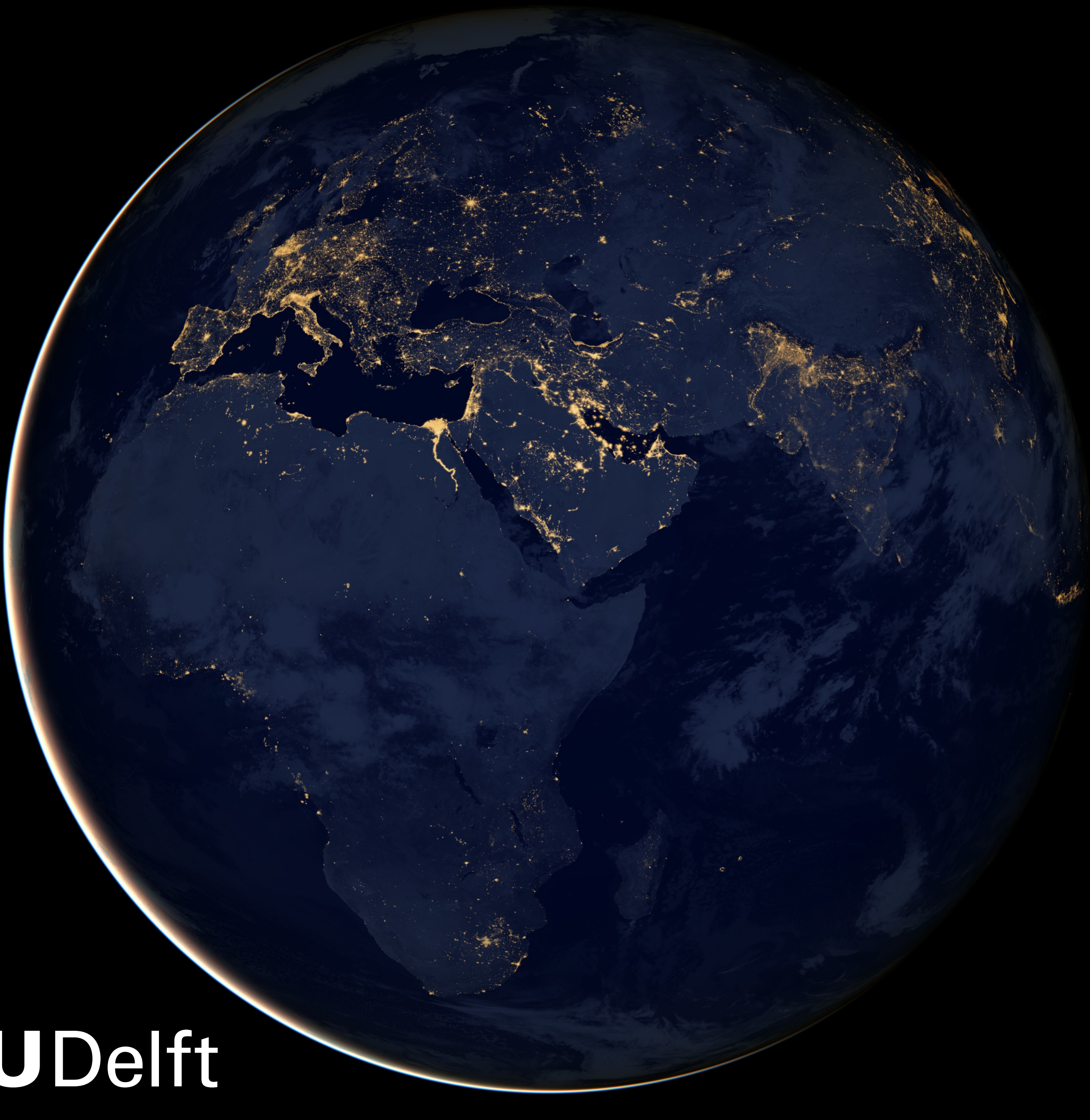
Dynamic Airspace Reconfiguration with Deep Reinforcement Learning

Timon Rowntree

Delft University of Technology



Dynamic Airspace Reconfiguration with Deep Reinforcement Learning

by

T.L. Rowntree

4544927

In partial fulfillment of the requirements for the degree of:

Master of Science
in Aerospace Engineering
at the Delft University of Technology

Date: Monday 10th of October, 2022

Instructors:	prof.dr.ir. J. M. Hoekstra, dr.ir. J. Ellerbroek, MSc M.J. Ribeiro
Institution:	Delft University of Technology
Place:	Faculty of Aerospace Engineering, Delft, The Netherlands

Preface

This work has undoubtedly benefited from the extensive knowledge of the field present within the ATM/CNS department of the faculty. I would like to thank Marta Ribeiro, Joost Ellerbroek and Jacco Hoekstra for their continued guidance in the completion of this MSc Thesis and the opportunity to contribute to ATM research.

Timon Rowntree
Delft, September 2022

Thesis Structure

This final thesis report consists of two parts:

1. **Scientific Paper:** summarises the research and contains the final findings and conclusions of the project.
2. **Preliminary report:** covers the background to the study and provides the rationale behind research activities and methodology.

Contents

Part I

Scientific Paper

Dynamic Airspace Reconfiguration with Deep Reinforcement Learning

Timon Rowntree, Marta Ribeiro, Joost Ellerbroek and Jacco Hoekstra
Control & Simulation, Faculty of Aerospace Engineering
Delft University of Technology (TU Delft), Delft, The Netherlands

Abstract—**For future operations of unmanned aviation, even higher traffic densities than previously seen in manned aviation are expected. Previous work has shown that a vertically layered airspace design performs best at improving safety metrics such as the total number of conflicts and Losses of Separation (LoSs). Furthermore, it has been shown that machine learning techniques are capable of selecting heading ranges for the vertically stacked layers in non-uniform traffic scenarios, in order to reduce the number of conflicts and LoSs compared to uniform structures. These works, however, set structures in an 'empty' airspace and do not take into account the necessary vertical deviations to get from one structure to the next. In this work reinforcement learning (RL) agents are used to select layer heading ranges, while taking into account the previous airspace structure. During this dynamic structuring, several challenges arise. First, it is not clear how to reduce the number of vertical conflicts when aircraft move into a new airspace structure. Second, specific structures should be selected that reduce the necessary vertical deviations from the old structure, while still minimising the cruising conflicts for the new traffic distribution. The present work is divided into three experiments. Experiment I focused on analysing the number of conflicts and LoSs that aircraft suffer during vertically moving towards their layer in the new structure. Experiment II tested whether a RL agent is capable of setting an aircraft structure in function of the expected future traffic scenario. Experiment III aimed to show the capability of a RL agent to select airspace structures, while taking into account the previous airspace structures, in order to decrease the number of vertical conflicts. The results of the research show that RL methods are capable of defining airspace structures appropriate for a given traffic scenario. For dynamic reconfiguration, it proved challenging to simulate traffic scenarios that cause an agent to select different structures to prevent the occurrence of vertical conflicts. Under the experimental conditions employed, analytical methods of structure selection performed better in terms of safety.**

Keywords— **Airspace Design, Airspace Structure, Unmanned Aerial Vehicles (UAVs), Reinforcement Learning, Dynamic Airspace Reconfiguration, BlueSky ATC Simulator**

I. Introduction

With the increasing demand for air traffic in recent years, the airspace capacity is reaching its limit [1]. Furthermore, the forecasts are that this demand will only continue to grow in the coming decades. For the future operations of unmanned aviation, which is the focus of this research, even higher traffic densities than previously seen in manned aviation are expected.

The main objective of Air Traffic Control (ATC) is to prevent collisions between aircraft. Because there is always uncertainty in the exact location of an aircraft, and there should always be enough space for aircraft to turn away from each other in the event of an imminent collision, a safety buffer is used in the form of separation criteria. When two aircraft are actually closer to each other than specified in the defined separation criteria, this is called an intrusion or a loss of separation (LoS). A conflict, on the other hand, is defined as a predicted, potential LoS within a specified prediction horizon, also referred to as the look-ahead time [2]. It is considered that, to ensure adequate safety in our future air spaces, not only will automated conflict detection & resolution (CD&R) become necessary, but there must also be a re-evaluation of coordination efforts that prevent conflicts [3]. In particular, the airspace structure, which is known to affect conflict probability and severity, should be further researched.

The Metropolis project [4] explored different types of distributed structures and found that a layered airspace concept, where aircraft are separated into vertical flight levels by their direction of travel, performed best in terms of safety metrics like the total number of conflicts and LoSs. This can be attributed to the fact that this imposes segmentation and an alignment effects. Aircraft are segmented per layer, and groups of aircraft remain separated from each other. Each layer has a limited heading range, thus aircraft are aligned in their headings, reducing the likelihood of conflict within a layer.

Previous research into layered airspace structures has investigated evenly distributed heading ranges per layer. This is adequate when the air traffic scenario is uniform. However, in reality, traffic can vary continuously. Recently, there has been research into using machine learning techniques to select the heading ranges per layer based on the expected traffic scenario [5], [6]. When doing so, the airspace structure is designed to accommodate a larger number of flight levels for frequently used travel directions. This results in a more uniform distribution of the aircraft altitudes for scenarios with non-uniform heading distributions [7]. Nevertheless, these previous works set the airspace structure in an 'empty' airspace and do not take into account the necessary vertical deviations to get from one structure to the next in the case of a dynamic airspace. It is unclear how safety can be guaranteed during airspace re-configurations and when and what configurations should be selected [7]. This work aims to answer this question by developing a reinforcement learning (RL) agent that is capable of defining heading ranges per vertical layer, while taking into account the previous airspace.

Sections II and III provide the necessary background information on layered airspace structures and the impact of vertical deviations on total conflict and LoS count, respectively. Thereafter, section IV presents an overview of the three experiments of this research. These experiments and their results are further shown in sections V through X. Finally, sections XI and XII present the discussion and conclusion of the research, respectively.

II. Layered Airspace Design

This section introduces the layered airspace concept (subsection II-A), the challenges of non-uniform traffic scenarios in a typical layered airspace (subsection II-B) and dynamic airspace reconfiguration (subsection II-C).

A. Introducing the layered airspace concept

The Metropolis Project [4] set out to investigate the influence of airspace structure on capacity, safety and efficiency for a high-density airspace. One of the concepts introduced in the research is the 'Layers' concept. Here, the airspace is segmented into vertically stacked bands, where each altitude layer limits the horizontal travel to within an allowed heading range. Figure 1 shows an illustration of a layered airspace structure that employs uniform heading range distribution per layer.

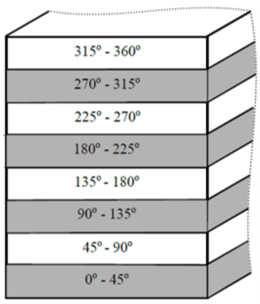


Fig. 1: Illustration of a layered airspace structure with uniform layers. [8]

The core conclusion of the Metropolis project was that vertical segmentation of the airspace results in a lower rate of conflicts, and thus enables higher airspace capacity. Two factors are thought to contribute to this result. First of all, by dividing the aircraft over separate layers of airspace, different groups of aircraft are created that remain separated from each other (segmentation effect). Secondly, within each layer, heading limitations enforce a degree of alignment between aircraft, thereby reducing the relative speed between aircraft cruising at the same altitude. This, in turn, reduces the conflict probability (alignment effect).

B. Non-uniform traffic scenarios

The Metropolis project shows the potential of the layered airspace concept for reducing the total number of conflicts and LoSs, while having minimal effect on the efficiency. However, there are limitations in the way the concept is used in this research. The use of uniformly distributed layers is adequate in cases where the traffic has a uniformly distributed heading distribution. In cases where the headings of the aircraft are not uniformly distributed, however, the aircraft are likely to accumulate in one of the layers, resulting in a non-optimal use of the defined layers. This effect is illustrated in figures 2 - 4. As a baseline, figure 2 shows a traffic situation with uniformly distributed headings (Ψ) in uniform layer structure. This shows how the aircraft are evenly distributed over the available layers. Note that the y-axis represents the heading, Ψ, rather than the height.

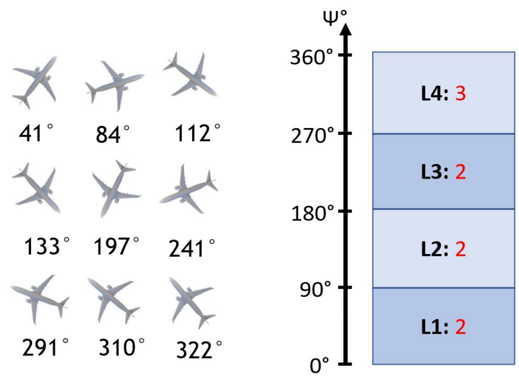


Fig. 2: Uniform traffic with uniform layers.

When a non-uniform traffic scenario is exposed to this same uniform structure, the challenge becomes apparent. Such a situation is shown in figure 3. As is clearly visible, the structure is not suitable for non-uniform traffic scenarios, as aircraft accumulate in one of the layers. In the non-uniform traffic scenario shown, the over-representation of aircraft flying in the 0° - 90° range, results in an unnecessarily full first layer. This has a negative effect on total conflict and LoS count.

Reconfiguring the layers to be able to divide the aircraft more suitably may solve this issue. Such a reconfiguration is shown in figure 4. For the same traffic scenario as presented in figure 3, it can now be seen that aircraft are divided almost equally over the layers again, which is expected to have a positive effect on the total number of conflicts and LoSs experienced. This has also been shown in previous research [5] [6].

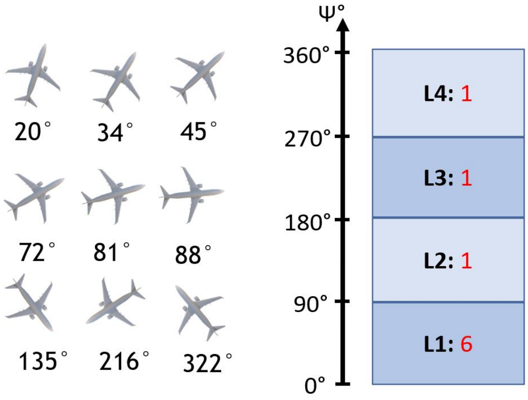


Fig. 3: Non-uniform traffic with uniform layers.

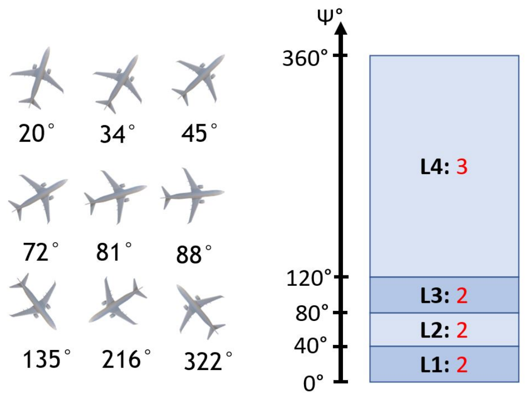


Fig. 4: Layer utilisation of non-uniform traffic for an uneven-layers design. Adapted from [5].

C. Dynamic Airspace reconfiguration

So far, studies have looked at (layered) airspace configurations that are static in time. However, to optimise the airspace utilisation, it may be beneficial to dynamically reconfigure the constraints imposed by a particular airspace design as the traffic scenario changes throughout the course of a day [7]. For the layered airspace concept, this could imply reconfiguring the airspace several times (in reaction to varying traffic scenarios) to achieve a more uniform distribution of aircraft among the layers, such as shown in figures 2 and 4 in the previous subsection. Such dynamic reconfigurations may be beneficial for the capacity of the airspace, but it is unclear how exactly these should happen and what structures should be selected. This research aims to answer this question.

III. Impact of Vertical Deviations on Total Conflict Count & LoS

In this research, there is a strong focus on the transitioning between layered airspace structures. Because such transitions will require aircraft in the airspace to move vertically to the a new correct layer, 'vertical conflicts' are expected to occur in the process. The term 'vertical conflicts' is used to indicate those conflicts in which at least one of the conflicting aircraft has a vertical velocity component that is nonzero. This type of conflict does not fully benefit from the segmentation and alignment effects, which mostly positively affects the number of 'cruising' conflicts and LoSs count. A 'cruising' conflict, on the other hand, occurs when both involved aircraft do have vertical velocity that is zero. In general, conflicts and LoSs do not necessarily scale proportionally (thus, double the number of conflicts does not imply double the number of LoSs). However, it is known from previous research that there is a strong correlation between these two safety metrics [9].

Section III-A illustrates how the rate of reconfiguration is likely to be an important variable in dynamic airspace reconfiguration. Section III-B introduces the challenge of selecting an appropriate manner of moving aircraft across vertical layers during an airspace reconfiguration.

A. Effect of reconfiguration rate on total conflict and LoS count

The reconfiguration rate directly affects the number of vertical manoeuvres and thus also the number of vertical conflicts. For a better understanding of these dynamics, consider a simple hypothetical scenario displayed in figures 5 and 6.

Figure 5 shows the cruising conflict rate for a traffic scenario that changes its predominant traffic direction every 15 minutes. By reconfiguring the airspace structure at that same rate, and with that keeping it suitable for the cruising traffic, the cruising conflict rate can be kept constant. Figure 6 shows the corresponding total number of conflicts, where the 'jumps' at the reconfiguration moments are the additional vertical conflicts that arise due to aircraft moving into the new structure, in line with equation 1.

$$Conflicts_{total} = Conflicts_{cruise} + Conflicts_{vertical} \tag{1}$$

From the example in figures 5 and 6, it becomes apparent that the reconfiguration rate affects the total time aircraft spend in a given airspace structure. In turn, the total time in cruise directly affects the total number of conflicts in cruise. Furthermore, the rate of reconfiguration will determine the number of times the total conflict count 'jumps' because of the vertical conflicts associated with the reconfiguration. All in all, it is clear that the reconfiguration rate affects the relative importance of cruising and vertical conflicts in the total number of conflicts.

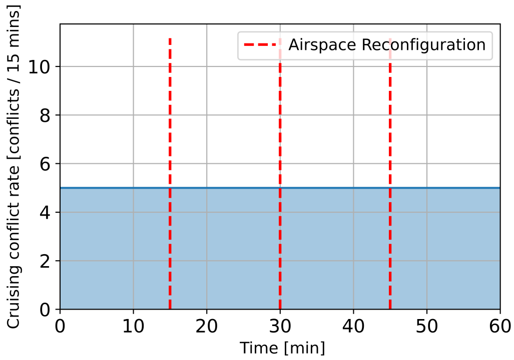


Fig. 5: Example - Cruising conflict rate.

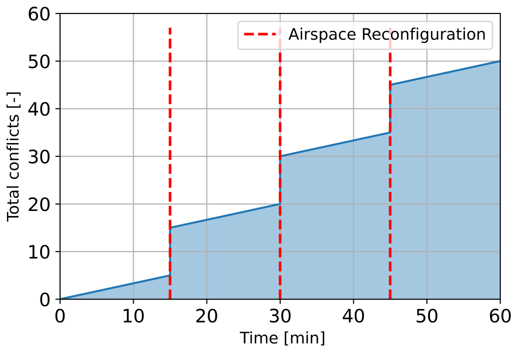


Fig. 6: Example - Total number of conflicts.

B. *Manner of transitioning aircraft*

As established, the restructuring of an airspace that is already filled with aircraft will require aircraft to move from one layer to another. However, exactly how this should be done is unclear. Several options are explored in experiment I, which is aimed at determining the most suitable way to move the aircraft into a new airspace structure during an airspace reconfiguration.

IV. Research Methodology

Three experiments are performed in this research. Experiment I is concerned with finding a suitable way of moving traffic between vertical layers when there is an airspace reconfiguration. In experiment II, a 'static' RL agent is developed and tested. This agent will only look at the current traffic in the scenario when selecting the airspace structures. Finally, in experiment III, the 'dynamic' RL agent is considered, which, in addition to the current traffic situation, also takes the previous airspace structure into account. For this experiment, the decision on the most suitable way of moving traffic between layers, as per the results of experiment I, will be implemented in the dynamic scenarios. After completing the experiments, the performance (in terms of safety, efficiency and stability) of the static and dynamic RL agents will be compared. Figure 7 shows a graphical representation of the aforementioned relation between the experiments.

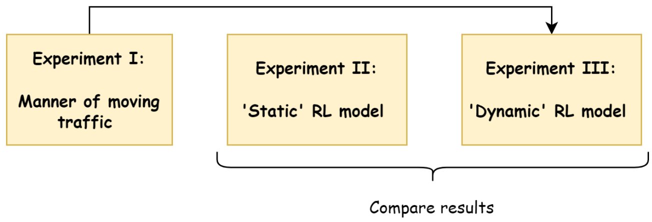


Fig. 7: Relation between the experiments of the research.

V. Experiment 1 - Manner of Moving Traffic

The first experiment aims to find a suitable way to move aircraft into a new structure in the event of an airspace reconfiguration. The method is selected by looking at the total conflicts, rather than the vertical conflicts only. Minimising the vertical conflicts alone is not sufficient if, for example, this happens at the expense of a long reconfiguration time. This will have aircraft remain in the 'wrong' structure for excessive lengths of time, increasing the cruising conflicts and thus also the total conflicts. Subsection V-A will give the outline of the experiment. Section V-B goes into further detail on the experiment setup. The main goal of this experiment is to determine the settings and variables for the control variables for experiment II and III.

A. *Outline of experiment I*

The first step is to define several ways of moving traffic between layers. The first option is to simply instruct all aircraft to move to their new layer at once (Option 1). Alternatively, it is possible to take an approach whereby aircraft move, sequentially, on a layer-by-layer basis (Option 2). Four variants are defined for this second option, which are given letters A-D. In 'Option 2A', the layers are moved through from bottom to top. That is, at the moment of airspace reconfiguration, the aircraft in the bottom layer are ordered to move to their new layer. When this has been completed, the aircraft in the second lowest layer are ordered to move, and so on until all the layers have been covered. 'Option 2B' is analogous to 'Option 2A', but passes through the layers from top to bottom. 'Option 2C' and 'Option 2D' also entail layer movements in a sequential manner. However, rather than going bottom - top (2A), or top - bottom (2B), these options pass through the layers based on the traffic density per layer. 'Option 2C' starts with the layer that has the lowest traffic density and ends with the layer with the highest traffic density. 'Option 2D', does the exact opposite by starting with the layer with the highest traffic density, while ending with the layer with the lowest traffic density. In summary, there are then the following five manners of moving traffic that are to be explored. For an overview, see table I.

TABLE I: Manners of moving traffic during a structure reconfiguration.

Manner of moving	Description
1	All at once
2A	Layer-by-layer: bottom - top
2B	Layer-by-layer: top - bottom
2C	Layer-by-Layer: low - high traffic density
2D	Layer-by-Layer: high - low traffic density

To account for the fact that the most suitable manner of moving aircraft to new layers may depend on the nature of the airspace reconfiguration, the structure before and after the reconfiguration is randomly selected during the experiment. Each manner of moving aircraft is run for 100 scenarios, thus yielding a total number of 500 simulations. Three dependent variables are recorded for analysis: 1) the total number of conflicts, 2) the number of vertical conflicts and 3) the reconfiguration time. Finally, the selected method based on this experiment will be used in experiment III, where the focus lies on dynamic airspace reconfiguration.

B. Experiment Setup

This section discusses the relevant parameters of the experiment. Specifically, it goes into the tools used for the simulations, the airspace structures, the aircraft types, traffic scenarios and finally the (CD&R) methods used.

1) Use of BlueSky Open Air Traffic Simulator: For the experiments, use is made of the BlueSky Open Air Traffic Simulator [10]. 'BlueSky' has been created by the ATM/CNS department of the Faculty of Aerospace Engineering at Delft University of Technology, in response to the need for comparing efforts and results in the field of ATM research. It is an open source and open data approach to air traffic simulation. The use of BlueSky makes it possible to take advantage of its performance library, which includes the specifications of many aircraft types, and many pre-programmed features like CD&R algorithms, a GUI and data logging. Furthermore, by using BlueSky, results will be obtained in a way that is easily verifiable, reproducible and can be extended upon in future research.

2) Airspace Parameters: As mentioned previously, this work builds upon the recommendation of the Metropolis project [4] to look further into the 'Layered' airspace concept, which has the airspace segmented into vertically stacked bands, where each altitude layer limits the horizontal travel to within an allowed heading range. To create such an airspace for experiments, the three-dimensional bounds of an airspace volume are defined in BlueSky. For the horizontal plane, a square airspace was selected, with sides of 1.8 Nm in length. The total experiment area then covers 1.8 x 1.8 = 3.24 Nm^2. The minimum altitude, alt_{min}, is set to $1100ft$, while the maximum altitude, alt_{max}, is set to $3500ft$. A total of eight vertical layers are defined that are distributed uniformly throughout the airspace and each have a height of $300ft$ each. Note that the layers are distributed uniformly and are fixed in terms of altitude, but the heading ranges are to be varied in this work. Figure 1 in section I contains a side view of what this looks like for a uniform airspace structure, where the heading ranges are of equal size.

It was chosen to simulate an airspace with unmanned (urban) aviation, as this poses several advantages in achieving the objective of the research. The first argument is the fact that future unmanned, urban aviation is expected to have higher traffic densities than manned aviation. The proof of concept for dynamically reconfiguring airspaces with RL techniques will be stronger if the experiments are set in a setting with very high traffic densities. The second argument for an airspace with unmanned aviation is that this type of aviation generally employs more trivial routes than commercial manned aviation [11], reducing the simulation development time. Thirdly, the idea of dynamically changing the vertically stacked airspace to improve safety, is founded on the principle of moving the flight altitude of aircraft to improve safety. The sacrifice in energy efficiency is relatively larger for manned aircraft (where flight altitude is a more dominant factor in efficiency) than it would be for unmanned aircraft, which could lead to an earlier adoption of RL techniques to enable dynamic airspace reconfigurations for the latter.

For unmanned aviation, two airspace types have been investigated in previous research: 1: 'very low level' traffic and 2) 'above-building' traffic. The Metropolis project [4], for example, looked at 'above-building' traffic, while [6] looks at air traffic that operates in a grid-like pattern that you would find in a 'very low level' structure. In this research, it is desirable to be able to analyse the airspace structures selected by the RL agent. In a low-level airspace, which includes the definition of directional streets, the aircraft may be forced away from the heading limits to not collide with a building. These are considerations that are not desirable to include in the results, as they will contaminate the simulation outcomes with decisions that are not solely a function of the airspace structure. Because of this, it was chosen to go with the more trivial simulation environment with traffic that is 'above-building'. Furthermore, this eases the definition of routes and enables them to be linear, greatly enhancing the feasibility of the research.

3) Aircraft type: For the experiments, it was chosen to simulate an airspace with a large amount of 'light-load' drones. This is a type of drone that is expected to be notably present in the skies of the future. They will likely be used for medical or lightweight industrial deliveries and for the completion of more traditional forms of delivering parcels of couriers to businesses and consumers [1]. For the sake of simplicity, the drones in the experiment will all be of the same type and will therefore have the same performance specifications. It was chosen to go with with the type of drone called 'DJI Mavic Pro', as its specifications resemble what could be a prominent aircraft type in our future airspace [1], as well as its availability within the BlueSky aircraft performance libraries.

4) Traffic scenarios: For the traffic in the scenarios, it was chosen to only consider the cruise phase of the flight and not to consider the take-off and landing operations. Therefore, each aircraft in the simulation is initialised on one of the four edges of the scenario. The aircraft spawn locations are chosen at random on the edges of the 'Experiment area' and aircraft spawn at a fixed rate of 12 aircraft per minute. This results in a traffic density of 50 aircraft/nm^2. On which edge an aircraft is spawned on is determined by a set of probabilities for the edges, which is fixed per episode. For example, if the probabilities are [North, East, South, West] = [0.85, 0.05, 0.05, 0.05], around 85% of the traffic gets initiated from the northern edge for that episode. On the other hand, if a [North, East, South, West] = [0.25, 0.25, 0.25, 0.25] setting was selected, the traffic in the scenario will be approximately uniform. For experiment I, the values of this [1 x 4] array are set at random for every simulated scenario.

At initialisation, the aircraft is given a random angle between 77.5° - 112.5° degrees from the edge. An illustrative example of an aircraft route is given in figure 8. The altitude at which the aircraft is spawned corresponds to its heading, in order to ensure that all aircraft are within the correct layer upon initialisation. The linear aircraft routes have three way-points, shown as green dots in figure 8 and an exit point, shown in red in the same figure, to guide the aircraft. The exit points naturally follow from the initial spawn location and heading, while the waypoints are added such that the aircraft will stick close to its intended route, even if the aircraft deviates from this to resolve a conflict. The climbing and descending happens in accordance with the specifications of the aircraft type used and are almost vertical. The speed at which the aircraft fly along the routes in experiment I is set to 4 *kts*.

5) Conflict detection: Both the term 'conflict' as well as 'LoSs' have been used extensively in previous research. For unmanned aviation, currently no standards are in place that define these. Furthermore, what is considered a safe separation distance is in reality a function of the (also currently unknown) traffic density and position uncertainty. For the detection of conflicts and LoSs in this

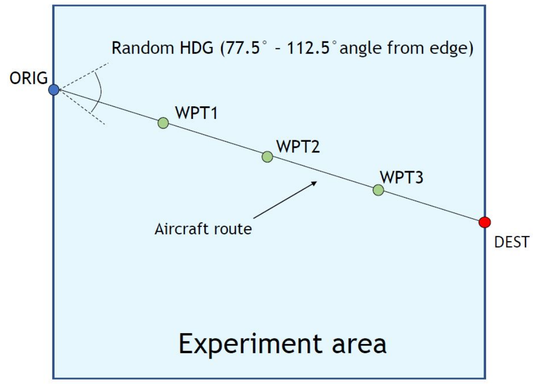


Fig. 8: Illustration of the experiment area, including an example of a flight route.

experiment, a minimum horizontal separation of 100m was selected. For the vertical separation, one airspace layer ($300 ft$) is taken. By default, aircraft cruise in the middle of their assigned layer, which means that aircraft cruising in adjacent layers will not be in conflict.

For the detection of conflicts, the experiment employs a state-based conflict detection method. This is a widely used method in the field [5]. It assumes linear propagation of the current state of the aircraft. The time to closest point of approach (CPA), t_{CPA}, is computed with equation 2:

$$t_{CPA} = -\frac{\vec{d}_{rel} \cdot \vec{v}_{rel}}{\vec{v}_{rel}} \tag{2}$$

where $\vec{d}_{rel}$ is the Cartesian distance vector between the aircraft (given in metres), and $\vec{v}_{rel}$ the vector difference between the velocity vectors of the involved aircraft (in metres per second). With this t_{CPA} known, the distance to CPA, d_{CPA}, can be computed by means of equation 3.

$$d_{CPA} = \sqrt{\vec{d}^2_{rel} - t^2_{CPA} \cdot \vec{v}^2_{rel}} \tag{3}$$

When this d_{CPA} is smaller than the radius of the pre-defined 'protected zone', R_{PZ}, within a set 'look-ahead time', $t_{lookahead}$, a horizontal conflict occurs. One may compute a time-interval during which horizontal separation will be lost. This is done through equation 4.

$$t_{in,hor}, t_{out,hor} = t_{CPA} \pm \frac{\sqrt{R^2_{PZ} - d^2_{CPA}}}{\vec{v}_{rel}} \tag{4}$$

The vertical time to conflict can be computed with equation 5 and 6:

$$t_{1,2} = \frac{-\Delta h \pm \frac{1}{2} h_{PZ}}{VS_{rel}} \tag{5}$$

$$\begin{aligned} t_{\text{in, ver}} &= \min\left(t_1, t_2\right) \\ t_{\text{out, ver}} &= \max\left(t_1, t_2\right) \end{aligned} \tag{6}$$

where Δh is the relative vertical distance, h_{PZ} is the height of the protected zone and VS_{rel} is the relative vertical speed. A conflict only occurs when the horizontal and vertical intervals overlap. The final times of entry and departure from the protected zone are as shown in equation 7:

$$\begin{aligned} t_{in} &= \max\left(t_{\text{in,hor}}, t_{in,ver}\right) \\ t_{\text{out}} &= \min\left(t_{\text{out,hor}}, t_{\text{out,ver}}\right) \end{aligned} \tag{7}$$

There is then a conflict if the conditions in equation 8 are met:

$$t_{\text{in}} < t_{\text{out}} \wedge t_{\text{out}} > 0 \wedge t_{\text{in}} < t_{\text{lookahead}} \tag{8}$$

For this work, a look-ahead time of 30s is implemented, as was done for [6] that also used unmanned aviation in a setting where conflict detection and resolution was implemented.

6) Conflict resolution: The resolution algorithm used is the MVP algorithm, which has proved effective in reducing the effect of resolution manoeuvres on flight efficiency while still guaranteeing minimal LoSs [9]. The geometric resolution corresponding to the MVP is shown in figure 9.

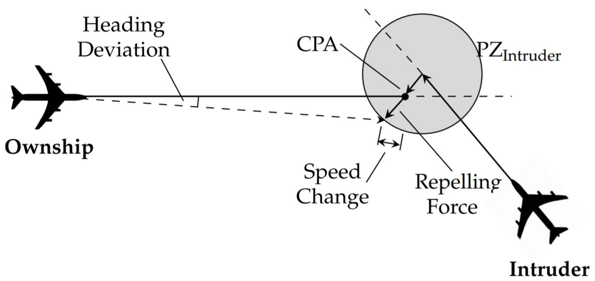


Fig. 9: MVP resolution [12]

In a layered airspace design, it could occur that the conflict resolution manoeuvre proposed by the MVP implies a violation of the heading range corresponding to the layer that the conflicting aircraft are currently in. When such a case occurs, this violation is accepted under the assumption that this option is more effective at reducing the total number of conflicts and LoSs than moving aircraft towards the vertical layers where the advisory heading is allowed. This would create vertical conflicts as aircraft traverse through vertical layers. Based on previous research, it is furthermore expected that the choice to violate the heading ranges of the current airspace structure for conflict resolution is more effective than limiting the output of the MVP algorithm [13].

C. Variables in experiment I

The independent variable in the experiment is the manner of moving traffic (method 1 and 2A-2D). The dependent variables related to safety are the reconfiguration time, vertical conflicts, vertical LoSs, cruising conflicts, and total conflicts and Loss. The time in conflict in conflict is also measured. This is defined as the time aircraft spend following the resolution advisory instead of their nominal path. After aircraft are no longer in conflict, they must redirect their course to the next active way-point. This, however, does count not count for the time in conflict. For efficiency analysis, flight times and flight distances (3D) are the dependent variables.

D. Hypotheses for experiment I

For experiment I, it is expected that the reconfiguration time is higher for options 2A-2D, due to the layer movements being sequential. It is furthermore hypothesised that fewer vertical conflicts occur for options 2A-2D (the layer-by-layer ones) than for the option 1 (all aircraft instantaneously). Fewer aircraft move vertically at any one time, which reduces the local traffic density of the moving aircraft and, with that, the probability of secondary conflicts. Lastly, it is hypothesised that more cruising conflicts occur with option 2C (starting with low traffic density layers), as aircraft will be cruising in layers where their headings are no longer allowed for a longer duration. This may temporarily decrease the alignment and segmentation effects in these layers.

VI. Experiment I - Results

In section V-A, five manners of moving traffic to investigate were introduced (options 1, 2A-2D). In this section, the outcomes of the dependent variables are given. They are shown in figures 10 through 17.

A. Reconfiguration time

Figure 10 shows the recorded reconfiguration times for each of the options. It reveals a shorter reconfiguration time for manner of moving 1 when compared to 2A-2D. This is expected, as the sequential manner of moving (as implemented in 2A-2D) has aircraft wait until the aircraft of a previous layer have arrived in their target layer. Simply moving all aircraft at once (option 1), is faster than doing it layer-by-layer (option 2).

B. Safety Analysis

In figure 11, the recorded vertical conflicts during experiment I are shown. It shows no observable difference in the recorded vertical conflicts. This is not in line with

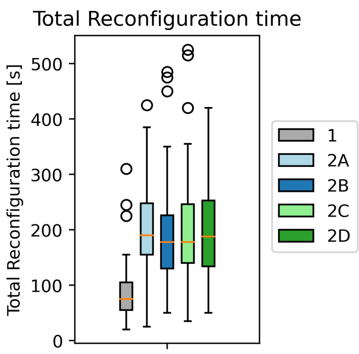


Fig. 10: Reconfiguration times for experiment I

the initial hypothesis. For a better understanding, figure 12 shows the mean vertical conflicts versus time (after reconfiguration).

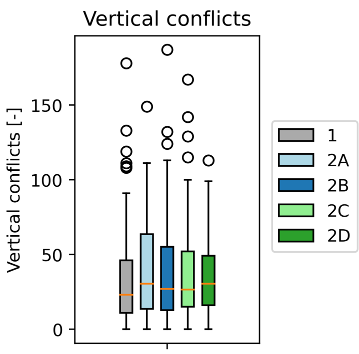


Fig. 11: Vertical conflicts for experiment I

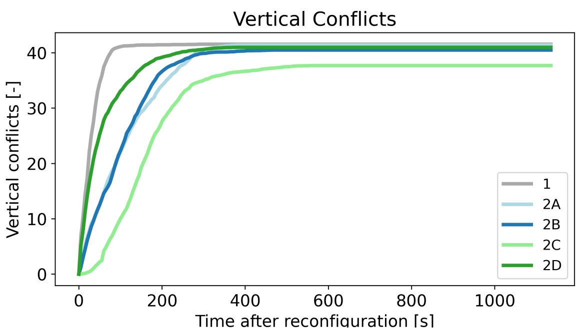


Fig. 12: Mean vertical conflicts vs. time after reconfiguration for experiment I

For figure 12, the slopes of the five curves in the initial timestamps after reconfiguration are as expected. There is a quick increase in vertical conflicts for the methods 1 and 2D, which move all aircraft and the aircraft in the highest density layer, respectively, just after reconfiguration. The more aircraft move at once, the higher the vertical conflict rate and the steeper the slope. The most shallow slope just after reconfiguration is seen for option 2C, which is due to this method moving the low traffic density layers first. This postpones the higher vertical conflict rate to the later timestamps. With their random number of moving aircraft just after reconfiguration, options 2A and 2B have an initial rate that falls in between options 1 and 2D, on the one hand, and 2C, on the other hand.

Figure 12 further reflects the similar number of vertical conflicts, also found earlier in figure 11, across the different options. For the experimental conditions at hand, the final number of vertical conflicts is not affected much by the number of aircraft moving at once. With the constant traffic density aircraft in each of the non-moving layers, there is little difference between enduring the vertical conflicts in a short time-frame or spread out over a longer duration. In higher traffic density scenarios, however, this may not hold. Moving more aircraft at once could result in more secondary conflicts.

Figure 12 also shows that option 2C finishes with fewer vertical conflicts than the other methods. This is due to the higher traffic density layers being moved only after a longer time, while some aircraft in these layers have already finished their flight. This happens for the other sequential reconfiguration options as well, but the effect is less visible as these start with higher traffic density layers than 2C. The situation of aircraft finishing their flight before reconfiguration finalises is not trivial to avoid in its entirety. Starting from the moment of reconfiguration, newly spawned aircraft are given spawn altitudes according to the new structure (or else a reconfiguration would be never-ending). A long reconfiguration time, however, 'dilutes' what was previously a layer of high traffic density, as aircraft leave the experiment area continuously. One could opt for a slower cruising speed or longer flights to reduce this effect. However, limited flight time must also be taken into account in real-life scenarios. Because of this, it was chosen to continue with the settings of the experiment at hand.

Figure 13 shows the 'time in conflict' for each of the five variants. Similarly to the vertical conflicts shown previously, the differences between the methods are negligible for the 'time in conflict'.

In figure 14, the mean vertical LoSs versus time after reconfiguration are displayed. It shows that the final values for the mean vertical LoSs are in line with the vertical conflict rates after reconfiguration (figure 12). In general, the higher the vertical conflict rate is (thus the steeper the increase in vertical conflicts), the more LoSs are seen in figure 14. This is expected, as with more aircraft moving at once, the local traffic density increases and MVP will be able to resolve fewer conflicts. Option 2C does not result in as many vertical LoSs as its counterpart 2D (high to low traffic density layer movement), because of the previously explained dilution of high traffic density layers.

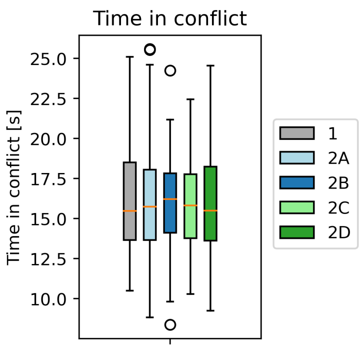


Fig. 13: Time in conflict for experiment I

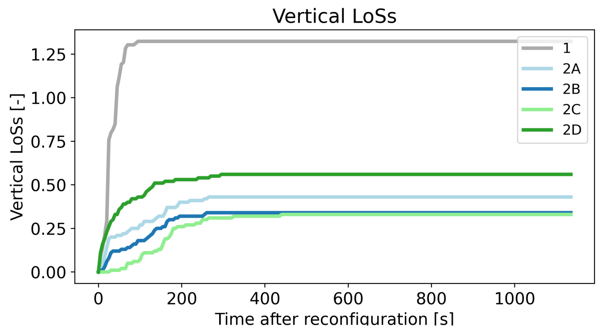


Fig. 14: Mean vertical LoSs vs. time after reconfiguration for experiment I

Figure 15 shows the mean cruising conflicts versus time after reconfiguration. It is seen that most cruising conflicts are found for option 2C. This is the direct effect of starting the sequential reconfiguration with the lowest traffic density layers. More than other options, 2C has aircraft fly around in the previous airspace structure for a long duration, increasing the cruising conflict rate and causing a steeper slope in the figure.

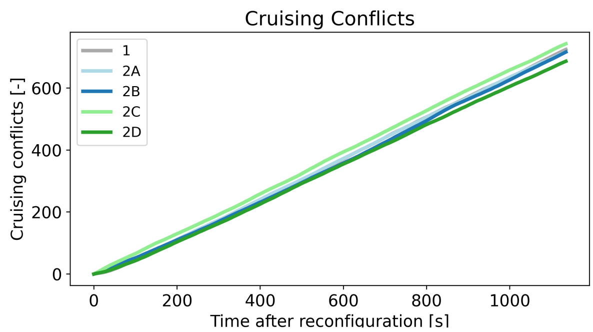


Fig. 15: Mean cruising conflicts vs. time after reconfiguration for experiment I

Figure 16 shows the total conflicts and LoSs for experiment I. It shows little noticeable difference between the total number of conflicts and LoSs for each of the different manners of moving. The null-hypothesis is setup that option 1 does not cause significantly more total conflicts and LoSs than options 2A-2D. Tables II and III show the p-values corresponding to this hypothesis.

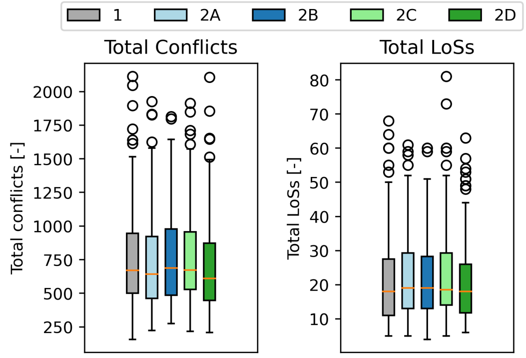


Fig. 16: Conflicts and LoSs for experiment I

TABLE II: P-values for total conflicts

Manners of moving traffic	1	2A	2B	2C	2D
1	1	0.85	0.98	0.81	0.47

TABLE III: P-values for total LoSs

Manners of moving traffic	1	2A	2B	2C	2D
1	1	0.98	0.94	0.66	0.69

As all p-values in tables II and III are larger than 0.05, any differences found are not statistically significant, indicating strong evidence for the null hypothesis. It is concluded that, safety-wise, and for the experiment conditions at hand, option 1 (moving all aircraft at once) can be used as the manner of moving traffic between vertical layers.

C. Efficiency Analysis

Figure 17 shows that no significant efficiency difference for the five methods of moving traffic into new airspace structures in terms of flight times and flight distances (3D). This is in line with the results of the safety analysis, that demonstrated no substantial difference in total conflict count or time in conflict. Similar to the safety metrics, the null-hypothesis is setup that option 1 does not cause significantly longer flight times and flight distances (3D) than options 2A-2D. Tables IV and V show the p-values corresponding to the this hypothesis.

TABLE IV: P-values for flight times

Manners of moving traffic	1	2A	2B	2C	2D
1	1	0.73	0.85	0.34	0.74

TABLE V: P-values for flight distances (3D)

Manners of moving traffic	1	2A	2B	2C	2D
1	1	0.37	0.34	0.07	0.24

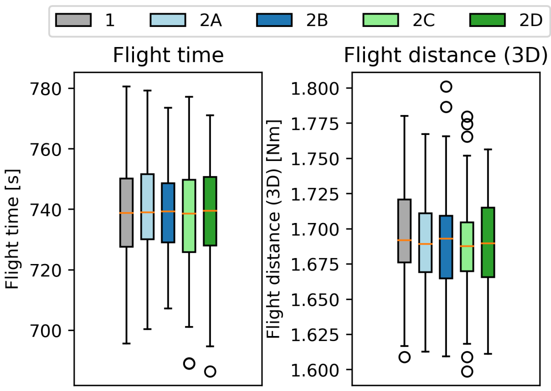


Fig. 17: Flight time and flight distance for experiment I

Similar to the safety metrics, all p-values in tables IV and V are larger than 0.05. This implies any differences found are not statistically significant, again indicating strong evidence for the null hypothesis. It is concluded that, for efficiency, in the experiment conditions at hand, option 1 (moving all aircraft at once) can be used as the manner of moving traffic between vertical layers.

D. *Relevance of results of experiment I*

As shown in subsections VI-B (safety analysis) and VI-C (efficiency analysis), no statistically significant differences were found between option 1 and the other manners of moving traffic. It was chosen to continue with option 1 for experiment III. This means that all aircraft will move to their new target layer at once during an airspace reconfiguration.

Initially, it was expected that moving all aircraft to a new layer configuration at the same time would create conflict 'hotspots' that considerably increase the number of LoSs. This was seen to some extent (figure 14), but not to the point that it greatly affected the overall safety. Moving all aircraft at once decreases safety during reconfiguration, but also moves aircraft to the new structure fast. The latter improves the safety during the cruising phase. As this phase is significantly longer than the reconfiguration phase, this has a positive effect on overall safety.

VII. Experiment II - Static RL Agent

In experiment II, an RL agent is developed to select airspace structures by looking at the traffic scenario. This experiment is performed with the objective to be able to compare the learning behaviour of this agent with those of experiment III, where vertically deviating aircraft are added. Subsections VII-A through VII-E elaborate on key elements of what we will name the 'Static RL' agent, such as the agent, learning algorithm and the state, action and reward formulations. The experimental setup for experiment II is the same as introduced previously in section V-B.

A. *Agent*

In this experiment, the agent has the objective of setting an airspace structure that is suitable for the expected traffic scenario. The agent is provided with knowledge on its environment, which in this case includes information on the headings of the aircraft, the airspace structuring and how the aircraft are currently spread over the layers. In a real-life application, the agent could be seen as the operator of the airspace and responsible for the defining the airspace structure.

B. *Learning Algorithm*

The learning algorithm used for this experiment is the soft-actor critic (SAC) algorithm. In this relatively new (2019) off-policy actor-critic algorithm 'the actor aims to simultaneously maximize expected return and entropy; that is, to succeed at the task while acting as randomly as possible' [14]. In this algorithm, the exploration/exploitation trade-off is such that the agent is explicitly pushed towards the exploration of new policies, while at the same time avoiding being stuck in sub-optimal behaviour. In general, an RL algorithm such as this one consists of an agent (see subsection VII-A) that interacts with its environment in discrete timesteps. It has the goal to learn a policy that maximises the sum of rewards, r_t, that is given to one or more action(s).

In an actor-critic architecture, there are two neural networks: one for the actor and one for the critic. The actor function, often named the policy, is usually written $\mu(s|\theta^{\mu})$ and specifies the output action a in regard to the input, the current state s of the environment in the direction proposed by the critic. The critic, on the other hand, is often denoted by $Q(s, a|\theta^{a})$ and tries to estimate the expected sum of rewards given a state, s, an action, a and the current actor policy, μ. It is updated from the gradients obtained from a temporal difference error signal each time step. The output of the critic drives learning in the actor by means of gradient descent with respect to the (negative) reward and action.

The activation functions used are 'relu' functions in the two hidden layers and a 'tanh' function in the output layer. The output of the network defined in this manner results in output values in the [-1,1] range. To convert these to positive values, which are needed for the setting of the heading ranges per vertical layer, use is made of a 'softmax' function, which converts it to values to a [0,1] scale. This final conversion, however, happens outside of the learning algorithm and is only done prior to the updating of the structures in the simulation. By doing so, any adverse affects on learning that could occur by changing the activation functions of a proven activation functions are mitigated [15].

C. State

It is key to select the state such that it provides sufficient information about environment, without becoming so large that it causes excessive computational effort. Though a bigger state array may allow for a more complete representation of the environment, increasing the dimension of the state increases the number of possible states and state-action combinations. As the state space of the problem grows, so will the training time [16]. In the case of this experiment, it was chosen to include three 'parts' of information in the state: 1) information on the headings of the aircraft, 2) information regarding the current airspace structure and 3) information on the number of aircraft per layer. Each of these is briefly discussed in the following.

1) Aircraft headings: For the heading information, the total aircraft heading range, 0° - 360°, is divided into 10 bins of equal size. The aircraft are then divided over these bins by their instantaneous heading to compute this part of the state array at a given time. The resulting array is then normalised before proceeding. This way, the value of each heading bin will be the fraction of the total number of aircraft that have a heading that falls within the heading range for a particular bin. A graphical representation of this is given below in figure 18, where h1, h2, ..., h10 represent the normalised number of aircraft of that bin. For example, the value of h1 is the number of aircraft with headings in the 0° - 36° range divided by the number of aircraft at that time, h2 is the number of aircraft that have headings from 36° - 72° divided by the number of aircraft.

h_1	h_2	h_3	h_4	h_5	h_6	h_7	h_8	h_9	h_{10}

Fig. 18: Aircraft heading information for the state formulation of the static RL agent. Each of the values h1, h2 ... h10 represent the (normalised) number of aircraft that have a heading that is within the heading range corresponding to that bin.

2) Current airspace structure: The information about the airspace structure is given in the form of an array of size [1x8]. Each of the values represents the portion of the full 360° heading range that the layer covers. A graphical representation of this formulation is given below in figure 19, where f1, f2, ..., f8 stand for the fractions of the full heading range. For example, if $f1 = 0.05$, the first layer will allow aircraft with headings 0° - 18°. Then, if $f2 = 0.15$, the second layer will allow aircraft with headings 18° - 72°, and so on, until the complete 360° heading range is covered.

f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8

Fig. 19: Representation of the information on the airspace structure. Each of the values f1, f2, ..., f8 (decimal, 0-1) represents the fraction of the total heading range that the layer covers.

3) Aircraft per layer: For the number of aircraft per layer, similarly to the current airspace structure, an array of size [1x8] is used. Each of the values in this array represents the fraction of the total number aircraft currently in that layer, see figure 20.

n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8

Fig. 20: Representation of the information on the number of aircraft per layer. Each of the values n1, n2, ..., n8 (decimal, 0-1) represents the fraction of the total aircraft that are located in that layer.

4) Final state array: To formulate the final state array, the three 'parts' of information introduced in sections VII-C1 through VII-C3 are combined to form one array of size [1x26].

The objective of experiment II, to develop an RL agent to set the airspace structure based on traffic, does not strictly require the information of parts 2) the airspace structure and 3) the aircraft per layer. These were, however, added in this experiment to enable future comparison between the decisions of this agent and the ones of experiment III, while having the same state formulation.

D. Action

An action is selected each time the state is given to the agent. The final output is a one-dimensional action array, such as shown to be part of the state-array in figure 19. It may be noted that the choice for eight layers was made in this research, but that this variable is in reality dependent on the environment. With fewer combinations of heading ranges to chose from, the resulting fewer altitude layers would likely lead to lower training times. However, with more layers the aircraft will be more dispersed throughout the airspace, possibly resulting in more optimal results in terms of the total number of conflicts/LoSs experienced.

E. Reward

A reward based on the total number of conflicts was chosen for experiment II. See equation 9 for the formulation.

$$\text{Reward} = -\frac{\text{Total number of conflicts}}{100} \tag{9}$$

The division by 100 was implemented upon finding better results during a number of initial training runs. Recent research has shown that reward re-scaling can indeed improve the stability of a RL agent [17].

F. Episode overview for experiment II

Figure 21 shows how the above states, actions and rewards are combined into the episodes of experiment II. It shows that the state of the environment is observed after 16 minutes. This length of time is line with the time it takes to achieve a constant traffic density for the selected experimental conditions. At this point in the episode, this state is passed to the RL agent, upon which an action to set the airspace structure is obtained. In this experiment, the logging of safety parameters (and thus also the conflicts for the reward) starts at 20 minutes. Note that this is after the traffic has settled into the new structure and that vertical conflicts are thus not yet taken into account for experiment II. This is because experiment II had the aim of training a 'static RL agent', which only looks at the traffic situation and selects a suitable structure for minimising the total number of conflicts and LoSs. The extension to a dynamic agent, which also takes into account the previous airspace structure (and the vertical deviations to get there) is made in experiment III.

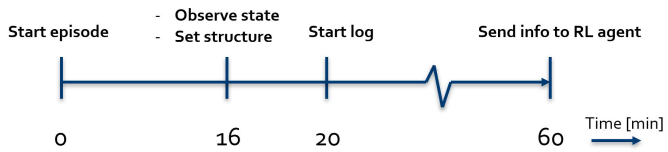


Fig. 21: Overview of episodes for experiment II

The settings in terms of airspace parameters, aircraft types and CD&R are the same as in experiment I. The simulation timestep, however, is updated from $dt = 0.05$s in experiment I to $dt = 5$s in experiment II. The reduction in the ability to resolve short-term conflicts that will be a consequence of this, is accepted in favour of a significant increase in simulation speed and the reduction in training time.

G. Benchmarking the performance of the RL agent

After training, the final agent is presented with 1000 random traffic scenarios to see how it performs at the task of reducing the total conflict and loss counts through suitable airspace structure selection. To be able to put its performance into perspective, the agent is compared to two benchmarks; a fixed uniform structure and an analytical method for selecting the airspace structure. To enable a valid comparison between these two alternative methods and the developed RL agent, they are also presented with the same set of 1000 random scenarios that the RL agent is tested with. For both the training and testing episodes, the traffic distribution is different for each episode, but within each of the scenarios the traffic distribution (thus the probability of an aircraft spawning on an edge) is fixed.

For the uniform structure benchmark, each of the 8 vertical layers covers exactly 45°, uniformly dividing the total 360° heading range.

For the analytical method, the airspace structure is selected that is most probable to lead to perfect segmentation of aircraft during the traffic scenario. To compute these structures, use is made of the information on the aircraft headings that are sampled for the state at $t = 16$ minutes (see figure 21). The 'analytical structure' is found by determining the structure that would divide the sampled aircraft evenly over the eight layers. Given a sample size of around 150 aircraft when the traffic has settled in the selected experiment conditions, it may reasonably be assumed that this structure will fit the traffic for the duration of the traffic scenario. The analytical method makes use of the same information as the RL agent to select structures, enabling a direct comparison of the performance of analytical and RL agents at the task of selecting airspace structures.

It must be mentioned that perfect segmentation is not necessarily optimal safety-wise. As mentioned in previous research, 'a wide heading range in a layer results in considerable heading differences between aircraft travelling in the same layer, leading to intercepting routes and large conflict angles. Adding more aircraft to a layer with a wide heading range comes at a higher cost than adding aircraft in a layer with a smaller range' [5].

H. Variables in experiment II

The independent variable in the experiment is the (manner of selecting) airspace structure: uniform, by means of RL agent or by means of an analytical method. The dependent variables in terms of safety are the total conflicts and total LoSs and the time in conflict. For the efficiency, the flight time and flight distance (3D) are investigated.

I. Hypotheses for experiment II

For experiment II, it is hypothesised that the RL agent will be able to outperform uniform structures when it comes to minimising the total number of conflicts and LoSs. It is thought that uniform structures will not 'fit' many of the non-uniform traffic scenarios that it will be presented with during training very well. Such non-uniform traffic scenarios will lead to conflict 'hotspots' in some of the layers, reducing any benefit of the separation and alignment effects on the total number of conflicts and LoSs.

It is furthermore hypothesised that the RL agent is not able to outperform an analytical method that perfectly segments the traffic in the current scenario. The advantage of RL methods is their ability to find optimal actions in environments where the state is affected affected by multiple variables. This extends to cases where the number of variables is far higher than a human can consider. In this particular experiment, however, a relatively simple environment is used. Human made rules, such as going for perfect segmentation, can be sufficient for this. In theory, RL is expected to be able to achieve the same performance as such rules. However, in practice, more training time and a greater variability of training scenarios than are are achievable in this work might be needed. The real question in this work is whether RL can outperform human made methods in experiment III, where both the vertical conflicts (due to reconfiguration) and the cruising conflicts in the new structure are under consideration.

VIII. Experiment II - Results

The results of experiment II are presented in this section. Firstly, subsection VIII-A shows the results of the training phase of the static RL agent. Thereafter, in subsection VIII-B, a safety analysis is provided.

A. Training results

In figure 22, the reward evolution over 45000 training episodes is displayed.

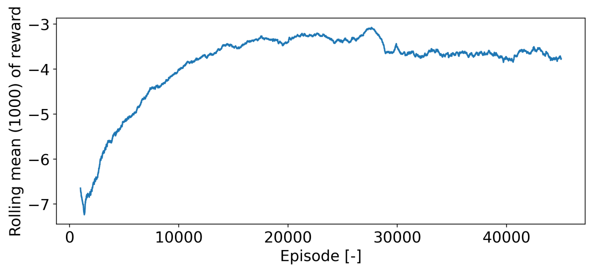


Fig. 22: Rolling mean (1000) of reward during training of the static RL agent for 45000 episodes.

Figure 22 shows that the RL agent is able to select actions that increase its reward over time. The training curve is similar those shown in the paper accompanying the release of the SAC algorithm [14], indicating that it has, in all likelihood, been successfully applied to the problem of selecting airspace structures.

B. Testing - Safety analysis

Figure 23 shows the total number of conflicts and LoSs for the 1000 scenarios respectively for the RL method and the two benchmarks.

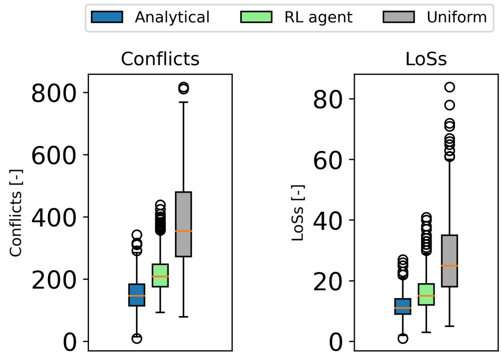


Fig. 23: Total conflicts and LoSs for various (methods of selecting) airspace structures

As seen in figure 23, most conflicts and LoSs occur when uniform structures are used for all 1000 testing scenarios. This is as hypothesised earlier in section VII-I, where it was stated that the non-uniformity of the testing scenarios and the uniform structure likely leads to a relatively large number of conflicts and LoSs. This result is also in line with findings from previous research [5], [6]. The RL agent is able to reduce the total number of conflicts and LoSs with respect to this first benchmark.

Figure 24 shows the time in conflict for each of the methods.

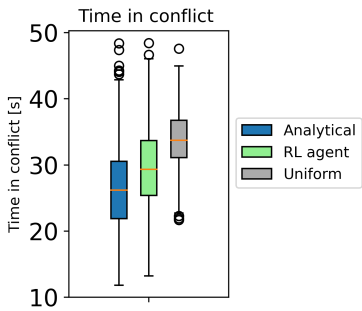


Fig. 24: Time in conflict for various (methods of selecting) airspace structures

Figure 24 shows similar results to the total conflicts and total LoSs: the RL agent reduces the time in conflict with respect to uniform structures, but is outperformed by the analytical method. Thus, in terms of safety, the analytical method performs best in all the presented metrics. These findings are due to the relative simplicity of the task at hand. Because the reward formulation is based on conflicts, and the logging of these only starts after the aircraft have settled into their new airspace structure (see subsection VII-F), it is fairly trivial to come up with this analytical method that yields results that already strongly reduces the total number of conflicts and LoSs, as well as decreases the time aircraft spend in conflict.

C. Testing - Efficiency analysis

The focus is now shifted towards the efficiency analysis. Figure 25 shows the results in terms of flight times and flight lengths, respectively.

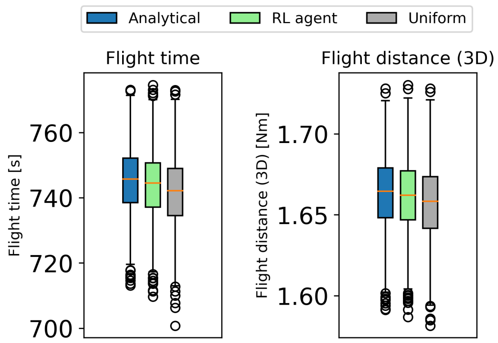


Fig. 25: Flight time and flight distance (methods of selecting) airspace structures

By comparing figure 25 to figure 23, it is seen that the analytical method of selecting airspace structures (which scored the best safety values), has slightly higher values for the flight times and efficiencies. This could be due to slightly different nature of the conflicts experienced in the simulations. It is thought that more conflicts (as experienced for the uniform structures) in combination with the MVP algorithm could lead to a wave-like patterns in the experiment area. This could, locally, lead to more space to resolve conflicts locally, and in turn a shorter flight time and flight distance, respectively.

IX. Experiment III - Dynamic RL Agent

In experiment II it was shown that an RL method can outperform uniform structures by looking at the traffic situation and selecting an airspace structure for the traffic scenario. However, it was also seen that analytical methods can still do a better job at minimising the total number of conflicts and LoSs for situations where only the cruising conflicts are to be minimised. In experiment III, the complexity of the problem is increased by including the vertical conflicts (and LoSs) in the total conflict counts of the episodes. An RL agent must then not only select airspace structures that fit the cruising traffic well, but must now also take into account the previous airspace structure and the associated vertical deviations to move from that structure to the next. Subsection IX-A gives an outline of the experiments, after which subsection IX-C gives the hypotheses.

A. Outline of experiment III

For experiment III, two different RL agents are trained. The traffic scenarios will still have a single airspace reconfiguration, yet, by changing the logtime after this reconfiguration, one can simulate the effects of different airspace structure reconfiguration rates. For the first agent, which will be named 'RL 1', a 'slow' reconfiguration rate is simulated by having a relatively long (44 mins) logtime after airspace reconfiguration. This agent has to deal with what is expected to be a problem of relatively low complexity, as the ratio of cruising conflicts to vertical conflicts will be high, given the long logtime. This makes for a situation where a suitable structure is likely one that is optimised for the cruising traffic. The other agent, 'RL 2' simulates dynamic airspace reconfiguration with a 'fast' reconfiguration rate by having a shorter logtime (8 minutes). This problem is thought to be a problem of relatively higher complexity, as the ratio of vertical to cruising conflicts is less trivial. This implies that the agent must, in some cases, find a balance between minimising the vertical conflicts and the cruising conflicts. The above discussion is summarised in table VI.

TABLE VI: Overview of RL agents for experiment III

Name	Simulating	Reconfig. rate	Complexity
RL 1	'Slow' reconfig. rate	44 min	Lower
RL 2	'Fast' reconfig. rate	8 min	Higher

Similarly to experiment II, the episode timelines for each of the agents are given in figures 26 and 27. Specifically note the 'Start log' events, which now happen at the moment of airspace reconfiguration. This is slightly different than was done for experiment II, where the log was started after all aircraft had settled into the new structure.

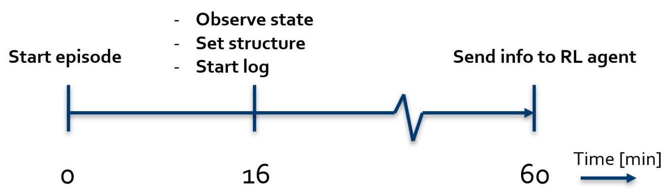


Fig. 26: Episodes for experiment III, agent RL 1

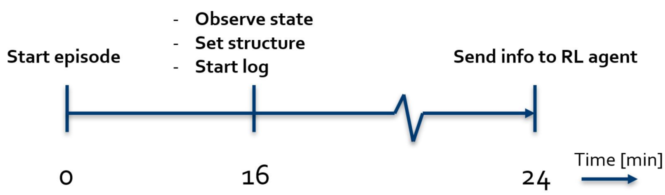


Fig. 27: Episodes for experiment III, agent RL 2

Also similar to experiment II, the agents will be evaluated by comparing their performance to uniform structures and the analytical method of selecting structures introduced earlier in section VIII-B. For this, the same

1000 testing scenarios introduced for experiment II will be used. Apart from only comparing the agents, the differences in structure selection by agents 'RL 1' (slow reconfiguration) and 'RL 2' (fast reconfiguration) are also investigated.

The settings in terms of airspace parameters, aircraft types and CD&R will be the same as experiment I and II. The simulation timestep is kept the same as in experiment II, at $dt = 5$s. The state, action and reward formulations for the RL agents are furthermore unchanged from experiment II.

B. Variables in experiment III

The independent variables in the experiment are the (manner of selecting) airspace structure and the reconfiguration rate. For the airspace structure (selection) there are 1) uniform, 2) by means of the RL agent and 3) by means of an analytical method. For the reconfiguration rate, both a 'slow' (44 minutes) and a 'fast' (8 minutes) version is used. The dependent variables are the total, cruising and vertical conflicts and total LoSs, as well as the time in conflict, flight time and flight distance (3D).

C. Hypotheses for experiment III

It is hypothesised that the results of the first agent, simulating 'slow' reconfiguration, are very similar to what was found in experiment II, which only considered the cruising phase. As the vertical conflicts in experiment III only cover a small fraction of the total conflicts for this episode length, it is expected that the RL agent outperforms uniform structures in terms of the total number of conflicts and LoSs, but will not match the performance of the analytical method aiming for perfect segmentation.

It is furthermore hypothesised that the RL2 agent will show considerably different results to what was previously seen for the RL agent of experiment II and the RL1 agent introduced in this experiment. The added complexity that is now introduced by the shorter episodes (representing 'fast' reconfiguration) is something which the analytical methods cannot grasp. Because of this it is thought that, this time around, the RL agent will outperform both the uniform structures and the analytical methods in terms reducing the total number of conflicts and LoSs.

Lastly, it is hypothesised that for the 'fast' agent, on some occasions, the selected structures will be tailored to minimising the vertical conflicts over the cruising conflicts. For identical traffic scenarios, this 'fast' agent should then minimise the number of vertical conflicts compared to the 'slow' agent. However, the structures selected by RL1 will likely result in fewer conflicts and LoSs during the cruising phase.

X. Experiment III - Results

The results of experiment III are presented in this section. Firstly, subsection X-A shows the results of the training phase of the dynamic RL agents. Thereafter, in subsections X-B and X-C, the safety and efficiency analyses of the testing results are provided, respectively.

A. Training results

In figure 28, the reward evolution for agent RL1 ('slow' reconfiguration) over 35000 training episodes is displayed. The same is done in figure 29 for the training of agent RL2 ('fast' reconfiguration).

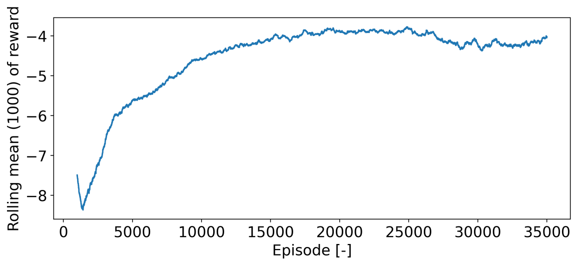


Fig. 28: Reward evolution during training of the first dynamic RL agent (RL1) during 35000 episodes.

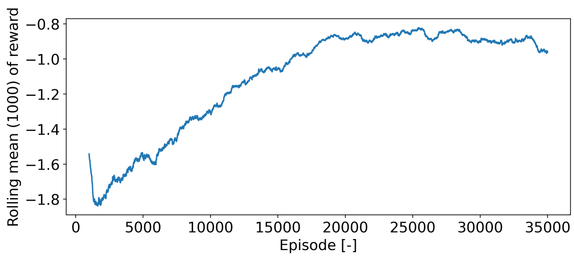


Fig. 29: Reward evolution during training of the second dynamic RL agent (RL2) during 35000 episodes.

The reward evolutions during training, shown in figures 28 and 29, look as expected. It can be seen that agent RL1 has a higher initial learning rate than agent RL2. This is due to the lower relative complexity of the task. As mentioned previously in section IX-C, the task of setting airspace structures for a 'slow' reconfiguration rate is thought to be more trivial than doing so for 'fast' reconfiguration. Besides that, the faster learning for RL1 is explained by the greater range of reward values. This is a consequence of the longer logging time, as it yields larger differences in rewards between suitable and unsuitable airspace structures.

The final values that the agent converges to during training are similar to the RL agent trained in experiment II (figure 22). Nevertheless, agent RL1 reaches an optimum around 20000 episodes, where this occurred after around 15000 episodes for the agent in experiment II. This shows that indeed considering dynamic structuring, where both cruising and vertical adjustment phases are considered, is a more complex task.

B. Testing - Safety analysis

The testing of the two agents is done in the same way as experiment II. Both are presented with the 1000 random traffic scenarios to see how they perform at the task of selecting suitable airspace structures. To put their performance into perspective, they are again compared to the two benchmarks used previously in experiment II: a fixed uniform structure and the analytical method of selecting the airspace structure (see section VII-G). Note that the analytical method does not take into account reducing the number of conflicts that occur during reconfiguration. This would result in a larger set of rules.

Figure 30 shows the vertical conflicts for RL1 and RL2. It reveals that both RL agents reduce the number of vertical conflicts with respect to the analytical method. It is also seen that the number of vertical conflicts for reconfigurations to the uniform structure is indeed zero. This is due to the initial structure being uniform as well. RL1 and RL2 show a similar number of vertical conflicts. It was initially hypothesised that RL2 would further reduce the number of vertical conflicts compared to RL1, as during its training, the reconfiguration phase is relatively longer (compared to the cruising phase). This suggests that the information available to RL2 might not have been enough for proper training. The reduced vertical conflicts for both agent RL1 and RL2 are further investigated by looking at the correlation between the starting airspace structure (uniform) and the selected structure. The results are shown in figure 31.

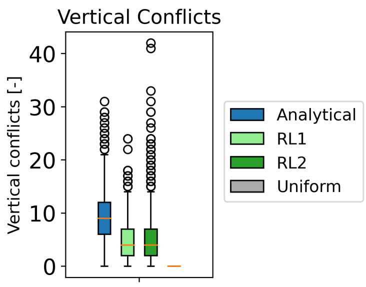


Fig. 30: Vertical conflicts experienced for agents RL1 and RL2, in relation to those from uniform or analytical structures.

Figure 31 shows that the least correlation between starting structure (uniform) and selected structure is found for the analytical method. Agents RL1 and RL2, on the other hand, show more correlation. As expected, the correlations found for the uniform - uniform reconfiguration are always 1. These results explain the earlier found

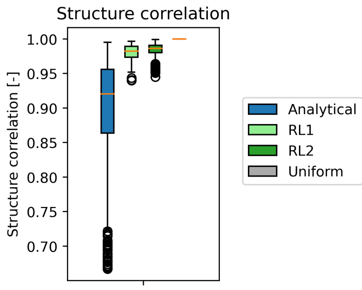


Fig. 31: Structure correlations between the starting structure (uniform) and the selected structures.

vertical conflicts (figure 30). The more that a selected structure is different from the starting structure, the more aircraft will have to perform vertical deviations to get into the new structure. As a direct consequence of this, more vertical conflicts are experienced.

Figures 32 and 33 show the total number of conflicts and LoSs, respectively. This is done for both the 'slow' reconfiguration setting (on the left-hand sides), as well as the 'fast' reconfiguration setting (on the right-hand sides).

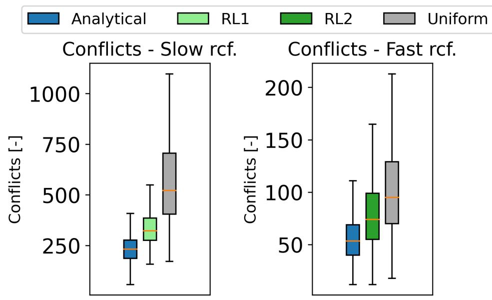


Fig. 32: Total conflicts for 'slow' and 'fast' reconfiguration settings

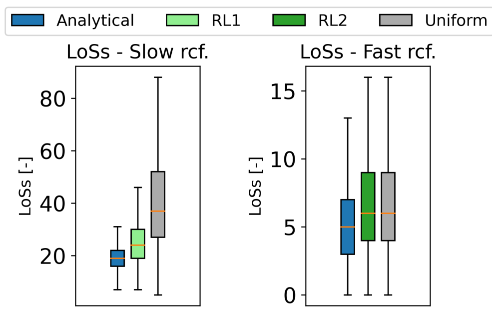


Fig. 33: Total LoSs for 'slow' and 'fast' reconfiguration settings

Figure 32 shows that the total number of conflicts is reduced most by using the analytical method for selecting airspace structures. This holds both for the 'slow' reconfiguration settings in which agent RL1 was trained, as well as for the 'fast' reconfiguration setting in which agent RL2 was trained. A similar pattern is seen in figure 33 for the LoSs, albeit with smaller values.

In general, the improvement that RL1 provides in terms of safety (with respect to uniform structures) is similar to what was found in experiment II (figure 23). Figure 33 shows that RL2 has final conflict and LoSs counts closer to the values obtained with uniform structures. This is a direct consequence of the short episode length for this RL agent. It suffers from the vertical conflicts (see figure 30) and LoSs during reconfiguration, plus the cruising conflicts in the limited 8 minutes of logging. The uniform structure has zero vertical conflicts and LoSs as no reconfiguration is required. For agent RL2, the available cruising time is too short to balance out the present vertical conflicts.

Figure 34 shows the time in conflict for the 'slow' and 'fast' reconfiguration times. The pattern is, as expected, in line with the results found for the total number of conflicts and LoSs.

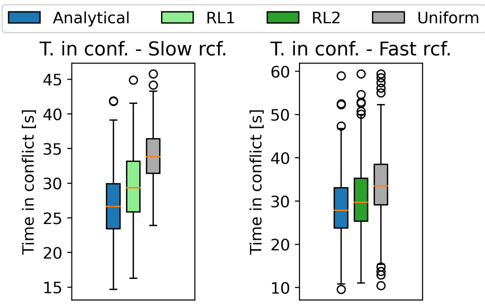


Fig. 34: Time in conflict for the 'slow' and 'fast' reconfiguration settings

From the findings in the safety analysis, it is deducted that, even in the presence of a number of vertical conflicts, the analytical method of selecting airspace structures outperforms the RL agent and uniform structures. Due to the way that the analytical method has been set-up (simply aiming for perfect segmentation), it is a method of selecting airspace structures that works well for a wide range of traffic scenarios (see figure 31). Even in 'extreme' traffic situations, where there is a strong dominant traffic direction, the analytical method is able to segment the aircraft and limit the number of total conflicts and LoSs. This is predominantly due to the experimental conditions, where cruising conflicts heavily outweigh vertical conflicts.

The RL agents, on the other hand, do not demonstrate this flexibility to the same extent. This prominently seen in figure 31, where the correlation between the structures selected by the RL agents and the starting (uniform) structure take on values over a narrower range than is the case for the analytical method. The effect of this is seen in the results of the vertical conflicts as well. Figure 30 showed reduced vertical conflicts for the RL agent structures in comparison to the analytical method.

At first glance, is could seem as though the RL agents have 'understood' the presence of vertical conflicts and learnt to minimise these to obtain a better rewards. This explanation, which comes down to the method getting stuck in a local optimum that reduces vertical conflicts, however, seems unlikely. Upon comparing the values of the vertical conflicts in figure 30 with the total conflicts in figure 32, it may be observed that the vertical conflicts are a very small fraction of the total conflicts. In the setting that RL1 was trained ('slow' reconfiguration), the total number of conflicts is around two orders of magnitude higher than the vertical conflicts, meaning that the conflicts in the scenarios were almost all of the 'cruising' type. The reduction of vertical conflicts for airspace structures selected by RL1 and RL2 can then hardly be seen as an optimum which the agents may strive for.

Rather, the reduced vertical conflicts are a bi-product of the RL methods' tendency to select structures that are similar to the uniform starting structure. As it happens, this also has a positive influence on the number of vertical conflicts, where fewer vertical deviations are needed to move aircraft into the new airspace structure. From this, fewer vertical conflicts result.

C. *Testing - Efficiency analysis*

Figures 35 and 36 show the total flight time and flight distances (3D), respectively.

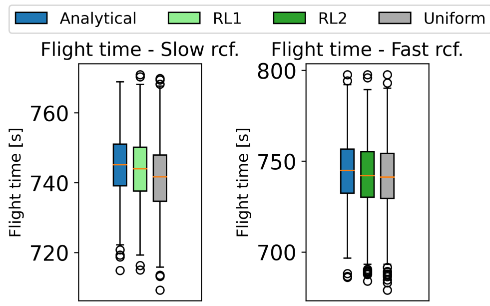


Fig. 35: Flight time and flight distance for various (methods of selecting) airspace structures

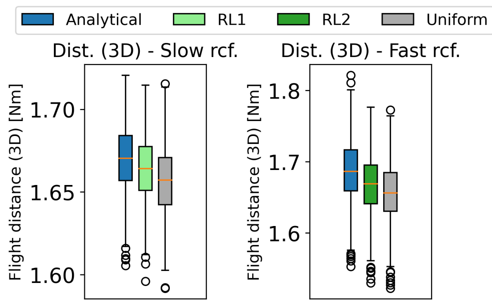


Fig. 36: Flight time and flight distance for various (methods of selecting) airspace structures

Figures 35 and 36 reveal no significant differences in flight times and flight distances (3D) for both the 'slow' reconfiguration rate and the 'fast' reconfiguration rate. The slight increase in flight distances (3D) for the analytical and RL methods, as compared with uniform structures, are a direct consequence of the vertical deviations that the aircraft make and different conflict resolution manoeuvres.

XI. Discussion

In this research, the objective has been to develop an RL agent that is capable of defining heading ranges per vertical layer, while taking into account the previous airspace structure. Such new structures must optimise the cruising phase for future traffic, while also reducing the vertical deviations that occur when adapting to this new structure. This is not trivial, as a suitable cruising structure could be one that leads to excessive vertical deviations, or vice versa: a structure that minimises the vertical deviations could be unsuitable for the future cruising traffic.

It was found that, for the selected experimental conditions, moving all aircraft at once during a reconfiguration does not significantly impact the total number of conflicts and LoSs. Furthermore, the work has shown that an RL agent can reduce the total number of conflicts and LoSs compared to uniform structures.

For dynamic airspace airspace reconfiguration, the results found in this work are greatly conditioned by the fact the number of conflicts/LoSs during the cruising phase are orders of magnitude higher than conflicts/LoSs during the reconfiguration phase. In this case, the challenge was relatively straightforward, as the RL agent simply needed to learn a policy that minimised the cruising conflicts in order to most effectively obtain higher rewards. From theory, it is known that when the task complexity increases, RL methods may start to outperform analytical ones. However, because of the relatively lower complexity in the current experimental conditions, it was found that an analytical method of selecting airspace structures outperforms the trained RL agent in terms of safety.

The fact that that analytical outperformed RL agents, raises the question what method should be used for dynamic airspace reconfiguration. An advantage of an analytical method is that it is strong over a wide range of traffic scenarios. Due to the way it is computed, there is no difference between presenting it with an (almost) uniform scenario and presenting it with an 'extreme' case, where aircraft fly predominantly in a certain direction. Furthermore, the analytical method has the advantage that it is made up of simple rules and therefore requires no training time. A disadvantage of analytical methods is their performance is expected to rapidly decline, or that the rules become hard to define, when the task complexity increases.

The main advantage of an RL agent is that it can perform at tasks that have a high complexity, such as cases where many variables play a role in the state of an environment. A disadvantage of an RL agent, is that it tries to learn a policy that suits a wide variety of states that it could be presented with. In experiment III, for example, RL methods demonstrated a more limited applicability to a diverse set of traffic scenarios. Difficulty in selecting very different structures from the starting (uniform) structure is observed from the correlation between the airspace structures before and after reconfiguration. One might say that it does not have the 'flexibility' that comes with the analytical method. Furthermore, an RL agent requires training time before it can start selecting structures, which is more of hurdle towards implementation than is the case for the analytical method. Nevertheless, this is a finding specific to these experimental conditions. An RL agent trained under different conditions, or with different state, action and reward formulations, could potentially achieve better results.

All in all, the decisions on what method should be used goes hand-in-hand with the nature of the problem. If the problem has a complexity which cannot be captured by analytical methods properly, any implementation will be limited in its performance. In more complex experimental settings, RL methods are expected to be more suitable. For dynamic airspace reconfiguration, the experimental conditions employed in this work yielded results in which analytical methods are sufficient. However, when the trade-off between minimising for vertical conflicts upon reconfiguration and cruising conflicts in the new structure becomes more more challenging, RL methods are expected to be more effective.

For future work, it is suggested that an RL agent is used for dynamic airspace reconfiguration in cases where the ratio of vertical/cruising conflicts is higher. The increases complexity that arises in such a setting makes better use of the previously mentioned advantages of RL methods. More vertical conflicts could occur in settings where the maximum vertical velocity of aircraft is lowered or differs per aircraft. This has aircraft spend relatively more time moving between layers. In turn, this could increase the vertical conflict rate and increase the ratio of vertical/cruising conflicts in the scenario. In creating these environments to train the RL agent, it must always be kept in mind that sound research adapts a tool (in this case RL techniques) to a given environment. Creating non-realistic environment to demonstrate the capabilities of a tool is, in most cases, of inferior value.

Besides a suitable environment for vertical conflicts, it is recommended that future research selects its manner of moving aircraft into a new airspace structure according to the experimental setup. For the research at hand, moving all aircraft at once did not significantly compromise the total conflicts or LoSs experienced during a reconfiguration. It was, however, found that the number of vertical conflicts and LoSs upon reconfiguring increases with the number of aircraft that are moved at once. This trend is an exponential one: as more conflicts occur, the flight path length increases, which in turn increases the chances of (secondary) conflicts. Because of this, a point is expected where sequential manners of moving traffic result in fewer total conflicts and LoSs than moving all aircraft at once during an airspace reconfiguration. Future research could also investigate other options of sequential movement than were explored in this work. For example, one could look at moving fractions of the traffic in a layer, rather than all the aircraft within a layer. One may also explore options that move traffic sequentially based on the free space for vertical deviations.

XII. Conclusion

This work has aimed at developing an RL agent capable of defining heading ranges per vertical layer, while taking into account the previous airspace structure. It showed that such an agent is capable of outperforming uniform structures in terms of safety metrics such as the total number of conflicts and LoSs. Analytical methods, on the other hand, showed better performance in the task of setting airspace structures in a relatively trivial environment, where little vertical deviations were present. In settings where the number of vertical deviations increases (and the complexity of the problem increases), however, the performance of analytical methods is expected to decline relative to an RL agent.

Future research should extend this work to different operational environments. Different performance limits or limitation of manoeuvres for conflict resolution will likely affect the number of conflicts/LoSs during the reconfiguration phase. Finally, it is of interest to see how RL agents behave in more complex environments, where the airspace reconfigurations result in a higher number of vertical conflicts. In these situations, where the development of human-made rules is not trivial, RL techniques could prove a valuable tool in improving the safety and efficiency of airspace structuring.

References

[1] SESAR-Joint-Undertaking, "European drones outlook study - unlocking the value for europe," 2016.

[2] J. Hoekstra, J. Maas, M. Tra, and E. Sunil, "How do layered airspace design parameters affect airspace capacity and safety?," *7th International Conference on Research in Air Transportation, Philadelphia, PA, USA*, 2016.

[3] SESAR-Joint-Undertaking, "U-space - supporting safe and secure operations in europe," 2020.

[4] E. Sunil, J. Hoekstra, J. Ellerbroek, F. Bussink, D. Nieuwenhuisen, A. Vidosavljevic, and S. Kern, "Metropolis: Relating airspace structure and capacity for extreme traffic densities," *11th USA/EUROPE Air Traffic Management R&D Seminar*, 2015.

[5] L. Caranti, M. Ribeiro, J. Ellerbroek, and J. Hoekstra, "Safety optimization of a layered airspace structure with supervised learning," *11th SESAR Innovation Days*, 2021.

[6] M. Ribeiro, J. Ellerbroek, and J. Hoekstra, "Using reinforcement learning to improve airspace structuring in an urban environment," *Aerospace*, vol. 9, no. 8, 2022.

[7] E. Sunil, *Analyzing and Modeling Capacity for Decentralized Air Traffic Control*. PhD thesis, 2019.

[8] E. Sunil, J. Ellerbroek, J. Hoekstra, and J. Maas, "Three-dimensional conflict count models for unstructured and layered airspace designs," *Transportation Research Part C Emerging Technologies*, 2018.

[9] M. Ribeiro, J. Ellerbroek, and J. Hoekstra, "Review of conflict resolution methods for manned and unmanned aviation," *9th SESAR Innovation Days, Athens, Greece*, 2020.

[10] J. Hoekstra and J. Ellerbroek, "Bluesky atc simulator project: an open data and open source approach," *International Conference for Research on Air Transportation, Philadelphia, PA, USA*, 2016.

[11] M. Doole, J. Ellerbroek, and J. Hoekstra, "Estimation of traffic density from drone-based delivery in very low level urban airspace," *Journal of Air Transport Management*, 2020.

[12] J. Hoekstra, R. van Gent, and R. Ruigrok, "Designing for safety: The 'free flight' air traffic management concept," *Reliability Engineering & System Safety, Volume 75, Issue 2, February 2002, Pages 215-232*, 2002.

[13] M. Giliam, "Incorporating geovectors into the process of conflict resolution for uav airspace," Master's thesis, Delft University of Technology, 2022.

[14] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, and S. Levine, "Soft actor-critic algorithms and applications," 2019.

[15] M. A. Mercioni and S. Holban, "The most used activation functions: Classic versus current," in *2020 International Conference on Development and Application Systems (DAS)*, pp. 141–145, 2020.

[16] G. Dulac-Arnold, R. Evans, H. van Hasselt, P. Sunehag, T. Lillicrap, J. Hunt, T. Mann, T. Weber, T. Degris, and B. Coppin, "Deep reinforcement learning in large discrete action spaces," 2015.

[17] Y. Duan, X. Chen, R. Houthooft, J. Schulman, and P. Abbeel, "Benchmarking deep reinforcement learning for continuous control," *Proceedings of the 33 rd International Conference on Machine Learning, New York, NY, USA*, 2016.

Part II

Preliminary Report (already graded)

Abbreviations

ATC air traffic control

ATCo air traffic controller

ATM air traffic management

CD&R conflict detection & resolution

CNS communication, navigation & surveillance

CPA closest point of approach

CR conflict resolution

DEP domino effect parameter

GUI graphical user interface

LoS loss of separation

MDP Markov decision process

MVP modified voltage potential

PAVs personal air vehicles

RAs research activities

RNAV area navigation

UAVs urban air vehicles

List of Symbols

α	heading segment size
γ	absolute flight path angle of climbing/descending aircraft
$\mu(s\|\theta^{\mu})$	actor function
π	policy
Ψ	heading
a_t	action
alt_{min}	minimum altitude
alt_{max}	maximum altitude
CR_{global}	global conflict rate
d_{CPA}	distance to closest point of approach
$\vec{d}_{rel}$	Cartesian distance vector
d_{sep_h}	horizontal separation minimum
d_{sep_v}	vertical separation minimum
f_n	fraction of total heading range
ft	feet
k_1	weighting term for number of cruising vs climbing/descending aircraft
k_2	weighting term for number of climbing/descending vs climbing/descending aircraft
kts	knots
lat_{centre}	centre latitude
lat_{max}	maximum latitude
lat_{min}	minimum latitude
lon_{centre}	centre longitude
lon_{max}	maximum longitude
lon_{min}	minimum longitude
L	number of layers
m	meter
n_{cfl}^{ON}	number of conflicts with CD&R ON
n_{cfl}^{OFF}	number of conflicts with CD&R OFF
N	number of aircraft
Nm	nautical miles
N_{cruise}	number of cruising aircraft
N_{CD}	number of climbing/descending aircraft
$Q(s,a\|\theta^{a})$	critic
r_t	reward
R_{PZ}	radius of protected zone
s	seconds
s_t	state
S_h	horizontal separation requirement
t	time
t_{CPA}	time to closest point of approach
t_{in}	time that aircraft enters protected zone
t_{out}	time that aircraft leaves protected zone
t_l	look-ahead time
v	aircraft velocity
$\vec{v}_{rel}$	relative velocity vector
V	total airspace volume
$^{\circ}$	degrees

List of Figures

Chapter 1

Introduction

With the increased demand for air traffic in recent years, the airspace capacity is reaching its limit [5]. Furthermore, the forecasts are that this demand will only continue to grow in the coming decades. For the future operations of unmanned aviation, which is the focus of this research, even higher traffic densities than previously seen in manned aviation are expected. The main objective of air traffic control (ATC) is to prevent collisions between aircraft. Because there is always an uncertainty in the exact location of an aircraft, a safety buffer is used in the form of separation criteria. The dimensions of the so-called 'protected zone' are set such that there is enough space for aircraft to resolve an imminent collision. When two aircraft actually are closer to each other than specified in the defined separation criteria, this is called a loss of separation (loss of separation (LoS)). A conflict, on the other hand, is defined as a predicted, potential loss of separation within a specified prediction horizon, also referred to as the look-ahead time [6]. It is thought that, in order to ensure adequate safety in our future air spaces, not only automated conflict detection & resolution (conflict detection & resolution (CD&R)) will become necessary, but there must also be a re-evaluation of coordination efforts that prevent conflicts. In particular, the airspace structure, which is known for decreasing conflict probability and severity, should be looked at. The Metropolis project [7] explored different types of distributed structures and found that a layered airspace concept, where aircraft are separated into vertical flight levels by their direction of travel, performed best in terms of safety metrics like the total number of conflicts and LoSs. This can be attributed to the fact that this imposes a 1) segmentation effect, where aircraft are grouped and remain separated from each other, thus reducing traffic density, and 2) an alignment effect within the layers, where aircraft that travel within a layer have a limited heading range leading to a reduced likelihood of conflict within a layer.

Previous research into layered airspace structures has investigated evenly distributed heading ranges per layer. This is adequate when the air traffic scenario is uniform. In reality, however, the traffic can vary continuously. The department (air traffic management (air traffic management (ATM)) / control, navigation & surveillance (communication, navigation & surveillance (CNS))) has recently started using machine learning techniques to change the heading ranges per layer based on the expected traffic scenario [3][8]. When doing so, the airspace structure is designed to accommodate a larger number of flight levels for popular travel directions. This results in a more uniform distribution of the aircraft altitudes for scenarios with non-uniform heading distributions [1]. Nevertheless, previous works set the airspace in an 'empty' airspace and do not take into account the necessary vertical deviations to get from one structure to the next in the case of a dynamic airspace. It is unclear how safety can be guaranteed during airspace re-configurations and when and what configurations should be selected [1]. This MSc Thesis aims to investigate this by developing a reinforcement learning model that is capable of defining heading ranges per vertical layer, while taking into account the previous airspace.

This midterm report starts off with a literature review in chapter 2. It discusses several key elements to this research, such as layered airspace design, setting such airspaces with reinforcement learning, the vertical deviations that are necessary for re-configuration and some considerations on conflict resolution. In chapter 3, the primary research objective and research activities are presented. Chapter 4, discusses the methodology used for the first research activity. It goes into more detail on the reinforcement learning method, the experiment set-up for the first research activity, the simplifying assumptions that are made and the hypotheses associated with the experiments. Chapter 5 presents the results of these experiments and their relevance, both in the larger ATM picture as well as for the rest of this MSc Thesis. Finally, chapter 6 explains the plans for the final experiments of this research. It will be shown how additional knowledge on the previous airspace structure is incorporated into the reinforcement learning model, what the final experiment scenarios will look like and what the hypotheses for the final experiments are.

Chapter 2

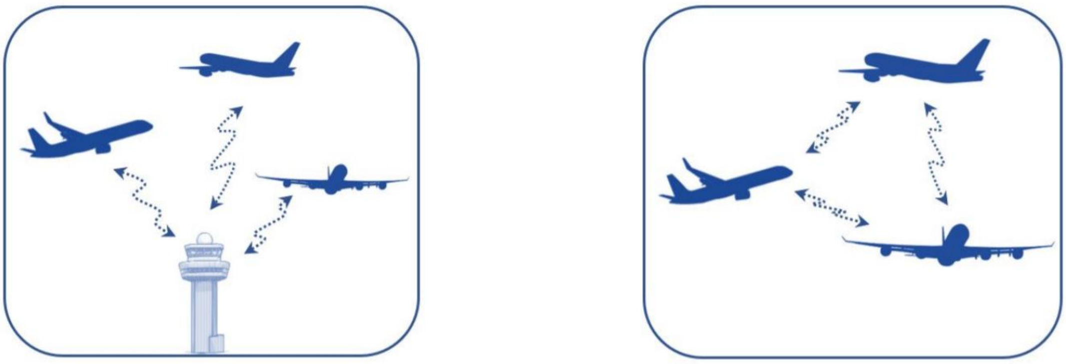
Literature Review

In this chapter, a review of relevant literature to this research is presented. It is split into several sections. Firstly, section 2.1 discusses centralised and decentralised airspace structures. Section 2.2 then presents some background information on layered airspace design. In section 2.3, the impact of vertical deviations on total conflict and LoS count when re-configuring between such layered airspaces is treated. After this, in section 2.4, a potential approach to airspace structuring problems is discussed, namely the use of reinforcement learning to assign airspace structures. Finally, section 2.5 presents some thoughts on conflict resolution (CR).

2.1 Centralised vs. Decentralised airspace structures

In the pioneering days of aviation, pilots relied on simple 'see-and-avoid' principles to prevent mid-air collisions, and navigated using landmarks such as roads, rivers and railway tracks [9]. From the 1930's however, more formal systems and procedures for ATC were deemed necessary and subsequently developed. With this, a new role came to exist in aviation: that of the air traffic controller (ATCo), who's primary task it is to control the aircraft within a predefined area. Though aviation has come long way since those pioneering days and it has seen many technological advancements, the principle of having a person on the ground controlling the aircraft from one central position is still today's reality for almost all air traffic. In ATM, this way of organising aircraft is referred to as 'centralised' ATC.

With the predicted increase in demand for air traffic, many studies have suggested shifting the responsibility of traffic separation back to the aircraft. This is generally referred to as 'decentralised' ATC. It is thought to have the potential to increase the capacity of an airspace beyond the limitations that a centralised system imposes [1]. To further clarify the difference between the centralised and decentralised ATC concept, consider figure 2.1 below.



(a) Centralised ATC

(b) Decentralised ATC

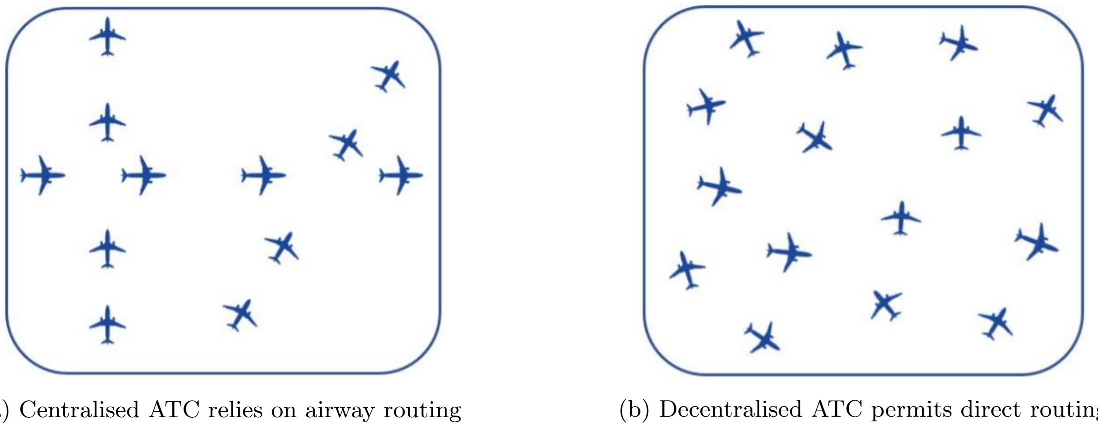
Figure 2.1: Difference between the conceptual designs of centralised and decentralised ATC. [1]

In figure 2.1a it is seen that centralised control has ATCo's that are responsible for the separation of aircraft from the ground. This means that all decisions regarding future trajectories or conflict resolution are made there and communicated to the aircraft. In figure 2.1b, on the other hand, this responsibility shifted towards the aircraft. In this case, individual aircraft are responsible for their own separation. Both of these methods rely on a communication network to pass messages of intent, trajectories or priorities [10]. Figure 2.1 shows manned aircraft, but similar communication links hold for unmanned aviation, albeit through non-verbal / digital communication.

The centralised airspace concept has as the greatest advantage that it can provide a global solution to a multi-actor problem. When a centralised agent tries to resolve (several) conflicting aircraft, it has information on all other aircraft in the airspace to compute a solution with. These centralised methods, however, are often trajectory based, where trajectories are sought that cross each other as little as possible. This generally comes at a high computational costs, as many aircraft routes need to be considered at once. With ever-higher traffic densities, this solution might no longer suffice [11]. Furthermore, there might be issues with availability of the information needed to perform the optimisation. This may be because of lacking technology for the information exchange between aircraft and ground station at all or fast enough, or because of limits on what is shared with the centralised agent. Aircraft operators, for example, are often commercial parties that are not keen to share more information than strictly necessary.

Decentralised separation in en route airspace, on the other hand, is expected to yield several advantages in terms of efficiency, safety and capacity [1]. For a discussion on these, consider figure 2.2. Here, it can be seen that the decentralised ATC concept permits direct routing. The effect on the efficiency, which is a measure of the flight distance or flight time required for aircraft to get from its origin to its destination, is evident: decentralised ATC has the potential to improve the efficiency of flights in an airspace. In addition to improving route efficiency, direct routing is also expected to distribute traffic more uniformly over the available airspace [12]. This increased utilisation of the available airspace has been shown to reduce conflict probability, thereby increasing the safety of decentralised airspace [12][13][14]. Even though the traffic patterns in the decentralised ATC concept, see figure 2.2b, can seem chaotic, distributing the task of separation among all aircraft increases the number of 'problem solvers' in the airspace. As each aircraft only takes into account its neighbouring aircraft when avoiding conflicts, each distributed avoidance system is expected to have only a fraction of the computational strain a centralised system would have [10]. A decentralised system also increases the overall system robustness to hardware failures when compared to centralised ATC [12][4]. If the CD&R system fails in a centralised system, all aircraft under its control will be affected. In a decentralised concept, on the contrary, a failing CD&R system on-board an aircraft still has the other aircraft in the airspace that can detect and avoid the non-nominal aircraft. Finally, the airspace capacity, which is determined by the desired levels of efficiency and safety, is no longer constrained by ATCo workload and therefore has the potential to be increased to levels that would be infeasible with the centralised airspace concept.

An important disadvantage of decentralised airspace concepts is, however, the fact that there is no guarantee of a globally optimal solution when more than two aircraft are involved [10]. Although current CR methods can guarantee implicit coordination in the case of two conflicting aircraft, the situation changes for multi-actor conflicts. In such a case, successive CR manoeuvres can result in unpredictable trajectories, which in turn increase uncertainty as to when and where intrusions occur. It is likely that this is then reflected in less effective CD&R globally and, finally, also a higher total number of conflicts and LoSs. Especially in an airspace with a high traffic density, multi-actor conflicts can be reasonably expected. The desire to return back to more globally optimal solutions in terms of safety, even when employing a decentralised airspace structure, is driving researchers to turn to other methods for reducing multi-agent conflicts. A more detailed discussion on such work is provided in section 2.4.



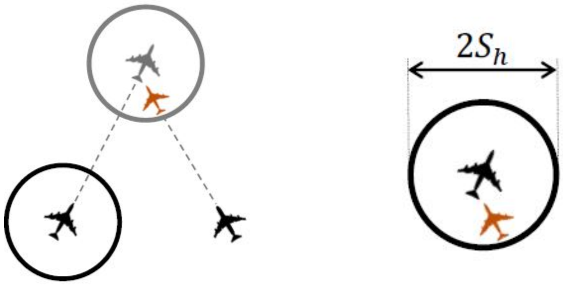
(a) Centralised ATC relies on airway routing

(b) Decentralised ATC permits direct routing

Figure 2.2: Difference between aircraft routing in centralised and decentralised ATC. [1]

The decentralised airspace concept is not new. In fact, the notion of distributing traffic separation tasks have been debated in literature since the introduction of automated ATC systems and area navigation (RNAV) in the mid-1970s [15]. Before continuing however, it is necessary to introduce the concepts of conflicts and LoSs in more detail than was done in the introduction. A proper understanding of how these differ will be needed when considering this and the following chapters of the report.

Referring to figure 2.3a, a conflict occurs if the horizontal and vertical distances between two aircraft are expected to be less than the prescribed separation standards within a predetermined 'look-ahead' time. They may be seen as predicted future LoSs. A LoS, depicted in figure 2.3b, occurs when separation requirements are violated at the present time [1].



(a) Conflict

(b) Loss of Separation (LoS)

Figure 2.3: The difference between conflicts and intrusions, displayed here for the horizontal plane. S_h is the horizontal separation requirement. Adapted from [1]

Most of the research into the decentralised airspace concept, also sometimes referred to as 'Free-flight', has focused on developing automated algorithms for airborne CD&R. An example is the development of the modified voltage potential (MVP) algorithm, which has proven to be effective in reducing the effect of resolution manoeuvres on flight efficiency while still guaranteeing minimal LoSs [10]. More information on this algorithm will be given in chapter 4. Outside the domain of CD&R, however, there remain certain open problems within ATM, such as airspace design, airspace safety modeling and airspace capacity modeling. These three have been tackled in [1]. One piece of work is of particular interest to this MSc Thesis, namely the Metropolis project, which found that a layered airspace concept, where aircraft are separated into vertical flight levels by their direction of travel, performed best in terms of safety metrics like the total number of conflicts and LoSs [7]. This layered airspace concept is discussed further in section 2.2.

2.2 Layered airspace design

This section discusses various aspects of layered airspace design. Subsection 2.2.1 introduces the layered airspace concept and the Metropolis project [7], which forms an important foundation for the work in this MSc Thesis with their recommendation to further investigate layered airspace structures in a decentralised ATM system. Subsection 2.2.2 discusses non-uniform scenarios, while section 2.2.3 introduces dynamic airspace re-configurations in more detail.

2.2.1 Introducing the layered airspace concept

Having established that airspace structures, which are known for decreasing conflict probability and severity, are an essential component of the design of future air traffic management systems, it becomes relevant to learn how different airspace structuring concepts compare to one another. The Metropolis Project [7] sets out to investigate the influence of airspace structure on capacity, safety and efficiency for a high-density airspace. The researchers argue that the rapid emergence of personal air vehicles (PAVs) and urban air vehicles (UAVs) over the last decade, and the fact that they are viewed as important components of the future air transportation system, makes it relevant to look at the airspace system that is required to accommodate these.

One of the concepts introduced in the research is the 'Layers' concept. Here, the airspace is segmented into vertically stacked bands, where each altitude layer limits the horizontal travel to within an allowed heading range. In their work, climbing and descending aircraft are allowed to maintain heading in their simulations, however cruising aircraft must stay within the airspace layer that belongs to their heading. Figure 2.4 shows an illustration of a layered airspace structure that employs uniform heading range distribution per layer. Though Metropolis used a variation to the layered structure in the figure (the researchers actually had two sets of layers that each spanned the full 360° heading, with which PAV and UAV traffic was separated), this example should clarify the concept. The total heading range (360°) is divided uniformly over the layers, resulting in a 45° range for each of the, in the case of this example eight, layers.

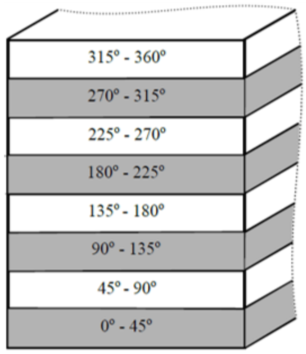


Figure 2.4: Illustration of the airspace structure [2]

The core conclusion of the Metropolis project, which also lays the foundation for the research at hand, is that a vertical segmentation of the airspace, by separating traffic with different travel directions into different flight levels (the layers concept), results in a lower rate of conflicts and thus enables higher capacity. Two factors are thought to contribute to this finding. First of all, by dividing the aircraft over separate layers of airspace, different groups of aircraft are created that remain separated from each other (segmentation effect). Second, within each layer, heading limitations enforce a degree of alignment between aircraft, thereby reducing the relative speed between aircraft cruising at the same altitude, which in turn reduces the likelihood of conflicts within a layer of airspace (alignment effect).

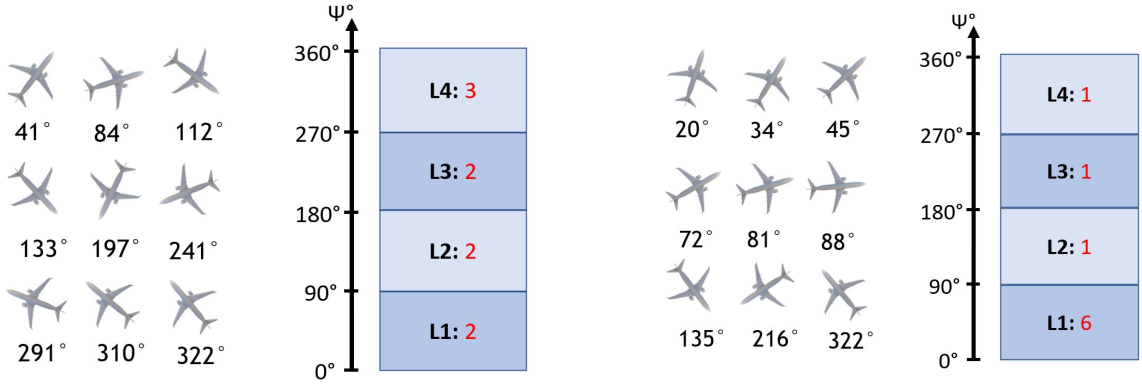
Contrary to performing large-scale simulation experiments, there have also been studies that have approached the problem of airspace design from an analytical standpoint. The researchers in [6] derive an analytical expression for why the layers concept works so well. It is given below in equation 2.1 below

$$CR_{\text{global}} = \underbrace{\frac{1}{2}N\left(\frac{N}{L}-1\right)}_{\text{spreading effect}} \cdot \underbrace{\frac{1}{\alpha}\left(1-\frac{2}{\alpha}\sin\frac{\alpha}{2}\right)}_{\text{reduced relative velocity effect}} \cdot \underbrace{k}_{\text{other influences}} \tag{2.1}$$

where CR_{global} is the global conflict rate, N stands for the number of aircraft, L for the number of layers and α for the heading segment size. This equation shows the distinct influence of the two beneficial effects of a layered airspace structure based on heading segments [6]. It shows the previously mentioned separation (or 'spreading' as the researchers call it) effect and the alignment (or 'reduced relative velocity effect') in a mathematical form. It may be seen how the number of layers (L) and a small heading segment (α) lead to a reduced global conflict rate CR_{global} .

2.2.2 Non-uniform traffic scenarios

The Metropolis project [7] shows the potential of the layered airspace concept to reduce the total number of conflicts and LoSs, while having minimal effect on the efficiency. There are, however, limitations in the way the concept is used in this research. The use of uniformly distributed layers is adequate in cases with traffic with a uniformly distributed heading distribution. In cases where the headings of the aircraft are not uniformly distributed, on the other hand, the aircraft are likely to accumulate in one of the layers, resulting in a non-optimal use of the defined layers. This effect is best understood by looking at figure 2.5.



(a) Uniform traffic with even layers

(b) Non-uniform traffic with even layers

Figure 2.5: Layer utilisation of uniform and non-uniform traffic for an even-layers design. Adapted from [3].

It can be seen in 2.5a that traffic with uniformly distributed headings (Ψ) has the aircraft nicely distributed over a also uniform layer structure. Looking to figure 2.5b, however, it may be noted that such a structure is not suitable for non-uniform traffic scenarios, as aircraft may accumulate in one of the layers. In this case, the over-representation of aircraft flying in the 0° - 90° range, results in an unnecessarily full first layer. This will have a negative effect on total conflict and LoS count. Re-configuring the layers to be able to divide the aircraft more suitably may solve this issue. Graphically, this would look as presented in figure 2.6.

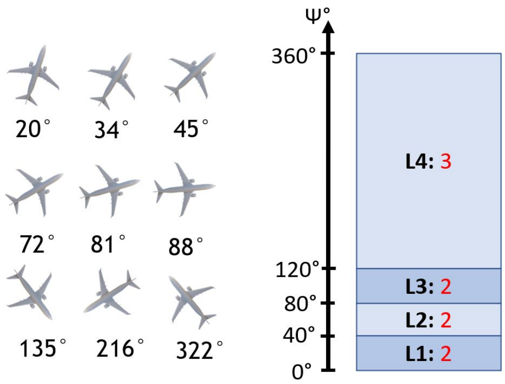


Figure 2.6: Layer utilisation of non-uniform traffic for an uneven-layers design. Adapted from [3].

For the same traffic scenario as presented in figure 2.5b, it can now be seen that aircraft are divided over the layers again, which is expected to have a positive effect on the total number of conflicts and LoSs experienced. This has also been shown previously in [3] and [8].

2.2.3 Dynamic Airspace reconfiguration

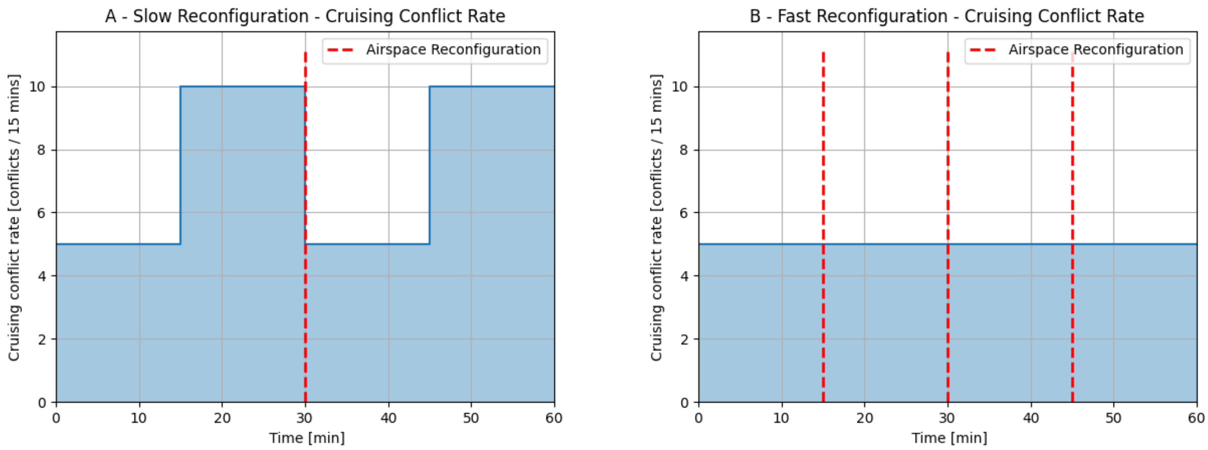
So far, most studies have looked at (layered) airspace configurations that are static in time. However, to optimise the airspace utilisation, it may be beneficial to dynamically reconfigure the constraints imposed by a particular airspace design as the traffic scenario changes throughout the course of a day [1]. For the layered airspace concept, this could imply re-configuring the airspace several times (in reaction to varying non-uniform traffic scenarios) to realise a uniform distribution of aircraft among the layers, such as shown in figures 2.5a and 2.6 in the previous section. Such dynamic re-configurations may be beneficial for the capacity of the airspace, but it is unclear how exactly these should happen in order to ensure safe operations during the change of structure. An important factor is likely to be the reconfiguration rate, as this will determine how often aircraft are required to change from one structure to the next (with all the vertical deviations that are associated with it). Furthermore, it is not known how the induced vertical conflicts play a role in the total conflict and LoS counts. Even given that it is known when and to which airspace one should configure, another question that remains is the ordering of the transition. Should all aircraft move instantaneously or is there some optimal logic for doing this? The above mentioned unknowns that arise in a research into dynamic airspace re-configurations will be investigated in more detail in the next section.

2.3 Impact of Vertical Deviations on Total Conflict and LoS count

In this MSc Thesis that concerns dynamic airspace re-configurations, there will be a strong focus on the transitioning between layered airspace structures. Because such transitions will require (some) aircraft in the airspace to move vertically to the a new correct layer, it is expected that 'vertical conflicts' occur in the process. The term 'vertical conflicts' is used to indicate those conflicts in which at least one of the conflicting aircraft has a vertical velocity component that is nonzero. These type of conflicts do not benefit from the segmentation and alignment effects, which only positively affect the number of cruising conflicts and LoSs count. A 'cruising' conflict, on the other hand, occurs when the involved aircraft do have vertical velocity that is zero. In general, conflicts and LoSs do not necessarily scale proportionally in the sense that double the number conflicts means double the number of LoSs as well. However, it is know from previous research that there is a strong correlation between these to safety metrics. In this section the discussions mainly use conflicts as examples, but the reasoning extends to a large extent to the case of LoSs.

2.3.1 Effect of reconfiguration rate on total conflict and LoS count

Depending on the 'reconfiguration rate' that is chosen for the scenario, the occurrence of vertical conflicts is expected to play a role in the decisions made by reinforcement learning agent on a suitable airspace structure for the traffic scenario. For a better understanding of these dynamics, consider a simple hypothetical scenario displayed in figures 2.7 and 2.8. The traffic scenario is equal for 'A' and 'B' and consists of a situation where the air traffic changes significantly every 15 minutes. The effect of the reconfiguration rate on the cruising conflicts may be as shown in figures 2.7a and 2.7b. A slow reconfiguration rate, as shown in figure 2.7a, may cause sub-optimality of the airspace for the cruising conflicts, as is reflected by the higher conflict rate in the 15-30 and 45-60 minute time-intervals. By increasing the reconfiguration rate, like in figure 2.7b, it would be possible to better optimise for cruising conflicts and keep them at a lower level than would be feasible with a slower reconfiguration rate.



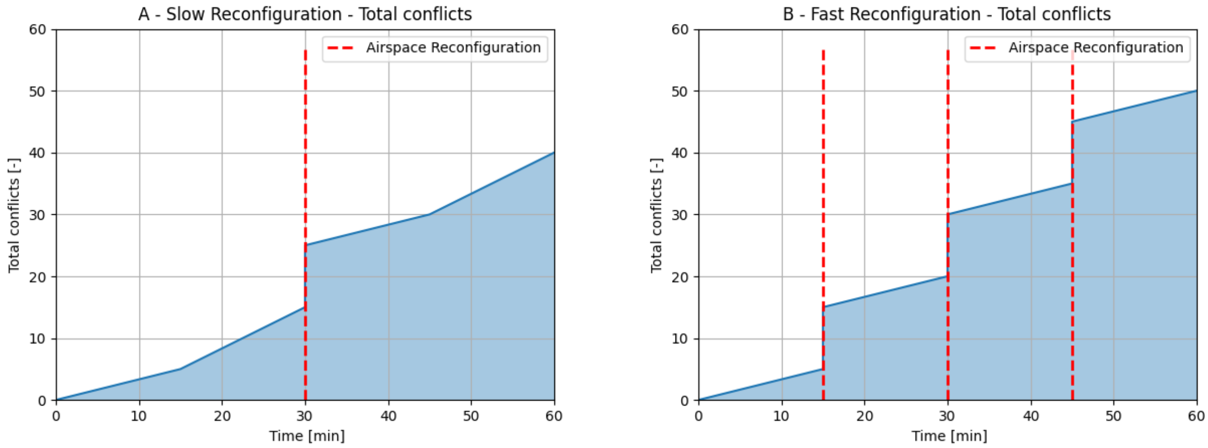
(a) Example cruising conflict rate for slow reconfiguration (b) Example cruising conflict rate for fast reconfiguration

Figure 2.7: Example of cruising conflicts for slow and fast reconfiguration rates in a hypothetical traffic situation that changes significantly every 15 minutes

At first glance this seems attractive, as it has the potential to reduce cruising conflicts. Studies [3][8] have even shown that an airspace that is configured for a traffic scenario has the potential to reduce the number of conflicts and LoSs. For the total number of conflicts in an airspace, however, equation 2.2 holds.

$$Conflicts_{total} = Conflicts_{cruise} + Conflicts_{vertical} \tag{2.2}$$

This shows that the vertical conflicts also count towards the total number of conflicts experienced for a given scenario. For the creation of the images in 2.8a and 2.8b, which show the total number of conflicts for both cases, the vertical conflicts during a single reconfiguration was set at an arbitrary number of 10. It is seen that although situation 'B' from before was best in terms of cruising conflicts, it ends up having a higher total number of conflicts due to the vertical conflicts experienced during the re-configurations (a total of 50 conflicts in situation 'B', while situation 'A' only has 40).



(a) Example of total conflicts for slow reconfiguration

(b) Example of total conflicts for fast reconfiguration

Figure 2.8: Example of total conflicts for slow and fast reconfiguration rates in a hypothetical traffic situation that changes significantly every 15 minutes, with 10 vertical conflicts per re-configuration

From this example, two things may be learnt. Firstly, the reconfiguration rate has an influence on the total number of conflicts, and with that on the airspace selection during this research into dynamic airspace re-configuration. Each time a re-configuration is made, there is a cost in terms of vertical conflicts that must be added to the total conflict count. Secondly, the vertical conflicts experienced during a single reconfiguration (set to 10 in the example) also influences the total number of conflicts. Were this to be, for example, 2 instead of 10 in the example, situation 'B' with the three re-configurations would be most optimal again (a total of 26 conflicts for 'B', while 'A' would have 32).

When the reconfiguration rate is to be determined for a real-life implementation of dynamic airspace reconfiguration, it is likely to depend for a large part on the traffic scenarios in the airspace. The rate of change of the global heading distribution of the aircraft in the airspace is thought to drive the 'ideal' reconfiguration rate for the set of traffic scenarios. If the direction of aircraft in the scenario rapidly changes throughout the day, the airspace may benefited by more re-configurations than a traffic scenario that stays relatively constant over time. It is likely to be a question of balancing the cruising and vertical conflicts. A 'too-slow' rate leads to selecting airspace configurations for a long time-horizon, leading to sub-optimal airspace selection and more cruising conflicts. It would be tough to 'fit' a large variety of traffic scenarios into a single selection of airspace layers. A 'too-fast' rate, on the other hand, leads to more optimal airspace selection in terms of cruising conflicts, but the apparent gain in terms of safety may be offset by a high number of vertical conflicts. In any case, it is thought that the reconfiguration rate must somewhat be reflected by the natural rhythm of the traffic demand. Whether this is indeed true must be shown by experimental simulations, where different rates and scenarios will be tested to better understand the impact of this variable. As a final note, the reconfiguration rate of a real-life setting must respect the constraints set by the technology used for information sharing and processing.

2.3.2 Vertical conflicts in analytical conflict count models

Having looked at the impact of vertical deviations arising from airspace re-configurations, the attention is now turned towards vertical conflicts in analytical conflict count models. The aim here is to show that 1) previous research has looked into these methods as opposed to only performing experimental simulations to study vertical conflicts and 2) that the number of vertical conflicts can be significant, especially as the number of climbing/descending aircraft increases.

In [16], the researchers look into the modeling of the intrinsic safety of both unstructured and layered airspace designs. When the layered airspace concept is considered, they start of with deriving an expression as already shown in equation 2.1, but also add that the extension to a 3D conflict rate (CR_{global}) may be made by adding two terms for the vertical conflicts. The terms they add are shown below in equation 2.3, where the first represents conflicts between cruising and climbing/descending aircraft, and the second represents conflicts which involve only climbing/descending aircraft.

$$\underbrace{N_{cruise} \cdot N_{CD} \cdot f\left(|\gamma|_{avg}\right) \cdot k_1}_{\text{Cruising vs. Climbing/Descending}} + \underbrace{\frac{N_{CD}\left(N_{CD}-1\right)}{2} \cdot f(|\gamma|) \cdot k_2}_{\text{Climbing/Descending vs. Climbing/Descending}} \tag{2.3}$$

Here, N_{cruise} is the number of cruising aircraft, N_{CD} stands for the number of climbing or descending aircraft, $f(|\gamma|)$ is given by equation 2.4 below, and the constants k_1 and k_2 are used in weighing them among the each other and the term for only 'cruising vs. cruising' conflicts.

$$f(\gamma) = \frac{2d_{sep_h} v t_l \left(2d_{sep_h}|\gamma| + 2d_{sep_v}\right)}{V} \tag{2.4}$$

In equation 2.4, d_{sep_h} is horizontal separation minimum, v is the average aircraft velocity, t_l is the look-ahead time, $|\gamma|$ is the absolute flight path angle of climbing/descending aircraft, $d_{sep_{hv}}$ is vertical separation minimum and V is the total airspace volume under consideration. Fixing all elements, but the number of climbing/descending aircraft, N_{CD}, reveals that part of the conflict rate that comes from climbing/descending aircraft increases as the number of climbing/descending aircraft increase. See figure 2.9 below. In this figure, the absolute value of the vertical conflicts is arbitrary. The reader is merely made aware of the exponential course of the vertical conflict as the number of climbing/descending aircraft increases. With that, the importance of these types of conflicts during dynamic airspace re-configuration is underlined.

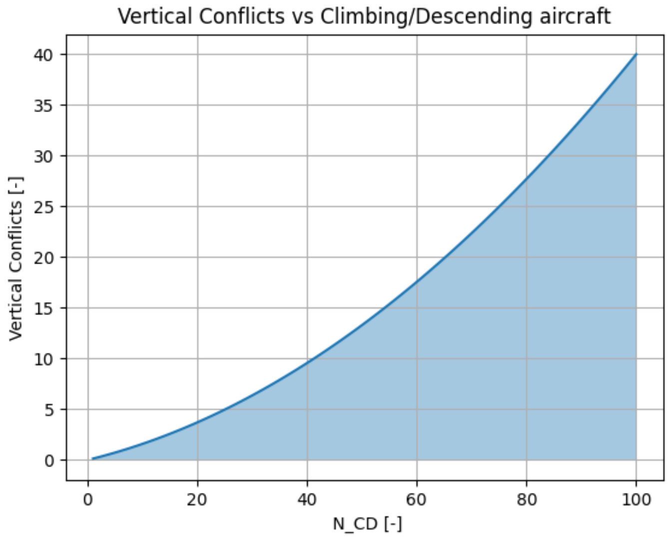


Figure 2.9: Vertical conflicts vs. climbing/descending aircraft

2.3.3 Order of transitioning aircraft

As stated before, the restructuring of an airspace that is already filled with aircraft will require (some) aircraft to move from one layer to another. Earlier in this section, it was also explained how the airspace structures selected by the RL agent might be influenced by the vertical conflicts that arise in these transition phases and how how the rate of reconfiguration is an important variable in this respect. On top of this, there is, however, also the order in which aircraft switch layers that will influence the total number of conflicts during a transition. Exactly how this should be done, however, is unclear. Several options are possible. Examples are 1) moving all aircraft at the same time, 2) performing the transitions on a layer-by-layer basis and going from top to bottom (or vice-versa), 3) transitioning the aircraft on a layer-by-layer basis, but doing this in an order corresponding to the number of aircraft in those layers. Many more variants would exists, but given the discussions in subsections 2.3.1 and 2.3.2, it is at least of importance that some attention goes to analysing or optimising the way in which aircraft move between airspace layers.

2.4 Reinforcement Learning to Assign Layered Airspace Structures

Previously there have been discussions on centralised and decentralised ATM concepts, layered airspace design and the impact of vertical deviations of on the total conflict and LoSs. In this section, the attention is shifted towards another core element of this MSc Thesis, namely that of setting airspace structures with RL methods. RL methods have proven successful in other ATM studies [8][17][18][19] and it was chosen here as it is believed that the nature of the challenge to select appropriate airspaces based on traffic scenarios is well suited to these methods. The section starts with a more in depth discussion as to what RL is in subsection 2.4.1. Subsection 2.4.2 then discusses a study into setting airspaces with RL and displays how its recommendations for future work form the basis for the work of this MSc Thesis. Finally, subsection 2.4.3 aims to balance the discussion by presenting some limitations of using RL methods for safety in aviation.

2.4.1 What is Reinforcement Learning?

RL methods are typically stated in the form of what is called a Markov decision process (MDP). Solving MDPs can be an effective method for determining actions for an agent in stochastic environments [20], such as an airspace with random traffic. In an MDP, an agent chooses action a_t at time t after observing some state s_t. The agent then receives reward r_t, and the state evolves probabilistically based on the current state-action pair. The assumption that the next state only depends on the current state-action pair is what is generally referred to as the Markov assumption [20]. The agent's behavior is defined by a policy, π, which maps states to a probability distribution over the actions. The goal is to learn a policy which maximizes the reward [17]. A graphical overview of the above definition of a RL problem is shown below in 2.10.

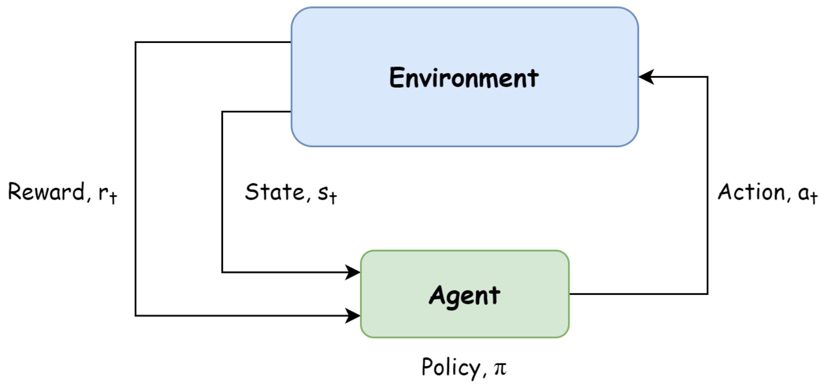


Figure 2.10: Definition of a typical RL problem

Like all methods, RL has its pros and cons. A strong point is its applicability to a variety of problems. Within ATM, it has been used in the domain of conflict resolution [17], but also for setting airspace structures [8]. Outside the ATM domain, studies have for example used RL for lane-following in autonomous driving applications [21]. RL methods have also been known to have good performance. When implemented successfully, these methods are capable of solving problems for which conventional algorithms fall short. On the other hand, some well-known drawbacks are non-convergence, high dependence on initial conditions, and long training times [17].

2.4.2 Reinforcement learning for setting airspace structures with unmanned aviation

RL has been used in previous research to improve airspace structuring in an urban environment [8]. Here, the researchers looked into the safe introduction of drones into the urban airspace by studying the effect of different airspace structures in relation to different traffic scenarios. It was found that the use of RL to set airspace structures had a positive effect by reducing the total number of conflicts and LoSs. As a main recommendation for future work, it is stated that it is still unclear how safety of operations can be guaranteed during configuration changes. Changing from one structure to another was not analysed. The research assumes that 'such transitions will entail several vertical deviations in order for cruising aircraft to adapt to the new structure'. It is furthermore said that 'increasing the number of vertical deviations may result in a rise of the number of conflicts' and that 'it is likely that, during a direct change of airspace structure, the RL agent must take into account the previous structure to reduce the number of vertical deviations'. A discussion on the expected impact of these deviations on the total conflict and LoSs count was already presented earlier, in section 2.3. It may be noted that this recommendation for further research in the area of re-configurations forms a direct basis for the work of this MSc Thesis project.

2.4.3 Limitations of using Reinforcement Learning for safety in Aviation

Though interesting results have been achieved recently with RL methods, it is of importance to keep some known limitations in mind. The first that is worth touching upon is the selection of the state of the model. In a study that tries to determine optimal conflict avoidance manoeuvres with a RL model [17], the researchers make some general statements about state-selection in a RL model that could also hold for one that is to assign airspace structures. It is said that the state should provide enough information to the agent to correctly respond to the behaviour, but that this is in reality limited by the availability of information to the agent and by the computational effort. The researchers have a preference for simplicity in the early stages of exploration, while stating that adding information to the model can always be done at a later stage. This is an approach that may be kept in mind for the development of the RL models for the work of this MSc Thesis. It is further stated that training the model is highly influenced by the reward structure. Also, the research discusses the balancing of conflicts and LoSs, where they say that the relation between the two is not trivial, making it hard to define how to weigh these in the reward formulation. A number of formulations were tried, including an option where only LoSs were counted, an option where both conflicts and LoSs are considered and an option LoSs and the time-in-conflict metrics were used. The magnitudes of the weighting between the (sometimes) multiple factors in the formulation was set empirically. Lastly, the rewards were set to be negative, rather than positive, to prevent stimulating the RL model to simply solve many conflicts (it might then learn to create many and then solve them to get the best score). Both the metrics considered for the formulation (conflicts, time-in-conflict and LoSs), as well as using negative rewards, are also aspects to keep in mind when defining the RL models for this MSc Thesis. Most of the justifications in this paper could hold true as well in a RL model that is to set assign a airspace structures.

In the results of the paper, another interesting point is made about the effect of the reward function formulation. It is stated that the number on conflicts indeed decreases when this is included in the reward formulation, but that this does not immediately lead to fewer LoSs. In fact, the opposite seems to occur: more conflicts lead to a decrease in LoSs. This has previously been observed in [4], where it is argued that that a moderately positive number of secondary conflicts can be beneficial on a global scale. The effect of sequentially running into a new conflict can then create a wave-like pattern, spreading the aircraft out in the available airspace thus 'creating' more airspace. This phenomenon highlights the importance of choosing appropriate reward formulations and that, for example, guiding the RL model with information on conflicts may lead to adverse effects when it comes to reducing LoSs. This is something to be mindful of when formulating the reward function for assigning airspace structures with a RL model. From the results it is also clear that a reward formulation based on LoSs is most effective at reducing the number of LoSs. It is, however, noted that the training progress is very slow. The researchers attribute this to the fact that there are very few occasions to improve, as the number of LoSs is relatively low in a given scenario.

2.5 Conflict Resolution

In this research, the focus is first and foremost on conflict prevention through setting appropriate airspace structures. This, however, doesn't imply that conflicts will not occur in the experimental simulations of this work. In fact, they will purposely be created to be able to train a reinforcement learning model to select an airspace configuration. This section provides a discussion on conflict resolution methods and as such forms a basis for proper handling of the conflicts in this work. Section 2.5.1 goes into conflict avoidance manoeuvres, after which section 2.5.2 considers conflict resolution in a layered airspace specifically.

2.5.1 Conflict Avoidance Manoeuvres

Conflict avoidance manoeuvres generally consist of an action to de-conflict the aircraft in question. Depending on the type of resolution algorithm that is implemented, these actions can be speed, heading or altitude alterations. The manoeuvres are just that; a variation in the speed, heading or altitude of one or both aircraft to get out of the conflict. Combinations of these manoeuvres may also be permitted to be able to achieve a more efficient solution.

2.5.2 Conflict resolution in a layered airspace

For layered airspace structures, some extra considerations for conflict resolution are needed to deal with the heading constraints imposed in each layer. In such a design, it could occur that the manoeuvre proposed by the resolution algorithm implies a violation of the defined airspace structure. Different approaches may be taken to tackle this issue. A first option is to simply allow the aircraft to break the limits as set by the airspace structure for the duration of the manoeuvre. In such a case, one would have airspace layers where aircraft generally travel within the predefined heading range, but are not strictly bound to it in the case that they are ordered out of it for a resolution manoeuvre. Alternatively, one could opt to have the aircraft adhere the heading ranges at all times. This could imply that aircraft perform additional vertical deviations to go to the correct layer where the avoidance heading is allowed. A third option for resolution manoeuvres in a layered airspace is to limit the avoidance solution to the heading ranges of its current layer.

It is expected that the third option, that limits the output of the resolution algorithm, reduces the effectiveness of conflict resolution when compared to the first option, which permits manoeuvres that 'break through' the heading range of a layer. On the other hand, limiting the resolution output to the defined heading range does hold in place the previously mentioned advantages of the alignment effect in terms of the reducing the conflict probability in a layer. In turn, this may have a positive effect on the total number of conflicts and LoSs if it offsets the increase in conflicts due to less effective resolution algorithm. The intricacies of these dynamics are best examined through experimental simulations, where it is possible that the answer depends on factors as traffic density or the type of resolution algorithm used. In a conflict resolution experiment explained in [1], the authors found that vertical conflict resolution maneuvers are more likely to trigger new conflicts in their experiment, which had the aircraft are more densely packed in vertical direction. Considering the packing density in horizontal and vertical directions for selecting a suitable of manner of conflict resolution in a layered airspace may prove a useful exercise. As the minimum vertical separation requirement is often set equal or close to the layer height, it will generally be the case that the vertical packing density is higher than the horizontal one, making vertical resolution manoeuvres an unattractive option for layered airspace designs.

Chapter 3

Problem Definition

This chapter formulates the problem definition of this MSc Thesis. It aims to make clear what the focus of the project is by defining a primary research objective in section 3.1. To achieve the primary research objective, six research activities have been layed out. These are explained in section 3.2.

3.1 Primary Research Objective

The primary research objective for this thesis is:

Primary Research Objective

To develop a reinforcement learning model, which can correctly assess and minimise the impact of airspace reconfiguration on safety, while guaranteeing that new structures are appropriate for future air traffic

3.2 Research Activities

To meet the objective stated in section 3.1, six research activities (RAs) have been defined. They are the following:

1. Create a RL model that selects an optimal airspace structure for a given traffic scenario
2. Define rules for moving traffic into new airspace structure
3. Define rules for conflict resolution in a layered airspace
4. Create a RL model that selects an optimal airspace structure for a future traffic scenario, while taking into account the previous airspace structure
5. Analyse the effect of the reconfiguration rate on the choice of an optimal airspace structure
6. Compare performance of the airspace structures picked by the RL models resulting from activities 1, 4 and 5

Each of these will be explained in more detail in subsections 3.2.1 through 3.2.6 respectively. First, however, consider figure 3.1 for an overview of how they relate to one another and the final research objective.

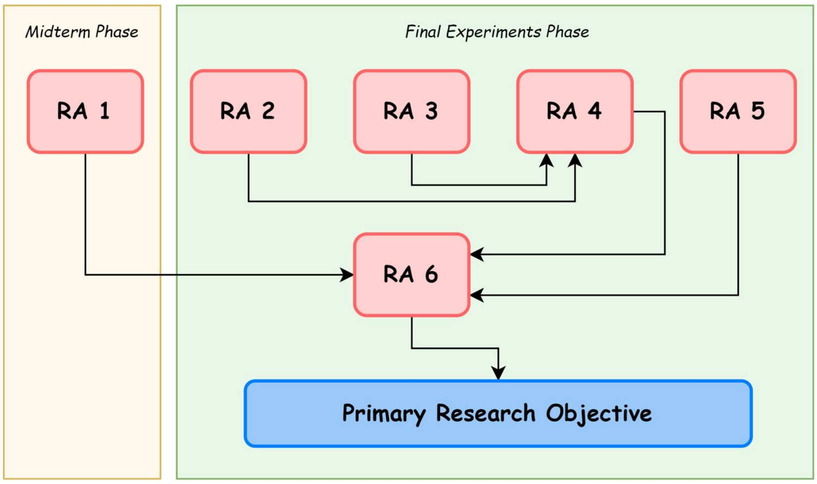


Figure 3.1: Relation of the six research activities to each other and the primary research objective

As shown in the figure, two phases are identified: the 'midterm phase' and the 'final experiments phase'. The results for the midterm phase, thus for research activity 1, will be included in this midterm report. The other activities will take place after completion of this phase and any results following from those will thus be presented at a later stage. Do note, however, that considerations have been added for almost all of the research activities in the literature review of the previous chapter. The reader will be referred to the relevant sections there in the following discussion of the individual research activities. As may also be seen in the figure, there are several dependencies between the research activities. Activities 1, 2 and 3 are stand-alone and can be completed without the results of any of the other activities. Activity 4 (creating the final RL model) will use the results of activities 2 and 3 for the rules of moving traffic between structures and conflict resolution. Activity 5 is similar to activity 4, but is aimed at creating and analysing the results of several models that are trained with different reconfiguration rates. It is therefore considered a separate research activity in this work. Finally, research activity 6 needs the results of the RL models defined in activities 1, 4 and 5 to be able make a proper comparison of the models (one without taking the previous airspace structure into account, one where the previous structure is taken into account, and a set that also takes previous structures into account but were trained using varying reconfiguration rates). By completing this final activity, a judgement can be made as to whether the primary research objective has been achieved.

3.2.1 Research activity 1: RL model for an optimal airspace structure

Research activity 1 is:

Research Activity 1

Create a RL model that selects an airspace structure that reduces the number of conflicts/LoSs with respect to a unform, fixed airspace structure

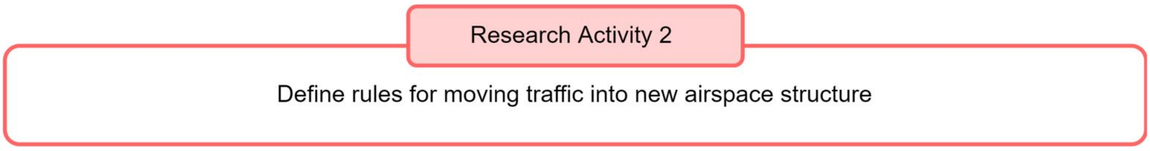
This research activity is centered around creating a first working RL model. The aim for this is that it is capable of setting the airspace structure for a single traffic scenario, while reducing the total number of conflicts/LoSs compared to a uniform, fixed airspace structure. It is thought that this forms a solid foundation for the development of the RL model that is to take the previous airspace structure into account (during a reconfiguration) as well. Several questions will need to be answered for this research activity, they are stated below.

- Which RL model would be suited for the problem at hand?
- What are the limitations of the chosen RL model?
- How should the traffic scenario be formulated to use it as an input for the RL model?
- How should the output from the RL model be formulated to define an airspace structure?
- How should safety be defined for the RL model (e.g., conflicts, LoSs)?

For some further considerations on the use of a RL model to assign layered airspace structures, the reader is referred to section 2.4 in the previous chapter. The experimental results for activity 1, which is independent of the activities that are to follow (see figure 3.1), will be included in this report. It is the only research activity for which this is the case. The results of the following five activities will be worked out for the final part of this MSc thesis, the final experiments phase.

3.2.2 Research activity 2: Rules for moving traffic into new airspace structure

Research activity 2 is:



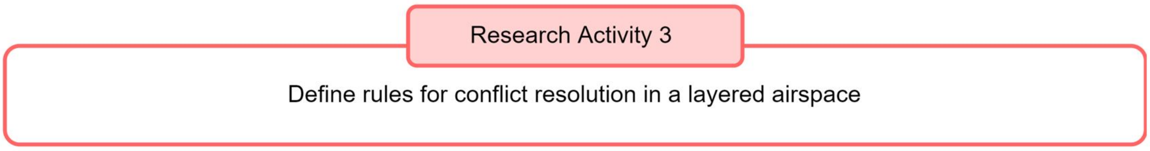
The intended addition of this MSc thesis project to previous research mainly lies in the dynamic transitioning from one airspace to the next and how this can be achieved with a reinforcement learning model that defines heading ranges. The transitions from one airspace to the next invokes the need for vertical deviations of aircraft to their (new) airspace layer. There is not one way that one might go about instructing the aircraft in a scenario to move their vertical position to get into their layer, however it is unclear what the effect is of any rules for doing this. This second research activity is to define rules for moving traffic into a new airspace structure. In subsection 2.3.3 of the previous chapter, it was already explained that the manner in which aircraft transition into a new structure during an airspace reconfiguration will influence the total number of conflicts and LoSs. With suitable rules, the number of vertical conflicts (conflicts for which at least one of the conflicting aircraft has a vertical velocity that is nonzero) may be limited. Several questions may again be posed for this activity, they are stated below.

- What are suitable rules for moving traffic into a new airspace structure?
- Should a maximum amount of time be allowed for aircraft to move into the new structure?
- Should aircraft that are transitioning between layers receive priority over cruising aircraft?

It is very well possible that the answers to these questions are a function of independent variables like traffic density (low, medium, high) or the setting of the conflict resolution (on/off). If this appears to be the case, the rules for moving the traffic would have to be adjusted to the context of the scenario for optimal safety. As is displayed in figure 3.1 at the beginning of the section, this research activity will take place in the final experiments phase, meaning that the results for this are not included in this report. Because it would be an unnecessarily lengthy process to investigate the questions for this activity with the RL model developed under research activity 1, separate experiments will be defined for investigating the rules for transitioning traffic.

3.2.3 Research activity 3: Rules for conflict resolution in layered airspace

Research activity 3 is:



The key characteristic of the layered airspace concept is that the aircraft in a layer must adhere to a set heading range (see figure 2.4 in the previous chapter). As explained as well in subsection 2.5.2, some extra considerations for conflict resolution are needed to deal with these heading constraints. This is as the manoeuvre proposed by a resolution algorithm could imply a violation of the defined heading range limits in the current vertical layer. There are roughly three possible ways of dealing with the issue.

1. Allow aircraft to break the heading limits set for their layer, reducing the benefits of the segmentation and alignment effects previously explained in subsection 2.2.1.
2. Strictly adhere to the defined heading ranges, implying aircraft might have to perform (additional) vertical manoeuvres for conflict resolution.

3. Limit the avoidance to the heading ranges of the layer, which is expected to reduce the effectiveness of the conflict resolution algorithm (see last paragraph of subsection 2.5.2).

This third research activity outlined here is concerned with defining rules for conflict resolution in a layered airspace with defined heading ranges per layer. The questions posed for this are the following:

- Which CR model can best be used?
- Which of the three aforementioned options for dealing with conflict resolution in a layered airspace results in the lowest total number of conflicts and LoSs
- When possible, should aircraft prioritise speed over heading variations for resolving conflicts?

For this research activity there are some similarities with research activity 2. Just as for that activity, the results for this third research activity will be completed in the second phase of the MSc project (after the midterm report). Also, separate experiments will be defined for the defining of the conflict resolution rules, as using the airspace structure set by the RL model would be unnecessarily complex. Lastly, it also holds for research activity 3 that the answers to the posed questions could depend on the settings for the independent variables like traffic density (low, medium, high).

For the continuation from research activities 2 and 3 into research activity 4 (see figure 3.1), the best results will be taken in terms of the rules for moving the traffic and conflict resolution. The RL model can then adapt to these rules during its training.

3.2.4 Research activity 4: RL Model for dynamic airspace reconfiguration

Research activity 4 is:

Research Activity 4

Create a RL Model that selects an optimal airspace structure for a future traffic scenario, while taking into account the previous airspace structure

This research activity gets to the core of the research and is centered at creating the final RL model that is to select suitable airspaces with information on the future traffic scenario, as well as the previous airspace configuration. A sequence of airspace selections, done with this information on the future air traffic and the previous structure, is the essence of the 'dynamic airspace reconfiguration', the main theme of this work. A relevant question for this research activity is shown below.

- How should the previous airspace structure be used as an input for the RL model?

Note that this research activity will consist mainly of adding new information on the current airspace structure to the model already created under activity 1. Many of the questions concerning formulations for the inputs and outputs of the RL model will have already been addressed there. It may be added that, with the larger state formulation in this activity (now including information on the previous structure), there will also be more possible state/action combinations. Due to this, it is expected that the training of the model takes longer.

As mentioned previously, the best results of research activities 2 and 3 will incorporated in the formulation of the experiments for this fourth research activity, see figure 3.1. The rules for moving the traffic into the new airspace structure (subsection 3.2.2) and for conflict resolution in a layered airspace (subsection 3.2.3) are then in principle lined up with the goal of optimal safety. It is, however, noted that research activities 2 and 3 will be relatively short investigations as they are not the main focus of this work. Furthermore, they will be explored individually and not at the same time. Their independence can, however, not be guaranteed, which means that the rules for moving traffic between layers, as well as the rules for conflict resolution, remain assumptions in the development of the final RL model.

3.2.5 Research activity 5: Effect of reconfiguration rate on the choice of optimal structure

Research activity 5 is:

Research Activity 5

Analyse the effect of the reconfiguration rate on the choice of an optimal airspace structure

In subsection 2.3.1 it was established that the reconfiguration has an influence on the total number of conflicts, and with that on selecting an optimal airspace structure. For this research activity, two questions are posed:

- How does reconfiguration rate affect the decisions taken by the RL model?
- Is there an optimal reconfiguration rate?

It is again likely that the answer to these questions depends on factors such as the traffic density and traffic scenarios. In the experiments of this work, the reconfiguration rate will be set as an independent variable. Several rates (slow, medium, fast) will be used, making it possible to examine the effect of this variable on the choices that the RL model makes. In the real world, there will be a minimum configuration rate, depending on the speed at which the environment is capable of gathering information and decides upon a new airspace configuration. This will be a function of the technology available for communicating the information between an agent for setting the airspace structure and the aircraft in the airspace. The maximum (useful) reconfiguration rate will depend on the rate at which the traffic scenario changes considerably.

Concerning the relation to other research activities, this activity will be completed after the RL model for dynamic airspace reconfiguration has been set up in research activity 4. With this in place, experiments can be run where the reconfiguration rate is varied to see what the effect is on the choices of the RL model. Because of the fact that new RL models will follow from this, it was chosen to view this as a separate reseach activity. Referring to figure 3.1, the results found with models developed here may then be used in the sixth and final activity.

3.2.6 Research activity 6: Compare performance of the airspace structures from activities 1, 4 and 5

Research activity 6 is:

Research Activity 6

Compare performance of the airspace structures picked by the RL models resulting from activities 1, 4 and 5

In section 2.4.2, it was explained how previous work has focused mostly on the application of reinforcement learning models to prescribe airspace structures without taking the previous airspace (and thus the need for re-configuring) into account. As explained, research activity 1 is concerned with this is well, while the step of adding information on the previous airspace is made in activity 4. In activity 5, the additional investigation into the effect of the reconfiguration rate on the selected airspace structures is made. That extension towards airspace reconfigurations and its impact on safety in activity 4 and 5, which is the crux of this research, must be analysed in depth in order to be able to extract the full value from the work. Specifically, the questions posed for this final activity are:

- How do the optimal structures of the RL model that is aware of the previous structure differ from structures as output by a RL model that only has information on the future traffic scenario?
- Does the RL model opt for structures that focus on reducing vertical conflicts during the transition period when information is added on the previous airspace or does it still focus on the reduction of cruising conflicts?

It can reasonably be expected that the airspace structure resulting from research activity 1 will be better at decreasing the number of cruising conflicts for every traffic scenario (as it only takes that into consideration). It will, however, likely also suffer from more vertical conflicts than the airspaces output by the RL model of activities 4 and 5.

For this final activity, it is envisioned that it forms the core of the results for this MSc project. A sound comparison between the outputs of the newly developed RL models that take the previous airspace structure into account and RL models that merely looks at the future traffic scenario, will be the best way of determining to what extent the research objective of section 3.1 has been achieved.

Chapter 4

Methodology for Research Activity 1

This chapter introduces the methodology for completing the first research activity: to create a RL model that selects an optimal airspace structure for a given traffic scenario. For a reminder as to how this activity ties in with the activities for the full MSc thesis, the reader is referred to section 3.2 of the previous chapter. The experiments to demonstrate the RL model will be referred to as the 'concept experiments' in this report, in order to be able to distinguish them from the experiments for the remaining parts of the research after the midterm. Firstly, the setting of the airspace structures for this research activity with RL is discussed in section 4.1. Here, the most important aspects of the RL model are discussed, such as the agent, the learning algorithm and the state, action and reward formulations. Following this, section 4.2 goes further into the details of the set-up for the concept experiments. The goals here is to present exactly how the experiments are defined for the first research activity. Lastly, in section 4.3 the hypotheses for the concept experiments are stated.

4.1 Setting the airspace structures with Reinforcement Learning

This section gets to the heart of the research by discussing how airspace structures will be set with a RL model. Subsections 4.1.1 through 4.1.5 elaborate on key elements of such a model, such as the agent, learning algorithm and state, action and reward formulations.

4.1.1 Agent

The agent has the objective of setting an airspace structure that is optimised for the expected traffic scenario. It is assumed that the RL agent has full knowledge of the future traffic density and trajectories. In a real-life application, the agent could be seen as the operator of the airspace and responsible for the defining configuration (changes).

4.1.2 Learning Algorithm

The type of learning algorithm used for the concept experiments is the soft-actor critic algorithm. In this relatively new off-policy actor-critic algorithm 'the actor aims to simultaneously maximize expected return and entropy; that is, to succeed at the task while acting as randomly as possible' [22]. In general, a RL algorithm such as this one consists of an agent (see subsection 4.1.1) that interacts with its environment in discrete timesteps. It has the goal to learn a policy that maximises a reward, r_t, that is given to an action.

In an actor-critic architecture, there are two neural networks: one for the actor and one for the critic. The actor function, often named the policy, is usually written $\mu(s|\theta^{\mu})$ and specifies the output action a in regard to the input, the current state s of the environment in the direction proposed by the critic. The critic, on the other hand, is often denoted by $Q(s, a|\theta^{a})$ and tries to estimate the correlation between the state and the action of the actor. It is updated from the gradients obtained from a temporal difference error signal each time step. The output of the critic drives learning in both the actor and the critic. The activation functions used for this first research activity are 'tanh' functions in the hidden layers, with a 'sigmoid' function in the output layer. [17]

4.1.3 State

It was chosen to use a state which contains information on the headings of the aircraft within the experiment area. For this, the total aircraft heading range, 0° - 360°, is divided into 10 bins of equal size. The aircraft are then divided over these bins by their instantaneous heading to compute a state array at a given time. The resulting array is then normalised before proceeding with the next steps. This way, the value of each heading bin will be the fraction of the total number of aircraft that have a heading that falls within the heading range for a particular bin. A graphical representation of this is given below in figure 4.1, where n1, n2, ..., n10 represent the normalised number of aircraft of that bin. For example, the value of n1 is the number of aircraft with headings in the 0° - 36° range divided by the number of aircraft at that time, n2 is the number of aircraft that have headings from 36° - 72° divided by the number of aircraft. It is key to select the state of the model such that it provides sufficient information about environment, without becoming so large that it causes excessive computational effort. Though a bigger state array may allow for a better representation of the heading differences between aircraft, increasing the dimension of the state increases the number of possible states and state-action combinations. As the solution space of the problem grows, so will the training time of the model [3].

n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}

Figure 4.1: State formulation for the RL model with 10 heading bins. Each of the values n1, n2 ... n10 represents the (normalised) number of aircraft that have a heading that is within the heading range corresponding to that bin.

4.1.4 Action

An action is selected each time the state is given to the agent. The state values are passed through the neural network, which contains network weights for every neuron and activation functions for every layer. The activation functions are such that they convert the output of the previous layer to a form that can be taken as an input for the next layer. The final output is a one-dimensional action array, which contains the information on the selected heading range each layer. Directly at the output, this array is of size [1 x 8] and filled with values ranging from 0 to 1 (due to the sigmoid function in the output layer). This action array is then normalised, such that the values add up to one and can trivially be used to define an airspace structure that covers the full 360° heading range. A graphical representation of the action formulation is given below in figure 4.2, where f1, f2, ..., f8 stand for the fractions of the eight layers that will be used in this research activity. For example, if f1 = 0.05, the first layer will allow aircraft with headings 0° - 18°. Then, if f2 = 0.15, the second layer will allow aircraft with headings 18° - 72°, and so on, until the complete 360° heading range is covered by the layers.

It may be noted that the choice for eight layers was made in this research, but that this variable is in reality dependent on the environment. With fewer combinations of heading ranges to chose from, the resulting fewer altitude layers would likely lead to lower training times. However, with more layers the aircraft will be more dispersed throughout the airspace, possibly resulting in more optimal results in terms of the total number of conflicts/LoSs experienced.

f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8

Figure 4.2: Formulation of the action that is given by the RL agent in response to a state for the 8-layer airspace design. Each of the values f1, f2, ..., f8 (decimal, 0-1) represents the fraction of the total heading range that the layer should cover

4.1.5 Reward

In this research, the focus lies on improving aviation safety through airspace design. To achieve this, it makes sense to give the RL model rewards based on safety metrics. Within this, however, there is still the option of going with a reward based on the total number conflicts or LoSs (or even combinations of these). It is not expected that there is a clear-cut answer as to which of these works best for the case of dynamic airspace reconfigurations. Likely, this is again dependent on various factors such as the traffic density in the experimental simulation, the nature of conflicts and LoSs ('cruising' or 'vertical') or the settings for conflict detection and resolution. Initially, the approach taken in this work will be to keep the reward function as close to the objective as possible, which means that a reward based on the total number of LoSs is implemented. Though it is the ultimate goal of the experiments to reduce the total number of LoSs, it is foreseen that a relatively low number of LoSs (at least, when compared to the total number of conflicts) might contain too little information for the model to learn a successful policy. In that case, one might consider methods to increase the number of LoSs in the experiments or trying reward formulations that include conflicts. First creating some results with a reward function based on the total number of LoSs, however, is thought to be a solid basis for further iterations. Later on in the MSc thesis, when completing research activity 4 (the RL model that also includes information on the previous airspace structure, see section 3.2.4), other reward functions may be tried to see if the results improve. For now, however, the reward for the concept experiments is formulated as follows:

$$\text{Reward} = -\frac{\text{Total number of LoSs}}{100} \tag{4.1}$$

The division by 1000 was implemented after better results were found during a number of initial training runs. Recent research has shown that reward re-scaling can improve the stability of a reinforcement learning model [23].

4.2 Concept Experiments Set-up

In this section, the concept experiments (the experiments for research activity 1) will be outlined. Before going into the elements of the experiments, subsection 4.2.1 is added to make clear how these concept experiments lead up to the final experiments and with what purpose they have been designed. For an additional overview of this, the reader is referred back to figure 3.1. Subsection 4.2.2 elaborates more on the use of the Bluesky Open Air Traffic Simulator in this work. Then, section 4.2.3 presents the choice of airspace type, whereas subsection 4.2.4 dives into the actual experiment design for the concept experiments.

4.2.1 Concept experiments as a step to the final experiments

For the concept experiments, a RL learning model is built that is capable of setting the airspace based on a single traffic scenario. As explained previously, no information is yet present on the previous airspace. It can in a sense be viewed as a recreation of the work performed in [8], that showed that using RL techniques to set the airspace structure has a positive effect on the conflicts and LoSs experienced in various traffic scenarios. The scenarios used for these concept experiments, however, will be different than what those researchers used in their work. This is due the decision to use unmanned aviation in an 'above-buildings' setting, a combination that wasn't investigated yet.

The questions posed in subsection 3.2.1 of chapter 3 form a solid guideline for the goals of the concept experiments. The first five that concern the selection of RL model type and the formulation of the state, action and reward have been addressed in section 4.1 earlier this chapter. The final question, which concerns the performance of the RL model's performance versus an airspace that employs a uniform, fixed structure, is to be determined by means of the concept experiments that are outlined in this section.

For the development process of the final RL model (research activity 4), having these intermediate concept experiments poses several advantages. Firstly, it forces the completion of a working RL model and all other code for running experiments at an early stage in the MSc project. This RL model will, furthermore, at its core be very similar to the final model in this work. It is already at the point where it can take the safety metrics of conflicts and/or LoSs into account to come up with an airspace structuring. At that point the reward formulation would be entirely based on the cruising conflicts and/or LoSs that occur in a given airspace structure. All that needs to be added into this reward formulation is the information on vertical deviations, which then functions as a metric of 'expense' to reconfigure the structure.

4.2.2 Use of BlueSky Open Air Traffic Simulator

For the concept experiments, use is made of the BlueSky Open Air Traffic Simulator [24]. 'BlueSky' has been created by the ATM/CNS department of the Faculty of Aerospace Engineering at Delft University of Technology, in response to the need to be able to compare efforts and results in the field of ATM research. It is an open source and open data approach to air traffic simulation written in Python and it was chosen to use BlueSky in this research for various reasons. Firstly, there is a substantial body of experience with this tool available within the department, that may be used to rapidly get going with the simulations that are needed to complete this work. Secondly, the use of BlueSky makes it possible to take advantage of its performance library, which includes the specifications of many aircraft types, and many other pre-programmed features like for example conflict detection & resolution algorithms, a graphical user interface (GUI) and datalogging. Lastly, by using BlueSky any results will be obtained in a way that is easily verifiable, reproducible and can be extended upon in future research (within or outside the ATM/CNS department).

4.2.3 Choice of Airspace Type

As this work is a research that builds on the layered airspace concept, an important variable to set to further define the airspace is the aircraft types that will be used in the experiments. In general, one could go for an airspace with manned, unmanned or mixed (both manned and unmanned) aviation. In this research, it was chosen to go with unmanned (urban) aviation, as this poses several advantages in achieving the research objective. The arguments are discussed in the following.

The first argument is that fact that future unmanned, urban, aviation is expected to have higher traffic densities than manned aviation. The proof of concept for dynamically re-configuring airspaces with RL techniques will be stronger if the experiments are set in a setting with very high traffic densities. The best would naturally be to check both manned and unmanned cases, but in the interest of delivering a work that is focused primarily on the transitioning of airspaces with a machine learning model, this option is omitted here.

The second argument for an airspace with unmanned aviation is that this type of aviation generally employs more trivial routes that commercial manned aviation [25], again reducing the scope of the research and leaving room for a good look at the dynamic re-configurations, the research objective of this work. It is mostly in the simulation development phase where the major gain of this choice will surface, as the time saved by creating more trivial scenarios may be used for the development of proper RL models.

The third argument for using manned aviation in this research into dynamic airspace reconfiguration with RL is the potential applications of these methods. The idea of dynamically changing the vertically stacked airspace to improve safety, is founded on the principle of moving the flight altitude of aircraft in such a way that conflicts & LoSs occur less often. The result of the new configurations would be that aircraft sacrifice an optimal altitude in terms of energy efficiency, in order to achieve better safety as measured by various metrics. This sacrifice in energy efficiency is relatively larger for manned aircraft (where flight altitude is a more dominant factor in efficiency) than it would be for unmanned aircraft, which could lead to an earlier adoption of RL techniques to enable dynamic airspace re-configurations. On top of that, it expected that it is more trivial to communicate an airspace configuration change to a system of live aircraft when they operate on an urban level than if they are performing flights over a longer distance like civil, manned, aviation generally does. This could make adoption of dynamic airspace configuring more attractive for an unmanned aviation setting.

4.2.4 Experiment Design

The experiment design for research activity 1 is done by making decisions on various aspects of the simulation. In the following, all of these aspects will be explained. The reader is reminded that the goal of the concept experiments is to be able to complete research activity 1 (see section 3.2.1). That is, to create a RL model that selects the optimal airspace structure, given only the information on the future traffic scenaro.

Simulated environment

As stated earlier in section 4.2.3, this research will look at airspaces for unmanned aviation. How the airspace is best configured for this type of aviation in the future is still under discussion within the ATM community. Generally, two possibilities are 'very low level' traffic and 'above-building' traffic, both of which have been investigated in previous research. The Metropolis project [7], for example, looked at 'above-building' traffic, while [8] looks at air traffic that operates in a grid-like pattern that you would find in a 'low-level structure. In this research, the main goal is to investigate the possibility of having a RL model that incorporates previous traffic information while outputting the most suitable airspace structure. To keep this focus, the choices for the configuration of the experiments, and as a part of that the simulated environment, should enable that. For this, it is desirable to be able to analyse the results with the airspace structure being the only affecting variable. In a low-level airspace, which includes the definition of directional streets, the aircraft may be forced away from the heading limits as to not hit a building. These are considerations that are not desirable to include in the results, as they will contaminate the outcomes with decisions that are not solely a function of the airspace structure. Because of this, it was chosen to go with the more trivial simulation environment with traffic that is 'above-building'. This furthermore eases the definition of routes and enables them to be linear, greatly simplifying the creation of the scenarios.

Airspace structures

The RL model is to set the airspace structure for a scenario. Before going any further into the experiments design and what such scenarios might look like, the airspace structures are discussed in greater depth. As mentioned previously, this work builds upon the recommendation of the Metropolis project [7] to look further into the 'Layered' airspace concept, which has the airspace segmented into vertically stacked bands, where each altitude layer limits the horizontal travel to within an allowed heading range. To create such an airspace for the (concept) experiments, several parameters need to be set. First of all the three-dimensional bounds of a cubic airspace are defined. For the horizontal plane this definition comes down to setting a minimum and maximum latitude and longitude ($lat_{min}, lat_{max}, lon_{min}, lon_{max}$). For the sake of simplicity in defining the area, the centre of this square was taken at the position of zero latitude and longitude ($lat_{centre} = lon_{centre} = 0$). The sides of the square were set to have a length of 0.3 degrees latitude and longitude respectively, which corresponds to around 18 Nm. The total experiment area then has an area of 1.5 x 1.5 = 2.25 Nm^2. The minimum and maximum altitude (alt_{min}, alt_{max}) then complete the dimensions of the airspace. The minimum altitude, alt_{min}, was set to $1100ft$, while the maximum altitude, alt_{max}, is set to $3500ft$. A total of eight vertical layers are defined that are distributed uniformly throughout the airspace and each have a height of $300ft$ each. Note that the layers are distributed uniformly and are fixed in terms of altitude, but that is the heading range that will be varied by the RL model. Figure 2.4 in chapter 2 contains a side view of what this looks like for a uniform airspace structure, where the heading ranges are of equal size. In the airspace that is to be set by the RL model, these heading ranges will vary according to the traffic scenario.

Aircraft type

For the experiments, it was chosen to simulate an airspace with a large amount of 'light-load' drones. This is a type of drone that is expected to be notably present in the skies of the future. They will likely be used for medical or lightweight industrial deliveries and for the completion of more traditional forms of delivering parcels of couriers to businesses and consumers [5]. For the sake simplicity, the drones in the experiment will all be of the same type and will therefore have the same performance specifications. It was chosen to go with with the type of drone called 'DJI Mavic Pro', as its specifications resemble what [5] describes as being a prominent aircraft type in our future airspace, as well as the availability within the BlueSky aircraft performance libraries.

Minimum Separation

In this work there is a lot of attention to the safety metrics in the analysis of the various airspaces. Both the term 'conflict' as well as 'LoSs' have been used extensively. For unmanned aviation, no current standards are in place at this point in time that define these. Furthermore, what is considered a safe separation distance is in reality a function of the (also currently unknown) traffic density. In previous research [26], however, a horizontal separation of $50m$ has been used. For the final experiments of the research, this could be a suitable value to use. For this first research activity, where a RL model for the Bluesky environment is set-up for the first time, however, it was chosen to increase the minimum horizontal separation to $200m$. This approach was thought to be best way to create a large number of LoSs for the model to learn with, while keeping the computational costs down. Alternatives such as higher traffic densities or longer episode times would have caused longer training times, reducing the speed at which iterations to the scenarios and the state, action and reward formulations could be tested. For the vertical separation, one airspace layer ($300ft$) is taken. By default, aircraft cruise in the middle of their assigned layer. This means that aircraft cruising in adjacent layers will not be in conflict.

Traffic scenarios

An important component of the experiments are the traffic scenarios that are run during training and testing. In the interest of a work that goes into depth on the dynamic airspace re-configuring, it was chosen to only consider the cruise phase of flight and not to consider the take-off and landing operation in this research. Each aircraft in the simulation is therefore initialised on one of the four edges of the scenario. The aircraft spawn locations are chosen at random on the edges of the 'Experiment area'. Within an episode, aircraft are spawned at a fixed rate. The decision for which edge an aircraft is spawned on is determined by a set of probabilities for the edges, which is fixed per episode. For example, if the probabilities are [North, East, South, West] = [0.85, 0.05, 0.05, 0.05], on average 85% of the traffic gets initiated from the northern edge for that episode. On the other hand if a [North, East, South, West] = [0.25, 0.25, 0.25, 0.25] setting was selected, the traffic in the scenario will be approximately uniform. During training, the values of this [1 x 4] array are set at random for every episode. This means that the agent sees different kinds of traffic scenarios, ranging from rather uniform traffic to traffic which comes predominantly from one or two of the four edges.

At initialisation, the aircraft is given a random angle between 45° - 135° degrees from the edge. The corresponding heading is computed to properly spawn it in the simulation environment. An illustrative example of an aircraft route is given in figure 4.3. The altitude at which the aircraft is spawned corresponds with their heading in order to ensure that aircraft are within the correct layer upon initialisation. The linear aircraft routes have three way-points, shown as green dots in figure 4.3 and an exit point, shown in red in the same figure, to guide the aircraft. The exit points naturally follow from the initial spawn location and heading, while the way-points are added such that the aircraft will stick close to its intended route, even if the aircraft deviates from this to resolve a conflict. The climbing and descending happens in accordance with the specifications of the aircraft type used and is almost vertical. The speed at which the aircraft fly along the routes is set to the maximum cruising speed of 18 kts. For this first research activity, the logging of parameters for analysis happens in the complete experiment area. In future research activities, a logging area that is offset by some fixed distance from the experiment area perimeter will be used to omit any edge effects from occurring. This will be explained in more detail in chapter 6.

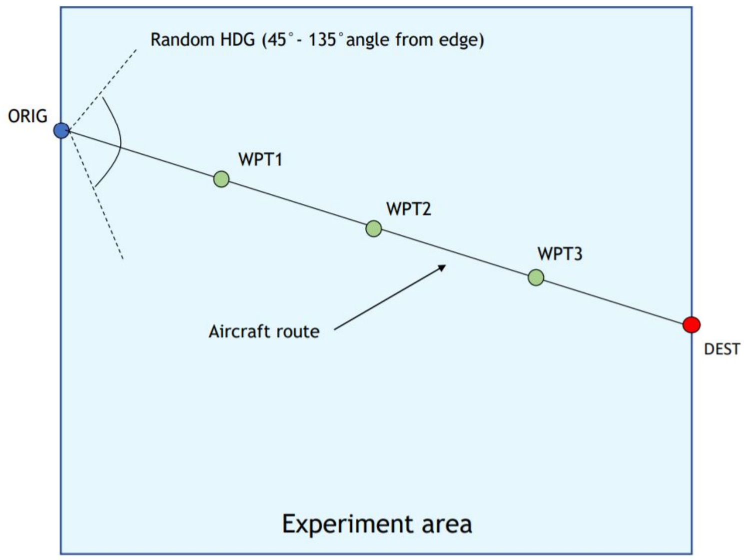


Figure 4.3: Illustration of the experiment area, including an example of a flight route

After reaching its destination on the edge of the experiment area, an aircraft gets deleted from the simulation. The data of an aircraft is saved whenever it leaves the experiment area. Data is collected on efficiency metrics like 2D & 3D distance flown as well as the flight time. Furthermore, and perhaps more relevant for this research, the safety metrics are logged, which save the number of conflicts and LoSs that an aircraft has experienced on its flight. To obtain the results of the experiment, a wind-up time of 8 minutes is taken before the state (with information on the headings of the aircraft in the area) is sampled. At this point, an action is requested from the RL model. After 12 minutes, when all aircraft have reached their new correct layer, the logging of the data starts. During training of the model for this research activity, episodes with a total duration of 40 minutes were simulated, which implies that there are 28 minutes of simulated traffic during which information on the LoSs is collected for the setting of the reward (see section 4.1.5 and equation 4.1).

Conflict detection

Even when coordination efforts are performed, like for example setting an appropriate airspace for the traffic scenario, it remains essential for safety to also have a conflict detection and resolution (CD&R) system in place. First, the conflict detection algorithm is elaborated upon. After that, a brief discussion on the chosen conflict resolution algorithm.

For conflict detection, the experiment employs a state-based conflict detection method equal to that used in [8]. This is a widely used method in the field, as is further solidified by its use in [3]. It assumes linear propagation of the current state of all aircraft involved. The time to closest point of approach (CPA), t_{CPA}, is computed with equation 4.2 below

$$t_{CPA} = -\frac{\vec{d}_{rel} \cdot \vec{v}_{rel}}{\vec{v}_{rel}} \tag{4.2}$$

where $\vec{d}_{rel}$ is the Cartesian distance vector between the involved aircraft (in meters), and $\vec{v}_{rel}$ the vector difference between the velocity vectors of the involved aircraft (in meters per second). With this t_{CPA} known, the distance to CPA, d_{CPA}, can be computed by means of equation 4.3 below.

$$d_{CPA} = \sqrt{\vec{d}_{rel}^2 - t_{CPA}^2 \cdot \vec{v}_{rel}^2} \tag{4.3}$$

When this d_{CPA} is smaller than some pre-defined separation distance, one may compute a time-interval during which separation will be lost. This is done through equation 4.4 below.

$$t_{in}, t_{out} = t_{CPA} \pm \frac{\sqrt{R_{PZ}^2 - d_{CPA}^2}}{\vec{v}_{rel}} \tag{4.4}$$

In other words, the above allows the computation of the already often used LoSs safety metric. Conflicts, on the other hand, are said to occur when $d_{CPA} < R_{PZ}$ and $t_{in} < t_{lookahead}$, where R_{PZ} is the radius of a protected zone (the minimum horizontal distance) in meters and $t_{lookahead}$ is the look-ahead time is seconds. For this work, a look-ahead time of 30s is implemented, as was done for [8] that also used unmanned aviation in a setting where conflict detection and resolution was implemented.

Conflict resolution

For experiments in research activity 1, it was opted to not make use of a CR algorithm. With this CR setting set to OFF, many LoSs occur per episode, reducing the need for more computationally expensive ways of creating this information, such as longer episode lengths or higher traffic densities. As this first research activity is intended mainly as a step up to the final RL model, which is to also take into account previous airspace structures, a shorter development time to a learning model was valued more than very realistic end-results.

4.2.5 Simplifying assumptions

Before proceeding, it is of importance to introduce all simplifying assumptions that define the gap between simulation and a real-life setting. First of all, non-linear routes are not considered. When an aircraft is spawned it is given several way-points that lie on a straight line towards its destination. Except for conflict resolution manoeuvres or manoeuvres to move to another airspace layer, all aircraft follow these routes with a constant speed, constant altitude and without heading deviations. The effects of wind are neglected. Though it is still unclear where future unmanned aviation will operate, an 'above-buildings' airspace is assumed in this work. Furthermore, only en-route scenarios are considered, which means there will be no aircraft taking-off or landing in the simulations. The aircraft used will all be of the same type and are assumed to have constant performance, implying that (component) failures are not simulated. For the airspace, constant traffic density is assumed for the duration of individual scenarios. The traffic density, however, will be varied between scenarios to be able to investigate its effect on the airspace configuration selected by the RL model. A fixed number of layers is further assumed, which seems reasonable as there is an argument that a RL model would converge towards as many layers as possible. As mentioned previously, it is assumed that the RL agent has full information on the future traffic density and trajectories. Also, no delay in communication of information such as aircraft states, conflict resolution advisories or airspace re-configurations is taken into account. Finally, concerning the optimisation of safety metrics, it is assumed that fewer conflicts lead to fewer LoSs. Although this relation is not as straightforward as assumed here, there is a strong correlation between the two that makes this assumption reasonable. Finally, no edge-effects due to a finite simulation space are assumed, as these are omitted through an experimental setup with sufficient flight space around the measurement area.

4.3 Hypotheses for concept experiments

For the experiments for research activity 1 as outlined in section 4.2, various hypotheses may be setup. They are presented in this section. Subsection 4.3.1 presents some hypotheses that relate to the training of the model. After that, in subsection 4.3.2, the expected results of testing the model are presented.

4.3.1 Hypotheses - training of the Reinforcement Learning Model

For the training phase, it is expected that the model is capable of learning a policy that allows the agent to collect better rewards over time. The airspace structures that are selected towards the final episodes should demonstrate that they have been set to fit the traffic scenario at hand. It is, however, expected that the reinforcement learning lacks some precision. For example, for a perfectly uniform traffic scenario, setting a perfectly uniform structure may achieve better results than the structure output by the RL model. Although it is expected that the model will output an almost uniform structure, it probably won't be as suitable as a perfectly defined uniform structure.

4.3.2 Hypotheses - testing of the RL model

For the testing of the model, it is expected that the performance of the model depends on the traffic scenario that it is presented with. For a uniform traffic scenario, a perfectly uniform structure is expected to be ideal. However, as mentioned previously, even after training the RL model is not expected to be perfect. It is therefore hypothesised that the airspace structure as output by the RL model will be close to uniform when it is tested with uniform traffic scenarios, but not exactly uniform.

For testing scenarios where there is traffic coming heavily from one or two of the edges, in essence very non-uniform structures, it is expected that the model is able to capture this at least to some extent and select a different airspace than it would for uniform structure. For example, for heading ranges containing a higher traffic density, the model should divide this heading range over multiple layers, as to decrease the local traffic density.

Chapter 5

Results of Research Activity 1

This chapter discusses the results that have been found in the completion of research activity 1. Firstly, in section 5.1, the results training process of the RL agent are presented. Next, in section 5.2, the performance of the RL model when presented with several scenarios for testing is showed. Here, it is discussed to what extent the model after training is capable of setting the a suitable airspace structure for several new scenarios that the model was not trained in. A look will be taken both at the resulting LoSs over the scenario, as well as the resulting decision for the airspace structure. Finally, section 5.3 discusses the relevance of the results for the rest of the research. This last section in particular, may be viewed as the bridge to the final experiments that are to performed after the midterm review. The plan for those will be explained more in chapter 6.

5.1 Training of the RL agent for Safety Optimised Airspace Structuring

After 90000 episodes, the training was terminated and the results were analysed. Firstly, the reward evolution was examined. This will be explained in subsection 5.1.1. Following this, a small set of states, actions and rewards of the model during the end of training were looked at. They latter will be explained in subsection 5.1.2.

5.1.1 Reward evolution during training

For a first indication of the results from training, the mean of the reward was tracked and plotted. Figure 5.3 below shows the rolling mean over the last 1000 episodes during training.

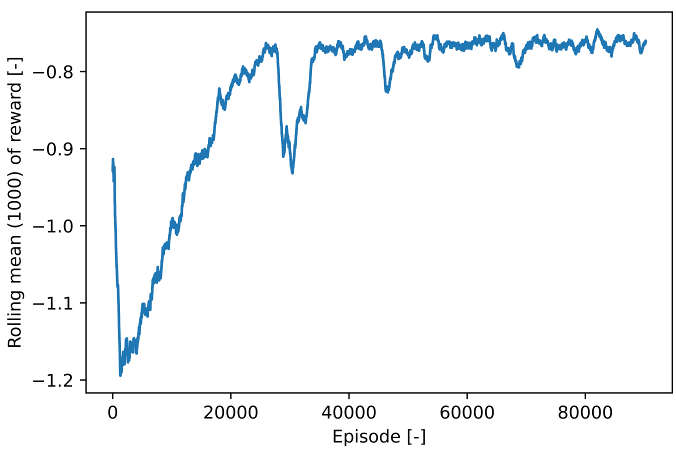


Figure 5.1: Training evolution over 90000 episodes

In figure 5.3, several things may be observed. Overall, it may be seen that the model (apart from the very beginning) start to obtain better rewards as training time goes on. At around 3000 episodes, the maximum rewards are reached. From that point it can also be observed that there is a decline in the mean reward that is received. The model, however, recovers from this and keeps a relatively constant mean reward around the -0.77 mark from that point onward.

5.1.2 Airspace structures in the final training episodes

Though the reward evolution shown in the previous subsection gives the indication that the model has learnt to obtain a better reward over time, this doesn't necessarily imply that it has learnt to perform the intended task of setting airspace structures for a given traffic scenario. Even before testing the model, more information on the model's ability to assign safety optimised airspace structures may be gathered by taking a look at the episodes during the final episodes of training.

Examples of selected airspace structures

Two episodes from the final ten episodes of training have been selected to illustrate the behaviour of the model after the first larger training run. The states, selected structures and corresponding rewards are shown below. The reader is briefly reminded that the state formulation is a normalised [1 x 10] array containing the number of aircraft per heading bin (see section 4.1.3), the action is a [1 x 8] array which sets the heading ranges for the eight layers (see section 4.1.4) and the reward is a negative value determined through a rescaling of the LoSs experienced in an episode (see section 4.1.5).

Example 1:

- **State:** [0.014 0.028 0.028 0.085 **0.371** **0.414** 0.042 0.000 0.014 0.000]
- **Selected airspace structure:**

Figure 5.2: Selected airspace structure for an episode at the end of training - Example 1

- **Reward:** -0.44 (44 LoSs)

At this episode close to the end of training, the model is not behaving as intended. Looking at the state, it can be seen that 78.5% of the aircraft have headings that fall in bin 4 and 5 (given in **bold** in the state formulation). This means that a majority of the aircraft had headings in the 144° - 216° range. The expected result was that an airspace structure would be selected with more layers for that range, such that the aircraft get spread over the layers better and, as a result of that, create fewer LoSs throughout the episode. The selected airspace shown in figure 5.2, however, clearly doesn't have extra layers for the more heavily occupied heading ranges. On the contrary, it looks rather uniform, raising the suspicion that the RL model might have learnt to use uniform structures regardless of the uniformity of the traffic scenario.

Further, it may be noted that the reward that is obtained is still 'good', as determined by comparing it to the mean of the rewards at the end of training (figure 5.3). Though the state doesn't immediately indicate how many aircraft there are in each layer during the complete scenario, but rather gives an indication at one point in time (which should be representative for the episode), it does show that having many aircraft within a limited heading range is not penalised heavily. The cause for this likely lies in the way that aircraft were created in the scenarios during this training routine. As explained in section 4.2.4, a new aircraft gets spawned on one of the four borders of the experiment area with a given probability (that changes every episode) and is subsequently given a random heading from that point from that point. That random heading was allowed to be a relatively wide 45° to either side of the aircraft in these simulations, which might have still caused a relatively high degree of uniformity in the traffic scenarios (for which uniform structures are indeed best). The wide spawning angle may have also contributed to a situation where states are less representative to the entire scenario than they could have been. The state is in full determined by 1) the edges where the aircraft have spawned and 2) the headings that they were given. A high degree of randomness for the latter decreases the consistency of traffic headings over the course of an episode. In essence, it may then seem from the state that the traffic in a scenario was very non-uniform, where in reality the degree of uniformity was higher. This factor may have also led to circumstances in which the RL model has difficulty learning a successful policy. Lastly, it could be that the traffic density, and thus the likelihood of aircraft creating conflicts, hasn't been high enough for the model to heavily penalise less optimal structures.

Example 2:

- **State:** [0.000 0.178 **0.329** 0.082 0.027 0.068 0.082 0.137 0.055 0.041]
- **Selected airspace structure:**

Figure 5.3: Selected airspace structure for an episode at the end of training - Example 2

- **Reward:** -0.83

This second example is presented to give backing to some of the statements made previously. Here too, it is seen that a non-uniform traffic state results in the RL model selecting an airspace structure that is very uniform-like. It is again expected that the manner of creating the aircraft in the scenario plays a role in this finding. It may also be noted that, even-though the scenario is more uniform than in example 1 (and the airspace structures are similar), the reward is worse than before. This indicates once more that the randomness creating the aircraft likely introduces a discrepancy between the state and the actual headings of the aircraft throughout the episode.

Future improvements to the RL model, traffic scenarios & analysis methods

From training results alone, a number of improvements for future runs can already be defined. The first follows from the suspected issue with the spawn headings explained previously. It is thought that creating the aircraft with headings in a narrower heading range (for example 22.5° to either side, as opposed to 45° to either side) could help the creation of more non-uniform scenarios. By doing this the model would be penalised more heavily for incorrectly choosing uniform structures. Another option could be to make a change in the activation functions, for example to use a 'sigmoid' function as opposed to a 'tanh' function in the output layer. During the training it was observed that the model had some issues exploring a wide range of different airspace structures. As the input (the state) is always positive, there are no negative input values to multiply that would make the output negative as well. Moving from a sigmoid (which covers the range from 0 to 1) instead of a tanh (which covers -1 to 1), could perhaps make the model explore a wider set of airspace structures and allow it to learn a policy that doesn't always select uniform structures. Lastly, increasing the traffic density to force a higher number of LoSs for non-optimal structures could also improve the results.

Besides from the aforementioned improvements to the setup of the RL model, there is also room for improvement in the way the data of the experiments is processed. Specifically, the possibility to quickly track the total or average number of aircraft that were present in a layer during an episode would be useful for rapidly making sense of structures that are outputted and the rewards that are given.

In any case, it is clear that the RL model can still be improved considerably. Based on what was presented in this section on the training of the RL model, it is not expected that the performance during testing will be too strong. Nevertheless, three tests will be performed to check this and, at a minimum, get some experience in the full cycle of training, testing and analysing the results of a RL model in this midterm phase. The results of the performed tests are presented in the next section.

5.2 Testing of the RL agent for Safety Optimized Airspace Structuring

Where the results from training the model were shown in the previous section, this section will display the results of testing. For the testing, the saved model (after 90000 episodes) was presented with some different scenarios. Doing so enables an analysis on the performance of the trained model in terms of selecting airspace structures for optimal safety. As previously touched upon in the introduction of this chapter, two factors will be looked at at this stage. The first is the resulting LoSs that are experienced during the test. Secondly, the chosen airspace structures will be shown as well in order to see if they make sense for the given traffic scenario.

Three testing scenarios were presented with the model. They are the following:

1. A uniform traffic scenario, with aircraft headings equally distributed in the 0° - 360° range.
2. A traffic scenario where almost all (94%) of the traffic comes from one border of the square experiment area. In this case, the left/west edge was selected, meaning that most of the aircraft headings were in the 45° - 135° range.
3. A traffic scenario where most (80%) of the traffic comes from two borders of the square experiment area. In this case, the north/top and east/right were selected, meaning that the traffic predominantly had headings in the 135° - 315° range.

The resulting rewards and selected airspace structures are presented in the following three subsections.

5.2.1 Test 1 - Uniform traffic scenario

For the uniform traffic scenario, with traffic coming equally from all four edges, the following results were found:

<u>**Test 1:**</u>

- **State:** [0.076 0.061 0.182 0.106 0.152 0.121 0.106 0.061 0.0152 0.121]
- **Selected airspace structure:**

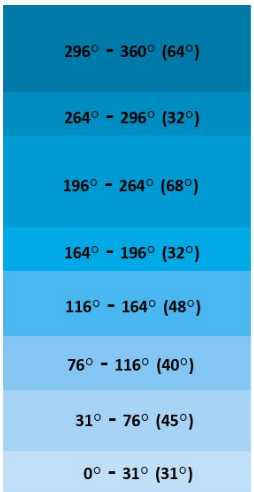


Figure 5.4: Selected airspace structure for test 1

- **Reward:** -0.52

For this it can be seen that the model actually behaves like it should. A uniform traffic scenario, as reflected relatively well in the state, is assigned an airspace structure that it is uniform as well. As discussed in section 5.1.2, however, it was seen during training that an almost uniform structure is often selected, regardless of the exact state or traffic scenario. This makes it hard to say with confidence that this structure was selected by the model based on it recognising a uniform scenario.

5.2.2 Test 2 - Traffic mostly from one border

For the one edge traffic scenario, where 94% of the traffic came from the left/western edge, the following results were found:

Test 2:

- **State:** [0.014 **0.178 0.411 0.315** 0.014 0.041 0.000 0.000 0.000 0.027]
- **Selected airspace structure:**

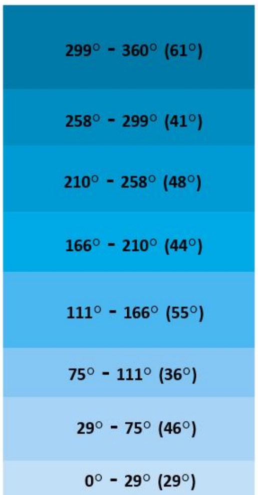


Figure 5.5: Selected airspace structure for test 2

- **Reward:** -1.14

For this second test, which has traffic so heavily coming from one side, the ideal solution should most likely feature an increased number of layers for the directions that are most prominent in the traffic scenario (45° - 135°). It can be seen in the state that this over-representation of aircraft in those direction is, at a minimum, capturing this with higher values in the 2nd, 3rd and 4th heading bins. As is seen, though, in figure 5.5, a structure with more layers for those is all but the case. The selected traffic scenario is still fairly uniform. An encouraging sign, however, is that the reward is indeed not great, as determined by looking at the mean rewards during the latter phase of training (see figure 5.3), meaning that an unsuitable airspace like this at least penalised to some degree. The measures explained previously concerning the decreasing of the randomness of initialisation headings of aircraft should improve that further in future training runs.

5.2.3 Test 3 - Traffic mostly from two borders

For the scenario with traffic mostly from two borders (north/top & east/right), it was also checked how the RL model would react. The following results were found:
Test 3:

- **State:** [0.044 0.029 0.088 0.074 **0.147 0.221** 0.029 **0.147 0.176** 0.044]
- **Selected airspace structure:**

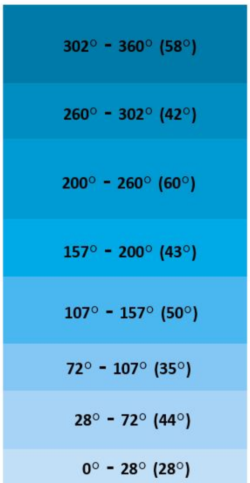


Figure 5.6: Selected airspace structure for test 3

- **Reward:** -0.71

For the third test, it is again seen that the state captures the fact that more aircraft are present in certain heading bins (the 5th, 6th, 8th and 9th). However, the airspace structure doesn't show any signs of reacting to that. A uniform structure is once more the result. The reward is close to the mean reward found during the final phase of training, which seems to make sense. The scenario is not-uniform, so will not get the best rewards, but also not so biased to a traffic direction that it causes an overly large number of LoSs.

5.3 Relevance of the results

This section contains a brief discussion on the relevance of the results obtained for research activity 1. They are discussed both the light of the work that is to follow in the next phase of this MSc thesis (subsection 5.3.1), as well as the relevance for ATM research in the broader sense in subsection 5.3.2. The latter will, naturally, be somewhat limited as the results obtained so far are preliminary results for larger research into dynamic airspace re-configurations by means a reinforcement learning model.

5.3.1 Relevance for current MSc Thesis

Looking at the results from training and testing in the previous two sections, it is seen that the setting airspace structures with a RL model leaves room for improvement. In terms of relevance for the rest of the thesis, the work done to create a model that is capable of learning (although not exactly as desired) in the Bluesky environment is deemed a solid basis for continuing into the final part of the MSc thesis. The fact that the model is improving, means that the model is constructed properly and that it can move in the direction of maximising the value of the rewards. What seems to be be happening is that the traffic scenarios that the model is trained in do not seem to sufficiently penalise incorrect structures. When this issue if resolved, the model and the traffic scenarios for the episodes may readily be iterated upon in the pursuit of a RL model that is capable of assigning airspace structures, while taking into account the previous airspace structure. For the second part of this MSc Thesis project, the focus will lie almost entirely on researching this possibility. Chapter 6 discusses the further research activities for this part of the work in more detail.

5.3.2 Relevance in larger ATM research

Finding suitable structures is relevant for future operations in terms of guaranteeing safety. Furthermore, previous research has found that there is no optimal structure for all situations and that this is directly depending on the current traffic scenario. The RL model, which has been developed for the first research activity, does not yet demonstrate the capability to set such airspace structures yet. It does, however, highlight some of the factors that are important to consider when looking into this. Mainly the method followed to create the traffic scenarios and the choices for state, action and reward formulations have been shown to greatly influence the learning capabilities of a RL model.

Chapter 6

Plan for final phase of the MSc Thesis

In this chapter, the plans for the final phase of the MSc Thesis are layed out. As already hinted upon in chapter 3, this phase will consist primarily of completing research activities 2-6 defined earlier. Chapter 3, which covered the problem definition, focused mainly on the defining the scope of the work. In essence, it looked at the 'what?' and 'why?' within this research. This chapter, on the contrary, will go a bit deeper into the 'how' of the defined activities. Sections 6.1 through 6.5 discuss each of the remaining research activities in greater detail and aim to specify what will be done to complete them to the extent that is currently known.

6.1 Research activity 2: Rules for moving traffic into structure

As defined in section 3.2.2, the second research activity is concerned with defining the rules for traffic moving into a new airspace structure. It is expected that these rules affect the number of vertical conflicts that occur during an airspace reconfiguration. Because it is envisioned that these type of conflicts form a part of the reward formulation of a RL model for dynamic airspace reconfigurations, it seems relevant to gain an insight in the manner of performing the movement of aircraft into new structures.

To gain this insight, some simple experiments will be performed. In these experiments, a number of traffic scenarios will be defined in which operating aircraft must adapt to a new structure midway through their flight. The independent variable in this experiment will be the manner in which aircraft transition from an initial airspace structure to the next. There are various ways that the transition may be done. At least the following will be investigated:

1. Moving all aircraft to their new correct layer at the same time
2. Moving aircraft on a layer-by-layer basis (from top to bottom, or vice-versa)

Two factors will be used to judge which manner is the most suitable for use in future simulations of this MSc Thesis. The first is the total number of vertical conflicts experienced during the reconfiguration. Obviously, a lower number here would be more favourable. The second is the 'reconfiguration time', which will be defined as the time needed from airspace structure change, until aircraft have settled into the new structure. Though this factor doesn't necessarily influence the number of conflicts in a traffic scenario directly, a large 'reconfiguration time' does imply that aircraft spend less time in a structure that has (in later research activities) been selected to suit the traffic scenario. In other words, aircraft could spend longer than necessary in an sub-optimal structure, which could compromise the benefits of performing the reconfiguration in the first place.

It is possible that the outcomes of this experiment depend on some of the settings of the traffic scenarios used in the simulations. The traffic density or for example the settings for conflict resolution (on/off), specifically, might affect the results of this experiment.

6.2 Research activity 3: Rules for conflict resolution in a layered airspace

For research activity 3, the focus lies on the rules for conflict resolution in a layered airspace. This is relevant as some heading manoeuvres can imply a violation of the defined heading range limits by the layered airspace structure. Section 3.2.3 already explained three options to deal with this issue: 1) allowing aircraft to break the heading limits set for their layer, 2) having aircraft strictly adhere to the defined structure or 3) limiting the avoidance to the heading ranges of the layer. It was chosen to resort to existing work done by the department to select suitable rules, as the department has recently done similar work that may readily be used in this research.

The resolution algorithm that will be used is the MVP algorithm, which has proved effective in reducing the effect of resolution manoeuvres on flight efficiency while still guaranteeing minimal losses of separation (LoSs) [10]. The geometric resolution corresponding to the MVP is shown in figure 6.1.

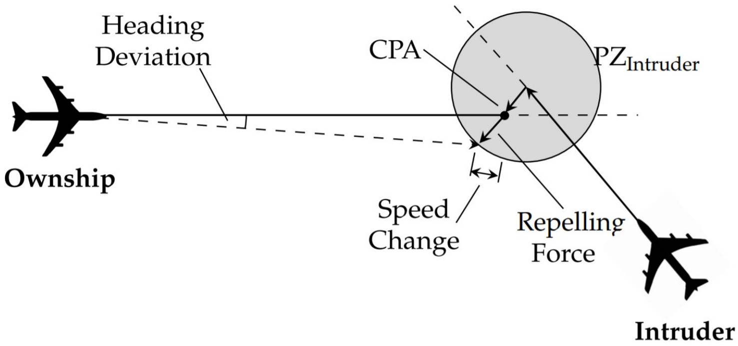


Figure 6.1: MVP resolution. Adapted from [4]

As also explained in [10], for aircraft in conflict, the predicted positions at CPA repel each other. The repelling force is converted to displacement of the predicted position at CPA, such that the minimum distance between the aircraft equals the required minimum separation. These displacements then imply a new advised heading and speed, such that it increases the predicted CPA. It also means that both aircraft take complimentary measures to evade each other, making the MVP implicitly coordinated. The resulting calculations are computationally light, and the geometric representation allows for taking into account any other constraints easily. A downside however is that it is solely based on conflict geometry and that it therefore may result in aircraft opposing the flight direction proposed by their flight plan.

6.3 Research activity 4: RL model for airspace reconfiguration

The fourth research activity consist of creating the RL model that selects an optimal airspace structure for a traffic scenario, while also taking the previous airspace structure into account. A sequence of such airspace selections, performed with information on the traffic and the previous structure is the essence of 'dynamic airspace reconfigurations', the main topic of this MSc Thesis. Firstly, the updates to the traffic scenarios for dynamic airspace reconfiguration are discussed in subsection 6.3.1. After that, in section 6.3.2, the updates to the RL model are discussed. Finally, in section 6.3.3 it will be explained how the methods of analysis will be improved from this research activity onwards.

6.3.1 Updating traffic scenarios for dynamic airspace reconfiguration

The idea of dynamically changing airspace structures stems from the desire to improve the safety in traffic scenarios that change considerably over time. Though airspace structures are by no means the only way of doing this, it is thought that an airspace structure that, as it were, evolves along with the traffic can have a positive effect on the safety. To be able to investigate this, new traffic scenarios will first have to be defined for this research activity. These scenarios will not have traffic that is created with a fixed heading distribution for the whole episode, like in research activity 1, but will feature changes in the traffic distribution within an episode. It is envisioned that, also like in research activity 1, the traffic in the episodes is still created by spawning aircraft on the edges of the experiment area with a certain probability per edge. As opposed to fixing these probabilities for a complete episode, this research activity will have non-constant traffic scenarios throughout the episode. The way in which the changes happen may be either continuously or in a discrete manner.

Logging area. Some further improvements to the way logging occurs will be made in this research activity as well. Specifically, a logging area (see figure 6.2), that lies within the square of the experiment area with a fixed offset, is implemented. This is to make sure that the experiment data is not affected by aircraft that experience a reduced likelihood of conflicts near the edge of the area due to the absence of aircraft in their vicinity. Furthermore, initialising the aircraft outside the logging area prevents recording immediate LoSs between a newly created and existing aircraft, for which there would be no chance to perform a conflict resolution manoeuvre.

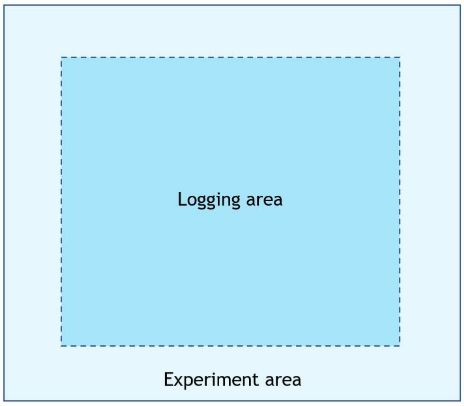


Figure 6.2: Updated logging area at fixed offset of experiment area borders to avoid edge effects

Traffic density. The traffic density will vary between the experiments in this fourth research activity. It was chosen to use three levels of traffic density (low, medium, high), for which the number of aircraft per scenario are chosen based on the metric of 'time-in-conflict'. If a scenario is run and the total 'time-in-conflict' is 2.5% of the total simulation time, the scenario is considered to have a low traffic density. For 'time-in-conflict's of 5.0% and 10.0%, the traffic scenario's are named 'medium' and 'high' in traffic density, respectively. The RL agent will be trained at a medium traffic density and then tested at low, medium and high traffic densities, to also be able investigate the efficiency of an agent performing in a traffic density other than it was trained for.

6.3.2 Updating the RL model for dynamic airspace reconfiguration

As previously mentioned, that RL model created under research activity 1 will form the basis for creating the model for this activity. Next to the changes in the traffic scenario, however, there will also be adaptations in the state, action and reward formulations. Though the best way of defining these is likely to be the crux of getting the model to work and it needs some careful consideration, some initial thoughts on how this could be done are shared here.

As a first iteration to the state variable, a [1x26] size formulation will be tried. This is broken down as follows:

- Index 0-9: the heading information as previously used for the first research activity. This is needed to have information on the future traffic scenario (which will change over time).
- Index 10-17: the current airspace structure. If the new selected structure (action) is very different from this part of the state, many vertical conflicts are expected. This is as there is a lot of need for transitioning in such a case.
- Index 18-25 to have information on where the aircraft are and where the problem with transitioning arises. For example, when a certain heading range is to be further divided into several layers, or if its assigned altitude changes with a new structure, the risk is directly related to how many aircraft are currently in that layer (i.e., how many aircraft will have to vertically deviate/merge into new layers).

As the intention is to stick to eight airspace layers, the action will stay the same in the next part of the research. The reward formulation, however, may also be used to improve the performance of the model. Specifically, experiments may be run with a heavier penalty on either cruising on vertical LoSs, to see if the the model starts selecting different structures that favour the least penalised type of LoS. In an ideal case however, the reward formulation would simply consist of the total number of LoSs in an episode, while the model figures out the best choice for reducing the sum of cruising and vertical LoSs.

Then some final remarks on the development process of the RL model under this activity. It will be started by having just a single reconfiguration throughout the episode. Once experience is gained through this exercise, multiple reconfigurations per episode may be considered to see if the model can handle it. Furthermore, it is expected that training times increase compared to the model created under research activity 1, as the larger state will lead to more state-action combinations to explore. The model will be trained at a 'medium' traffic density, but the testing will be done at 'low', 'medium' and 'high' traffic densities as well, to investigate the capability of the model to generalise its solutions.

6.3.3 Updated methods of analysis

Though the method of analysis of the results presented for research activity 1 in chapter 5 gives a simple overview of the performance during training and testing, more information may be extracted to discuss the performance of the model. From research activity 4 onwards, several extra steps will be taken in the form of safety, stability and efficiency analyses. They are each touched upon in the following.

Safety analysis. For the safety analysis the focus wil be on both the number and duration of conflicts and LoSs. Naturally, fewer and shorter is considered to be safer. Additionally, one may look at the severity of the LoSs, a metric that further specifies how 'bad' or 'close' the LoSs actually became. Its calculation is shown in equation 6.1 below [8], where R_{PZ} stands for the radius of the protected zone and d_{CPA} represents the distance to the closest point of approach.

$$LoS_{sev} = \frac{R_{PZ} - d_{CPA}}{R_{PZ}} \tag{6.1}$$

Stability analysis. Stability refers to the tendency for tactical conflict avoidance manoeuvres to create secondary conflicts [8]. In previous work, it has often been 'measured' with the so-called domino effect parameter (DEP) [27], the calculation of which can be performed by equation 6.2 below:

$$DEP = \frac{n_{cfl}^{ON} - n_{cfl}^{OFF}}{n_{cfl}^{OFF}} \tag{6.2}$$

where n_{cfl}^{ON} and n_{cfl}^{OFF} represent the number of conflicts with CD&R ON and OFF, respectively.

Efficiency analysis. The efficiency analysis consists of two metrics primarily: the distance flown and the duration of flight. Naturally, if both if these grow large this is considered inefficient.

6.4 Research activity 5: RL model(s) for airspace reconfiguration with extra independent variable (reconfiguration rate)

The fifth research activity is concerned with investigating the effect of the reconfiguration rate on the choice of airspace structure. As explained previously in section 2.3.1, this is thought to be an important variable for this research as it may influence the decisions made by the model considerably.

The plan for completing this research activity is to create three additional RL models, each of which will be trained at a different reconfiguration rate. For example, the following models could be defined:

1. 1x reconfiguration at 50% of scenario time (slow reconfiguration rate)
2. 2x reconfiguration at 33% and 66% of scenario time (medium reconfiguration rate)
3. 3x reconfiguration at 25%, 50% and 75% of scenario time (fast reconfiguration rate)

To ensure that more reconfigurations indeed imply a faster reconfiguration rate, the scenario times for the above will be fixed. Specifically, each will be trained with a medium traffic density scenario that has a changing heading distribution that is in sync with the reconfiguration rate. To analyse the effect of the rate on the decisions made by the RL model, the models will then be tested on scenarios with:

- The same reconfiguration rate at which it was trained

- The other two reconfiguration rates (to see what happens if you train at the 'wrong' rate, but still use the model to perform decisions for dynamic airspace reconfiguration).

The results from testing could give additional insight into effects of the 'reconfiguration rate' variable and the ability of a RL model to set airspace structures when it also has information on the previous structure. It will be interesting to see the differences in choices for a model trained at a 'slow' rate versus those trained for a 'fast' rate. In an ideal case, one would see the model understanding the relative importance of cruising and vertical conflicts for each of the rates. The model trained with a 'slow' reconfiguration rate should result in airspace structures that are more suitable for the relatively long phase of cruising aircraft ahead, whereas a model trained with a 'fast' rate may be more focused on limiting the negative effects (= costs in terms of vertical conflicts) of the multiple reconfigurations. For resulting airspace structures the latter could mean that they are less different from the previous structure than would be ideal if traffic would spend a long time cruising in that structure.

6.5 Research activity 6: Comparing results of activities 1, 4 & 5

The sixth and final research activity will be to compare the results from the models created under activities 1, 4 and 5. In essence the interest here is how well these models are capable of reducing the total number of LoSs/conflicts with respect to a given baseline. The plan would be to present all outcomes of the test runs performed under the activities. The scenarios used for these tests of the model may also be run with structures not chosen by one of the RL models, but structures that are thought to be ideal for the given scenario. Defining such structures 'manually' can be trivial for some cases, like for example uniform traffic scenarios, where it can reasonably be expected that a uniform gives optimal performance. It can also be somewhat complex for cases which have more varying traffic. In any case, it is thought that a comparison between the performance of the RL models with a baseline can be insightful and help determine how well the models have performed. For comparison the safety (LoSs / conflicts), efficiency (flight path lengths) and stability (DEP) will be investigated. Furthermore, this activity will entail an analysis of the resulting airspaces that are selected by the models, to how they differ from each other.

Chapter 7

Conclusion

This report has presented the work done so far in a research into dynamic airspace reconfiguration with reinforcement learning (RL). Specifically, it has discussed the development of a RL model that was to set airspace structures based on a given traffic scenario. Though the model showed an evolution during an extended training run, the performance of this model was found to be sub-optimal in terms of its capability to select airspace structures for safety. Both the selected airspace structures at the end of training, as well as the results found through the testing of various traffic scenarios, showed that some iterating will be needed to arrive at a model that can select suitable airspace structures that minimise the number of loss of separations (LoSs).

The findings described above were defined as the first of six research activities that will be completed in this research. The other five activities consist of defining rules for moving traffic into new airspace structures, defining rules for conflict resolution (CR) in a layered airspace, extending the already made RL model to the point where it also takes into account the previous airspace structure, investigating the effect of the reconfiguration rate on the selected airspace structures and finally analysing, comparing and presenting the results found. It is thought that these activities will contribute to the primary research objective, which is to 'develop a RL model, which can correctly assess and minimise the impact of airspace reconfiguration on safety, while guaranteeing that new structures are appropriate for future traffic'.

Previously developed models for setting layered airspace structures only take into account a future traffic scenario and include no information on the previous airspace structures in their decision making. This makes the applicability to more real-life situations somewhat limited. To improve on this, this MSc thesis project aims to add information on the previous airspace structure to such models, bringing it a step closer to the real-life situation where an airspace would already be filled with aircraft when a decision may be made on the next (more suitable) airspace structure. Currently, almost all Air Traffic Control (ATC) is done by humans. However, looking at ways to have this controlled by an agent that has learnt to set airspace structures optimised for safety (or for efficiency for that matter, if this would be desired) poses several advantages as well. For example, the use of a machine learning (ML) agent for this task could reduce the need for humans in the operation. Some advantages of this are readily imaginable; a ML model could at some point perform better (and cheaper) than its human counterpart, while also having better availability.

With the increased demand for air traffic in recent years and the projected growth in the decades to come, solutions are needed to accommodate safe flight in the future. The fact that future operations of unmanned aviation are also expected to operate at even higher traffic densities than seen previously with manned aviation only emphasises the need for finding suitable ways of controlling the air traffic. The overarching idea of dynamic airspace reconfiguration is that it could increase the capacity of our future airspaces to levels required by the aforementioned developments.

Bibliography

[1] E. Sunil, *Analyzing and Modeling Capacity for Decentralized Air Traffic Control.* PhD thesis, 2019.

[2] E. Sunil, J. Ellerbroek, J. Hoekstra, and J. Maas, "Three-dimensional conflict count models for unstructured and layered airspace designs," *Transportation Research Part C Emerging Technologies*, 2018.

[3] L. Caranti, M. Ribeiro, J. Ellerbroek, and J. Hoekstra, "Safety optimization of a layered airspace structure with supervised learning," *11th SESAR Innovation Days*, 2021.

[4] J. Hoekstra, R. van Gent, and R. Ruigrok, "Designing for safety: The 'free flight' air traffic management concept," *Reliability Engineering & System Safety, Volume 75, Issue 2, February 2002, Pages 215-232*, 2002.

[5] SESAR-Joint-Undertaking, "European drones outlook study - unlocking the value for europe," 2016.

[6] J. Hoekstra, J. Maas, M. Tra, and E. Sunil, "How do layered airspace design parameters affect airspace capacity and safety?," *7th International Conference on Research in Air Transportation, Philadelphia, PA, USA*, 2016.

[7] E. Sunil, J. Hoekstra, J. Ellerbroek, F. Bussink, D. Nieuwenhuisen, A. Vidosavljevic, and S. Kern, "Metropolis: Relating airspace structure and capacity for extreme traffic densities," *11th USA/EUROPE Air Traffic Management R&D Seminar*, 2015.

[8] M. Ribeiro, J. Ellerbroek, and J. Hoekstra, "Using reinforcement learning to improve airspace structuring in an urban environment," *(Under review)*, 2021.

[9] M. Nolan, *Fundamentals of Air Traffic Control, Chapter 1 History of ATC, 5th Edition.* Delmar Cengage Learning, 2010.

[10] M. Ribeiro, J. Ellerbroek, and J. Hoekstra, "Review of conflict resolution methods for manned and unmanned aviation," *9th SESAR Innovation Days, Athens, Greece*, 2020.

[11] K. Bilimoria and H. Lee, "Comparison of centralized and decentralized conflict resolution strategies for multiple-aircraft problems," *18th Applied Aerodynamics Conference*, 2000.

[12] J. Hoekstra, R. Ruigrok, D. van Gent, and R. Van Gent, "Free flight in a crowded airspace?," *3rd USA/Europe Air Traffic Management R&D Seminar, Napoli*, 2000.

[13] M. Jardin, "Air traffic conflict models," *AIAA 4th Aviation Technology, Integration and Operations (ATIO) Forum, Chicago, Illinois*, 2004.

[14] M. Jardin, "Analytical relationships between conflict counts and air-traffic density," *Journal of Guidance, Control, and Dynamics*, 2005.

[15] G. Gilbert, "Historical development of the air traffic control system," *IEEE Transactions on Communications 21 (5) 364–375*, 1973.

[16] M. Tra, E. Sunil, J. Ellerbroek, and J. Hoekstra, "Modeling the intrinsic safety of unstructured and layered airspace designs," *Proceedings of the 12th USA/Europe Air Traffic Management Research and Development Seminar*, 2017.

[17] M. Ribeiro, J. Ellerbroek, and J. Hoekstra, "Determining optimal conflict avoidance manoeuvres at high densities with reinforcement learning," *10th SESAR Innovation Days*, 2020.

[18] R. Isufaj, D. Sebastia, and M. Piera, "Towards conflict resolution with deep multi-agent reinforcement learning," *Fourteenth USA/Europe Air Traffic Management Research and Development Seminar*, 2021.

[19] M. Ribeiro, J. Ellerbroek, and J. Hoekstra, "Velocity obstacle based conflict avoidance in urban environment with variable speed limit," *Aerospace*, 2021.

[20] H. Ong and M. Kochenderfer, "Markov decision process-based distributed conflict resolution for drone air traffic management," *Journal of Guidance, Control, and Dynamics*, 2016.

[21] H. Rui, L. Haipeng, Z. Sumin, Z. Dong, and Z. Hang, "Lane following method based on improved ddpg algorithm," *Sensors 2021, 21(14), 4827*, 2021.

[22] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, and S. Levine, "Soft actor-critic algorithms and applications," 2019.

[23] Y. Duan, X. Chen, R. Houthooft, J. Schulman, and P. Abbeel, "Benchmarking deep reinforcement learning for continuous control," *Proceedings of the 33 rd International Conference on Machine Learning, New York, NY, USA*, 2016.

[24] J. Hoekstra and J. Ellerbroek, "Bluesky atc simulator project: an open data and open source approach," *International Conference for Research on Air Transportation, Philadelphia, PA, USA*, 2016.

[25] M. Doole, J. Ellerbroek, and J. Hoekstra, "Estimation of traffic density from drone-based delivery in very low level urban airspace," *Journal of Air Transport Management*, 2020.

[26] D. Alejo, R. Conde, J. Cobano, and A. Ollero, "Multi-uav collision avoidance with separation assurance under uncertainties," *IEEE International Conference on Mechatronics*, 2009.

[27] K. Bilimoria, K. Sheth, H. Lee, and S. Grabbe, "Performance evaluation of airborne separation assurance for free flight," *18th Applied Aerodynamics Conference*, 2000.

Gantt chart MSc Thesis - Dynamic Airspace Reconfiguration with Reinforcement Learning

Legend: = Timeline of an item; = Timeline of a phase; = End of phase

Calendar week nr.: 24 25 26 27 28 29 30 31 32 33 34 35 36 37
Project week nr.: 27 28 29 30 31 32 33 34 35 36 37 38 39 40

Phase	Start date	End date	Item	Status	Comments
					Comments *(Final deadline = Friday of week 37 = 16/09/2022)*
Midterm phase	**02/12/2021**	**23/05/2022**	**Working towards midterm**	Done	This is done now!
	23/05/2022	23/05/2022	Deliverable: **draft** midterm report	Done	
Research Activities (RAs)	**23/05/2022**	**12/08/2022**	**Working towards green light**	In Progress	Writing of final paper happens concurrently in this phase
	13/06/2022	19/06/2022	RA3: 'Define rules for CR in a layered airspace'	In Progress	
	13/06/2022	26/06/2022	RA2: 'Define rules for moving traffic into new structure'	In Progress	
	13/06/2022	10/07/2022	RA4: 'Final RL model'	In Progress	
	11/07/2022	24/07/2022	RA5: 'Analyse the effect of the reconfiguration rate'	To start	
	25/07/2022	07/08/2022	RA6: 'Comparing performance of models'	To start	
	07/08/2022	07/08/2022	Deliverable: draft of final paper	To start	
	08/08/2022	12/08/2022	Finalising final paper	To start	
	12/08/2022	12/08/2022	Deliverable: final paper	To start	
Defence phase	**15/08/2022**	**16/09/2022**	**Working towards MSc Thesis defence**	To start	
	15/08/2022	15/09/2022	Preparing presentation & defence	To start	
Presentation	**16/09/2022**	**16/09/2022**	**MSc Thesis presentation & defence**	To start	**Presentation & defence** on **Friday 16/09**