also a page step angle of $\pi/180$ rad is sufficient.

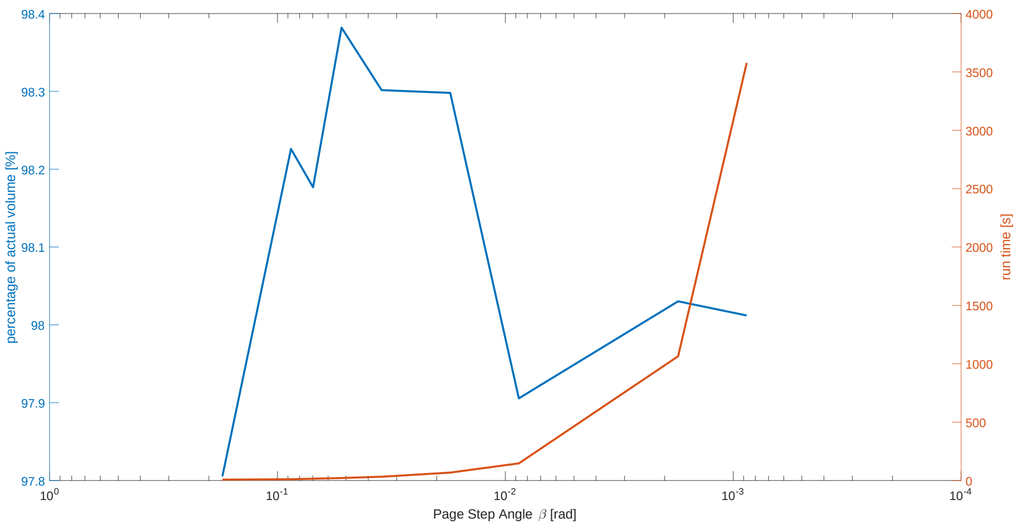



**Figure 4-20:** Sensitivity Analysis - Step Angle $\beta$ (Grid size=0.05, generated points = 10000)

### Number of generated points

As discussed in the previous section, to improve the computation time, first the arc lines are numerically approached by generating many data points in between the boundary points of the vertical cross section, afterwards grid points are matched to them. Hence, the number of generated points is highly correlated with the grid size. The lower the grid size, the more data points have to be generated. As a reference here, a grid size of 0.05 is taken. As can be seen in Figure 4-21, from 10000 points the maximum reachable estimation of the volume with a grid size of 0.05 is reached. More data points will not be able to get a better estimation with the same grid size, but will only increase the running time.

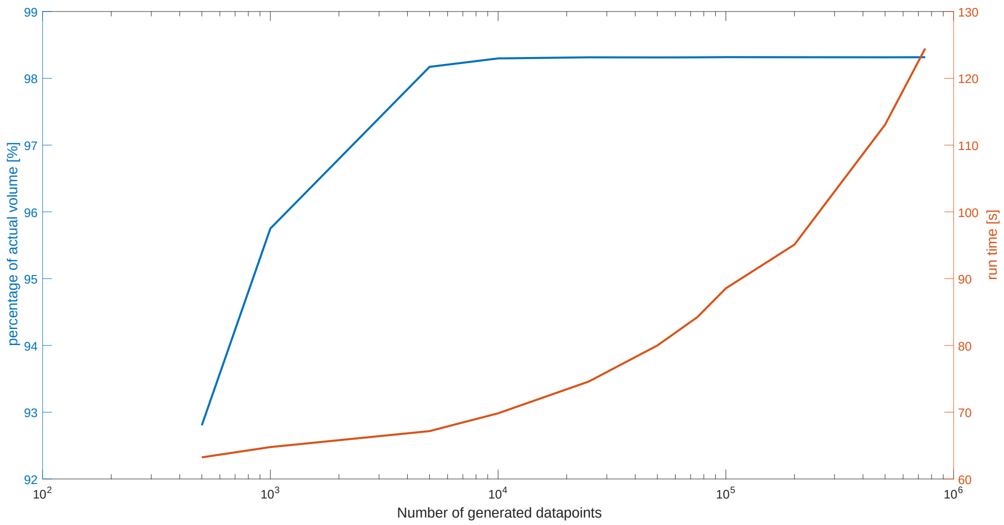



**Figure 4-21:** Sensitivity Analysis - Generated points (Step Angle = $\pi/180$, grid size=0.05)

### Averaging the volumes

To account for the fact that the volume is obtained by rotating vertical cross sections around step angle '$\beta$', volume or page areas could be averaged over multiple pages to get a better approximation. However, it was found that this only gives a very small improvement compared to the current calculated volumes. Which is related to the fact that a very small step angle also only gives a small improvement.

### Conclusions

It can be concluded that the most important parameter that influences the metric is the grid size. Therefore, the grid size should be selected carefully. What also should be taken into account is, the smaller the grid size taken, the more data points have to be generated to get a good estimation. In addition, the step angle '$\beta$', can be taken quite roughly, where $\pi/180$ rad already is sufficient for good estimations. More over, averaging the volumes from different pages to get a more smooth estimation does not give a big improvement, which is related to the fact that a smaller step angle also does not give large improvements, most likely due to the shape of the cone and the forbidden beam zones. This is, therefore, unnecessary. Finally, taking all this into account, close to 1% of the actual volume can be reached, which was set as a requirement for acceptable values.

## Assumptions

For simplification certain assumptions were done while setting up the design of the metric.

- No wind velocity or weather conditions are taken into account
- Changes in velocity can be made instantaneously
- Aircraft have to respect the protected zone
- Current and future intended state of the aircraft are known
- Air traffic controllers don't use the full performance envelope to look for conflict solutions

The first assumption is made for simplification to make this initial metric design less complex. The second assumption is also done for simplification, but can also be validated since the time it takes for the pilot to respond and make a change in velocity compared to other times and events involved in the whole air traffic control communication is negligible. The third and fourth assumption have to do with the way the 3D forbidden beam zone is constructed, since it uses the future path to predict future conflict. Here, the fact is used that intruding the protected zone also implies a loss of separation. The final assumption was done as both a simplification and to try to get closer to reality for better metric performance.

## Preliminary Conclusions

This research started out to fill in the gap of the need for an objective, reliable ATC workload metric. Which would be done by analytically evolving the solution space to the third dimension. Technically a complete analytical approach has been given. However, when encountering the problem of overlap areas, caused by the presence of multiple aircraft, a completely analytical approach is less time efficient than a hybrid approach. Therefore, the latter was chosen. Especially if the metric will have to be applied in real-time, a completely analytical approach seems not feasible. The main slow down occurs when using complicated integrals. As a side note, it should be mentioned that programs will never be able to completely analytically determine the outcome, since they have to use a numerical approach to determine integrals in the first place, due to their inability for analytical determinations.

Therefore, for the calculation of the 3D solution space volumes, as part of the 3D solution space metric, a hybrid approach is suggested where the initial 3D shapes and 2D vertical cross sections are analytically determined, afterwards by means of a grid approach is determined which area is inside or outside the solution space. Which again is rotated analytically to determine the final volume. Due to the numerical side of this approach, the final metric will have to be tuned for reasonable values for grid size, amount of spacing between pages, number of generated points versus computation time as was demonstrated in the sensitivity analysis. Here, it was found that the grid size has the biggest influence on the accuracy of the volume estimation. Furthermore, the grid size and number of generated data points for time efficiency are correlated. Hence, the smaller the grid size, the more data points should be generated to keep reasonable volume estimations. Finally, a step angle '$\beta$' of $\frac{\pi}{180}$ rad was found to give good approximations within a reasonable computation time.

By the means of the verification and sensitivity analysis, it was also shown that the proposed metric is able to approach the actual volume very closely. Especially since in the final experiment due to other noise getting as close to the actual volume within a 1 percentage range will be accurately enough. The next steps will be actual validation of the metric by using an ATC simulator and comparing it to actual workload.

# Chapter 5

# Research Proposal

To test the performance of the designed metric a human-in-the-loop experiment will be done. This chapter will explain the further research steps and the experimental design.

The main underlying research goal is to analyze the correlation of the metric with workload to determine if the 3D solution space metric is a good predictor for workload. Additionally, it will be compared with a 2D layered SSD to determine if it is a better predictor than already available metrics. This will be done with a human-in-the-loop experiment where participants will have to behave as real life air traffic controllers, by controlling traffic in a certain sector.

## Workload estimation

During the experiment the workload will have to be determined. This will, first of all, be done by using the metric. However, to evaluate these calculated workload values with the 3D Solution Space Metric it will have to be tested against the actual workload as experienced by the participating air traffic controllers. This will be measured by using the Instantaneous Self Assessment (ISA). Though there are many techniques to measure perceived workload during a human-in-the-loop experiment, ISA is one of the simplest tools to quickly get acceptable workload ratings. (Abdul Rahman et al., 2011) The ISA method is developed by the UK Civil Aviation Authority to get immediate subjective workload ratings during air traffic control tasks. This is done by requiring the controller to give feedback on his perceived workload by rating it on a scale from 1 to 5. (Tattersall & Foord, 1996; Farmer, Brownson, & Eurocontrol, 2003) However, to get more specific feedback, other scales, such as from 1 to 10 are also used.

The advantages of ISA are, first of all, as mentioned before, that it is very easy and low-key to use. No specific equipment is required. Even more, it can gather time-based workload ratings, which suits a dynamic experiment with changing task demands. Therefore, it is more accurate than post-task methods. Furthermore, it is very practical and has shown to be consistent with other subjective workload measures. (Tattersall & Foord, 1996; Farmer et al., 2003)

A disadvantage is that it can interfere with the primary task performance. However, according to Farmer (2003), it still has the lowest intrusiveness compared to other dynamic metrics. The main drawback and risk, however, is that the ISA workload method is very subjective and therefore, only reliable if the participants rate their perceived workload accurately.

Therefore, it is important that participants are trained with low and high workload scenarios, to calibrate their workload ratings beforehand. For the ISA workload bar a scale from 1 to 10 will be used to get more precise feedback. Furthermore, has to be decided what the time interval of the measurements is going to be. If it is measured too often, it will interfere with the primary task. However, if the intervals are too big, the trend of the metric can be less evaluated. Finally, the ISA workload rating bar has to pop up clear enough that is gets noticed in time. This can be done by adding sound for example. Though if it gets too annoying, it might accidentally also increase the stress of the participant and start acting as a confounding factor.

## Participants

As mentioned, one of the disadvantages of using the ISA workload is, it is very dependent on how accurately the participants rate their workload. Therefore, because the experiment stands or falls with the workload ratings the participants give, the selection of participants is very important. The ideal situation would be to have experienced air traffic controllers, because they have the most experience with air traffic control and can give a good indication if a certain scenario has a high or low workload. However, it is quite likely that it will not be possible to find enough experienced air traffic controllers to participate in the experiment. Using novices instead, though widely available, is unfortunately not a realistic option. There is too much risk of inaccurate self-assessed workload ratings. Especially since they might experience even more effect of the ISA ratings interfering with the primary task of controlling traffic, which increases their workload as a confounding variable. For example the ISA workload bar that will pop up, might be too easily clicked aside by novices, due to having to spend more energy in grasping the air traffic situation, resulting in unreliable workload ratings.

Therefore, another option might be to use 'medium experienced' participants. This implies participants who have followed an extensive ATC course and have experience with ATC projects and experiments. Especially since realistic workload ratings do not only come with experience in ATC scenarios. Experience with using the ISA scale can also add to the accuracy of the reported workload ratings. Medium experience controllers do have an idea about how Air Traffic Control works, are more familiar with it and are preferably also more familiar with the ISA workload bar.

Finally, also a mix of experienced air traffic controllers and medium experienced participants can be taken to test the sensitivity of the metric to the expertise level of the controller.

## Scenario design

The scenario design will be the main factor in the search for the behavior and strengths/weaknesses of the metric. It can determine if high or low workload will occur and with which rate the workload changes during the scenario. Hence, the more can be seen how the 3D solution space metric correlates with the ISA ratings.

A scenario will consist of one particular sector for which the controller will have to control the air traffic present. This can be either resolve conflicts, merge traffic on a certain route or guide aircraft such that they exit the sector at a certain way point and flight level. Future incoming air traffic will already be visible and therefore, taken into account for solution space calculations; however, they will only be controllable when they enter the sector.

Because the metric is designed specifically to take into account the third dimension, the scenarios will have to use this third dimension. Furthermore, including different types of aircraft can provide some interesting feedback on the metric, because different aircraft have different performance envelopes, resulting in different 3D solution spaces. It can be interesting to see how the metric behaves in those cases.

To increase the reality of the experiment, ideas are for example that every aircraft entering the sector will have to be instructed/welcomed into the air sector, even when they don't need adjustments. This is done in real life as well and additionally takes into account the part of ATC workload added by aircraft that are not in conflict, but still are in contact with Air Traffic Control and give more load to the air sector. Moreover it can be included that the participant will have to give actual commands to the aircraft instead of just resolving conflicts as was done in earlier experiments.

Since not every possible scenario and sector can be tested, a more specific choice for first feedback on the metric will have to be done. For the sector the ACC AMS South Sector can be used as a reference. Aircraft come into this sector and have to be merged and leave at a certain flight level. While other aircraft are flying over at different flight levels, to add load and reality to the sector.

## Validation

To validate if the metric does indeed provide objective and reliable workload predictions, besides comparing it to the ISA ratings, it will be compared to the 2D layered SSD. This method uses the 2D SSD projected at different flight levels. As illustrated in Figure 5-1, it looks at the flight levels that will be crossed and takes the corresponding 2D solution spaces and projects them on one plane. This can also be used to see if the trouble that was done by taking into account the complete 3D situation is worth it.

The outcome can give insights into how air traffic controllers think. If they see the air space as layered and control it in that way, a layered SSD might predict the workload more accurately. Or if their is no significant difference, the ease of the layered SSD might be preferred.

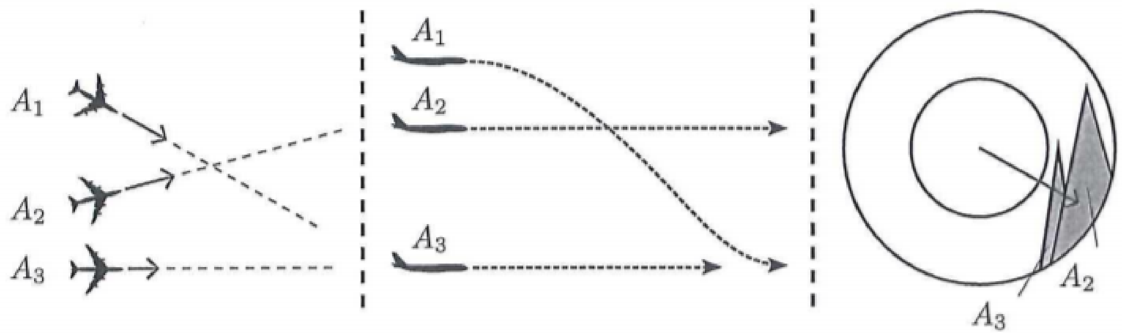



**Figure 5-1:** 2DSSD layered approach (Lodder, 2011)

Finally, also the difference in viewing angles '$\phi$' can be tested to see if it can add to the reality of the solution space metric.

## Hypotheses and Expectations

The main hypothesis and underlying hypotheses for the experiment are as follows:

- If the 3D solution space metric is inversely related with ATC workload and it is an better predictor for ATC workload than existing metrics, then:
  - The 3D solution space metric will show significant correlation with subjective ATC workload indications
  - The 3D solution space metric will significantly show better correlation with the subjective workload ratings than the 2D layered SSD

The keeping or rejecting of these hypotheses in the final conclusion of the project will indicate if the designed 3D metric has indeed potential to be a good objective Air Traffic Control workload metric.

Finally, is expected that their might be a time shift between the workload calculated by the metric and the indicated ISA workload ratings. Participants might also take into account situations and workload from the past when rating their current status. Or the opposite way around, upcoming complexity might already influence current ratings. If this occurs, which can be quickly noticed by visualizing the results, it has to be decided how this will be taken into account.

# Bibliography


Abdul Rahman, S. M. B. (2014). *Solution Space-based Approach to Assess Sector Complexity in Air Traffic Control.* Unpublished doctoral dissertation, Delft University of Technology.

Abdul Rahman, S. M. B., Borst, C., Mulder, M., & Paassen, M. M. v. (2010). Measuring Sector Complexity : Solution Space-Based Method. *Advances in Air Navigation Services*, 11–34.

Abdul Rahman, S. M. B., Paassen, M. M. van, & Mulder, M. (2011). Using the solution space diagram in measuring the effect of sector complexity during merging scenarios. (August), 1–25.

Abdul Rahman, S. M. B., Borst, C., Mulder, M., & Paassen, M. M. van. (2015). Sector complexity measures: A comparison. *Jurnal Teknologi*, *76*(11), 131–139.

Athènes, S., Averty, P., Puechmorel, S., Delahaye, D., & Collet, C. (2002). ATC Complexity and Controller Workload: Trying to Bridge the Gap. *AAAI*.

Bilimoria, K. D. (2000). A Geometric Optimization Approach to Aircraft Conflict Resolution. In *Aiaa guidance, navigation, and control conference and exhibit.*

Boag, C., Neal, A., Loft, S., & Halford, G. S. (2006). An analysis of relational complexity in an air traffic control conflict detection task. *Ergonomics*, *49*(January 2014), 1508–1526.

Brantegem, K. (2011). *Ecological 2-D Coplanar Airborne Separation Assurance System.* Unpublished master's thesis, Delft University of Technology.

Carbone, C., Ciniglio, U., Corraro, F., & Luongo, S. (2006). A Novel 3D Geometric Algorithm for Aircraft Autonomous Collision Avoidance. In *Proceedings of the 45th ieee conference on decision and control* (pp. 1580–1585).

Chakravarthy, A., & Ghose, D. (1998). Obstacle avoidance in a dynamic environment: A collision cone approach. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, *28*(5), 562–574.

Chakravarthy, A., & Ghose, D. (2011). Collision cones for quadric surfaces. *IEEE Transactions on Robotics*, *27*(6), 1159–1166.

Chakravarthy, A., & Ghose, D. (2012). Generalization of the collision cone approach for motion safety in 3-D environments. *Autonomous Robots*, *32*(3), 243–266.

Crutchfield, J. M. (2005). Predicting subjective workload ratings: A comparison and synthesis of theoretical models. *ProQuest Dissertations and Theses*, *3178305*(March), 65–65 p.

Damas, B., & Santos-Victor, J. (2009). Avoiding moving obstacles: The forbidden velocity map. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009*, 4393–4398.

d'Engelbronner, J. (2010). The Use of the Dynamic Solution Space to Assess Air Traffic Controller Workload. In *Aiaa guidance, naviagtion and control conference.*

Djokic, J., Lorenz, B., & Fricke, H. (2010). Air traffic control complexity as workload driver. *Transportation Research Part C: Emerging Technologies*, *18*(6), 930–936.

Ellerbroek, J., Brantegem, K. C. R., van Paassen, M. M., & Mulder, M. (2013). Design of a coplanar airborne separation display. *IEEE Transactions on Human-Machine Systems*, *43*(3), 277–289.

Ellerbroek, J., Visser, M., van Dam, S. B. J., Mulder, M., & van Paassen, M. M. (2011). Design of an airborne three-dimensional separation assistance display. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, *41*(5), 863–875.

Farmer, E., Brownson, A., & Eurocontrol. (2003). Review of Workload Measurement , Analysis and Interpretation Methods. *Methods*, *47*(25), 33.

Fiorini, P., & Shiller, Z. (1993). Motion planning in dynamic environments using the relative velocity paradigm. *1993 IEEE International Conference on Robotics and Automation*, *1*, 560–566.

Fiorini, P., & Shiller, Z. (1998). Motion Planning in Dynamic Environments using Velocity Obstacles. *I. J. Robotic Res*, *17*, 760–772.

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

Goss, J., Rajvanshi, R., & Subbarao, K. (2004). Aircraft Conflict Detection and Resolution using Mixed Geometric and Collision Cone Approaches. In *Aiaa guidance, navigation, and control conference* (pp. 1–20).

Hermes, P., Mulder, M., van Paassen, M. M., L. Boering, J. H., & Huisman, H. (2009). Solution-Space-Based Complexity Analysis of the Difficulty of Aircraft Merging Tasks. *Journal of Aircraft*, *46*(6), 1995–2015.

Heylen, F., Dam, S. van, Mulder, M., & Paassen, M. van. (2008). Design and Evaluation of a Co-planar Separation Assistance Display. In *Aiaa guidance, navigation and control conference and exhibit* (pp. 1–6).

Hilburn, B. (2004). Cognitive Complexity in Air Traffic Control a Literature Review. *EEC note*(04).

Hilburn, B., & Jorna, P. G. A. M. (2001). Workload and air traffic control. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload and fatigue.* Lawrence Erlbaum Associates.

Jenie, Y. I., van Kampen, E.-J., Visser, C. C. de, & Chu, Q. P. (2013). Selective Velocity Obstacle Method for Cooperative Autonomous Collision Avoidance System for Unmanned Aerial Vehicles. In *Aiaa guidance, navigation, and control (gnc) conference* (pp. 1–20).

Jenie, Y. I., van Kampen, E.-J., Visser, C. C. de, Ellerbroek, J., & Hoekstra, J. M. (2016). Three-dimensional velocity obstacle method for UAV uncoordinated avoidance maneuvers. In *Aiaa guidance, navigation, and control conference* (pp. 1–16).

Kopardekar, P., & Magyarits, S. (2002, October). Dynamic density: Measuring and predicting sector complexity. In (Vol. 1, p. 2C4-1 - 2C4-9).

Lodder, J. (2011). *Altitude-extendend solution space diagram for air traffic controllers.* Unpublished master's thesis, Delft University of Technology.

Loft, S., Sanderson, P. M., Neal, A., & Mooij, M. (2007). Modeling and Predicting Mental Workload in En Route Air Traffic Control : Critical Review and Broader Implications. *Human Factors*, *49*(3), 376–399.

Luongo, S., Carbone, C., Corraro, F., & Ciniglio, U. (2009). An Optimal 3D Analytical Solution for Collision Avoidance Between Aircraft. In *2009 ieee aerospace conference* (pp. 1–9).

Luongo, S., Corraro, F., Ciniglio, U., Di Vito, V., & Moccia, A. (2010). A novel 3D analytical algorithm for autonomous collision avoidance considering cylindrical safety bubble. In *Ieee aerospace conference proceedings.*

Mercado Velasco, G. A., Borst, C., Ellerbroek, J., van Paassen, M. M., & Mulder, M. (2015). The Use of Intent Information in Conflict Detection and Resolution Models Based on Dynamic Velocity Obstacles. *IEEE Transactions on Intelligent Transportation Systems*, *16*(4), 2297–2302.

Mercado Velasco, G. A., Mulder, M., & van Paassen, M. M. (2010). Analysis of Air Traffic Controller Workload Reduction Based on the Solution Space for the Merging Task. In *Aiaa guidance, navigation, and control conference* (pp. 1–18).

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* (Tech. Rep. No. July).

RTCA. (2002, March). *Airborne conflict management: Application description v2.5* (Tech. Rep. No. RTCA SC- 186). Federal Aviation Authorities.

Shiller, Z., Large, F., Sekhavat, S., & Laugier, C. (2007). STAR 35 - Motion Planning in Dynamic Environments. *Autonomous Navi. in Dyn. Environ. STAR*, *35*, 107–119.

Snape, J., Berg, J. v. d., Guy, S. J., & Manocha, D. (2011). The hybrid reciprocal velocity obstacle. *IEEE Transactions on Robotics*, *27*(4), 696–706.

Tattersall, A. J., & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, *39*(5), 740–748.

Tychonievich, L., Zaret, D., Mantegna, J., Evans, R., Muehle, E., & Martin, S. (1989). A Maneuvering-Board Approach to Path Planning with Moving Obstacles. In *Proceedings of the 11th international joint conference on artificial intelligence-volume 2* (pp. 1017–1021).

van Dam, S. B. J., Mulder, M., & Paassen, M. M. van. (2008). Ecological interface design of a tactical airborne separation assistance tool. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, *38*(6), 1221–1233.

Zhou, W. (2011). *The 3d solution space.* Unpublished master's thesis, Delft University of Technology.

# Part III

# Appendices

# Part III-A

# Preliminary Report

# A Preliminary Report - Derivation volume calculation

The simplified equation to determine the volume of a grid point rotated around step angle $\beta$, can be found as follows.

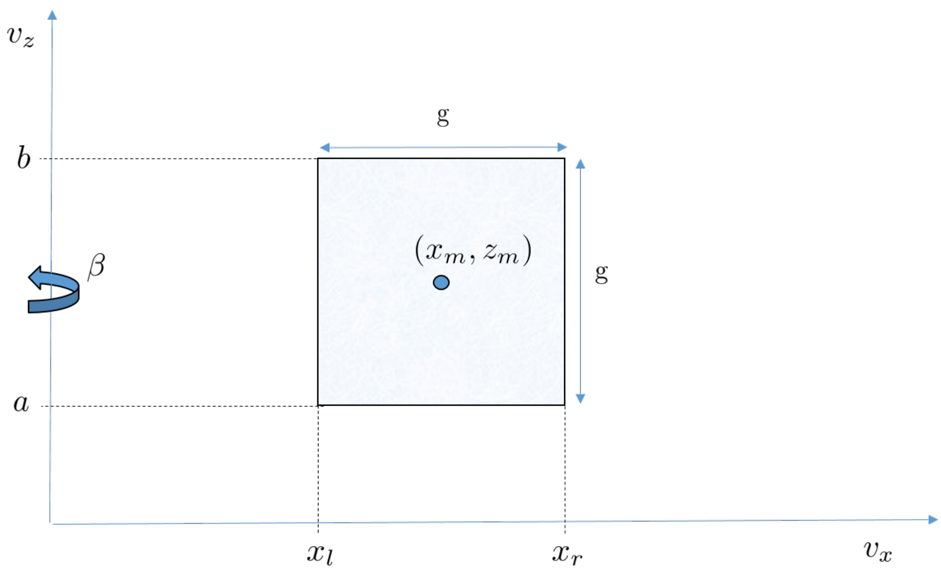



**Figure 2:** A single grid point with dimensions to determine the rotational volume

Equation 1 defines how the volume is calculated of the rotational figure that is obtained by rotating an area around the $v_z$-axis over step angle $\beta$. Since the goal is to determine the rotational volume of a single grid point, the area of the grid point as illustrated in Figure 2 is taken as input and rotated as illustrated in Figure 3. Where $x_m$ is the $v_x$ coordinate of the grid point, which is defined as the middle point of the single grid square. $x_l$ and $x_r$ are defined as respectively the left and right $v_x$ coordinate of the grid point.

$$VOL = \frac{\beta}{2\pi} \int_a^b AREA(v_z) \, \mathrm{d}v_z \tag{1}$$

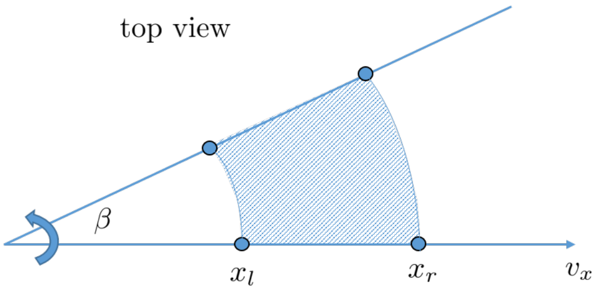



**Figure 3:** Rotating a single grid point over angle $\beta$ - top view

- $x_l = x_m - \frac{1}{2}g$
- $x_r = x_m + \frac{1}{2}g$

The upper and lower limits of the integral are defined as follows, where $z_m$ is the $v_z$ coordinate of the grid point, see Figure 2.

$$\left.\begin{array}{l} \bullet\ a = z_m - \frac{1}{2}g \\ \bullet\ b = z_m + \frac{1}{2}g \end{array}\right\} b - a = g$$

Using this, Equation 1 can be written out:

$$\begin{aligned} VOL &= \frac{\beta}{2\pi}\int_a^b (\pi x_r^2 - \pi x_l^2)\ \mathrm{d}v_z \\ &= \frac{\beta}{2}(b-a)(x_r^2 - x_l^2) \\ &= \frac{\beta}{2}\cdot g\cdot \underbrace{((x_m + \frac{1}{2}g)^2 - (x_m - \frac{1}{2}g)^2)}_{2x_m g} \\ &= \beta\cdot g^2\cdot x_m \end{aligned}$$

Hence, the final equation is defined as Equation 2. Note that the larger $x_m$, that is the further away the grid point from the rotation axis $v_z$, the bigger the rotation volume will be.

$$\therefore \boxed{VOL = \beta\cdot g^2\cdot x_m} \tag{2}$$

# Part III-B

# Experiment Set-up

# B Experiment Set-up pictures

The experiment set-up consisted of a computer with a main screen and additional touchscreen on which the command window was displayed. Also a headset was used to audit the sounds accompanying the ISA workload bar and possible warnings and conflicts. The complete experiment set-up is visible in Figure 4.

It should be noted that the person visible in these pictures did not participate in the experiment.

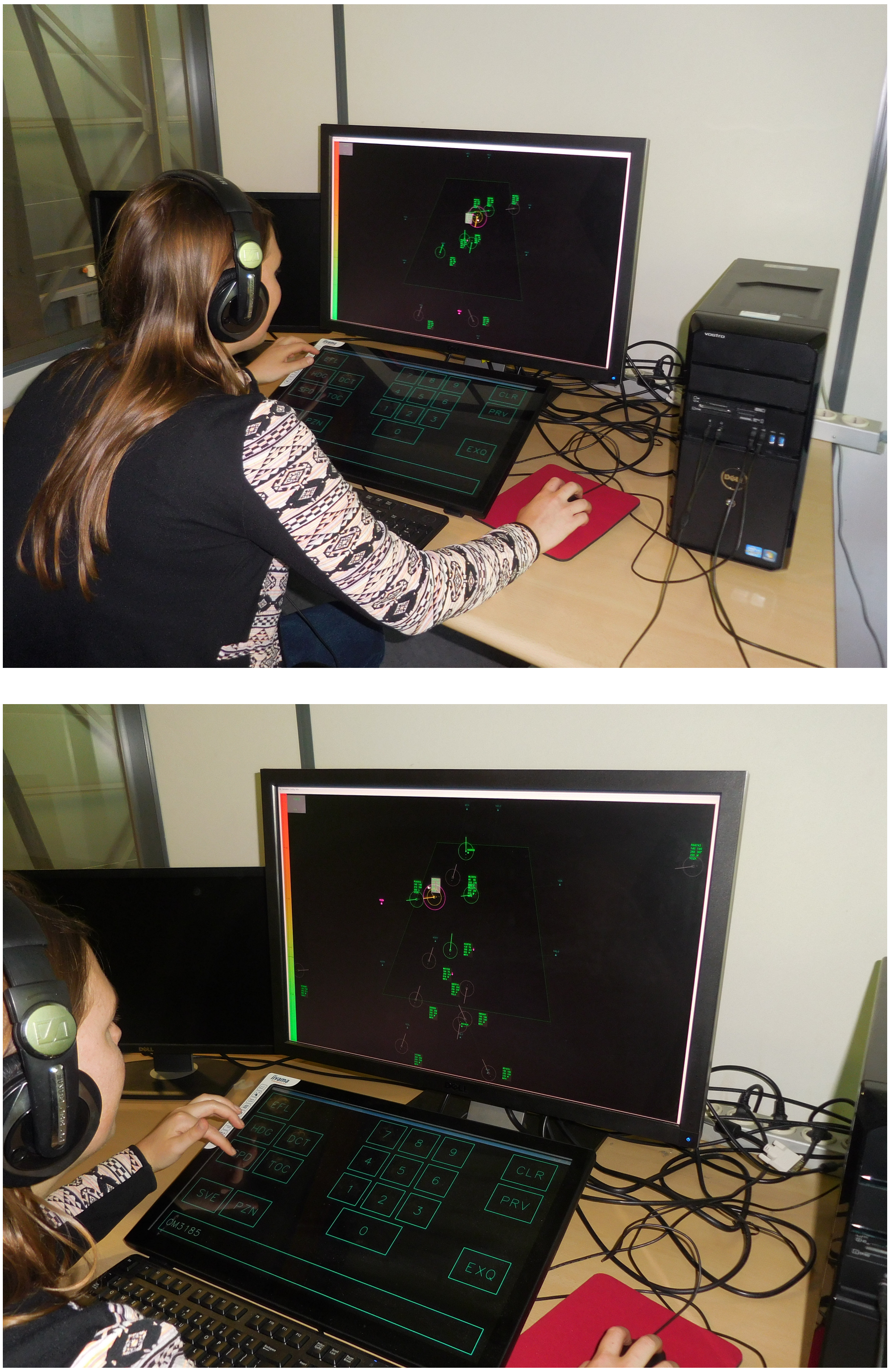



**Figure 4:** The experiment set-up with main screen, touchscreen and headset

# C Pre-Questionnaire

Before the experiment participants were required to fill out the following pre-questionnaire to determine their level of expertise.

Subject ID: ____________________

Date: ____________________

**Pre-Questionnaire**

1. Wat is uw Air Traffic Controller ervaring?
   - ◯ ______ jaren ervaring, ga verder naar vraag 2
   - ◯ ATC cursus, ga verder naar vraag 3

*vul nu enkel vraag 2 of vraag 3 in*

2. Waar heeft u gewerkt?
   - ◯ TWR
   - ◯ ACC
   - ◯ APP/DEP
   - ◯ Anders, nl ____________________
3. Wat voor cursus heeft u gedaan en hoe lang duurde deze?

# D Instruction Manual

After the pre-questionnaire participants were instructed both with the instruction manual that can be found in this Appendix and with additional verbal explanation on the goal of the experiment and the experiment set-up. It was made clear to participants that accurate workload ratings were very important for the experiment and ISA ratings should be given carefully.

# Instruction Manual

Dear participant,

First of all thank you for participating in this experiment regarding workload of Air Traffic Control. This experiment will start with 9 short training scenarios which this instruction manual will guide you through to get familiar with the sector, the aircraft and the commands. Afterwards the final 4 x 20 min test scenarios will take place. During the entire experiment, subjective workload ratings will have to be done in the form of ratings from 0 to 100 %. Please keep in mind, since this experiment will focus on workload, that these workload ratings should be given as accurately as possible. More detailed instructions will follow later.

## 1 Sector

The sector that will be used is shown below, in Figure 1.

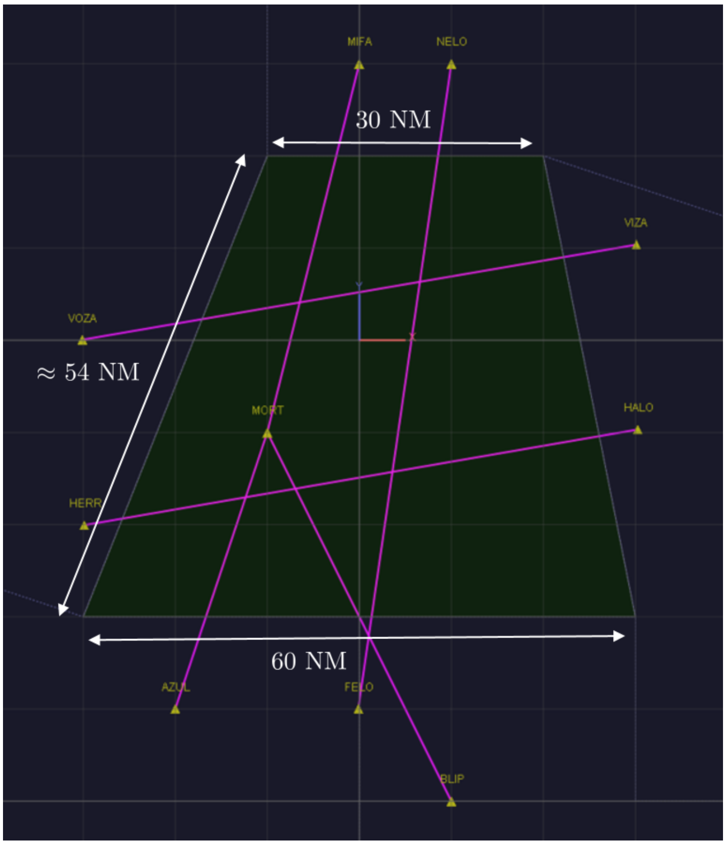



Figure 1: Experiment Sector

To give an idea about the size of the sector, a grid is shown here, where a square is 10 x 10 nautical miles. Also the lengths of the boundaries of the sector are indicated.

The rules that apply to the sector are the following:

- Inbound traffic coming from AZUL and BLIP and going to the northern waypoing MIFA, has to be merged and leave the sector between FL 70 - 100
- Outbound traffic from NELO to FELO has to leave the sector at FL 200
- Over flights towards HALO have to be handed over at FL210
- Over flights towards VOZA leave the sector at the same flight level as they enter (FL140)
- Aircraft have to be given a transfer of control before they leave the sector.
- When aircraft are given a transfer of control they have to be separated (at least 1000 ft vertically and 5NM horizontally) from each other and should not be involved in any conflicts

## 2 Control Task

During the experiment different traffic scenarios will have to be controlled where the rules of the sector, as explained before, apply. Safely guide aircraft through the sector: prevent conflicts from happening, separate the aircraft, guide them to their target waypoint and hand aircraft over to the next sector at the predefined flight levels. Though aircraft have a destination waypoint, they don't have a programmed flight plan, so if no instructions are given they will keep flying in a straight line following their current heading.

## 3 Aircraft

The scenarios will feature 3 different type of aircraft: light, medium and heavy. Their performance overview is given on the next page.

Table 1: AIRCRAFT TYPE: LIGHT

- Label: L

Performance

| IAS MIN | 180 kts | IAS MAX | 250 kts |
|---|---|---|---|
| ROC (FL 0 - 100) | 2370 fpm | ROD (FL 0 - 100) | 2400 fpm |
| ROC (FL 100 - 200) | 2375 fpm | ROD (FL 100 - 200) | 3600 fpm |
| ROC (FL 200 - 300) | 1455 fpm | ROD (FL 200 - 300) | 3285 fpm |

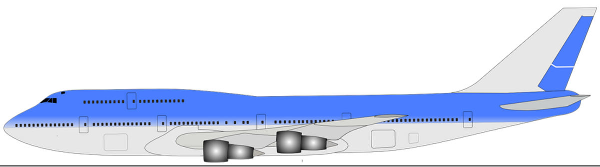
Table 2: AIRCRAFT TYPE: MEDIUM

- Label: M

Performance

| IAS MIN | 200 kts | IAS MAX | 290 kts |
|---|---|---|---|
| ROC (FL 0 - 100) | 2490 fpm | ROD (FL 0 - 100) | 1570 fpm |
| ROC (FL 100 - 200) | 2265 fpm | ROD (FL 100 - 200) | 1985 fpm |
| ROC (FL 200 - 300) | 1530 fpm | ROD (FL 200 - 300) | 2700 fpm |

Table 3: AIRCRAFT TYPE: HEAVY



- Label: H

Performance

| IAS MIN | 230 kts | IAS MAX | 350 kts |
|---|---|---|---|
| ROC (FL 0 - 100) | 3160 fpm | ROD (FL 0 - 100) | 1345 fpm |
| ROC (FL 100 - 200) | 2770 fpm | ROD (FL 100 - 200) | 1755 fpm |
| ROC (FL 200 - 300) | 2040 fpm | ROD (FL 200 - 300) | 2205 fpm |

# 4 Display and commands

Aircraft in the sector are labeled as shown in Figure 2. The arrow next to the label indicates if the aircraft is climbing or descending as here is the case.

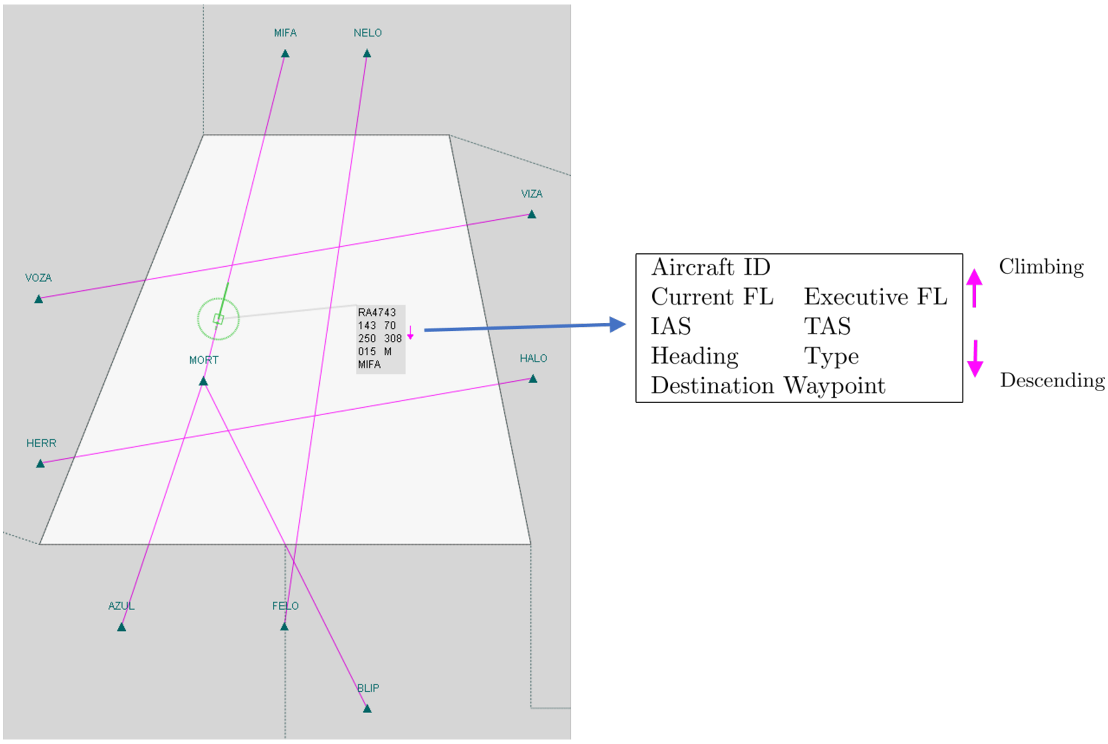



Figure 2: Aircraft in sector with label explanations

If an aircraft is selected the destination waypoint will be highlighted as shown in Figure 3. They can only be controlled by giving them commands using the command display, shown in Figure 4. Multiple commands can be given at the same time.

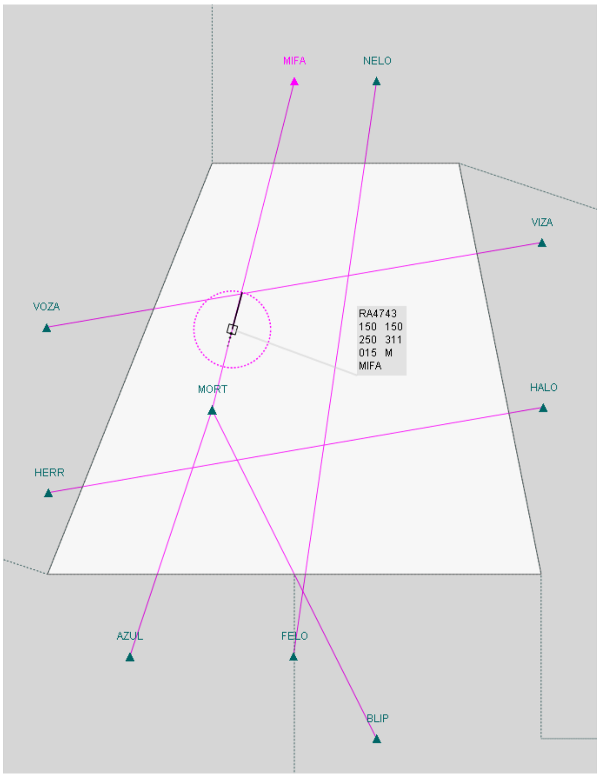



Figure 3: Selected Aircraft in sector

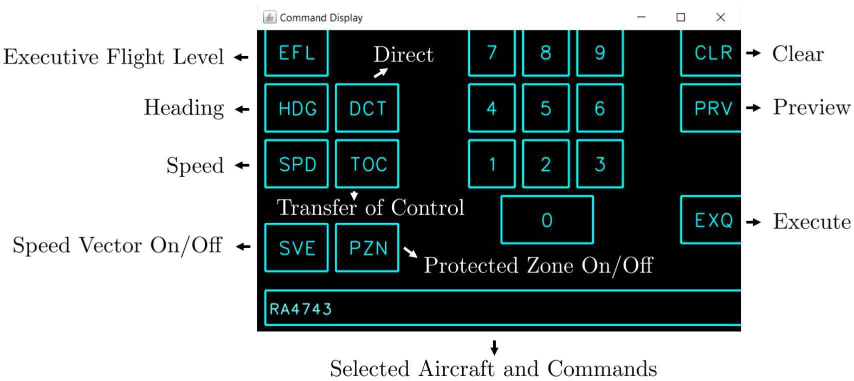



Figure 4: Command Window

A more detailed explanation of the different buttons on the command window is given below:

- EFL: Define the executive or target flight level
- HDG: Select a heading between 0 - 360 degrees
- DCT: Direct to predefined destination waypoint, no other waypoint can be selected
- SPD: Give the target speed within the performance envelope of the aircraft
- TOC: Transfer of Control to adjacent sector. No control of the aircraft is possible after a TOC is given
- SVE: Show/Hide speed vector of all aircraft
- PZN: Show/Hide the protected zone for all aircraft
- CLR: Clear current selected commands to correct for errors, only possible if the commands are not yet executed
- PRV: Preview the validity of selected commands
- EXQ: Execute current selected commands

## 5 Workload measurements

During the experiment, every 30 seconds the ISA (Instantaneous Self Assessment) workload bar will pop up on the left, as shown in Figure 5. Also a sound will be audible as a reminder. Please, try to give feedback on the current workload experienced as soon as possible by clicking with the left mouse button on the corresponding number (from 0-100) on the workload bar. The previous rating will be indicated on the bar by a black horizontal line.

At the end of every scenario also an overall workload feedback is required, this will be done by filling in the sheet provided after each scenario.

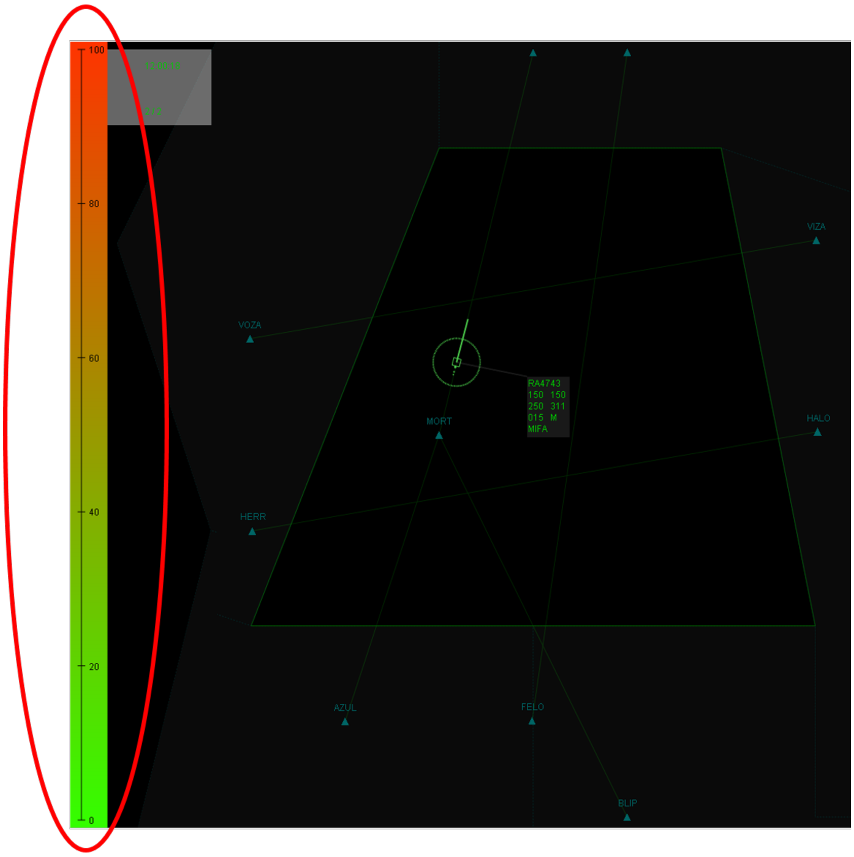



Figure 5: The ISA workload bar is circled in red

Remember these workload ratings are very important for the experiment and therefore please take these ratings seriously and focus on an accurate feedback when clicking on the bar.

## 6 Training Scenarios

If there are still any questions, please don't hesitate to ask.

### 6.1 Training 1

Now let's start with training 1 to get an idea of how the controls work. Please let me know you are ready for training 1 and control the heading of the aircraft as instructed. As you will notice already the ISA workload bar will pop up.

## 6.2 Training 2

The next scenario will be similar to the previous one, but now altitude and speed commands will be added. Select training 2.

## 6.3 Training 3

In case you don't completely feel comfortable with the commands or controlling the aircraft, first repeat training 2, before continuing with training 3.

In Training 3 the merge of 2 aircraft in the sector will be practised. Please select training 3.

## 6.4 Training 4

The next scenario will introduce more aircraft into the sector and is meant to practise with the established exit flight levels as mentioned on page 2. For clarification they are repeated below:

- Inbound traffic coming from AZUL and BLIP and going to the northern waypoing MIFA, has to be merged and leave the sector at FL 70 - 100
- Outbound traffic from NELO to FELO has to leave the sector at FL 200
- Over flight towards HALO have to be handed over at FL210
- Over flights towards VOZA leave the sector at the same flight level as they enter (FL140)
- When aircraft are given a transfer of control they have to be separated (at least 1000 ft vertically and 5NM horizontally) from each other

Start training 4

## 6.5 Training 5

If all the rules of the sector are clear, please select training 5.

## 6.6 Training 6

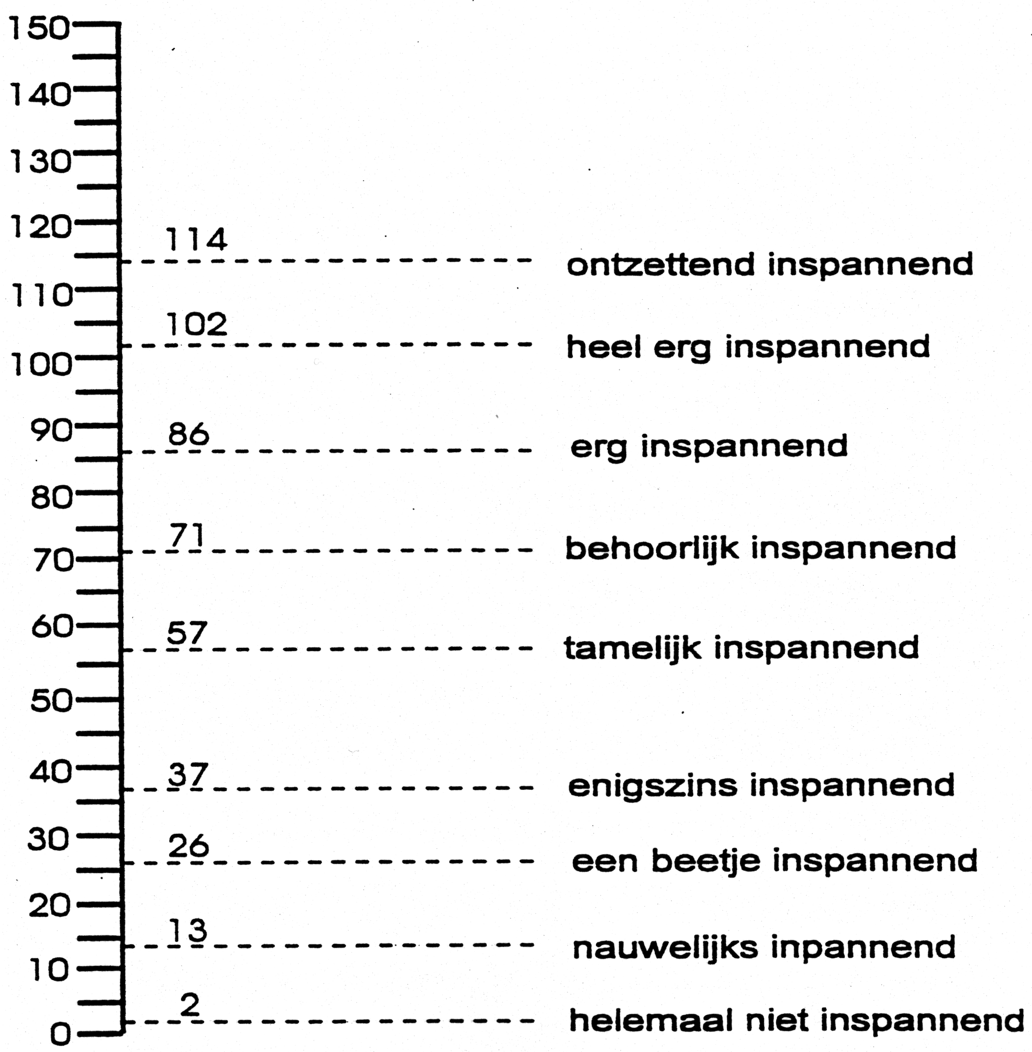
Now even more aircraft will be added to the sector, keep in mind the rules of the sector as practised before and select training 6.

## 6.7 Training 7, 8 and 9

Before starting the experiment, three more complete scenario training runs will be done to get familiar with the sector and possible scenarios.

# E Rating Scale Mental Effort

All participants had Dutch as their native language, therefore, the following Rating Scale Mental Effort (RSME) was used.



# F Post-Questionnaire

At the end of the experiment, participants were asked to fill out the following Google post-questionnaire to obtain feedback on the experiment.

## End-Questionnaire

* Required

1. **Subject ID** *

2. **Date** *

   *Example: December 15, 2012*

3. **Wat vond u van het realisme van de simulator?** *

   *Mark only one oval.*

   | | 1 | 2 | 3 | 4 | 5 | |
   |---|---|---|---|---|---|---|
   | Zeer onrealistisch | ◯ | ◯ | ◯ | ◯ | ◯ | Zeer realistisch |

4. **Heeft u nog suggesties voor verbetering van de simulator?** *

5. **Vond u de moeilijkheidsgraad van de scenario's representatief voor de werkelijkheid?** *

   *Mark only one oval.*

   | | 1 | 2 | 3 | 4 | 5 | |
   |---|---|---|---|---|---|---|
   | Veel te makkelijk | ◯ | ◯ | ◯ | ◯ | ◯ | Veel te moeilijk |

6. **Hoe zeker bent u van uw eigen workload ratings? Zou u ze als betrouwbaar omschrijven?** *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Erg onzeker | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | Zeer zeker |

7. **Kunt u dit verder toelichten?** *

8. **Loste u de simulaties anders op dan normaal?** *

# Part III-C

# Experiment Results

# G Experiment curiosities

During the experiment several observations were made. First of all, it was considered if participants had the protected zone and speed vector turned on or off. Furthermore, loss of separations and warnings were logged. These observations can be found in Table 1. In this table P1 - P4 indicate the professional group, whereas C1 - C6 represent the ATC course/research group. A checkmark indicates no remarks. It can be noted that loss of separation only occurred in the ATC course/research group. Turning the protected zone and speed vector on or off deemed to be personal preference and not group specific.

**Table 1:** Settings and conflicts per participant

| Participant | Protected Zone | Speed Vector | Loss of Separation | Warnings |
|---|---|---|---|---|
| P1 | on | on | - | 3 |
| P2 | on | on | - | 1 |
| P3 | on | on | - | - |
| P4 | on/off | on/off | - | 1 |
| C1 | off | on | 1 | 2 |
| C2 | off | on | 1 | 1 |
| C3 | on | on | - | - |
| C4 | on | on | 1 | - |
| C5 | off | on | 1 | 1 |
| C6 | on | on | - | 4 |

Next, for each participant specific remarks and observations were written down during both the training and the experiment phase. T1 - T9 indicate the training runs and A - D the different scenarios, written down in the order they were executed.

## Participant P1

The observation and remarks for participant P1 can be found in Table 2.

Additional observations/remarks:

- Adding a mouse pad to the set-up for the experiment scenarios solved issues experienced with the mouse during training phase.

**Table 2:** Observations and remarks for participant P1

| | |
|---|---|
| | Training |
| T1 | ✓ |
| T2 | skipped |
| T3 | ✓ |
| T4 | ✓ |
| T5 | selection of aircraft does not always go smoothly |
| T6 | ✓ |
| T7 | EXQ button does not always respond |
| T8 | many speed separations |
| T9 | many speed separations |
| | Experiment |
| A | ✓ |
| B | ✓ |
| C | ✓ |
| D | ✓ |

- Currently no communication available with adjacent sector, now done by communicating with me.
- Malfunctioning equipment caused frustrations and imposed at certain moments more work load than actual conflicts in the scenarios.
- Participant switched to using the mouse to operate the command window instead of the touch screen.
- Physical buttons are more desirable than a touch screen. They are more reliable and easier for touch-typing, because they give pressure feedback if you hit the actual button.

## Participant P2

The observation and remarks for participant P2 can be found in Table 3.

Additional observations/remarks:

- Selection of aircraft does not go smoothly.
- Participant knows when the ISA workload bar appears and anticipates on that.
- Participant finds it difficult to rate workload and ponders on the concept of what workload is.
- Participant is very engaged with the concept of human-factors and workload in daily life and during the experiment.
- Equipment imposed at times more workload than the scenarios.
- Participant is singing during experiments: low workload.

**Table 3:** Observations and remarks for participant P2

| | |
|---|---|
| | Training |
| T1 | ✓ |
| T2 | skipped |
| T3 | ✓ |
| T4 | ✓ |
| T5 | delay touchscreen and touchscreen does not always respond |
| T6 | ✓ |
| T7 | ✓ |
| T8 | ✓ |
| T9 | skipped, due to sufficient level of participant |
| | Experiment |
| B | ✓ |
| A | two aircraft narrowly missed each other, during stress aircraft handled careless; for example no DCTs given |
| D | ✓ |
| C | participant's phone rang at the end of the scenario |

- Professional controllers, i.e. the participant, look at solving patterns. Every merge is solved in the same specific way.

## Participant P3

The observation and remarks for participant P3 can be found in Table 4. There were no additional observations or remarks.

**Table 4:** Observations and remarks for participant P3

| | |
|---|---|
| | Training |
| T1 | ✓ |
| T2 | skipped |
| T3 | ✓ |
| T4 | ✓ |
| T5 | ✓ |
| T6 | ✓ |
| T7 | ✓ |
| T8 | ✓ |
| T9 | skipped, due to sufficient level of participant |
| | Experiment |
| C | aircraft selection does not go smoothly, couple of workload ratings forgotten, orange warning upper part outside of the sector |
| D | ✓ |
| A | execute button and selection of aircraft does not go smoothly |
| B | participant got a phone call half way the scenario |

## Participant P4

The observation and remarks for participant P4 can be found in Table 5.

**Table 5:** Observations and remarks for participant P4

| | |
|---|---|
| | Training |
| T1 | ✓ |
| T2 | skipped |
| T3 | ✓ |
| T4 | ✓ |
| T5 | ✓ |
| T6 | ✓ |
| T7 | ✓ |
| T8 | ✓ |
| T9 | skipped, due to sufficient level of participant |
| | Experiment |
| D | TOCs given early |
| C | almost a conflict, selection does not go smoothly, no DCTs given to aircraft that go roughly in the right direction |
| B | commands to aircraft given very early |
| A | ✓ |

Additional observations/remarks:

- The sound and head phones did not work during the training scenarios, this was fixed for the experiment scenarios.
- Command window was set to half size.
- Mouse was used instead of the touch screen functionality for the command window.
- PZN and SVE were switched on for the first two scenarios and off for the second half of the experiment.

## Participant C1

The observation and remarks for participant C1 can be found in Table 6. There were no additional observations or remarks.

## Participant C2

The observation and remarks for participant C2 can be found in Table 7.

Additional observations/remarks:

- Conflicts are not always indicated if aircraft cross each others flight level.

**Table 6:** Observations and remarks for participant C1

| | |
|---|---|
| | Training |
| T1 | ✓ |
| T2 | skipped |
| T3 | ✓ |
| T4 | ✓ |
| T5 | ✓ |
| T6 | ✓ |
| T7 | ✓ |
| T8 | ✓ |
| T9 | ✓ |
| | Experiment |
| A | couple of conflict situations |
| C | ✓ |
| D | TOC given too early, caused conflict bottom part of the sector |
| B | ✓ |

**Table 7:** Observations and remarks for participant C2

| | |
|---|---|
| | Training |
| T1 | ✓ |
| T2 | skipped |
| T3 | ✓ |
| T4 | TOCs given early |
| T5 | conflict was not indicated (top part of the sector two aircraft both crossed FL150) |
| T6 | ✓ |
| T7 | ✓ |
| T8 | ✓ |
| T9 | ✓ |
| | Experiment |
| C | incorrect commands given; increase in stress/workload, conflict top part of the sector, workload ratings forgotten or clicked away, makes the scenario more complicated |
| A | aircraft does not go exact to waypoint after DCT command (top part of the sector) |
| B | more strategies/routine developed for solving conflicts |
| D | ✓ |

## Participant C3

The observation and remarks for participant C3 can be found in Table 8. There were no additional observations or remarks.

## Participant C4

The observation and remarks for participant C4 can be found in Table 9.

**Table 8:** Observations and remarks for participant C3

| | |
|---|---|
| Training | |
| T1 | ✓ |
| T2 | skipped |
| T3 | ✓ |
| T4 | ✓ |
| T5 | ✓ |
| T6 | ✓ |
| T7 | ✓ |
| T8 | ✓ |
| T9 | many HDG solutions |
| Experiment | |
| D | ✓ |
| B | ✓ |
| A | ✓ |
| C | ✓ |

**Table 9:** Observations and remarks for participant C4

| | |
|---|---|
| Training | |
| T1 | ✓ |
| T2 | skipped |
| T3 | with difficulty |
| T4 | difficulty with merges |
| T5 | difficulty with merges |
| T6 | ✓ |
| T7 | ✓ |
| T8 | ✓ |
| T9 | ✓ |
| Experiment | |
| B | ✓ |
| D | ✓ |
| C | ✓ |
| A | ✓ |

Additional observations/remarks:

- For selecting aircraft, clicking and dragging the mouse at the same time does not work.
- Command window was set to a smaller size
- Participant thinks more from a pilot's perspective

## Participant C5

The observation and remarks for participant C5 can be found in Table 10.

**Table 10:** Observations and remarks for participant C5

| | |
|---|---|
| | Training |
| T1 | ✓ |
| T2 | skipped |
| T3 | ✓ |
| T4 | ✓ |
| T5 | ✓ |
| T6 | skipped, due to participant being familiar with the simulator and rules of the sector |
| T7 | ✓ |
| T8 | ✓ |
| T9 | ✓ |
| | Experiment |
| A | conflict occurred, TOC command almost forgotten, difficulty with merges |
| D | ✓ |
| B | ✓ |
| C | ✓ |

Additional observations/remarks:

- Participant used the strategy to make everything a 2D problem by putting aircraft on the same flight level as soon as possible.
- Protected zone was sometimes turned on to check for possible future conflicts.

## Participant C6

The observation and remarks for participant C6 can be found in Table 11.

Additional observations/remarks:

- The experiment was done in 3 sessions. The first session ended after Scenario A due to limited time availability of the participant. The second run was done 5 days later and included T9, C, B. This data was lost. The final run, included T8, T9, C and B and was done 3 weeks later.
- A DCT command does not always actually provide an exact direct to waypoint to the aircraft.
- The mouse was used instead of the touchscreen to interact with the command window.
- Conflict warnings are not displayed if aircraft cross each others flight level.

**Table 11:** Observations and remarks for participant C6

| | |
|---|---|
| Training | |
| T1 | ✓ |
| T2 | skipped |
| T3 | ✓ |
| T4 | ✓ |
| T5 | ✓ |
| T6 | ✓ |
| T7 | ✓ |
| T8 | ✓ |
| T9 | ✓ |
| Experiment | |
| D | ✓ |
| A | TOC often given outside the sector |
| C | TOC two times given too late, participant got a phone call |
| B | ✓ |

# H Additional Experiment Results

In this Appendix additional experiment results will be presented and shortly discussed. These results include a training effect analysis, all correlation values, additional time traces and scatter plots, additional correlation boxplots, exact cross correlations values and additional mean absolute errors for the metric predictability tests with a random forest regressor.

## Training effect

To investigate whether there existed a training effect during the experiment, the given ISA ratings are plotted for the training phase. Some participants finished certain experiments early and certain training runs were skipped as can be found in Appendix G. Therefore, not for every training run the ISA data were logged. The resulting boxplots can be found in Figure 5.

T7, T8 and T9 were training scenarios representable for the experiment runs. In general, the spread of given ISA ratings narrowed towards the end and/or the ISA ratings stabilized. Combining this with the careful decision at the end of the training phase if participants were ready for the actual experiment scenarios or that extra training was necessary, it was decided not to take training effect into account for further analyses.

## Correlation analysis - All data

In Part I not all correlations could be discussed. Therefore, an extension and more results are presented here. The individual results of all metrics per scenario per participant can be found in Tables 12 - 21. The correlations are split out for taking the mean and maximum of the metric values to compare with the workload data.

## Correlation Analysis - Additional time traces and scatter plots

An example of a good and weak correlation was given in Part I for the 3D $SSD_{sum}$ metric for participant C3 and C5 for scenario A. The time traces and scatter plots of the aircraft

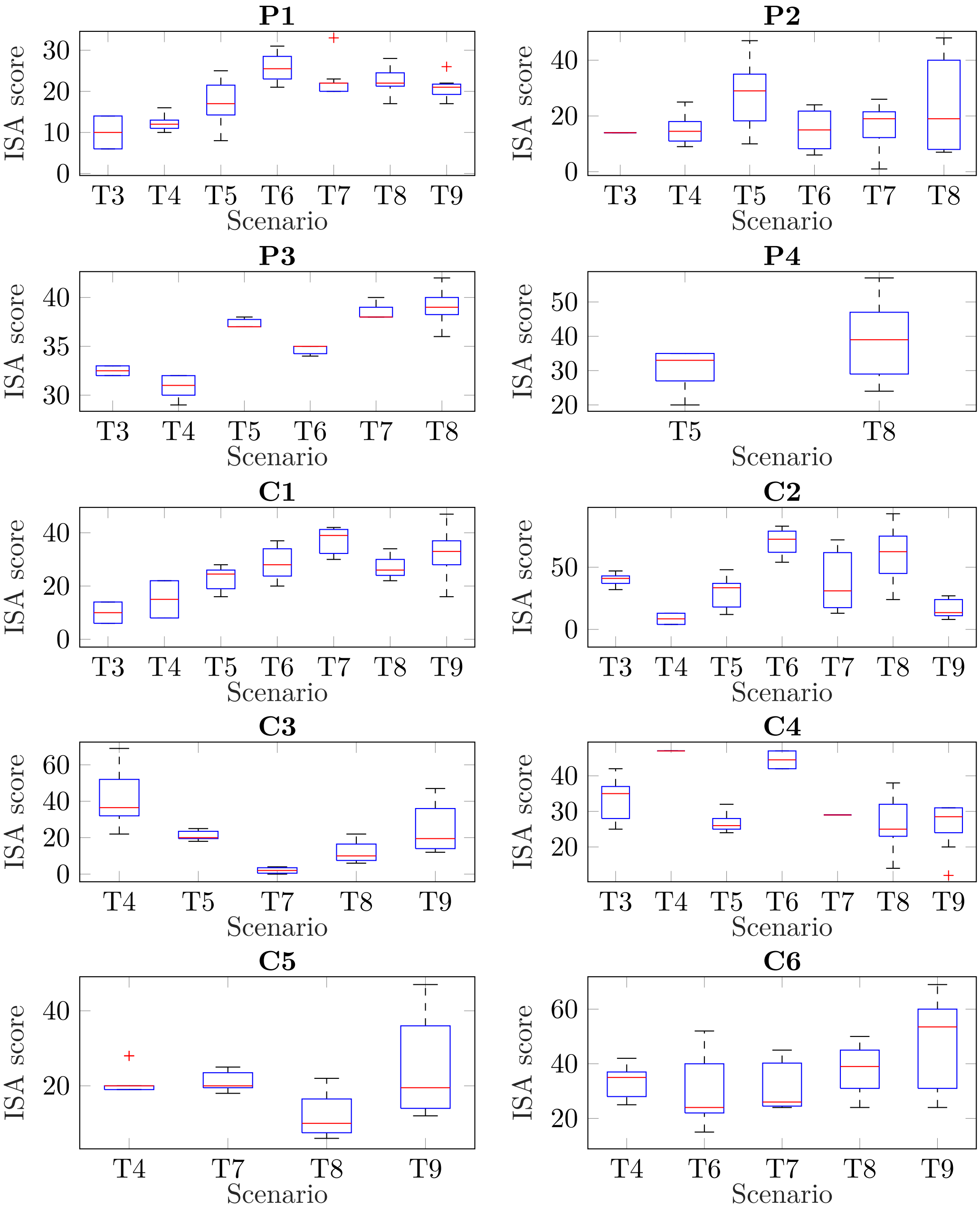
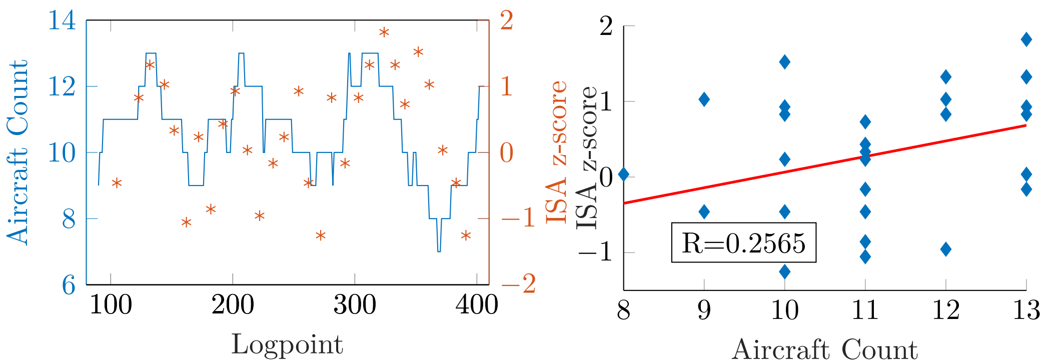



**Figure 5:** ISA ratings per participant per training scenario

count and 2D layered $SSD_{sum}$ of the same scenario for the same participants are visualized in Figures 6 - 8. In all these examples the aircraft count did not have a good correlation with the ISA z-scores and the data were quite scattered. Therefore, an additional figure displaying a good correlation for the aircraft count is visualized in Figure 10. It should be noted that the aircraft count values can only be in the set of natural numbers, resulting in the aircraft count scatter plots always looking slightly different compared to the SSD metrics plots.

**Table 12:** Correlations for participant P1

| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
|---|---|---|---|---|---|---|---|---|
| **Mean values** | | | | | | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | -0.2539 | 0.2350 | 0.2047 | 0.2272 | -0.0096 | 0.2625 | 0.2864 | 0.1995 |
| 2D Layered $SSD_{mean}$ | -0.3485 | -0.0558 | -0.2741 | 0.1390 | -0.0213 | -0.0024 | 0.0164 | -0.0048 |
| 2D Layered $SSD_{max}$ | -0.2991 | -0.1382 | -0.1569 | 0.2829 | -0.0498 | 0.0412 | -0.2319 | -0.0384 |
| 2D Layered $SSD_{sum}$ | -0.3627 | 0.0121 | -0.1007 | 0.2158 | -0.0213 | 0.1334 | 0.2694 | 0.0671 |
| 3D $SSD_{mean}$ | -0.0782 | 0.2837 | -0.4286 | 0.2493 | 0.1114 | 0.2352 | -0.3818 | 0.1055 |
| 3D $SSD_{max}$ | -0.1588 | 0.3079 | -0.3865 | 0.2973 | 0.0498 | 0.2885 | -0.3209 | 0.1295 |
| 3D $SSD_{sum}$ | -0.1541 | 0.2594 | -0.2319 | 0.3069 | 0.0830 | 0.2691 | -0.3068 | 0.1678 |
| **Maximum values** | | | | | | | | |
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | -0.3189 | 0.1821 | 0.2139 | 0.2410 | -0.0233 | 0.2441 | 0.4145 | 0.0918 |
| 2D Layered $SSD_{mean}$ | -0.3631 | -0.0946 | -0.3049 | 0.2445 | -0.0972 | -0.0364 | -0.0516 | 0.0505 |
| 2D Layered $SSD_{max}$ | -0.2804 | -0.0897 | -0.3177 | 0.2733 | -0.0095 | -0.0073 | -0.2673 | 0.1678 |
| 2D Layered $SSD_{sum}$ | -0.3959 | 0.0218 | -0.0305 | 0.3021 | -0.0545 | 0.0946 | 0.2017 | 0.1128 |
| 3D $SSD_{mean}$ | -0.0499 | 0.2530 | -0.3846 | 0.2376 | 0.1735 | 0.2670 | -0.3537 | 0.1608 |
| 3D $SSD_{max}$ | -0.0523 | 0.2600 | -0.3658 | 0.2520 | 0.1899 | 0.3225 | -0.3537 | 0.1612 |
| 3D $SSD_{sum}$ | -0.1756 | 0.2287 | -0.2111 | 0.2589 | 0.1044 | 0.3010 | -0.2319 | 0.1416 |




**(a)** Time Trace

**(b)** Scatter Plot

**Figure 6:** Correlation aircraft count with ISA z-scores for participant C3, scenario A

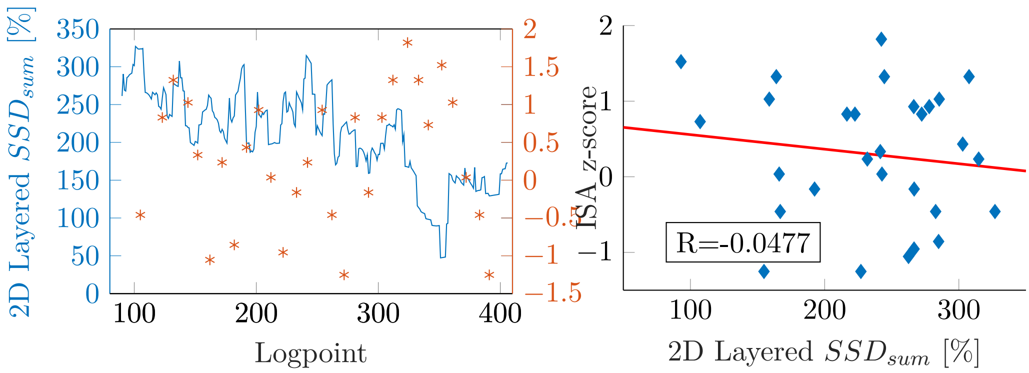
## Correlation Analysis - Additional box plots

In the results section of Part I, the boxplots are discussed of the correlations between the workload and taking the maximum values of the aircraft count and $SSD_{sum}$ values. The other correlation results are presented here, sorted by the different SSD metrics.

First, the correlations of the $SSD_{mean}$ metrics for taking the mean values compared to the

**Table 13:** Correlations for participant P2

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Mean values | | | | | | | | |
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | -0.2308 | 0.3266 | 0.2988 | 0.3730 | 0.3058 | 0.2281 | 0.2167 | 0.3038 |
| 2D Layered $SSD_{mean}$ | -0.1842 | -0.0254 | -0.1422 | 0.3968 | 0.0898 | -0.0102 | 0.0117 | 0.0775 |
| 2D Layered $SSD_{max}$ | 0.1398 | -0.1372 | 0.0163 | 0.4712 | 0.0614 | -0.2287 | 0.1329 | 0.2932 |
| 2D Layered $SSD_{sum}$ | -0.3213 | 0.1271 | 0.1049 | 0.4578 | 0.1559 | 0.1169 | 0.0723 | 0.1714 |
| 3D $SSD_{mean}$ | -0.2457 | 0.2846 | -0.3474 | 0.1714 | 0.1370 | 0.2338 | -0.3567 | -0.0258 |
| 3D $SSD_{max}$ | -0.3071 | 0.3659 | -0.3240 | 0.3170 | 0.0331 | 0.3151 | -0.3847 | 0.0587 |
| 3D $SSD_{sum}$ | -0.3496 | 0.3405 | -0.2308 | 0.2418 | 0.1701 | 0.2541 | -0.1795 | 0.0681 |
| Maximum values | | | | | | | | |
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | -0.2224 | 0.3466 | 0.3118 | 0.2636 | 0.2678 | 0.3377 | 0.1314 | 0.3032 |
| 2D Layered $SSD_{mean}$ | -0.0544 | -0.0483 | -0.1472 | 0.3864 | 0.1206 | -0.0432 | -0.0653 | 0.2338 |
| 2D Layered $SSD_{max}$ | 0.2759 | -0.0586 | -0.0023 | 0.4342 | 0.0142 | -0.1476 | 0.0676 | 0.3971 |
| 2D Layered $SSD_{sum}$ | -0.1752 | 0.0789 | 0.0795 | 0.4871 | 0.1679 | 0.0738 | -0.0023 | 0.3086 |
| 3D $SSD_{mean}$ | -0.2294 | 0.2239 | -0.2901 | 0.2398 | 0.1348 | 0.2033 | -0.4837 | 0.0258 |
| 3D $SSD_{max}$ | -0.1209 | 0.2999 | -0.2617 | 0.2398 | 0.1301 | 0.3027 | -0.5514 | 0.1081 |
| 3D $SSD_{sum}$ | -0.3098 | 0.3053 | -0.1986 | 0.3009 | 0.1963 | 0.2442 | -0.3108 | 0.1057 |




**(a)** Time Trace **(b)** Scatter Plot

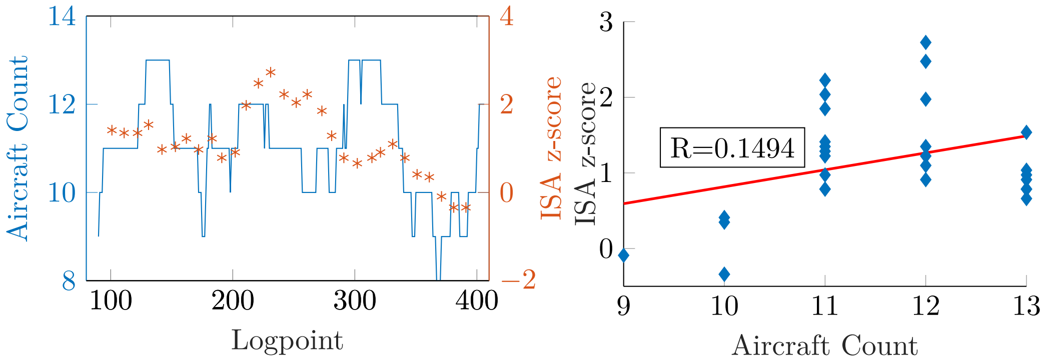
**Figure 7:** Correlation 2D layered $SSD_{sum}$ with ISA z-scores for participant C3, scenario A

ISA ratings including and excluding TOC aircraft can be found in respectively Figures 11 and 12. The correlations for these same metrics, but now taking the maximum values compared to the ISA ratings are visualized in Figure 13 and 14.

Next, the correlations of the $SSD_{max}$ metrics for taking the mean values compared to the ISA ratings including and excluding TOC aircraft can be found in respectively Figures 15 and

**Table 14:** Correlations for participant P3

| Mean values | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | 0.2891 | 0.1544 | 0.0122 | -0.0687 | 0.2891 | 0.1544 | 0.0122 | -0.0687 |
| 2D Layered $SSD_{mean}$ | 0.2687 | 0.0078 | 0.2265 | 0.2852 | 0.1459 | 0.0182 | 0.3373 | 0.1711 |
| 2D Layered $SSD_{max}$ | 0.0077 | -0.1276 | 0.0626 | 0.4750 | 0.1768 | 0.0599 | 0.4192 | 0.3279 |
| 2D Layered $SSD_{sum}$ | 0.3301 | 0.0859 | 0.1783 | 0.3422 | 0.1971 | 0.1172 | 0.4000 | 0.3007 |
| 3D $SSD_{mean}$ | 0.1305 | 0.2526 | 0.0530 | 0.3163 | -0.0384 | 0.1849 | 0 .0000 | 0.1452 |
| 3D $SSD_{max}$ | 0.2636 | 0.2630 | -0.0241 | 0.2178 | -0.0128 | 0.1849 | -0.1397 | 0.1348 |
| 3D $SSD_{sum}$ | 0.2482 | 0.2474 | 0.0145 | 0.2074 | -0.0282 | 0.1953 | 0.0096 | 0.2696 |
| **Maximum values** | | | | | | | | |
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | 0.2415 | 0.2081 | 0.0234 | -0.0523 | 0.3213 | 0.0620 | 0.3038 | 0.2670 |
| 2D Layered $SSD_{mean}$ | 0.2847 | 0.0417 | 0.2099 | 0.3059 | 0.1821 | 0.0235 | 0.3139 | 0.1765 |
| 2D Layered $SSD_{max}$ | 0.0515 | -0.0104 | 0.1157 | 0.4662 | 0.2026 | 0.0679 | 0.3478 | 0.3276 |
| 2D Layered $SSD_{sum}$ | 0.3243 | 0.0782 | 0.2361 | 0.3319 | 0.2178 | 0.1018 | 0.3739 | 0.3374 |
| 3D $SSD_{mean}$ | 0.1305 | 0.2532 | 0.0097 | 0.2622 | -0.0435 | 0.1877 | -0.0846 | 0.1377 |
| 3D $SSD_{max}$ | 0.1998 | 0.2734 | -0.0048 | 0.2051 | -0.1100 | 0.2057 | -0.1159 | 0.1013 |
| 3D $SSD_{sum}$ | 0.2431 | 0.2688 | -0.0387 | 0.2287 | 0.0256 | 0.2138 | 0.0821 | 0.2313 |




**(a)** Time Trace **(b)** Scatter Plot

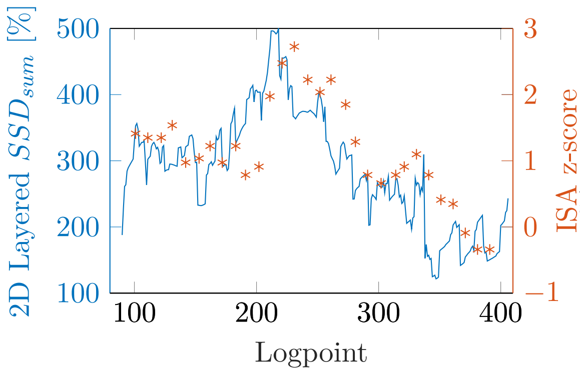
**Figure 8:** Correlation aircraft count with ISA z-scores for participant C5, scenario A

16. The correlations for these same metrics, but now taking the maximum values compared to the ISA ratings are visualized in Figure 17 and 18.

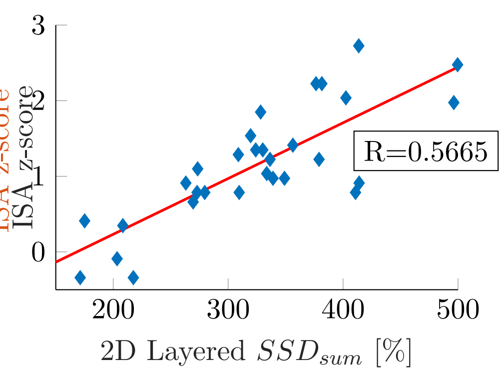
Finally, the correlations of the $SSD_{sum}$ metrics for taking the mean values compared to the ISA ratings including and excluding TOC aircraft can be found in respectively Figures 19 and 20. The correlations for these same metrics, but now taking the maximum values compared to the ISA ratings can be found in the results section of Part I.

**Table 15:** Correlations for participant P4

| Mean values | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | 0.0409 | 0.0073 | 0.3553 | 0.4129 | 0.0192 | 0.1402 | 0.1603 | 0.6043 |
| 2D Layered $SSD_{mean}$ | 0.1320 | 0.0567 | 0.3856 | 0.1429 | -0.0189 | 0.0567 | 0.5632 | -0.1970 |
| 2D Layered $SSD_{max}$ | 0.1272 | 0.0425 | 0.4510 | 0.5065 | -0.1178 | 0.1181 | 0.3529 | 0.0991 |
| 2D Layered $SSD_{sum}$ | 0.1225 | 0.0614 | 0.5118 | 0.3350 | 0.0613 | 0.0945 | 0.4557 | 0.2157 |
| 3D $SSD_{mean}$ | -0.0566 | 0.1370 | 0.1612 | 0.0492 | -0.1602 | 0.2126 | -0.0164 | 0.2413 |
| 3D $SSD_{max}$ | -0.0236 | 0.1654 | 0.3856 | 0.0539 | -0.2215 | 0.1984 | 0.1566 | 0.4005 |
| 3D $SSD_{sum}$ | -0.0283 | 0.1701 | 0.3856 | 0.1382 | -0.1367 | 0.1654 | 0.0070 | 0.4427 |
| Maximum values | | | | | | | | |
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | -0.0181 | 0.0217 | 0.3577 | 0.3180 | -0.0162 | 0.2463 | 0.1844 | 0.6513 |
| 2D Layered $SSD_{mean}$ | 0.1181 | 0.0237 | 0.4117 | 0.1643 | 0.0094 | 0.1161 | 0.5334 | -0.1109 |
| 2D Layered $SSD_{max}$ | 0.1037 | 0.0378 | 0.5282 | 0.4953 | -0.1086 | 0.1752 | 0.3149 | 0.4205 |
| 2D Layered $SSD_{sum}$ | 0.0402 | 0.0852 | 0.5082 | 0.3662 | 0.0094 | 0.1588 | 0.4585 | 0.3557 |
| 3D $SSD_{mean}$ | -0.0518 | 0.1726 | 0.1799 | 0.0516 | -0.1274 | 0.2220 | -0.0070 | 0.2111 |
| 3D $SSD_{max}$ | -0.0047 | 0.1181 | 0.2854 | -0.0094 | -0.2027 | 0.2031 | 0.0608 | 0.3818 |
| 3D $SSD_{sum}$ | -0.0189 | 0.1795 | 0.4211 | 0.1523 | -0.1321 | 0.1795 | 0.0328 | 0.4620 |




**(a)** Time Trace




**(b)** Scatter Plot

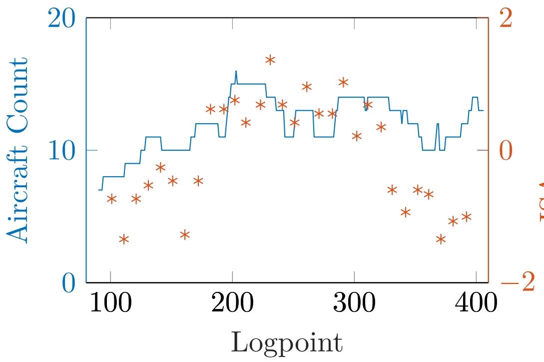
**Figure 9:** Correlation 2D layered $SSD_{sum}$ with ISA z-scores for participant C5, scenario A

## Cross correlations

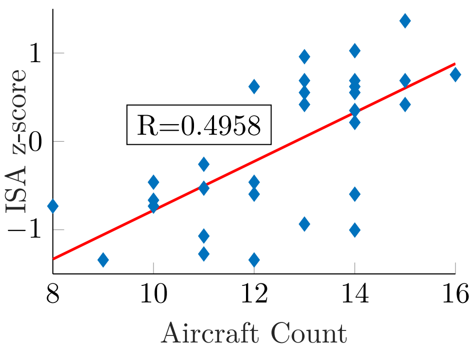
The detailed cross correlations between the 2D Layered SSD and 3D SSD metric per scenario can be found in Table 22. The cross correlations per scenario between the metrics including and excluding TOC aircraft are in Table 23.

**Table 16:** Correlations for participant C1

| Mean values | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | -0.0475 | 0.1769 | 0.1807 | 0.2725 | 0.1134 | 0.1921 | 0.5209 | 0.2561 |
| 2D Layered $SSD_{mean}$ | -0.0511 | 0.1984 | -0.0973 | 0.1982 | -0.0418 | 0.1937 | -0.0358 | 0.3520 |
| 2D Layered $SSD_{max}$ | 0.2973 | 0.0850 | 0.0256 | 0.3734 | 0.0743 | 0.1842 | 0.1740 | 0.5989 |
| 2D Layered $SSD_{sum}$ | -0.0929 | 0.2551 | 0.0819 | 0.2494 | 0.1487 | 0.2835 | 0.2252 | 0.4826 |
| 3D $SSD_{mean}$ | 0.3345 | 0.1323 | 0.1638 | 0.1142 | 0.4274 | 0.0472 | 0.2099 | 0.0070 |
| 3D $SSD_{max}$ | 0.2277 | 0.0992 | 0.2662 | 0.2634 | 0.3485 | 0.0614 | 0.2815 | 0.0583 |
| 3D $SSD_{sum}$ | 0.2927 | 0.1606 | 0.2508 | 0.1329 | 0.3996 | 0.0661 | 0.4505 | 0.1842 |
| **Maximum values** | | | | | | | | |
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | -0.1243 | 0.1749 | 0.2377 | 0.2091 | 0.1130 | 0.2343 | 0.6054 | 0.1960 |
| 2D Layered $SSD_{mean}$ | -0.0023 | 0.2584 | -0.0999 | 0.3240 | 0.0046 | 0.2604 | 0.0205 | 0.5461 |
| 2D Layered $SSD_{max}$ | 0.4247 | 0.1705 | 0.1153 | 0.4101 | 0.2049 | 0.2226 | 0.2771 | 0.6414 |
| 2D Layered $SSD_{sum}$ | -0.0581 | 0.3010 | 0.0692 | 0.3380 | 0.0442 | 0.3267 | 0.2001 | 0.5584 |
| 3D $SSD_{mean}$ | 0.3668 | 0.0639 | 0.2383 | 0.1005 | 0.4173 | 0.0850 | 0.1996 | 0.0771 |
| 3D $SSD_{max}$ | 0.2422 | -0.0095 | 0.2720 | 0.2287 | 0.3605 | 0.0284 | 0.3357 | 0.0958 |
| 3D $SSD_{sum}$ | 0.2894 | 0.1395 | 0.3101 | 0.0889 | 0.4144 | 0.1276 | 0.4351 | 0.2126 |




**(a)** Time Trace




**(b)** Scatter Plot

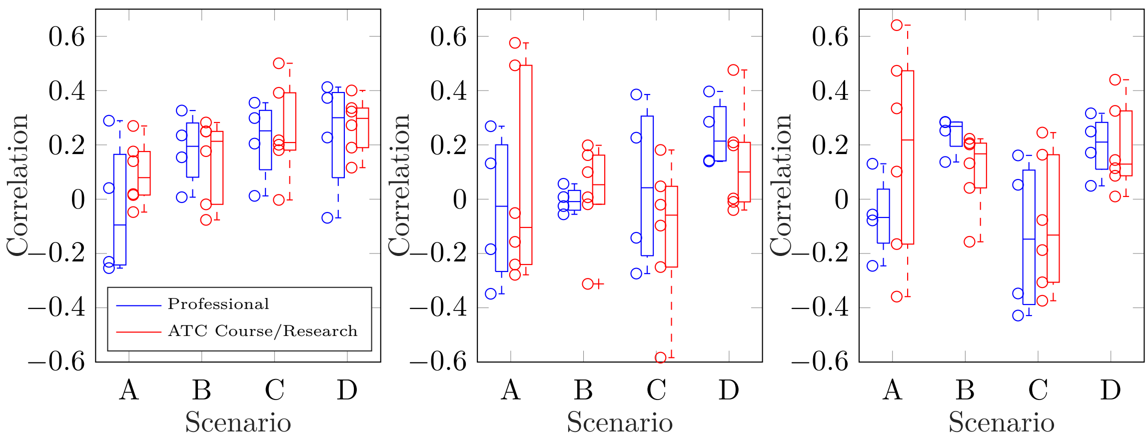
**Figure 10:** Correlation aircraft count with ISA z-scores for participant C6, scenario C

## Metric Predictability

In Part I the mean absolute errors for the aircraft count and $SSD_{sum}$ metrics were discussed. The results for the $SSD_{mean}$ and $SSD_{max}$ metrics can be found in respectively Figures 21 and 22.

**Table 17:** Correlations for participant C2

| Mean values | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | 0.0187 | 0.2497 | -0.0027 | 0.1899 | 0.0420 | 0.2830 | 0.1635 | 0.4365 |
| 2D Layered $SSD_{mean}$ | -0.1572 | 0.1622 | 0.1813 | -0.0400 | 0.0231 | 0.0324 | 0.0107 | -0.0900 |
| 2D Layered $SSD_{max}$ | -0.1803 | -0.2039 | 0.2667 | 0.2250 | -0.0046 | -0.2271 | 0.2560 | 0.3000 |
| 2D Layered $SSD_{sum}$ | -0.0786 | 0.2457 | 0.1440 | -0.0050 | -0.0139 | 0.1808 | 0.1120 | 0.2850 |
| 3D $SSD_{mean}$ | 0.1017 | 0.2225 | 0.2453 | 0.4400 | 0.2358 | 0.1993 | 0.1547 | 0.4400 |
| 3D $SSD_{max}$ | 0.0971 | 0.2086 | 0.2187 | 0.3750 | 0.3283 | 0.1808 | 0.1013 | 0.4801 |
| 3D $SSD_{sum}$ | 0.2035 | 0.2642 | 0.1333 | 0.3900 | 0.2543 | 0.2086 | 0.1280 | 0.5301 |
| **Maximum values** | | | | | | | | |
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | 0.0490 | 0.3611 | -0.0408 | 0.1795 | 0.1487 | 0.2932 | 0.1021 | 0.4597 |
| 2D Layered $SSD_{mean}$ | -0.0023 | 0.2584 | -0.0999 | 0.3240 | 0.0046 | 0.2604 | 0.0205 | 0.5461 |
| 2D Layered $SSD_{max}$ | -0.0486 | 0.1673 | 0.1551 | 0.1350 | 0.0462 | 0.0186 | 0.1148 | 0.0150 |
| 2D Layered $SSD_{sum}$ | -0.1307 | -0.2747 | 0.3666 | 0.4111 | -0.0278 | -0.2695 | 0.3829 | 0.4262 |
| 3D $SSD_{mean}$ | -0.0810 | 0.2927 | 0.1794 | 0.1377 | 0.0832 | 0.2390 | 0.3360 | 0.3137 |
| 3D $SSD_{max}$ | 0.3171 | 0.1879 | 0.2781 | 0.3630 | 0.3866 | 0.1647 | 0.1682 | 0.4361 |
| 3D $SSD_{sum}$ | 0.2497 | 0.2483 | 0.2139 | 0.4512 | 0.3330 | 0.2184 | 0.1736 | 0.4241 |

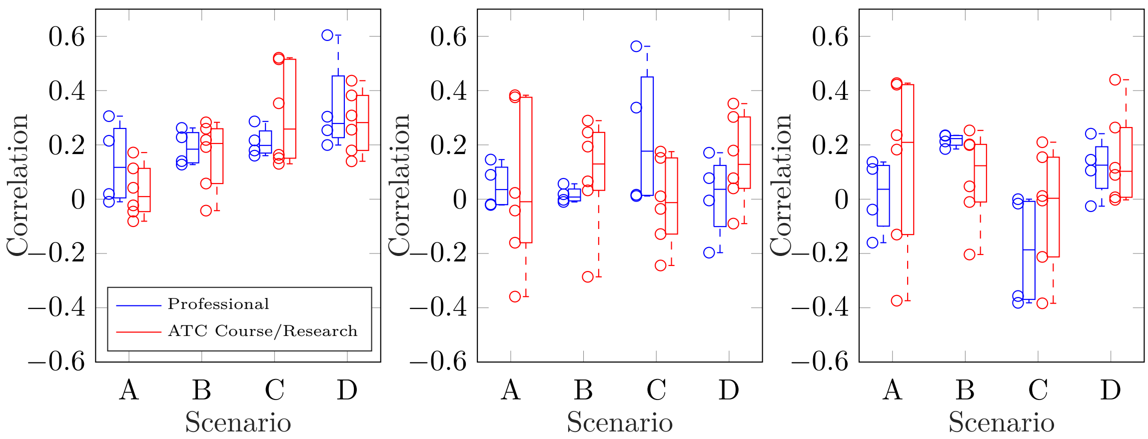



**(a)** Aircraft Count **(b)** 2D Layered $SSD_{mean}$ **(c)** 3D $SSD_{mean}$

**Figure 11:** Correlations for taking the mean values compared to the ISA z-score for the aircraft count and $SSD_{mean}$ metrics

**Table 18:** Correlations for participant C3

| Mean values | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | 0.2697 | -0.0763 | 0.2005 | 0.3227 | -0.0812 | 0.0576 | 0.1303 | 0.1795 |
| 2D Layered $SSD_{mean}$ | -0.2786 | 0.0070 | -0.2500 | -0.0100 | -0.3589 | 0.2459 | -0.1285 | 0.0398 |
| 2D Layered $SSD_{max}$ | -0.0929 | 0.1569 | -0.1846 | 0.2040 | -0.0878 | 0.2178 | 0.0164 | 0.1940 |
| 2D Layered $SSD_{sum}$ | -0.1431 | 0.0117 | -0.0397 | 0.0846 | -0.3740 | 0.1991 | 0.0117 | 0.0547 |
| 3D $SSD_{mean}$ | -0.3589 | -0.1569 | -0.3061 | 0.1443 | -0.3740 | -0.2038 | -0.2126 | 0.0896 |
| 3D $SSD_{max}$ | -0.3890 | -0.1101 | -0.3529 | 0.1343 | -0.4442 | -0.1054 | -0.3388 | 0.0945 |
| 3D $SSD_{sum}$ | -0.1682 | -0.1569 | -0.1005 | 0.2289 | -0.3639 | -0.1241 | -0.0631 | 0.1891 |
| Maximum values | | | | | | | | |
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | 0.2565 | -0.1475 | 0.1753 | 0.2728 | -0.1118 | 0.0752 | 0.1387 | 0.1532 |
| 2D Layered $SSD_{mean}$ | -0.2384 | 0.1970 | -0.2480 | 0.1121 | -0.2886 | 0.3049 | -0.1241 | 0.0473 |
| 2D Layered $SSD_{max}$ | -0.0907 | 0.2215 | -0.1385 | 0.2818 | -0.0806 | 0.2659 | 0.0891 | 0.3909 |
| 2D Layered $SSD_{sum}$ | -0.0477 | 0.0869 | -0.0515 | 0.1619 | -0.3238 | 0.2962 | 0.0000 | 0.0822 |
| 3D $SSD_{mean}$ | -0.2886 | -0.1313 | -0.2761 | 0.1719 | -0.4051 | -0.2204 | -0.2059 | 0.0872 |
| 3D $SSD_{max}$ | -0.2915 | -0.0610 | -0.2854 | 0.1723 | -0.4543 | -0.0610 | -0.2834 | 0.1423 |
| 3D $SSD_{sum}$ | -0.1330 | -0.1667 | -0.1054 | 0.2587 | -0.3238 | -0.1266 | -0.0140 | 0.2117 |

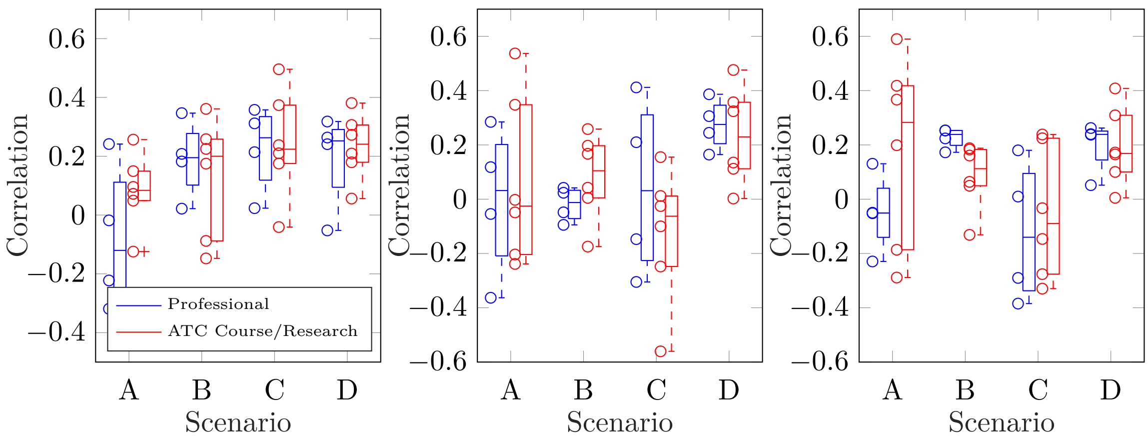



**(a)** Aircraft Count **(b)** 2D Layered $SSD_{mean}$ **(c)** 3D $SSD_{mean}$

**Figure 12:** Correlations for taking the mean values compared to the ISA z-score for the aircraft count and $SSD_{mean}$ metrics excluding TOC aircraft

**Table 19:** Correlations for participant C4

| Mean values | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | 0.1401 | -0.0191 | 0.2169 | 0.4006 | -0.0228 | -0.0422 | 0.3533 | 0.3083 |
| 2D Layered $SSD_{mean}$ | 0.4930 | -0.3120 | -0.0203 | 0.2093 | 0.3829 | -0.2860 | 0.1520 | 0.3029 |
| 2D Layered $SSD_{max}$ | 0.4780 | -0.0780 | -0.0608 | 0.3768 | 0.4580 | -0.0988 | 0.1368 | 0.3422 |
| 2D Layered $SSD_{sum}$ | 0.5431 | -0.2652 | 0.1318 | 0.3128 | 0.3078 | -0.2444 | 0.3244 | 0.3473 |
| 3D $SSD_{mean}$ | 0.4730 | 0.0416 | -0.1875 | 0.0862 | 0.1827 | -0.0104 | -0.0051 | 0.1158 |
| 3D $SSD_{max}$ | 0.4680 | 0.0936 | -0.1875 | 0.1010 | 0.2578 | 0.0468 | -0.0051 | 0.1453 |
| 3D $SSD_{sum}$ | 0.4480 | 0.0312 | 0.0355 | 0.2290 | 0.1727 | 0.0104 | 0.1520 | 0.2290 |
| **Maximum values** | | | | | | | | |
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | 0.0957 | -0.0879 | 0.2102 | 0.3810 | -0.0725 | -0.0182 | 0.4031 | 0.2176 |
| 2D Layered $SSD_{mean}$ | 0.3479 | -0.1746 | -0.0253 | 0.3571 | 0.2430 | -0.2114 | 0.1522 | 0.4606 |
| 2D Layered $SSD_{max}$ | 0.4244 | -0.0808 | 0.0051 | 0.4162 | 0.3922 | -0.0808 | 0.0509 | 0.5109 |
| 2D Layered $SSD_{sum}$ | 0.4880 | -0.1486 | 0.1115 | 0.4014 | 0.2283 | -0.0833 | 0.3856 | 0.4162 |
| 3D $SSD_{mean}$ | 0.4180 | 0.0495 | -0.0330 | 0.1650 | 0.2105 | 0.0469 | 0.0178 | 0.1652 |
| 3D $SSD_{max}$ | 0.2928 | 0.1302 | -0.0737 | 0.1208 | 0.2982 | 0.0677 | 0.0584 | 0.1111 |
| 3D $SSD_{sum}$ | 0.4285 | 0.0130 | 0.0685 | 0.2192 | 0.1353 | 0.0312 | 0.1446 | 0.2219 |

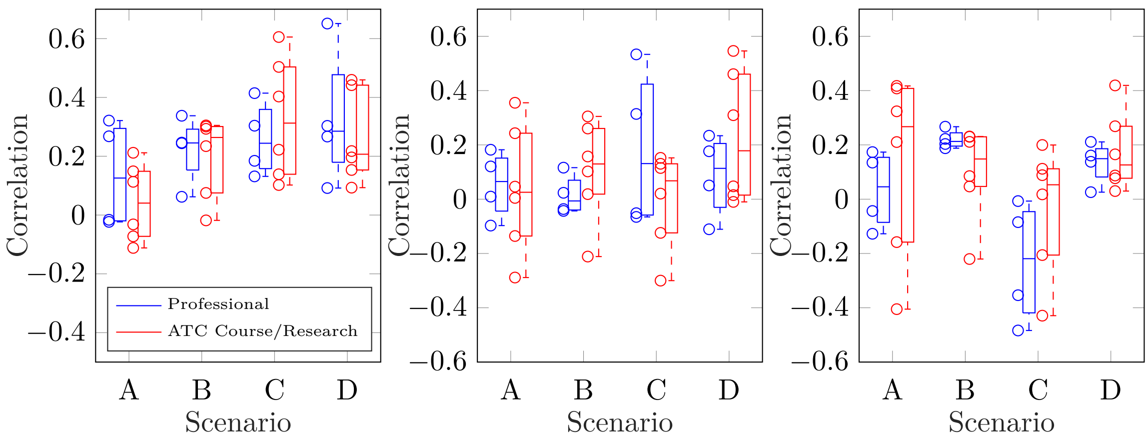



**(a)** Aircraft Count **(b)** 2D Layered $SSD_{mean}$ **(c)** 3D $SSD_{mean}$

**Figure 13:** Correlations for taking the maximum values compared to the ISA z-score for the aircraft count and $SSD_{mean}$ metrics

**Table 20:** Correlations for participant C5

| Mean values | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | 0.1756 | 0.2825 | 0.3920 | 0.3357 | 0.1718 | 0.2182 | 0.1509 | 0.3822 |
| 2D Layered $SSD_{mean}$ | 0.5758 | 0.0997 | 0.0474 | 0.4760 | 0.3753 | 0.2893 | 0.1752 | 0.1791 |
| 2D Layered $SSD_{max}$ | 0.5478 | -0.1434 | 0.1894 | 0.6692 | 0.4080 | 0.1677 | 0.1279 | 0.4383 |
| 2D Layered $SSD_{sum}$ | 0.5851 | 0.1726 | 0.4452 | 0.5561 | 0.3427 | 0.2941 | 0.2605 | 0.3535 |
| 3D $SSD_{mean}$ | 0.6411 | 0.2018 | -0.3741 | 0.3252 | 0.4220 | 0.2018 | -0.3836 | 0.2639 |
| 3D $SSD_{max}$ | 0.4266 | 0.2309 | -0.0521 | 0.3158 | 0.3240 | 0.2698 | -0.1373 | 0.3393 |
| 3D $SSD_{sum}$ | 0.6551 | 0.2698 | -0.0426 | 0.4666 | 0.3707 | 0.2407 | -0.2463 | 0.4194 |
| Maximum values | | | | | | | | |
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | 0.1494 | 0.2250 | 0.3738 | 0.3064 | 0.2117 | 0.3010 | 0.2222 | 0.4417 |
| 2D Layered $SSD_{mean}$ | 0.5368 | 0.0049 | 0.0119 | 0.4760 | 0.3552 | 0.1022 | 0.1304 | 0.3094 |
| 2D Layered $SSD_{max}$ | 0.4174 | -0.0827 | 0.1493 | 0.7039 | 0.3977 | 0.0414 | 0.1258 | 0.5102 |
| 2D Layered $SSD_{sum}$ | 0.5665 | 0.0633 | 0.4438 | 0.5231 | 0.3151 | 0.2677 | 0.2300 | 0.4341 |
| 3D $SSD_{mean}$ | 0.5898 | 0.1608 | -0.3295 | 0.3090 | 0.4080 | 0.2309 | -0.4291 | 0.2686 |
| 3D $SSD_{max}$ | 0.4948 | 0.1872 | -0.0687 | 0.3496 | 0.3594 | 0.2844 | -0.1825 | 0.3346 |
| 3D $SSD_{sum}$ | 0.6038 | 0.2095 | -0.0546 | 0.4506 | 0.3986 | 0.3018 | -0.3350 | 0.4006 |



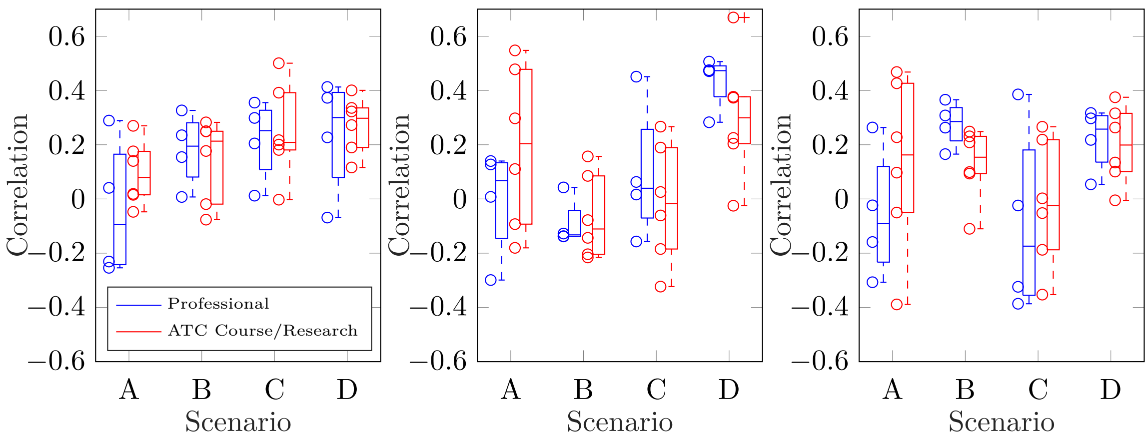

**(a)** Aircraft Count **(b)** 2D Layered $SSD_{mean}$ **(c)** 3D $SSD_{mean}$

**Figure 14:** Correlations for taking the maximum values compared to the ISA z-score for the aircraft count and $SSD_{mean}$ metrics excluding TOC aircraft

**Table 21:** Correlations for participant C6

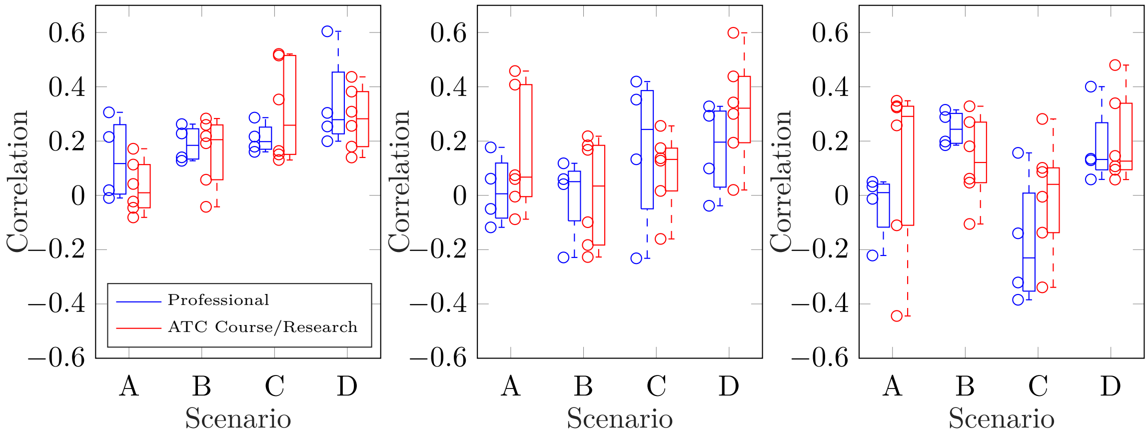
| Mean values | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | 0.0153 | 0.2494 | 0.5006 | 0.1162 | -0.0457 | 0.2595 | 0.5153 | 0.1396 |
| 2D Layered $SSD_{mean}$ | -0.2406 | -0.0188 | -0.5838 | 0.0025 | -0.1604 | 0.0657 | -0.2442 | 0.0772 |
| 2D Layered $SSD_{max}$ | 0.1103 | -0.2158 | -0.3233 | -0.0249 | 0.0602 | -0.1829 | -0.1605 | 0.0200 |
| 2D Layered $SSD_{sum}$ | -0.2306 | 0.0797 | -0.1186 | 0.0224 | -0.1504 | 0.2861 | -0.0488 | 0.1320 |
| 3D $SSD_{mean}$ | -0.1654 | 0.2064 | -0.0767 | 0.0100 | -0.1303 | 0.2533 | 0.0116 | -0.0025 |
| 3D $SSD_{max}$ | -0.0501 | 0.2486 | 0.0023 | -0.0050 | -0.1103 | 0.3283 | 0.0861 | 0.1071 |
| 3D $SSD_{sum}$ | -0.1153 | 0.2158 | 0.1884 | 0.0798 | -0.0953 | 0.2861 | 0.2209 | 0.1021 |

| Maximum values | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Including TOC aircraft | | | | Excluding TOC aircraft | | | |
| | Scenario | | | | Scenario | | | |
| Metric | A | B | C | D | A | B | C | D |
| Aircraft Count | 0.0722 | 0.2581 | 0.4958 | 0.0560 | -0.0318 | 0.3048 | 0.5037 | 0.0931 |
| 2D Layered $SSD_{mean}$ | -0.2038 | 0.0422 | -0.5605 | 0.0025 | -0.1357 | 0.1573 | -0.3000 | -0.0100 |
| 2D Layered $SSD_{max}$ | 0.0427 | -0.0729 | -0.3047 | -0.0678 | 0.0100 | -0.0235 | -0.1257 | -0.0201 |
| 2D Layered $SSD_{sum}$ | -0.1983 | 0.1126 | -0.1630 | 0.0375 | -0.1129 | 0.2794 | -0.0885 | 0.0449 |
| 3D $SSD_{mean}$ | -0.1862 | 0.1829 | -0.1465 | 0.0050 | -0.1581 | 0.2298 | 0.0885 | 0.0299 |
| 3D $SSD_{max}$ | -0.1531 | 0.1573 | -0.0280 | 0.0500 | -0.1404 | 0.2673 | 0.0793 | 0.0950 |
| 3D $SSD_{sum}$ | -0.1159 | 0.2158 | 0.1515 | 0.0426 | -0.1558 | 0.2298 | 0.2474 | 0.0798 |



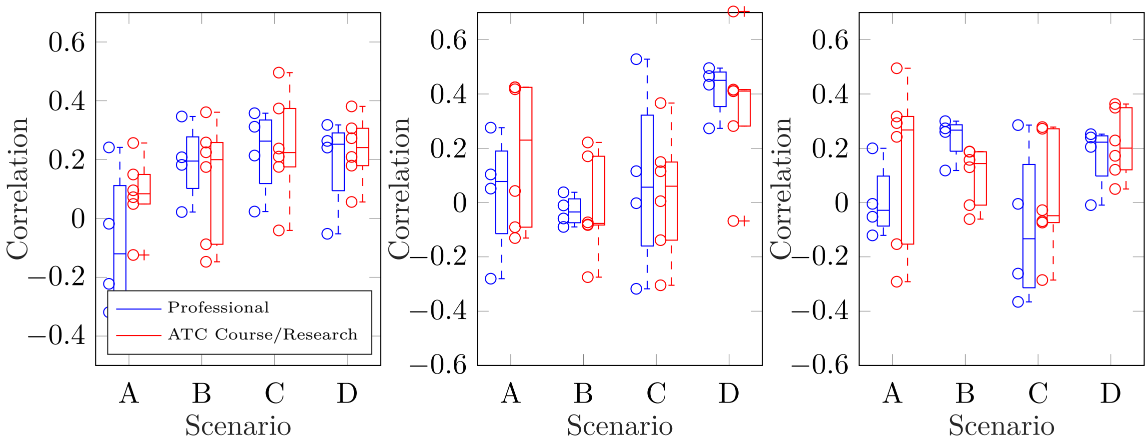

**(a)** Aircraft Count **(b)** 2D Layered $SSD_{max}$ **(c)** 3D $SSD_{max}$

**Figure 15:** Correlations for taking the mean values compared to the ISA z-score for the aircraft count and $SSD_{max}$ metrics




**(a)** Aircraft Count **(b)** 2D Layered $SSD_{max}$ **(c)** 3D $SSD_{max}$

**Figure 16:** Correlations for taking the mean values compared to the ISA z-score for the aircraft count and $SSD_{max}$ metrics excluding TOC aircraft

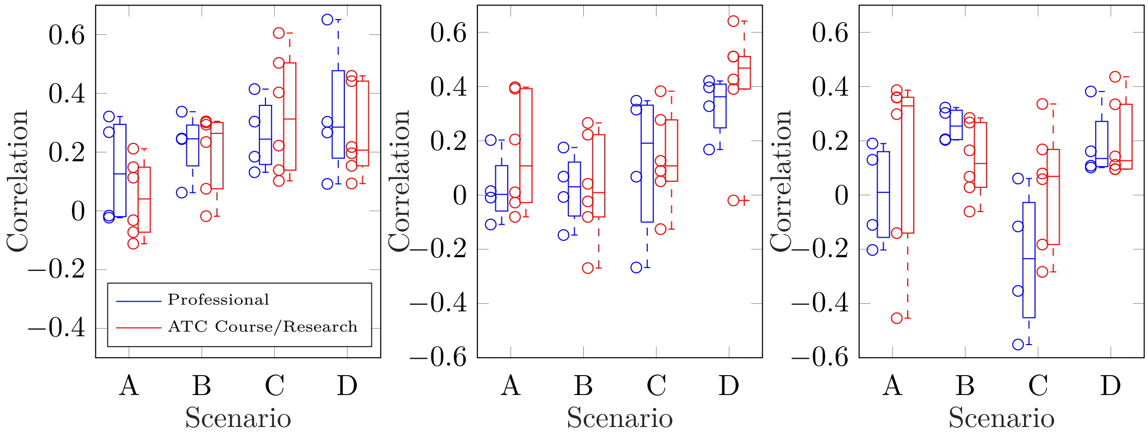



**(a)** Aircraft Count **(b)** 2D Layered $SSD_{max}$ **(c)** 3D $SSD_{max}$

**Figure 17:** Correlations for taking the maximum values compared to the ISA z-score for the aircraft count and $SSD_{max}$ metrics

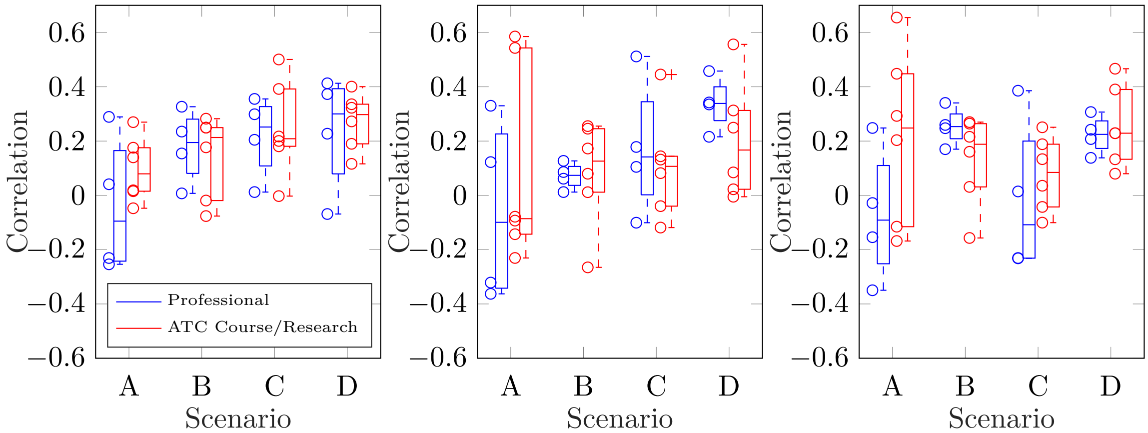
**Table 22:** Cross correlations 2D Layered SSD with the 3D SSD per scenario

| | Scenario | | | |
|---|---|---|---|---|
| Metric | A | B | C | D |
| $SSD_{mean}$ | 0.3116 | -0.0008 | 0.0823 | 0.3871 |
| $SSD_{max}$ | 0.3290 | -0.0459 | 0.1956 | 0.4129 |
| $SSD_{sum}$ | 0.3860 | 0.1866 | 0.2290 | 0.4468 |




**(a)** Aircraft Count **(b)** 2D Layered $SSD_{max}$ **(c)** 3D $SSD_{max}$

**Figure 18:** Correlations for taking the maximum values compared to the ISA z-score for the aircraft count and $SSD_{max}$ metrics excluding TOC aircraft

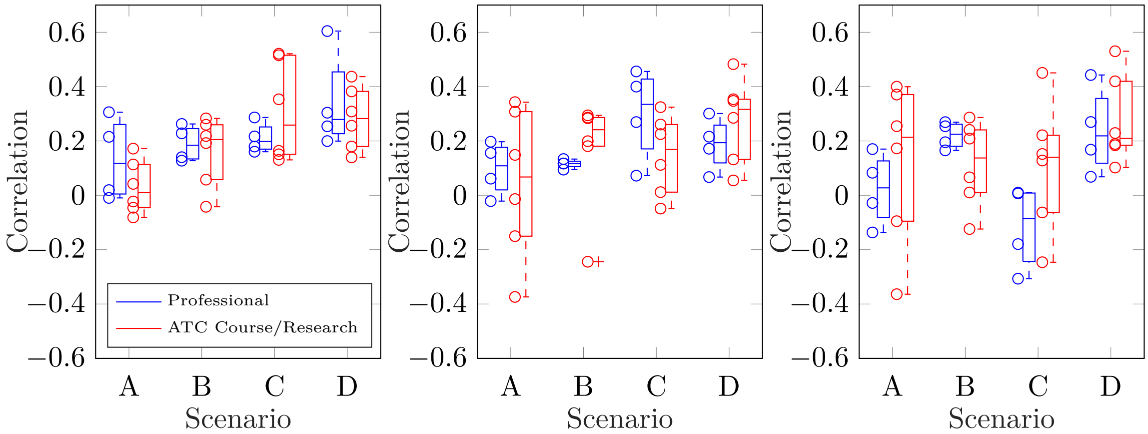



**(a)** Aircraft Count **(b)** 2D Layered $SSD_{sum}$ **(c)** 3D $SSD_{sum}$

**Figure 19:** Correlations for taking the mean values compared to the ISA z-score for the aircraft count and $SSD_{sum}$ metrics

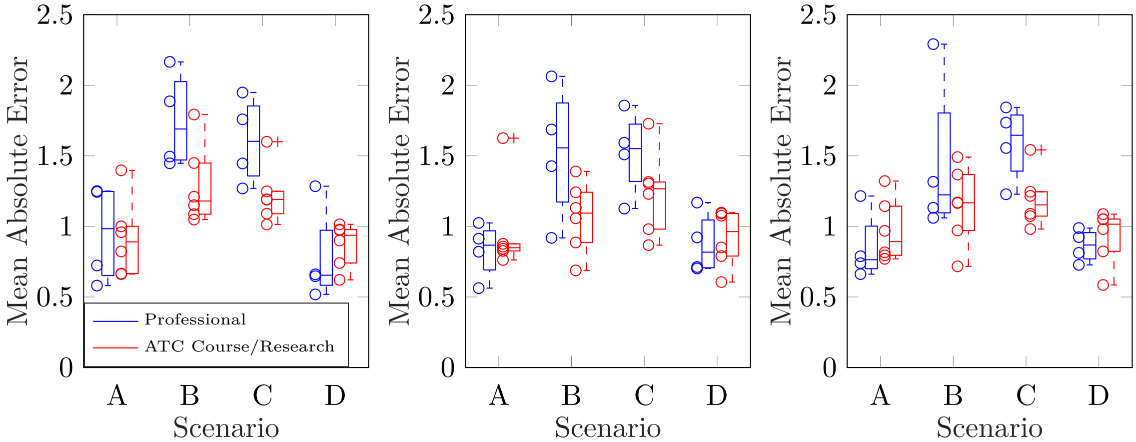
**Table 23:** Cross correlations including and excluding TOC aircraft

| | Scenario | | | |
|---|---|---|---|---|
| Metric | A | B | C | D |
| Aircraft Count | 0.2846 | 0.5396 | -0.0237 | 0.0242 |
| 2D Layered $SSD_{mean}$ | 0.5382 | 0.6042 | 0.5358 | 0.4812 |
| 2D Layered $SSD_{max}$ | 0.4328 | 0.5533 | 0.3980 | 0.4503 |
| 2D Layered $SSD_{sum}$ | 0.3860 | 0.1866 | 0.2290 | 0.4468 |
| 3D $SSD_{mean}$ | 0.5781 | 0.7149 | 0.6299 | 0.5644 |
| 3D $SSD_{max}$ | 0.4369 | 0.7648 | 0.6324 | 0.5087 |
| 3D $SSD_{sum}$ | 0.4849 | 0.7138 | 0.4766 | 0.5137 |



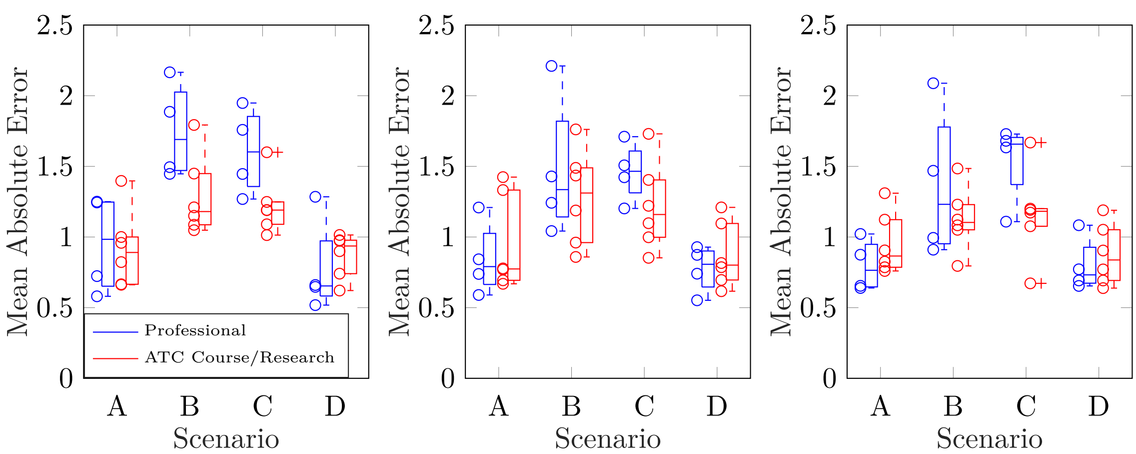

**(a)** Aircraft Count **(b)** 2D Layered $SSD_{sum}$ **(c)** 3D $SSD_{sum}$

**Figure 20:** Correlations for taking the mean values compared to the ISA z-score for the aircraft count and $SSD_{sum}$ metrics excluding TOC aircraft




**(a)** Aircraft Count **(b)** 2D Layered $SSD_{mean}$ **(c)** 3D $SSD_{mean}$

**Figure 21:** Mean absolute errors for the predictions made for the aircraft count and $SSD_{mean}$ metrics per scenario per group




**(a)** Aircraft Count **(b)** 2D Layered $SSD_{max}$ **(c)** 3D $SSD_{max}$

**Figure 22:** Mean absolute errors for the predictions made for the aircraft count and $SSD_{max}$ metrics per scenario per group

# I Questionnaire Results

In this Appendix the result of the post-questionnaire can be found. Participants were, first of all, asked to rate the realism of the simulator. As can be seen in Figure 23, participants deemed the simulator as quite realistic.

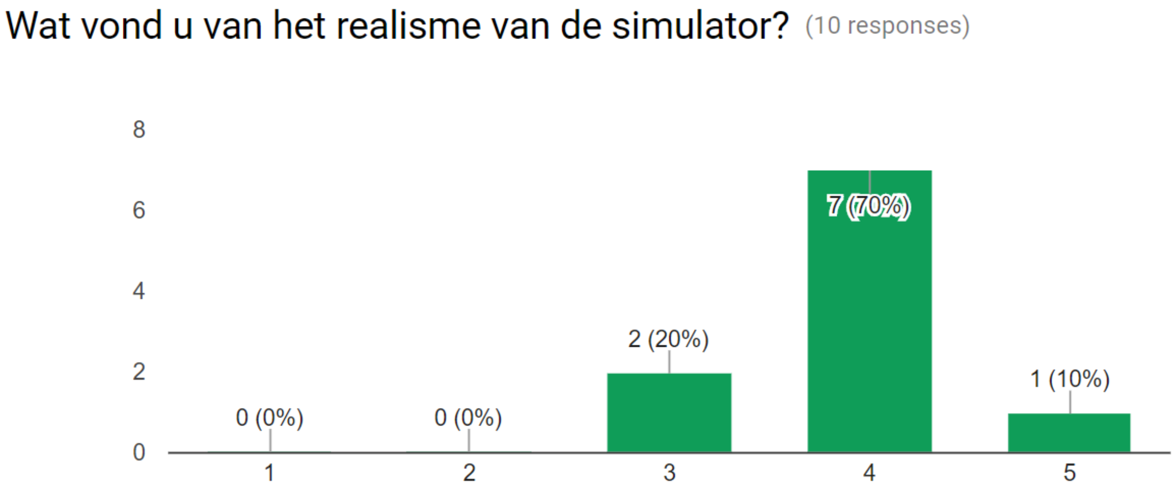



**Figure 23:** Participants' response to realism of the simulator

Next, improvement suggestions for the simulator were asked. These are visible in Figure 24. Additional improvement suggestions, feedback, comments and explanations on them can be found in Appendix K.

The average difficulty of the scenarios as rated by the participants can be found in Figure 25. Though opinions differed, most participants rated that the scenarios on average were neither too hard nor too easy with four participants indicating the scenarios were too easy and one participant indicating that the scenarios were too hard compared to reality.

To get additional feedback on the realibitily of the ISA ratings, participants were asked how confident they felt about their own ratings. The results are visualized in Figure 26, with additional comments in Figure 27. It can be concluded that almost all participants felt very confident about their ability to correctly rate their experienced workload. For the participant that felt less confident, no remarkable outliers or abnormalities were detected for the ISA

Heeft u nog suggesties voor verbetering van de simulator? (10 responses)

| |
|---|
| Werking touch screen en muis moet vloeiend zijn. Touch screen moet goed werken. Als de commandos niet goed pakken is dat erg iritant (werkdruk schiet omhoog) |
| descent performance te sterk |
| touchscreen "miste" soms mijn commando's, moest ik het opnieuw invoeren |
| TID wat kleiner. DCT in label. Flightstrips (kan intended speed zien). Deelnemers niet toestaan om lvl/hdg te veranderen buiten sector. |
| Descent performance aanpassen |
| Muis verspringt van Radar scherm naar touchscreen als je het touchscreen gebruikt (die zouden beter los van elkaar kunnen zijn). Vliegtuigen worden gedeselecteerd bij het aanklikken van de ISA score, dat was soms hinderlijk. |
| zoals besproken |
| specifieke waypoints kunnen selecteren in DCT command |
| knoppendoos kleiner, voice |
| Geen mogelijkheid bieden om speeds te increasen |

**Figure 24:** Participants' suggestions for improvement of the simulator

Vond u de moeilijkheidsgraad van de scenario's representatief voor de werkelijkheid?
(10 responses)

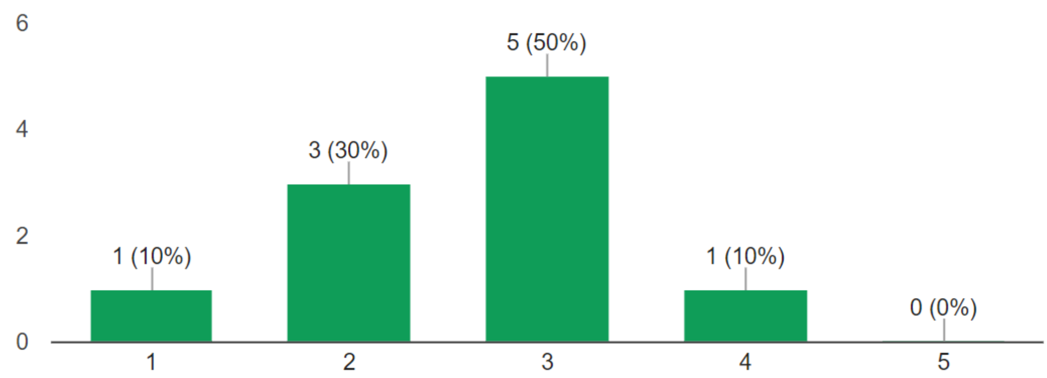



**Figure 25:** Participants' response to the difficulty level of the scenarios

ratings. Therefore, all participants and ISA ratings were taken into account for further analysis.

For further insight into the realism of the experiment, participants were asked if they solved the traffic situations in a different way than is common due to the set-up of the experiment

Hoe zeker bent u van uw eigen workload ratings? Zou u ze als betrouwbaar omschrijven?
(10 responses)

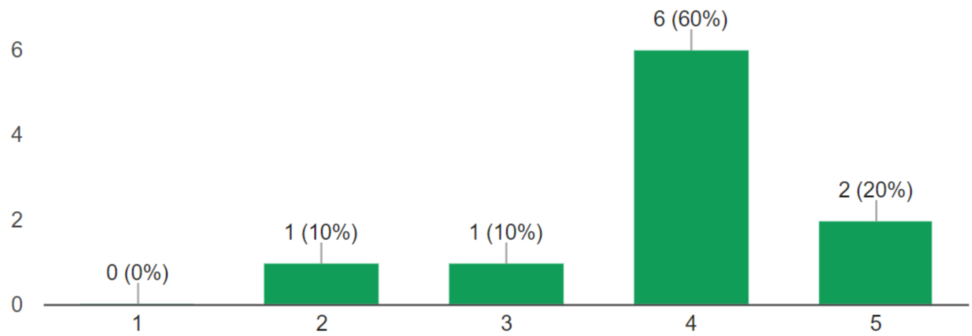



**Figure 26:** Participants' confidence level of their ISA ratings

Kunt u dit verder toelichten? (10 responses)

Moeilijk om je eigen workload te kwantificeren, soms moet je ineens twee versnellingen omhoog om boven het werk te blijven staan ,maar als je dat doet is er niks aan de hand gevoelsmatig.

Een verkeersleider laat zich niet gemakkelijk kennen dus scoort mogelijk vaker op Lage Workload(?)

Er waren duidelijk twee niveau's qua workload

Bij task demand load heb je te weinig tijd voor zelfreflectie.

Te weinig differentiatie in opkomstgegevens

Ik heb de ISA rating naar waarheid ingevuld, ik denk dat responstijd icm absolute waarde ook veel zegt (soms was het scenario druk, waardoor de ik de ISA even links liet staan)

soms wel veel WL queries achter elkaar

Ik heb wel serieus de ratings proberen in te vullen. De workload varieerde wel tussen scenarios, maar ook binnen de scenarios. Daardoor kon ik aardig goed mijn workload inschaten tussen "laag, gemiddeld, en hoog". Ik vond geen enkel scenario te moeilijk, dus heb nooit tot 100% ge-rate.

relatieve waarden wel OK, absoluut is moeilijk te beoordelen

Relatief korte oefeningen; workload belasting voel je vooral na wat langere tijd.

**Figure 27:** Additional comments to participant ISA ratings confidence

and scenarios. From Figure 28 it can be concluded that this is not the case.

Loste u de simulaties anders op dan normaal? (10 responses)

| |
|---|
| Nee |
| Nee |
| Nee. coodinatie over oplossingen met adjacent centre (Vera) |
| negative |
| meestal op dezelfde manier |
| Dit is de eerste tactische (HDG/SPD/ALT) ATC simulator waar ik mee heb gewerkt, dus ik heb geen referentiekader |
| nee |
| Mijn strategie was voornamelijk de complexiteit eruit halen door zsm de a/c die naar dezelfde FL moesten dezelfde snelheid gegeven en ze gelijktijdig te laten dalen, zodat het in principe een 2D probleem wordt. Top experiment!! |
| verder buiten de sector begonnen. |
| Ja; vaker gebruik gemaakt van speed increase. Ook is "gedrag"van de sim voorspelbaar. Alle veranderingen gaan bij ieder vliegtuig exact gelijk. |

**Figure 28:** Participants' controller strategy compared to reality

# J Additional Participant feedback

During the experiment and post-questionnaire also additional feedback and comments were gathered. They can be categorized in three categories: air traffic control workload, experiment scenario design and simulator improvement suggestions.

## Comments on workload and air traffic control

The professional retired air traffic controllers told a lot about their work and workload. The main points are summarized below.

- For controllers a tunnel vision can occur where they only focus on one conflict.
- In the Netherlands TWR/APP controllers are not separated groups.
- TWR and ACC are completely different regarding conflict situations and how they manage the air traffic.
- ATC is a consideration between efficiency and safety.
- Air traffic controllers are a certain type of person, they rate workload lower than it will actually be. They will always have an attitude that of course they can handle more.
- Air traffic controllers are a certain elite, very proud group. They are in general not interested in something new or different than they are used to (concerning apparatus, displays etc.).

These comments provide an insight into the controller job and are a necessity for everyone doing air traffic control related research. Furthermore, these comments indicate the difficulty and complexity of the air traffic control job and group. Especially, they indicate that measuring their actual experienced workload is due to their attitude and strategy quite hard.

## Scenario feedback

The participants also provided feedback on the different scenarios. These comments can be used to further improve future experiments.

- The earlier commands can be given to aircraft, the easier the task gets.
- The scenarios were quite easy for a professional, partly due to absence of radio contact and delays.
- Being able to control aircraft outside the sector makes the scenarios easier.
- No fly zones could give an extra dimension to the scenarios.
- The workload bar popped up quite often, perhaps workload ratings only every 60 seconds instead of every 30 seconds.

## Simulator improvement suggestions

Finally, many suggestions were given to improve the current used 3D simulator.

- Normally if you click on HDG, a menu appears with all possible waypoints.
- The command window should also respond to for example EFL 7 instead of only to EFL 70. The same holds for SPD.
- Normally radio contact or non-listening pilots can cause possible conflicts. Adding 1 to 2 seconds delays in input from the command window to execution might be an option to implement this.
- Descending performance of aircraft is too fast.
- Normally the command would be HDG DCT instead of only DCT.
- Display the given SPD command in the aircraft label.
- The difference between white and green aircraft is not clear. Aircraft that fly within the sector are both white and green.
- ‘Upspeeden’ (going faster after slowing down the speed of the aircraft), should not be possible. This is a no-go in practice.
- Aircraft call-signs are IATA instead of ICAO.
- Show HDG/DCT command given in aircraft label. Now not clear if a DCT has been given or not.
- Command window should be half the size, so touch-typing can be done.

- Integrate the command window with the sector and put the ISA workload bar on an external (touch)screen.
- Have two different mouse cursors to separate touchscreen and ISA/sector screen. This can prevent the jumping of the mouse cursor when giving commands.
- Improve touchscreen performance and response.
- It was quite hard to select aircraft, make it possible to also select them by clicking not exactly on the right spot.

# K Concluding Remarks and Recommendations

In this Appendix, some final remarks and recommendations concerning the experiment and metric will be discussed.

## 3D Solution Space Metric

No final conclusions could be made about the performance of the 3D solution space metric in its correlations with workload due to the weak correlations found for all tested metrics and the non-significance between the metrics. However, an exact calculated 3D solution space volume metric does not seem to be the way to go for further research. Besides the weak correlations found, this mainly also has to do with the complex calculations involved with the analytical 3D solution space volume. Trying to calculate this partly analytically, but programming it with numerical based programs, caused long calculation times and many exceptions that had to be taken into account. Especially, since there does not seem to be an exact relation between the 3D solution space volume and the workload, an approximation of the volume with a brute force method or an updated 2D layered solution space metric is more advisable.

Furthermore, it became clear that external factors such as equipment can have a huge influence on the experience workload. Therefore, trying to measure only the taskload caused by the traffic is almost impossible and more research into external factors adding to the workload is advisable and especially, research into how an experiment can be designed for them.

## Experiment set-up

Many improvements have been suggested in Appendices I and J. Of these suggestions, certain were mainly personal preferences compared to current systems, and with a proper training, do not necessarily have to influence the experiment. However, the different performance compared to reality of aircraft did significantly have an influence on the experiment and made it easier for the participants to control the traffic. Furthermore, the absence of radio contact and or delays was another factor that made the controlling of the traffic easier compared to reality. Finally, the interaction of the mouse, touchscreen and main screen is advised to improve upon to prevent annoyance that can cause additional workload for participants.

## Scenarios

The scenarios could have been designed with more variations and deviations to make it harder for the participants and get more information of the correlation between the workload and the metrics. This can be done for example by adding no fly zones or flight level restrictions in certain areas. Even more, prohibiting controlling the aircraft outside the sector can be introduced. However, this currently imposes a problem when controllers forget to give a transfer of control in time. When this functionality is improved, this can be integrated. Another idea might be to take wind velocity into account. Especially, because it would be interesting to see how this influences the solution space calculations and correlations with workload. Finally, more variations in incoming flight levels can be taken. However, it should be avoided that the scenarios get too unrealistic.

## Participants

There is a clear difference in strategy between actual controllers and people who have only followed an extended ATC course. Therefore, only using controllers is recommended. Another difference, that so far has not been investigated, is the difference between tower approach and area control center controllers. Their job is significantly different and this can also influence the experienced workload when a certain sector is used for the experiment. In this experiment a sector typical for area control was used, which could have made it harder for tower controllers, who indicated they have different strategies for separation than their colleagues at area control.

Another difficulty that was experienced during the experiment, is that air traffic controllers are a difficult group to measure workload for. They are very proud of their job and do not want to indicate their workload is that high to a (student) researcher. Even more, they will always have the strategy to keep their workload as low as possible, within the boundaries of safety, instead of handling the air traffic as perfect as possible.