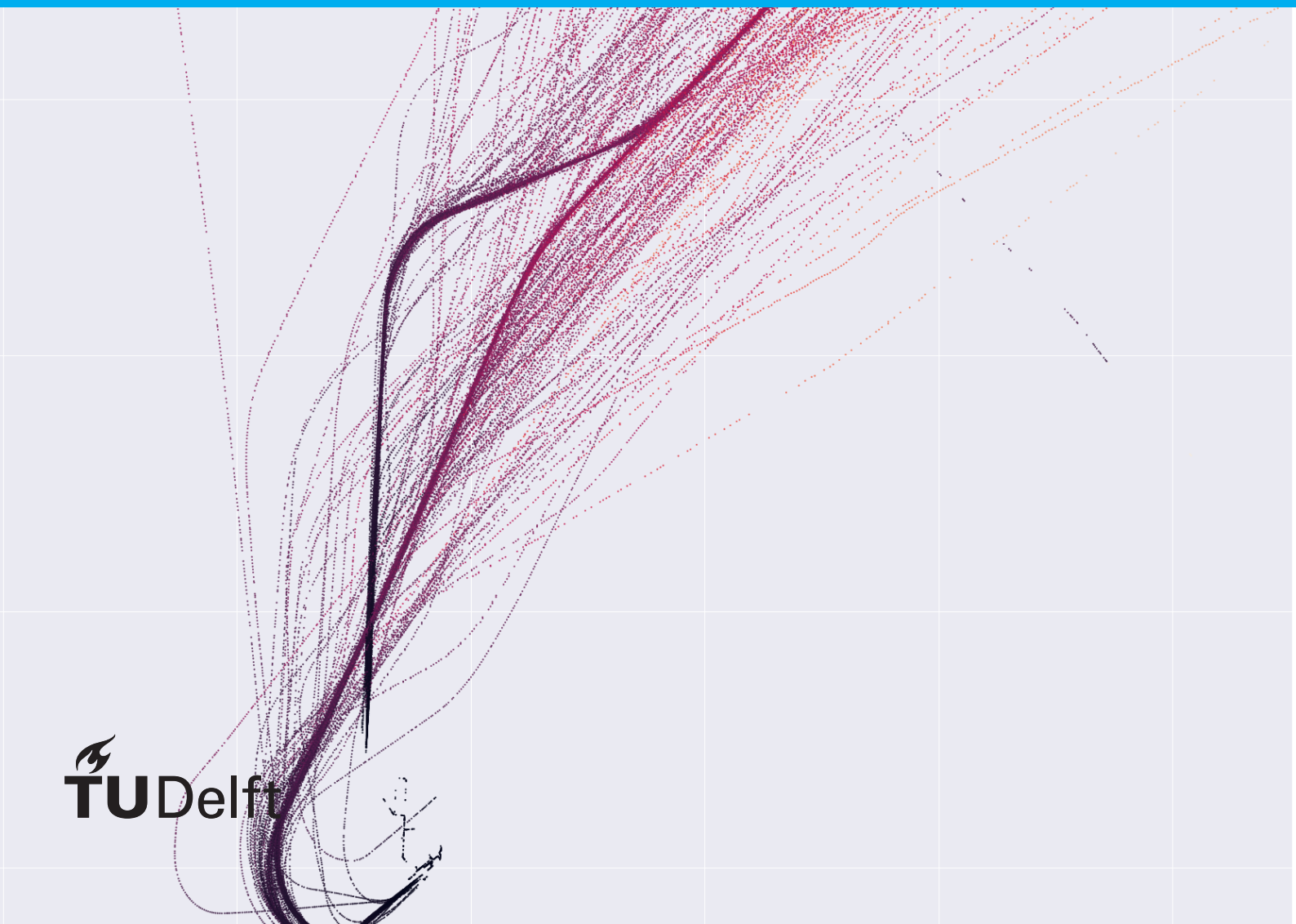
Thesis Report

A Data-Driven comparison of Trajectory Prediction methodologies using ADS-B data

V.A. Pereboom

Department of Control & Simulation
Faculty of Aerospace Engineering



TU Delft

Thesis Report

A Data-Driven comparison of Trajectory Prediction methodologies using ADS-B data

by

V.A. Pereboom

Student number: 4150554
Date: July 2019
Thesis supervisors: Prof. Dr. J.M. Hoekstra, TU Delft
Dr. J. Ellerbroek, TU Delft

TU Delft

Abstract

Better predictability of future aircraft locations will result in more efficient air-traffic management operations and therefore an increase in airspace capacity. However, flight paths tend to have a stochastic character causing their predictability to decrease with the look-ahead time.

In this research a quantitative analysis is made of the statistical properties of flight trajectory deviations and their dependency on the look-ahead time based on historical ADS-B flight segments and corresponding flight plan data. An empirical boundary on the prediction horizon for future aircraft locations in en-route flights is determined based on a deterministic state propagation method. Using these results, an analysis is made of how these uncertainty distributions affect conflict detection performance. A comparison is therefore made between state-deterministic, state-probabilistic and intent-deterministic conflict detection methods when applied to medium-term conflict scenarios.

From a large historic set of ADS-B transponder messages and ETFMS flight plan data from Eurocontrol, a data set of flight trajectories is created. From this set intrusion scenarios are constructed on which the three mentioned conflict detection methods are benchmarked and compared based using an F-measure score based on a binary conflict detection classification.

This statistical analysis on state-deterministic projection errors shows that the error distribution of the cross-track, along-track and total track error all show the presence of fat tails. This peakedness of the distributions diminishes slightly for larger look-ahead times as the chance of trajectory changes grows with time. However, in none of the scenarios the error distributions are normally distributed. This has to be taken in mind for any future work based on these data sources.

Based on the combination of total track errors an empirical upper limit for the look-ahead time can be found in the context of state-deterministic projection. This is based on the look-ahead times where the TTE distribution percentiles grow larger than the IPZ. A 70th percentile confidence bound grows beyond the IPZ range around 9 minutes (540 seconds). When using a 90th percentile bound, this look-ahead time reduces to 5 minutes and 20 seconds (320 seconds).

Altitude largely affects the error distritutions. As flights at lower altitudes are more prone to vectoring directed by ATC, the presence of abrupt changes in the flight path is significantly higher. This reflects in larger variance in the error distributions and a higher presence of outliers. Therefore, the predictability of trajectories is much lower at altitudes below 30.000 feet compared to flights above this flight level with mean absolute deviations of up to a factor three larger.

From the general comparison of the three conflict detection methodologies the performance shows to be very similar for the observed look-ahead time range to 1200 seconds ahead. For lower look-ahead times a state-deterministic projection approach yields the best results based on model accuracy up to a look-ahead time of approximately 600 seconds. Thereafter, intent-deterministic conflict detection performs better for look-ahead times beyond 600 seconds. Although a state-probabilistic approach has lower accuracy, the number of false negatives is significantly lower, which in this context might be preferred as false negatives might result in unwanted situations when applied in practice. This lower number of false negatives is however balanced by a larger number of false positives. The state-probabilistic approach is mainly useful as a risk-averse methodology where a lower number of false negatives is traded for a larger number of false positives but performs worse in general regarding accuracy.

Preface

This research is performed as part of the completion of the master degree in Aerospace Engineering at the Delft University of Technology.

I would like to sincerely thank my supervisors Jacco Hoekstra and Joost Ellerbroek for their good feedback, support and flexibility during my research work.

Furthermore I would like to thank Junzi Sun for his support with decoding the ADS-B information and making the data sets available, Julia Rudnyk for her help in accumulating the ETFMS data sets from Eurocontrol for use in this research and both for taking time to provide feedback and good discussions on the topic.

Victor Pereboom
Delft, July 2019

Contents

Nomenclature

Acronyms

ACAS	Airborne Collision Avoidance System
ADS-B	Automatic Dependent Surveillance Broadcast
ATC	Air Traffic Control
ATE	Along-Track Error
ATM	Air Traffic Management
CD&R	Conflict Detection & Resolution
CPA	Closest Point of Approach
CPR	Correlated Position Report
CTE	Cross-Track Error
CTFM	Current Tactical Flight Model
ETFMS	Enhanced Traffic Flow Management System
FAA	Federal Aviation Administration
FIS	Flight Information System
FIS-B	Flight Information Services Broadcast
FL	Flight Level
FN	False Negative
FP	False Positive
GNSS	Global Navigation Satellite System
ICAO	International Civil Aviation organization
IMM	Interacting Multiple Model
IPZ	Intruder Protected Zone
KDE	Kernel Density Estimate
KDE	Kernel Density Estimate
MDP	Markov Decision Process
MTCD	Medium Term Conflict Detection
NM	Nautical Mile
RA	Resolution Advisory

SO6	SegOut6 Data
STCA	Short Term Conflict Alert
TA	Traffic Alert
TCAS	Traffic Collision Avoidance System
TCP	Trajectory Change Point
TIS-B	Traffic Information Services Broadcast
TMA	Terminal Manouvering Area
TN	True Negative
TP	True Positive
TTE	Total Track Error
TTI	Time to Intrusion

Greek Symbols

α	Clockwise angle between two headings in degrees
β	The smallest angle, in degrees, between the aircraft heading and the line connecting two aircraft
γ	The angle in degrees between two flight paths
λ	Longitude in radians
ϕ	Latitude in radians
σ	Standard deviation
Θ	Compass bearing in degrees
Θ_{c1_c2}	Bearing in degrees of the line from a/c 1 to a/c 2

Roman Symbols

$\tilde{D}_{ac}$	The change in distance between two aircraft resulting from the speed vector of one of the aircraft
b	The KDE kernel bandwidth
$D_{ac1/ac2}$	The distance between two aircraft in meters
P	The cumulative distribution function
p	Probability density
q	The q-th quantile of the data distribution
R_{earth}	Radius of the earth in meters
t_{la}	Look-ahead time in seconds
V	Airspeed in meters per second
V_{conv}	The convergence speed between two aircraft in meters per second
V_x	The component of the airspeed vector in x direction in meters per second
V_y	The component of the airspeed vector in y direction in meters per second
x_q	The value from a data set belonging to quantile q

List of Figures

List of Tables

1 Introduction

Air Traffic Control (ATC) operations experience increasing pressure from the rise in use of the available airspace. The European Air Traffic Management system is currently responsible for the daily handling of around 26.000 flights. This number is likely to double over the next years. [12] Also, the diversification of air traffic with the introduction of new airspace users, like UAVs, will require a larger workload from existing systems. [36]

New digital transformations in the aviation industry will help in the transition to overcome existing challenges. More data creation, interoperability of existing systems for system-wide data sharing, stronger analytical capabilities and automation provide new opportunities in this field for modernization. [36]

Valuable information in airspace organization and management relates to the prediction of future air traffic movements. Both for safety, in the form of conflict detection, as well as capacity planning. Better trajectory predictions and CD&R systems will support the increase of traffic density and maneuvering freedom of aircraft which is a direct benefit to ATM .

As information exchange between aircraft is becoming more advanced with the introduction of the Automatic Dependent Surveillance Broadcast (ADS-B) system and initiatives like System Wide Information Management (SWIM), better data driven Conflict Detection & Resolution (CD&R) methodologies can be developed. [36] The interest for new methods related to data driven trajectory predictions expresses itself as well through the initialization of larger research projects in this field, like DART and datAcron. [9],[41]

Aircraft trajectory prediction however is a challenging problem. Flight paths have a stochastic character due to different external influences like wind, weather areas and interaction between aircraft and ATC. In literature, many approaches have been proposed to model these behaviors and combine them into trajectory prediction models with corresponding conflict detection methods. A distinction is often made between three different approaches to conflict detection: Deterministic (Nominal) conflict detection, probabilistic conflict detection and worst case conflict detection. [23]

One of the main questions is how far into the future one can still make a reliable estimation of the aircraft position based on certain input information. This boundary on the look-ahead time will provide insight in potential gains in anticipation time for air traffic control and the reduction of spacing between aircraft for en-route traffic.

The objective of the proposed research is to perform a data driven evaluation of three different trajectory prediction and conflict detection methodologies: state-based deterministic conflict detection (state-deterministic), state-based probabilistic conflict detection (state-probabilistic) and state-based deterministic conflict detection with intent (intent-deterministic). This will provide insight in the empirical relation between the detection performance and the look-ahead time and thereby as well a limit on how far into the future it is feasible to detect conflicts. For this research historical Automatic Dependent Surveillance Broadcast (ADS-B) data sets and flight plan data from Eurocontrol will be used. The focus in this research will be short-term to medium-term conflict detection considering a look-ahead time horizon of up to 20 minutes.

This report includes a literature study conducted on reviewing the current state-of-the-art in trajectory pre-

diction and conflict detection for air traffic. This background information will be presented in chapter 2. The research questions and objectives are discussed in detail in chapter 3. Technical background on the modelling and data processing techniques used is provided in chapters 4 and 5. An overview of the experimental setup is discussed in chapter 6. A discussion of the results obtained in this research is provided in chapter 7 after which the conclusion can be found in chapter 8. Recommendations for future research are discussed in chapter 9.

2

Background on Trajectory Prediction and Conflict Detection

This chapter discusses literature on trajectory prediction and conflict detection methods, both for currently applied methods as well as recent research. Also, an overview will be given of current airspace management approaches which provide context to the discussed methods.

2.1. Airspace Organization

In air traffic management and control, most of the tactical planning, independent of its complexities, lies with human controllers. With the aim to reduce any form of collision risk and potentially dangerous situations, a strict set of flight rules and management systems are in place to control the air traffic flow. Separation criteria of aircraft are defined by an intruder protected zone (IPZ) . This virtual zone is a cylinder around the aircraft with a radius of 5 nautical miles and a height of 2000 feet, inducing a horizontal separation minimum between the own aircraft and other traffic of at least 5 nautical miles and a vertical separation criterium of 1000 feet as shown in figure 2.1. These criteria generally hold for all aircraft in en-route airspace. In terminal airspace, the separation criteria are lower, usually 3 nautical miles of radius. Where in areas with less reliable radar coverage, as in oceanic regions, the horizontal separation criterium is usually increased to a 10 nautical mile radius.

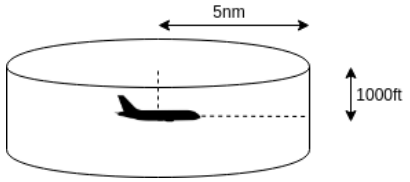


Figure 2.1: Schematic of intruder protected zone (IPZ)

The Conflict Detection & Resolution process consists in general of three different steps:

- **State Propagation (Trajectory Prediction)**
  This step involves estimating the future location of an aircraft using its current state information and additional data like intent. The performance of the applied state propagation methods are the key component to reliable conflict detection.
- **Conflict Detection**
  Using a combination of propagated states from two aircraft, an assessment can be made of the likelihood of a conflict. The detection method is based on a set of metrics and thresholds. These metrics could be for instance a closest point of approach, time to conflict or expected minimum separation. These

metrics are applied to the combined propagated aircraft states which could result in a conflict detection if implied thresholds are met.

- **Conflict Resolution**
  When a conflict is detected, conflict mitigation is applied either automatically or by input from a human controller. This depends on how imminent a conflict is and what the time to conflict is. Generally, air traffic control will mitigate any conflict risks by redirecting the potentially conflicting aircraft. As a conflict is more imminent, the Traffic Collision Avoidance System (TCAS) on board of aircraft will automatically communicate the best resolution strategy to both involved aircraft.

2.2. Trajectory Prediction methods

In recent years, much research has been performed in the field of trajectory prediction and state propagation for aircraft and vessels. Not only is trajectory prediction a critical component in in-flight CD&R applications, but it is also applied in decision support tools for separation assurance and flow management used by ATM. A distinction can be made between four fundamental approaches to trajectory prediction [23][42]:

- **Deterministic (Nominal)**
  Current aircraft states are projected into the future along a single trajectory without consideration of uncertainties in the aircraft states or flight path. For instance a position can be extrapolated based on the velocity and heading vector at an earlier point. Uncertainties are managed in this method by incorporating a safety buffer to account for unexpected behavior. This method is most suited for trajectory predictions over a short time window as the prediction uncertainty grows with time. One of the issues with this approach is the risk of undetected conflicts in the face of uncertainty.

- **Worst Case**
  This approach assesses a wide range of possible trajectories resulting from the current state of the aircraft, for instance based on a dynamic model of the aircraft. An alert will be raised if any of the assessed trajectories leads to a conflict with another aircraft, even if this maneuver is unlikely. This approach might lead to a large amount of unnecessary warnings to the pilot and therefore a reduction in airspace capacity if enforced throughout.

- **Probabilistic**
  In a realistic scenario, aircraft and their flight paths are subject to various influences with a level of randomness in their behavior. Weather, wind fields, other air traffic and pilot errors are all possible factors inducing deviations from the expected flight path. In a probabilistic approach to trajectory prediction, uncertainties in different state variables are modeled to describe potential variations in the aircraft flight path. This results in a set of trajectories which are weighed by their respective probability of occurrence. Conflict detection in this approach can be applied by calculating the intersection of the position probability distributions of two aircraft and raising an alert if the probability of conflict rises above a certain threshold. Compared to deterministic projection and worst case projection, this approach provides a balance between the exclusion of randomness and the nuisance warning effect of directly acting upon possible highly unlikely conflict events.[5][31]

- **Intent Based**
  Intent based projection will use a shared navigation intent in the form of the next waypoint or trajectory change point (TCP) as an indication for the expected heading of the aircraft. In newer versions of the ADS-B system, intent in the form of the next waypoint will be broadcasted which makes this information useful for other aircraft and ground stations. In most research, aircraft intent is combined with other approaches like deterministic or probabilistic projection to create a more accurate model.

All mentioned methodologies are depicted in figure 2.2. Three of the above mentioned methodologies, being Deterministic, Probabilistic and Intent Based projection will be used as fundamental methodologies applied in this research. Combinations of the described methods are also found throughout literature. One of the reasons for this is that each method relies on different input data with varying complexity. When more information can be added and captured by a model, its accuracy will logically increase. As Deterministic and Probabilistic trajectory prediction are mainly based on the current aircraft states, Intent Based prediction takes into account communicated next waypoints for which it relies on additional data exchange. Combining the approaches is

expected to improve prediction accuracy for both short term as well as medium term.

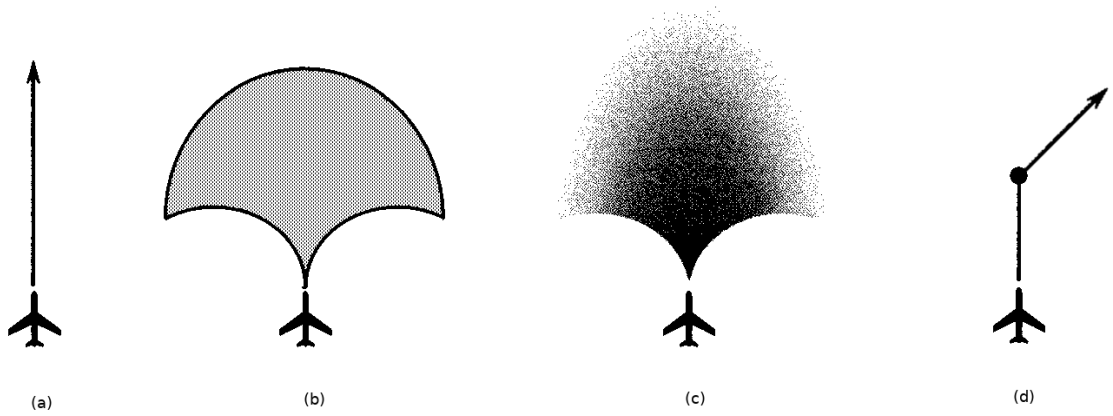


Figure 2.2: Representation of Deterministic state projection (a), Worst Case state projection (b), Probabilistic state projection (c) and Intent Based state projection (d). [23]

2.3. Research initiatives on trajectory prediction

This section will discuss research based on the mentioned four fundamental trajectory prediction approaches and combinations thereof.

2.3.1. Deterministic Projection

Deterministic state propagation projects the current aircraft state into the future based on a dynamic model. The most basic application is the linear propagation of the heading and velocity vector resulting in a straight single-path trajectory model for the aircraft. However, the method can also be extended using more complex dynamical models making use of multiple state variables.

deterministic state projection itself does not account for uncertainty in the aircraft states, resulting in a single track estimation which is a straight line segment. However, this might result in erroneous state predictions, especially when propagating trajectories further into the future as uncertainty grows with time. To overcome this, when combined with conflict detection, usually designed security buffers are implemented in order to leave room for randomness in the future estimated aircraft location. Because the method uses relatively few state variables and also the computational complexity of the dynamical model is limited, deterministic projection is efficient in implementation. When implementing, a trade-off needs to be made between look-ahead time and consistency. Also, because no intent information is used, the use of deterministic projection might cause nuisance warnings for pilots. Deterministic state propagation does not necessarily have to be a straight-line projection based on heading and velocity. Some research improves deterministic trajectory prediction by including a dynamic model of the physical system which lowers the prediction uncertainty. One way to achieve this is by combining measured state information with existing aircraft dynamic models creating a hybrid estimation. This enables the more accurate extrapolation of coordinated turns, level-offs and other maneuvers.

Several examples for deterministic projection with more complicated dynamics can be found: Baek et al.[2] applies an Interacting Multiple Model (IMM) approach to estimate what flight mode the aircraft is currently in. It is a hybrid approach relating the continuous state of the aircraft to a discrete set of flight modes. A set of dynamic models for different horizontal and vertical flight modes are defined. In these state definitions, zero-mean Gaussian process and measurement noise is included to model uncertainty. An overarching estimator is used that selects the most likely model based on the respective estimated flight mode derived from the current aircraft state. A set of Kalman filters is used to match each mode and bases the final mode selection on the fusion of the filter estimate outputs, which leads to the derived mode probability. This also tends to a probabilistic approach. Trajectory prediction is then performed by propagating the selected mode along a

forecast window T_f. The model input is based on information that can be derived from ADS-B data, making the approach future proof and applicable for ADS-B equipped aircraft.

In terms of conflict detection, deterministic state propagation approaches using a single path predicted trajectory result in a binary detection of conflicts. Either a conflict between two aircraft happens, or it does not. This approach provides no room for determining any continuous risk measures which is usually compensated for by applying buffer zones and assumed uncertainty distributions in the conflict detection and resolution steps.

2.3.2. Worst Case Projection

One approach to overcome the lack of incorporation of uncertainty in the deterministic trajectory projections is to assess the extremes of the dynamical model output as described earlier. Thereby covering uncertainty in aircraft state measurements and their propagation. A worst case model would compare the edge outcomes of the projection of the current aircraft state into the future and indicate a risk of collision if these extremes of two approaching aircraft overlap, providing the worst-case scenario indication. The main goal of this approach is to ensure highly safe operations by having aircraft operate outside of each other's dynamical range. The downside is however a high number of warnings, especially with larger look-ahead times. Therefore, worst-case scenario approaches are mostly considered and applied in a very short-term time range.

2.3.3. Probabilistic Projection

Probabilistic state propagation is of most interest in current research, driven by a rise in data availability to infer uncertainty distributions. This is an important step in the improvement of currently applied deterministic models in order to overcome mentioned shortcomings by increasing the accuracy of conflict detection and making the measure more continuous. Uncertainty modeling becomes more accurate with the increased availability of data and processing capabilities. The actual uncertainty which needs to be taken into account in trajectory prediction results not only from the airspace context but also from the data measurements itself. Therefore, information accuracy also needs to be assessed when regarding probabilistic state propagation approaches.

Flight paths are naturally influenced by some level of variability in the aircraft state, surrounding airspace and traffic control mechanisms. This leads to uncertainty and randomness in the flight path. This uncertainty increases with time when the flight path is extrapolated using a model. Basically, all conflict detection models rely on trajectory prediction for which the input data is limited to the state variables available to the system in flight. Accuracy of prediction models is logically improved if more explanatory variables are added to the model, increasing the amount of information used for modeling and thereby reducing the overall uncertainty. However, a perfect model accuracy is infeasible as the available information is limited to the applied sensoring systems and infrastructure. Both in-flight and on the ground. Also, data resulting from these systems are subject to measurement noise.

The following categories of possible available in-flight information sources in a modern aircraft, can be defined:

1. Dynamics of the own aircraft and corresponding state variables
2. Pilot intent which includes the flight-plan and corresponding waypoints as well as its deviations as commanded by Air Traffic Control. Intent information can also be inferred from standard rules like instrumental flight rules and standard terminal arrival routes. In the future, flight plan sharing might be enabled as well.
3. Weather information from radar and ground-to-air weather reports
4. The measured aircraft position and altitude based on GPS and internal sensoring equipment.
5. The airspace structure and imposed limits
6. The location, altitude, heading, speed and type of surrounding air traffic from transponder data like ADS-B

Trajectories of aircraft are subject to various influences introducing uncertainty and randomness in the execution of flight plans. Therefore, trajectory prediction models that take into account uncertainty seem more

natural compared to deterministic approaches. Also, working with probability distributions for future aircraft states will refine setting thresholds for conflicts.

Figure 2.3 provides a simplified overview of trajectory prediction and the locations where uncertainty is introduced. Two feedback loops can be distinguished in the dynamic system of aircraft trajectories as presented in figure 2.3. The first one is the inner loop which covers the interaction between aircraft dynamics and the pilot as human controller. This loop is nested in a larger loop which includes the flight plan information and waypoints provided by air traffic control, taking into account movements of the own aircraft and other aircraft. In both loops, uncertainty is introduced to the system through the factors as listed above. [19]

The following main error sources, which introduce uncertainty in trajectory estimation, can be defined:

- **Intent changes due to weather or traffic**
  The intent is driven by both the flight plan and feedback from ATC. The pilot acts according to this intent and provides the corresponding input to the aircraft or the flight management system. Intent changes are mainly present in medium to long term trajectory forecasts. Depending on the viewers point, they mostly occur unexpected as a response to other traffic, bad weather areas or ATC redirects. ADS-B allows broadcasting of the next trajectory change point, which is sufficient for short to medium term trajectory estimations, but not for long term estimations.
- **Navigation errors**
  Navigation errors are introduced by errors made by the pilots themselves or false entries in the flight management system. The pilot also responds to the state of the aircraft resulting in variance around the actual planned trajectory.
- **Atmospheric disturbances and weather**
  Regarding weather, mainly wind is a source of disturbance to the aircraft trajectory. As the measured velocity is the airspeed and not the ground speed, incorrect modeling of wind fields causes errors in the estimated trajectory. Also, bad weather areas might result in ad-hoc intent changes.
- **Measurement errors for aircraft states and location**
  The trajectory estimations rely on kinematic and dynamic models of the aircraft based on a set of measured aircraft states. These measured state variables which are used as inputs contain sensor noise and sometimes measurement errors which results in erroneous predictions. Also, other variables related to the aircraft type or configuration might have been discarded or uncertain.
- **Model errors for aircraft dynamics**
  The models used to describe the aircraft dynamics and kinematics are simplifications of the real world problem. This introduces errors in the estimated trajectories.

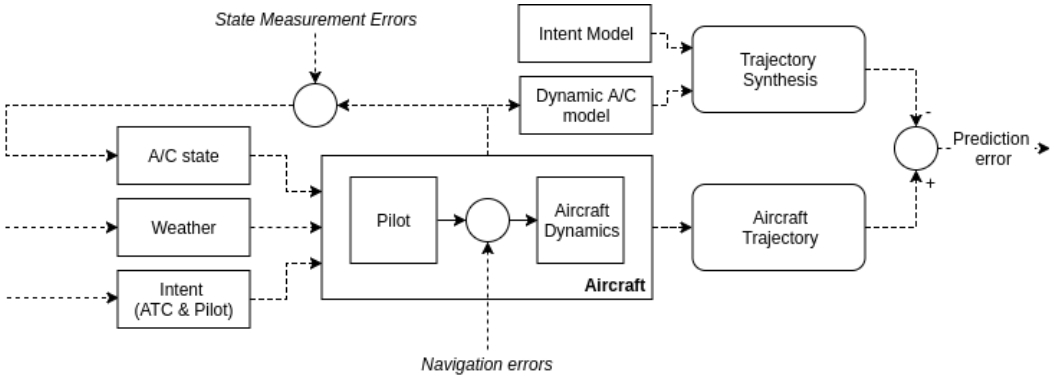


Figure 2.3: Schematic of flight path modeling errors

In a certain way, the airspace is a dynamic system that responds to predictions that are made about its future state. ATC and pilots anticipate on estimated trajectories from the own aircraft and surrounding aircraft. When pilots respond to potential conflicts, as given by a state extrapolation model, they affect the predictions they are responding to. ATC consumes information from both the aircraft flight path, model based extrapolation (prediction) of this flight path and has this same information for all other aircraft flying in the area.

ATC accordingly determines if any adjustments to the existing flight path are required and communicates this information with the aircraft by providing it with a new route intent.

Also, in all communication and control loops in the diagram there is a certain degree of lag involved. Systems need time to handle and adjust to new information. Transfer of a new route intent from ATC to the aircraft also takes time because this needs to be done manually. Lag is not a significant problem in most situations as the time delays are low compared to the general agility of the system and the required response times which are often in the order of minutes. [24] However, when a possible conflict approaches, required response times can be much lower. Therefore, different systems are used like the automatic TCAS system for very imminent conflicts instead of, or in combination with, response from ATC.

Lauderdale et al.[25] performed a study on the sensitivity of an airborne separation assurance algorithm to a set of six different types of trajectory prediction errors, assessing variance in: 1) Wind speed, 2) cruise speed, 3) aircraft weight, 4) top-of-descent location, 5) descent speed and 6) maneuver initiation timing. The prediction model uses dynamic aircraft performance models using the Base of Aircraft Data (BADA). This research indicates that the main contributing factors to losses of separation are errors in descent speed and top-of-descent location. Within the 2d plane mainly maneuver initiation timing and wind speed are the most important factors.

Research by Paielli et al. [28] empirically addresses the statistical distribution of cross-track and along-track errors based on actual flight data. The analysis is based 16 days of raw flight tracking and flight plan data from the Air Route Traffic Control Center in Denver. Only level flights above FL290 are considered and aircraft type or availability of a flight management system is not taken into account. The results show that assuming a normal (Gaussian) distribution for trajectory prediction errors is valid for at least a look ahead time up to 20 minutes. Also is shown that the along-track error increases linearly with approximately 0.22 nmi/minute look ahead time. The cross track error has a similar growth rate but is shown to level off for higher look-ahead times. Furthermore, the error components for the cross-track and along-track are shown to have a very low correlation which support assumptions on orthogonality. Other research by Ballin et al. [4], based on a study of historical flight data to Dallas/Forth Worth airport, indicates that distance and time based separation errors closely resemble a Gaussian distribution as well.

Several different approaches can be found to probabilistic trajectory prediction and conflict detection. The ways to include uncertainty in a trajectory prediction can be done either by the superposition of assumed or calculated distributions on top of the aircraft motion model result, or it can be achieved by directly incorporating stochastic variables into the equations of the aircraft motion dynamics. The latter method leads to more accurate results, but is more difficult to determine analytically. Therefore it is common to simulate a set of directories using a Monte Carlo approach and generate a distribution of future states from which the conflict probability can be inferred.

An approach related to the method of superposition is proposed by Paielli et al.[29]. The deterministic trajectory prediction method is extended by assuming normally distributed (Gaussian) cross-track and along-track root mean squared prediction errors. These errors are used to determine a resulting error covariance for the aircraft trajectory. Visually, this can be viewed as an uncertainty ellipse centered around the aircraft where usually the along-track axis tends to be longer than the cross-track axis. Also, the size of the ellipsoid grows with time as predictions become more uncertain further into the future, however, the correlation stays the same. By combining the two horizontal position errors for a pair of aircraft into a combined error covariance matrix using the heading angle between the two flight paths, a conflict probability can be determined. The combined error covariance matrix describes the relative position error distribution between the two aircraft. In order to relate the covariance matrix back to the known separation criteria the error covariances are translated into the covariance matrix for a single aircraft which will be referred to as the stochastic aircraft. Now, the overlap between this combined covariance matrix and a separation zone can be determined for a certain point in time and related to a conflict detection probability. This method can also be implemented in a three dimensional space, resulting in an error ellipsoid.[5][31] Assuming a Gaussian distribution for the track errors makes combining the covariances for a pair of aircraft analytically easier and thereby more efficient.

Examples of including stochastic variables directly into the dynamic equations for the aircraft motion are proposed in [30] where the trajectory estimation is performed by using Monte Carlo simulation for a stochastic differential equation describing the aircraft motion. This results in a reachable set of stochastic trajectories from an initial state. This set can then be used for conflict detection. In order to avoid reinitiating an entire set

of Monte Carlo simulated trajectories for every update of the measured aircraft state, a sequential Monte Carlo method is applied for higher computational efficiency.

The concept of reachable sets is also used by Sahawneh et al.[34]. Here a kinematic model of the intruder aircraft is based on the turn rate and speed used to determine a reachable set of potential intruder positions over a certain time horizon. Probability of reaching a specific state in the set is based on inference from the Uncorrelated Encounter Model as developed by the MIT Lincoln Laboratory.[20] Using the resulting probabilistic reachable set, conflict probability between the intruder aircraft and the ownship can be derived.

2.3.4. Intent Based Projection

Encountered difficulties in trajectory prediction usually result from the lack of contextual information from the surrounding airspace to create a reliable forecast of the flight path. This missing contextual information is often correlated with pilot intent which is driven by, among others, the following factors; ATC commands, weather and surrounding traffic. The accuracy of models which are solely based on current aircraft states and flight modes tend to degrade for larger look-ahead times. Intent inference is therefore an important prerequisite for accurate predictions. Part of the available intent information relates to the a-priori filed flight plan, which can be considered available information for at least the own aircraft. In the future, flight plan sharing between aircraft might become enabled through Mode-S communication. This is important regarding future operational concepts like NextGen and SESAR as flight plans will become increasingly dynamic throughout the flight, making a-priori shared information less reliable.

Models for intent inference have been extensively studied in literature. Hwang et al.[18][44] approaches trajectory prediction by applying a combination of an aircraft dynamics model for short term prediction and an intent inference model for medium and long term prediction. For the intent inference model, the intent is defined as a set of waypoints for the aircraft, mapping the intended flight path. Besides flight plan information and weather, static information about the airspace is included as well. From a finite set of predefined intent models, the respective likelihood is calculated based on the current aircraft state and the estimated flight mode using an interacting multiple model approach. Krozel et al. [22] generalizes this approach by reasoning from the fact that in a future free flight environment trajectory change points (TCPs) will likely be broadcasted and exchanged. However, when such a broadcast or information exchange fails, the system still needs to be able to infer the pilot intent of the own aircraft and other aircraft. A distinction is made between two feedback loops; outer-loop navigation and inner-loop navigation. The outer-loop navigation relates to navigation commands that change infrequently, like a predetermined flight path which sets the context for the pilot intent. The inner-loop includes the more direct pilot control on a shorter time interval, for which the outer loop state can be assumed constant. Using this model, a distinction can be made between local intent correlations and global intent correlations, where the latter sets the context of the behavior in the inner loop. When insufficient information is present to imply one of the TCP based intent inference models, the approach falls back on using an Interacting Multiple Model (IMM) based approach using Kalman filters and the current continuous aircraft state. Prerequisites for all intent based models is the availability of sufficient information about the flight plan, airspace and weather. Sources where this can be derived from are the flight plan, the flight information system (FIS) , static information about the airspace in the flight management system and transponder messages from other aircraft, for instance using ADS-B.

Liu et al.[26] extends the Interacting Multiple Model approach and intent inference techniques proposed by [18] with a backwards smoothing algorithm in order to preserve the built up cross-track error to reach more accurate predictions. The solution applies a check on the controlled time of arrival (CTA) to input pseudo measurements in the future flight path to maintain the course deviation. This is under the assumption that in a natural flight environment, the course deviation will not converge to zero between two trajectory change points.

Yokoyama[45] approaches the problem for intent inference by means of inverse optimal control. This implies that the intent is derived from the current and past aircraft state, assuming that the aircraft is approximately following the local optimal path corresponding to this intent. The intent then corresponds to the unknown cost function that is being optimized. Different intent models need to be defined beforehand. The use of direct heading to a waypoint, circling a waypoint and holding a reference heading are examined. The motion modeling of the aircraft is based on a set of state equations based on airspeed, wind velocity, heading angle, bank angle and axial acceleration. In reference [46] this approach is extended by applying Model Predictive

Control to the conflict detection and resolution phase. Furthermore, the unscented transform is applied to cover the propagation of non-linear uncertainty dynamics related to wind, control accuracy and intent uncertainty.

Intent inference models are closely related to conformance monitoring models used by ATC to detect anomalous behavior of the aircraft with respect to the latest flight plan. These type of models assist ATC in executing strategic conflict detection and resolution and assess the validity of the applied logic.[32] All intent inference approaches in literature assume the existence of a discrete set of intents in different dimensions. Available models in literature map the continuous aircraft state to one of the known intents or flag the aircraft state as abnormal. This is however prone to unexpected behavior flags when aircraft are considered that operate in different state envelopes, like UAVs. These might therefore be more difficult to map to the intent set used for larger general aviation airliners. Also the independent multiple model algorithms described before have their roots in mapping continuous states to discrete sets of predefined and known flight modes. As this mapping can be approached in different ways, the discrete nature of the problem outcome is prone to become hard to maintain as the diversity of air traffic vehicles and air traffic management continues to increase.

More classical model-based trajectory prediction approaches are often resource intensive to compute, especially when dynamic models of the aircraft are incorporated. Furthermore they can be affected by changes in model parameters as aircraft weight and environmental factors. Data driven approaches are less affected by these uncertainties as they could better adapt to the current state and contextual features. Also for these cases where it is infeasible to construct models for a large diverse set of vehicles, more data driven approaches can be used. Yang et al.[43] applied a data driven intent inference approach by modeling the aircraft intent based on trajectory clustering in the terminal airspace. As in the terminal airspace predefined flight plans are often not valid anymore and the set of possible intents is much larger than in an en-route environment. An online trajectory clustering approach is used to account for the real-time updating of the available routes. The method relies on both static and dynamic data related to the aircraft type and state and is less dependent on external information about weather, special use airspace or planned trajectories.

2.3.5. Additional approaches

The discussed methods for trajectory prediction are largely based on predetermined dynamic or kinematic models of the aircraft. Uncertainty is modeled into the equations or superimposed based on derived or assumed error distributions. However, recently, more data driven approaches have appeared in studies which infer models directly from historical state measurements and flight track data. These research initiatives are moving away from deterministic prediction methods towards more stochastic models and data driven approaches. Machine learning is often mentioned in this context as well.

One example of this type of research initiative is the DART project, founded by the EU and Eurocontrol as part of the SESAR program, is focused on Data Driven Aircraft Trajectory Prediction Exploratory Research. The objective of the program is to explore the application of machine learning and agent-based modelling techniques for predicting single and multiple aircraft trajectories. Taking a pure data-driven approach makes the use of these models less resource intensive compared to computing dynamical aircraft models as described before. Within the scope of the DART project, a distinction is made between "single trajectory prediction" and "collaborative trajectory prediction". Trajectories might interfere with each other, resulting in significant prediction errors. Therefore, collaborative trajectory prediction takes into account other trajectories resulting in globally optimized predictions.[9] Also DatAcron is a research project focusing on threat and abnormal activity detection in data related to fleets of moving entities. This project focuses on large-scale heterogeneous data integration, management and analysis. Their defined work packages include "Mobility pattern detection and forecasting" and "Complex event recognition and forecasting".[41]

In the context of purely data driven approaches, the concept of trajectory clustering as mentioned is studied quite extensively in literature. Eckstein[11] defines a taxonomy for different flight track segments based on TRACON radar tracks. First, filtering of the raw measurements is applied based on non-parametric least-squares filters. After, principle component analysis and k-means clustering is used to group flight tracks from the past into clusters of representative traffic flows. These traffic flows can be used to measure conformance of flights to these established paths. Any deviations can accordingly be categorized. The method of combining PCA and k-means is applied to a large set of historical track data. The approach as applied in the reference is not made suitable for on-line adaptation of these clusters. Schelling[35] applies trajectory based similarity

metrics in historic flight data to propose optimal Traffic Management Initiatives to ATM based on information derived from past situations. This is applied in the context of avoidance of bad weather areas. Thirteen similarity metrics are compared qualitatively by applying them on historic flight track data and weather data. From this it was concluded that an Euclidean distance metric performed best compared to its computational efficiency where the use of Dynamic Time Warping produced more accurate results but at a higher computational cost.

Gariel et al.[16] applies trajectory clustering in the context of monitoring the average real-time conformance of all aircraft to their planned paths in the airspace at a given point in time. This is done by first identifying the turning points in the flight path and clustering them using k-means or DBSCAN. These clusters then represent respective waypoints which are used to create clusters of trajectories. Trajectory matching is performed using the Longest Common Subsequence metric. Also time series clustering is used for extracting specific flight phases based on ADS-B data using DBSCAN.[38]

Approaches that are more tailored to changing dynamic environments can also be found. Zhou et al.[47] propose a method for adaptive probabilistic path prediction with the use of reference trajectories which are selected on the fly from a registry. Using previous states from the own vehicle, a set of matching recent trajectories of other vehicles is derived from the system memory. Then a path prediction algorithm is applied which is based on a generalization of the Hidden Markov Model including derived state spaces from the selected trajectories. The state space clusters are build using an agglomerative clustering process. Hidden Markov Models are more often applied in this context. They are well suited to capture state changes in an environment with a finite number of possible states as is the case with aircraft flight paths which can be viewed as a sequence of a limited set of maneuvers.[1] Ayhan[33] approaches trajectory prediction by handling the airspace as a three dimensional grid network modeling the trajectory as a sequence of four-dimensional cubes based on latitude, longitude, altitude and time. A trajectory prediction model is proposed based on a Hidden Markov Model trained on historical data. Different data sources are considered including weather data.

Monreale et al.[27] applies similar trajectory pattern mining to path prediction by building a tree of trajectory patterns. One of the additional features here is to represent a trajectory by a small number of patterns that represent frequent movements. These frequent movement patterns are connected in a hierarchical tree structure where the nodes represent shared frequently visited locations.

Recently, neural networks are often proposed as they form very good self-optimizing non-linear mappings between the inputs and outputs. Recurrent variants like Recurrent Neural Networks and Long-Short-Term memory networks are often applied to the forecasting of time series. De Leege et al.[10] apply Generalized Linear Models to forecasting the arrival time of an aircraft in Continuous Descent Operations using aircraft type and weather parameters as inputs to describe the variance. The use of neural networks for regression and support vector regression provide similar results in this case.

Fernando et al.[15] proposes the use of a Tree Memory Network for trajectory prediction. This type of neural network could better capture both short and long term dependencies in sequence to sequence mappings by using a hierarchical tree structure for memorizing sequence features.

2.4. Conflict detection and alerting methods

The discussed trajectory prediction methodologies are a preliminary for airborne and ground based conflict detection and collision avoidance. Three important systems can be distinguished:

- **Medium Term Conflict Detection (MTCD)** MTCD is a general term for ground based conflict detection systems used by air traffic control for separation management. MTCD warns about potential conflicts between aircraft up to 20 minutes ahead.
- **Short Term Conflict Alert (STCA)** STCA indicates systems for ground based conflict detection for seperation violations within a short look ahead time of two minutes. The system provides a visual warning to the air traffic controller.
- **Airborne Collision Avoidance System (ACAS)** ACAS is term for airborne conflict detection systems for very short look ahead times up to one minute. The current implementation of ACAS is the Traffic Alert & Collision Avoidance System (TCAS).

Figure 2.4 provides an overview of the mentioned systems relative to the relevant detection horizon.

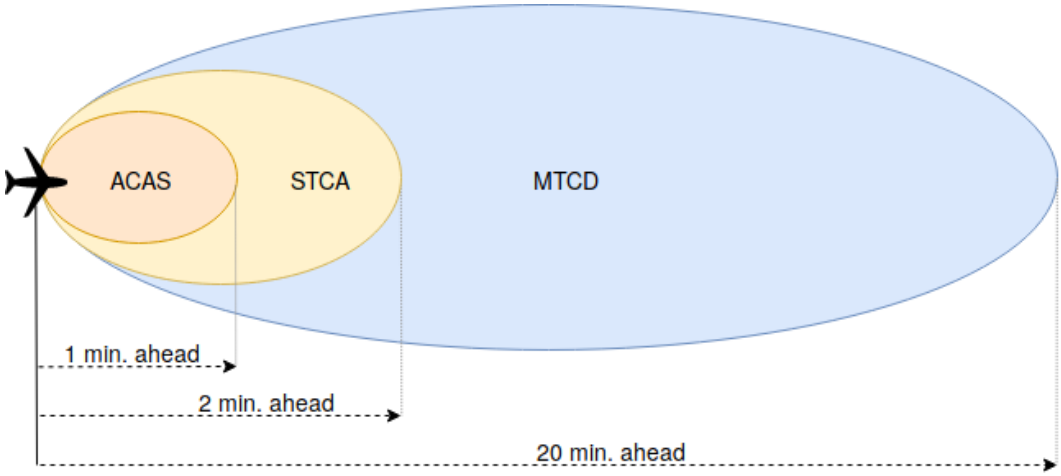


Figure 2.4: Overview of CD implementations and respective look ahead time horizons

The purpose of the MTCD system is both for safety as well as strategic planning for ATC. MTCD is viewed as a decision support tool which aims at providing means for more pro-active air traffic control. This will reduce the Tactical Controller workload as potential conflicts can be detected earlier and mitigated by the Planning Controller.[7] In more imminent conflict situations, STCA is a concept which acts as a safety net for mitigating short term conflict situations. When a conflict is projected to happen within a time horizon of two minutes, the controller is alerted so he can take appropriate action. As MTCD, STCA is more a concept than an actual system of which different implementations can be found.[13] The purpose of STCA is primarily to enhance safety. In comparison, MTCD is seen more as a decision support tool which helps in mitigating risks and workload away from the tactical controller. Both MTCD and STCA are ground based systems.

The Traffic alert and Collision Avoidance System (TCAS) is a mandatory system to be installed in all aircraft with a take-off mass of over 5.700kg or that allowed to carry more than 19 passengers.[14] It is an implementation of ACAS and has been introduced in 1989 to reduce the risk of mid-air collisions. The current version which is in use throughout is TCAS-II. The TCAS system consists of a combination of surveillance functionality with collision avoidance logic by providing resolution advisories (RA) when two aircraft are prone to be in conflict with each other. These resolution advisories are based on vertical movements only. The TCAS system is not able to also incorporate horizontal resolution advisories. Therefore, the TCAS system will tell a pilot to climb, descend or stay at the current flight level. Before a RA is emitted, first a traffic advisory (TA) alert is raised to warn the pilot about a nearby aircraft. The surveillance functionality of TCAS is primary based on radar but switches to active Mode-C or Mode-S transponder interrogation below a certain range. When another aircraft is not equipped with a TCAS system, a resolution advisory can not be generated. The inner workings of the TCAS system are based on deterministic propagation of the own and other aircraft relative position using the slant range, altitude and relative bearing. Based on this information, the time to the Closest Point of Approach (CPA) between two aircraft is calculated. Alerts are respectively raised based on this calculated time. the TCAS system takes into account several buffers in order to take into account flight path uncertainty. As the uncertainty grows with the projection time into the future, usually a look-ahead time buffer of five minutes is used for raising alerts. One downside of the TCAS system is that it often results in nuisance warnings to the pilot as no other intent information or unexpected maneuvers are taken into account in the calculations. Therefore there will always be a trade-off between the false alarm rate and the robustness of collision detection of the system.[14][21][6]

In order to overcome the limitations of the current TCAS system, the FAA initiated research in 2008 into a new system called the Airborne Collision Avoidance System X (ACAS-X). ACAS-X mainly tackles the problem of nuisance warnings and is usable in new operational concepts, making it more future-proof. The new ACAS-X system is also more suited for different types of aircraft, like unmanned vehicles, as it is more agnostic to the type of sensory input data it receives. Its workings are based on probabilistic state estimation models instead of the deterministic models for determining the closest point of approach used in TCAS. The state estimation

of an aircraft is both based on a dynamic aircraft model as well as a sensor model, which are both probabilistic. ACAS-X uses a pre-generated look-up logic table which is based on optimized action logic for different aircraft states which is created by modeling the resolution advisory problem as a Markov Decision Process (MDP) .

Whether an alert should be given depends on a combination of the conflict probability and a set of criticality measures.[30] This conflict probability depends on the chosen trajectory prediction method. In the case of deterministic single trajectory propagation, the conflict probability can be set by superimposing error distributions and calculating the probability based on the convolution of the future location distribution of both approaching aircraft. The same holds when a probabilistic propagation method is used.[28][3] When using worst case state propagation, the maximum probability of conflict is used implicitly for evaluation of the scenario. A collision risk measure can also be based on different methods by placing a set of thresholds on relative future position and closing rate as in the TCAS system.

Often, the probability of conflict is difficult to determine analytically. An analytical approach can be applied when for instance Gaussian error distributions are assumed, but these assumptions might not always hold in every scenario, resulting in unreliable probabilities. When an analytical approach is not feasible, an often used method is making use of real-time Monte Carlo method simulations. By generating many different simulations of a situation based on the determined probability distributions, a close approximation of the actual conflict probability can be determined. However, depending on the complexity of the trajectory propagation model, simulating many different possibilities is not always feasible in terms of computational efficiency in a real-time scenario. The conflict probability can also be approximated computationally by discretizing both the aircraft probability distributions for their trajectories and calculating the overlap in the distributions which could serve as a metric.

These resulting probabilities are translated into a binary classification for triggering an alert or resolution advisory. Setting the corresponding thresholds leads to a trade-off between the missed detection (false negative) rate of the algorithm and the number of nuisance warnings to the pilot (false positives). Usually different thresholds are being used for alarms which are often lower than the thresholds for a resolution advisory.

2.5. Model validation and performance metrics

The validation of the trajectory prediction and conflict detection methods can either be done using synthesized aircraft trajectories or by using actual trajectories from radar or ADS-B data. The challenge in using actual aircraft data however is the fact that conflict resolution maneuvers are present in the data which makes these scenarios unusable for validation of CD&R methodologies. An approach would be to simulate conflicts by combining flights from different flight levels which would have resulted in a conflict if the same flight level would have been used.[30] A source of error to account for is the varying wind direction and speed for different altitudes which might affect the results. However, flight levels which are close together are assumed to have less difference in wind and weather conditions, especially at higher altitudes.

Trajectories can also be emulated using various approaches. Existing models like for instance the Uncorrelated Encounter Model as described in[20] developed by the MIT Lincoln Laboratory could be used. Or trajectories can be simulated from an existing dynamical or kinematic model for aircraft.

2.6. Automatic Dependent Surveillance Broadcast (ADS-B)

Automatic Dependent Surveillance Broadcast is one of the newer systems in use for surveillance purposes in air traffic management. Instead of relying on primary and secondary surveillance radars for aircraft state information, the ADS-B system enables a direct information link between two aircraft or between an aircraft and the ground. One of the important features of ADS-B is that it makes use of GNSS data for position determination and data from the on-board systems for velocity, altitude and heading. This results in more accurate state information than when this information should be derived from radar systems. ADS-B does not require active interrogation by other transponders as the broadcast is automatic and omnidirectional. This enables it's use by a variety of systems, both in the air and on the ground. Aircraft equipped with an ADS-B IN system are able to receive ADS-B messages, Flight Information Services Broadcast (FIS-B) messages and Traffic Information Services Broadcast (TIS-B) messages. Aircraft with an ADS-B OUT system are also able to send messages.

The term Automatic Dependent Surveillance refers to the fact that the position information is determined by the aircraft itself (dependent) and that the system does not need external input to function (automatic). Also no active link or contract needs to be established between the ADS-B transmitter and a receiver. ADS-B operates at 1090MHz as a Mode-S Extended Squitter.[37]

A message broadcast takes place with an interval of a few seconds. An ADS-B message consists of 112 bits split up in five parts. Important are the encoded ICAO address, a message type code and the data payload. Several different message types can be distinguished, encoding at least the following information:

- Surface position message
- Airborne position message (with either barometric altitude or GNSS altitude)
- Airborne velocities message:
  - East-West or North-South velocity. Heading could be derived from this.
  - Vertical climb rate
  - Intent change flag
- Aircraft operation status message

Every message also holds a time stamp from the epoch time for the creation of the message.

The use of ADS-B technology provides several advantages over the legacy infrastructure. The system increases situational awareness by providing coverage of the movements of other aircraft at a high update rate and with high position accuracy. Because of less expensive ground equipment compared to primary and secondary radar, surveillance will become cheaper and coverage will increase. ADS-B is an important element in the scope of Future Air Navigation Staregies (FANS) expecting to increase airspace capacity and making more autonomous operations possible. ADS-B data becomes more widely used in research to do data driven modeling. Sun et al.[39] for instance makes use of ADS-B data to estimate aircraft performance parameters using maximum likelihood estimations.

In the scope of this research, for which ADS-B messages are the main data source, it is important to be aware of its limitations and system accuracy. Previous studies have been performed on the quality and usability for ADS-B state information in CD&R applications.[40][24]

Verbraak et al.[40] studied the quality and accuracy of ADS-B data in the context of primary surveillance by assessing the latency and data quality of historical ADS-B messages. This was done by comparing ADS-B messages with reference radar based ground tracks obtained from Eurocontrol. The research shows that the main sources of inaccuracy arise from the quality of the installed ground and airborne equipment. Latency in the position acquisiton and broadcast, for both data processing as well as data transfer introduce errors because of the time offset which can not be fully known to the receiving party. Also, the GNSS position accuracy is directly translated into the message, making it hold at least a position accuracy equal to the capability of the GNSS system and its coverage at that moment. It can be concluded that for the majority of aircraft, the quality of the ADS-B data is sufficient for surveillance purposes taking into account the size of the Intruder Protected Zone (IPZ) but that the use as a primary means of surveillance still needs improvement. [40]

A different study from Langejan et al.[24] assesses the performance degrading elements of ADS-B in the context of CD&R. Both situation related disturbances, which include a degraded reception probability due to range or signal inference, as well as system related disturbances are described.

Considering system related properties, mentioned are the effect of the truncation of number digits because of the limited number of bits in one message. This effectively reduces the number of digits to be transmitted to six for latitude and longitude. The other mentioned system inaccuracies are GNSS accuracy and latency, as discussed earlier. For situation related properties, the range and inference affect the ADS-B data transfer. A larger range decreases the reception probability of a message due to the lower received signal power. The research uses a physical model to assess this effect. Interference also effects the reception probability of a message. When there are more aircraft in range, the chance of inference increases, lowering the reception accuracy of the ADS-B data link. Inference of messages is modeled based on a Poisson distribution for the number of sent messages at a certain point in time, depending on the number of aircraft in range.

The accuracy of the measured state on the GPS accuracy related also to the on-board measurement equipment. This leads to an ± 8 meter accuracy with 95% confidence. The truncation effect resulting from the the availability of only six significant digits for position results in 9-17 m error. There is also system latency present which is approximately 20ms resulting in a ± 5 meter accuracy in cruise. Considering situation related accuracies. The interference probability of a message in a realistic scenario decreases approximately below 0.8 above a reception range of 150 nm. This same threshold of 0.8 reception probability is reached when the number of aircraft in range is larger than 400. These two parameters influence eachother in a real scenario, so the probabilities need to be multiplied in this case.

These studies show that mainly the situation related elements affect the ADS-B data quality. The range and the number of surrounding aircraft largely affect the signal quality due to interference and reception probability. However, also here it is shown that the accuracy of ADS-B state information is sufficient for CD&R purposes by comparison to a perfect state model.

2.7. Eurocontrol ETFMS Data

This data source is made available by Eurocontrol. For the purpose of this research, use is made of data originating from the Enhanced Traffic Flow Management System (ETFMS) .[8] There are two versions of this data available:

- **ETFMS Model 1 (M1) data**
  This data corresponds to the 4D trajectory of the latest filed flight plan before the start of the flight.
- **ETFMS Model 3 (M3) data**
  This data holds the information about the 4D trajectory after the flight is operated and integrated with the Correlated Position Report (CPR) data. This data is intended to be closest to the actual trajectory that the aircraft has flown. Its position determination is based on ground radar data. Only deviations from the flight plan larger than five minutes, seven flight levels or 20 nautical miles are taken into account. Otherwise, the trajectory from the original flight plan is shown.

Provided from the system are historical SegOut6 (SO6) files which hold the described 4-dimensional trajectory data. A flight is split up into different segments, each with start and end coordinates and a corresponding time stamp. The important data fields are shown in table 2.1.

Field	**Description**
Segment identifier	States the start and end waypoints of the segment
Flight Origin	The identifier of origin airport of the flight
Flight Destination	The identifier of the detination airport
Aircraft Type	
Segment start/end latitude/longitude	For both the start and end location of the flight segment, the respective latitude and longitude are provided
Segment start/end time and date	For both the start and end location of the segment, the time and date are provided corresponding to the 4D trajectory
Segment start/end Flight Level	The FL for the start and end location of the segment
Flight Status	0=climb, 1=descent, 2=cruise
Callsign	The callsign of the aircraft

Table 2.1: Overview of variables available in the SO6 data set

This data will be used as reference tracks for the ADS-B data sets to derive initial aircraft intent. For comparison of the actual flown track with ADS-B data, the mentioned thresholds should be taken into account.

3

Research Objectives

The main goal of this research is to assess the uncertainty in flight trajectories with respect to short to medium-term conflict detection. This is approached by comparing three trajectory prediction methodologies with the use of historic ADS-B data sets and flight plan data. The data-driven performance benchmarking of the different models not only provides insights in how the approaches compare, but also aims to provide an empirical boundary on the look-ahead time for conflict detection.

The following sub-goals are defined:

- Performing a statistical analysis of flight path uncertainty with respect to the look-ahead time based on available historical ADS-B data sets
- Creating a data driven assessment of the performance of state-deterministic, state-probabilistic and intent-deterministic conflict detection methodologies with the use of historical ADS-B and flight plan data.

The expected outcomes of the research are therefore twofold: First to create a data-driven, statistical view on the relationship between uncertainty in a flight trajectory and the look-ahead time. Second, to apply this knowledge to the assessment of three different conflict detection methodologies by using historical ADS-B data. For the latter, state-deterministic projections, state-state-probabilistic projections and intent-deterministic projections are considered where for intent-deterministic projection, flight plan data will be used as a data source. The proposed research will provide insight in how well the different methodologies mitigate uncertainty and how they compare on detection capability based on historic flight data.

3.1. Research Questions

The following questions and corresponding sub-questions have been formulated for the proposed research.

1. How does the uncertainty in a flight path relate to the look-ahead time based on state-deterministic projection?
   (a) Regarding the separation limits in ATM, what would be the empirical upper bound on look-ahead time for conflict detection using this state-deterministic approach?
   (b) How to extract conflict situations from historical ADS-B data while filtering out the actual resolutions and mitigations of potential conflicts present in the data?
2. How do state-deterministic projection, state-probabilistic projection and intent-deterministic projection compare in terms of conflict detection capability when applied to intrusion scenarios from historical ADS-B data?
   (a) How does the conflict detection performance compare between state-deterministic, state-probabilistic and intent-deterministic approaches?

(b) How does the conflict detection performance change with respect to the look-ahead time for these three methodologies?

(c) What is the empirical upper bound on look-ahead time for conflict detection using these methodologies?

The first part of the proposed research is focused on quantifying the relationship between the variance, or uncertainty, in aircraft trajectories and the look-ahead time using state-deterministic trajectory propagation. The expected outcome of the analysis will show how the distribution of the errors from the state propagation estimates will behave for different values of the look-ahead time. Historical flight paths derived from ADS-B data sets will be used for the analysis. It is to be expected that the variance is increasing with the look-ahead time but the relationship is up to now only quantified in a limited way. Initially a two-dimensional approach will be used, taking only horizontal deviations in level flights into account. This choice is made first of all to be able to derive a sufficiently large set of conflict scenarios from the historic ADS-B data. Assuming a two-dimensional approach enables combining trajectories from different flight levels into one conflict scenario without compromising the reliability of the data resulting from mitigation measures taken in the actual conflict scenarios. The respective look-ahead time horizon for the analysis is chosen to be 20 minutes which will cover both short and medium-term conflict detection approaches.

The second research question aims at addressing the lack of quantitative comparisons between different trajectory prediction and conflict detection approaches. Prevalent in literature are qualitative assessments of methodologies or the implementation in simulated environments. This research will address the comparison of three global methodologies for trajectory prediction and assess them based on their conflict detection performance in scenarios built from historical flight data. The methods which will be compared are state-deterministic trajectory prediction, state-probabilistic trajectory prediction and intent-deterministic trajectory prediction. The look-ahead time will be taken into account by means of a time-to-conflict measure to be able to compare the methodologies on both their short term as well as medium or long term detection capabilities. By comparing different methodologies, their prediction capability and their input data, an assessment can be made of how well the model makes use of the input information and if increasing model complexity also results in better predictions.

The performance comparison will be based on a measure which includes false positives and false negatives of conflict detections. An important challenge in using actual historic data is that conflict avoidance maneuvers are already present in the data resulting from the present air traffic management systems. An approach is proposed where the representative statistical distribution of a flight trajectory is preserved while filtering out deviations corresponding to conflict avoidance.

Improved trajectory prediction enables new methods to cope with current estimated capacity issues of the current air traffic control systems in place by enabling a higher traffic density with a smaller safety buffer size and better capacity planning capabilities. Furthermore, by developing a first step towards a standardized framework for the evaluation of trajectory prediction approaches based on historical ADS-B data, future research and modeling endeavors can be benchmarked in a more realistic way.

3.2. Research Methodology

An overview of the different steps taken in order to achieve the research goals is provided in figure 3.1. A distinction is made between data processing, model development and statistical analysis. The outcomes of the research follow from the latter, however, first development of data processing and analytical methods are required. The swimlane model provided in the figure shows a breakdown of the work involved in the different parallel tracks and their respective sequence in time.

The first steps of the research aims at providing a quantitative evaluation of the distribution of deviations between a state based trajectory propagation and the actual trajectory. This will include mainly a statistical analysis of a large set of historical flight data from ADS-B data sets and corresponding ETFMS flight plan information. This research phase also includes setting up a proper data infrastructure to work with the large amount of available data. Thereafter, a state-probabilistic trajectory propagation method and an intent-deterministic propagation method will be developed and compared based on their conflict detection capability in different scenarios based on historic ADS-B data.

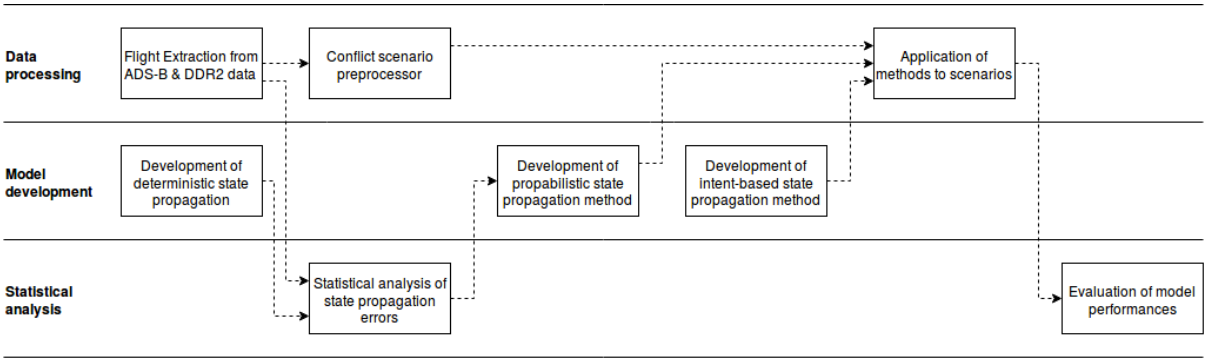


Figure 3.1: Swimlane model of the proposed research methodology

3.3. Hypotheses

As elaborated in section 2.3.3, the expected largest influences on a planned flight path result from changes in intent from the pilot or from ATC. These can be induced by traffic, weather or redirects for timing purposes. These intent changes are likely to cause large deviations from planned trajectory.

Considering deterministic state propagation, the variance in deviations from the expected flight trajectory will be expected to grow continuously with the look-ahead time as the chance on trajectory changes increases. It is also likely that errors build up over time and cascade, creating increasingly larger errors compared to the projection. Current limits on the look-ahead time incorporated in implemented systems are around 5 minutes in order to hold the right balance between anticipation time and any nuisance warnings from any implemented conflict detection system.

The respective flight level will likely influence the quantity of observed track deviations as well. Deterministic projections made for flights at higher altitudes likely have less deviations than those made at lower altitudes. Especially when aircraft reach their cruise flight level, their behavior tends to be more stable and less affected by vectoring procedures and the presence of other traffic.

In the scope of conflict detection methods, a distinction will be made between state-based and intent-based approaches. The state based methods, consisting of state-probabilistic projection and state-deterministic projection, only use information from the recent aircraft states without looking at intent information like waypoints. Intent-deterministic methods use next waypoints as indicators for the aircraft trajectory, in combination with a state based model. State based methods are expected to be accurate in the short term as they are based on the extrapolation of the current state, not the expected one. Thereby current biases or deviations from the intended track to not affect the projections. Intent-deterministic methods are expected to perform better in the medium to long term range as part of the future trajectory changes are already known. Therefore, the upper bound on look-ahead time is expected to be higher for intent-deterministic methods compared to probabilistic or deterministic methods although the accuracy of the latter methods is likely higher, especially in the short to medium term. The reason for this is that precision gained in the short term by looking at the current aircraft motion is likely to be made insignificant in the medium to long term as intent-deterministic trajectory changes take place which can for a large part be forecasted from the flight plan information.

4

State Propagation & Conflict Detection methods

This research considers three methods for the detection of conflicts. A state-deterministic method which uses a linear extrapolation of the states of the aircraft pair. A state-probabilistic method which takes into account the distribution of convergence speeds of both aircraft. And an intent-deterministic method which uses flight plan data to create a deterministic state extrapolation based on the expected heading instead of the actual heading. The three approaches share the fact that they are all based on the same linear model but differ in the way they are enriched with additional information. The state-deterministic propagation serves as the main benchmark as it only relies on the latest aircraft state.

The global outcome metric for the three detection methods is an estimated time to intrusion (TTI_{est}) . This can accordingly be compared to an actual time to intrusion (TTI_{act}) to validate if an intrusion was detected timely, too late, or not at all.

The time to intrusion metric is calculated based on a propagation of the initial convergence speed (V_{conv}) in m/s between the two aircraft depending on the look-ahead time t_{la}. Meaning the rate in meters per second at which the two aircraft come closer to eachother. In the linear propagation applied in all three detection methods it is assumed that the initial airspeed and heading of both aircraft remains constant.

The initial convergence speed is calculated as follows. First an indication is needed if both aircraft fly at an angle in which they approach eachother. Otherwise the flight paths are assumed to diverge. This angle has to be derived from both the coordinates of the aircraft as well as their headings. The convergence angle γ of the paths as shown in figure 4.1 is calculated as described below.

The compass bearing of the line from the location of aircraft 1 to aircraft 2 ($\Theta_{c1_{c}2}$) is used. Then the angle between the headings of both aircraft and $\Theta_{c1_{c}2}$ are determined in degrees on a range from $0^o - 360^o$ as in $\alpha_{ac1} = (360 + \Theta_{c1_{c}2} - \Theta_{ac1}) \pmod{360}$ and $\alpha_{ac2} = (360 + \Theta_{c1_{c}2} - \Theta_{ac2}) \pmod{360}$.

Now if $\alpha_{ac1} > \alpha_{ac2} > 180^o$ or $\alpha_{ac1} < \alpha_{ac2} < 180^o$ it can be assumed that the flight paths form a triangle and therefore converge.

To calculate the convergence speed accordingly, use is made of the speed vectors of both aircraft with respect to the line between aircraft 1 and aircraft 2. Based on the angles β_{ac1} and β_{ac2} the airspeed is decomposed in two perpendicular speed vectors V_x and V_y for both aircraft as shown in figure 4.2 using the mentioned angles as in:

$$V_{y_{ac}} = \sin(\beta) \cdot V_{ac} \tag{4.1}$$

$$V_{x_{ac}} = \cos(\beta) \cdot V_{ac} \tag{4.2}$$

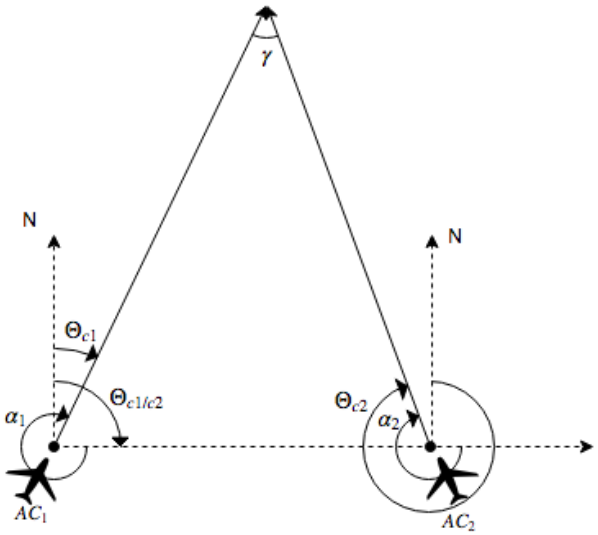


Figure 4.1: Visual representation of considered angles between flight paths

The decomposed distance vectors as a function of t_{la} based on the initial speed vectors are:

$$\bar{D}_{ac1}(t_{la}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} dx_{ac1} \\ dy_{ac1} \end{bmatrix} \cdot t_{la} \tag{4.3}$$

and

$$\bar{D}_{ac2}(t_{la}) = \begin{bmatrix} D_{init} \\ 0 \end{bmatrix} + \begin{bmatrix} dx_{ac2} \\ dy_{ac2} \end{bmatrix} \cdot t_{la} \tag{4.4}$$

The simplified equation for the decomposed distance between the two aircraft becomes:

$$\bar{D}_{ac1/ac2}(t_{la}) = \begin{bmatrix} (dx_{ac1} + dx_{ac2}) \cdot t_{la} - D_{init} \\ dy_{ac1} - dy_{ac2} \end{bmatrix} \tag{4.5}$$

To calculate the total distance at look-ahead time t_{la} the equation 4.6 is used.

$$D_{ac1/ac2}(t_{la}) = \sqrt{\bar{D}^T_{ac1/ac2}(t_{la}) \cdot \bar{D}_{ac1/ac2}(t_{la})} \tag{4.6}$$

4.1. Development of state-deterministic propagation and conflict detection method

Deterministic state propagation implies that the current velocity vector is linearly extrapolated over a certain look-ahead time. This is a basic form of trajectory prediction and independent of external information or intent. A more detailed description of the method and its applications is given in chapter 2. To determine the resulting coordinate based on a starting point and initial heading and airspeed, the following formulas are applied:

$$\phi_{end}(V, \Theta, t, \phi_{start}) = \sin^{-1}(\sin(\phi_{start}) \cdot \cos\left(\frac{V \cdot t}{R_{earth}}\right) + \cos(\phi_{start}) \cdot \sin\left(\frac{V \cdot t}{R_{earth}}\right) \cdot \cos(\Theta)) \tag{4.7}$$

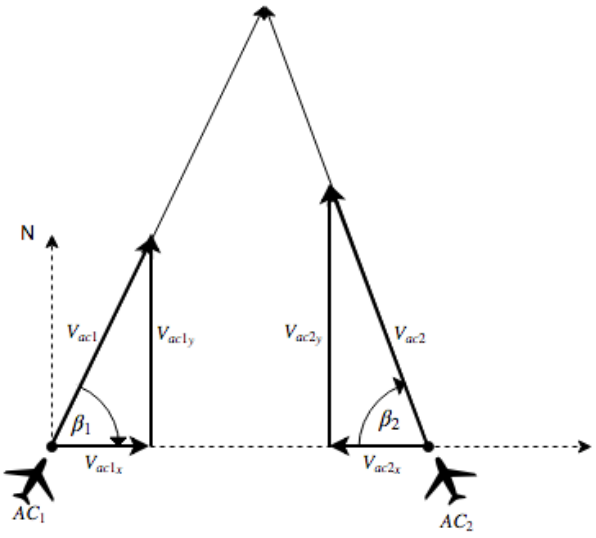


Figure 4.2: Visual representation of the speed decomposition

$$\lambda_{end}(V,\Theta,t,\lambda_{start},\phi_{start},\phi_{end}) = \lambda_{start} + \tan^{-1}\left(\frac{\sin(\Theta)\cdot\sin(\frac{V\cdot t}{R_{earth}})\cdot\cos(\phi_{start})}{\cos(\frac{V\cdot t}{R_{earth}}) - \sin(\phi_{start})\cdot\sin(\phi_{end})}\right) \tag{4.8}$$

Here, ϕ is the latitude in radians, λ is the longitude in radians, Θ the initial heading, V the initial airspeed and R_{earth} the radius of the earth in meters.

In the context of conflict detection, use is made of equation 4.6 to calculate the future distance between the pair of aircraft while the assumption is made for state-deterministic projection that the initial speeds and headings remain constant.

The estimated time to intrusion is calculated by applying equation 4.6 for a range of look-ahead times where an intrusion would imply that $D_{ac1/ac2} < d_{IPZ}$.

4.2. Statistical analysis of state propagation errors using extracted historic flights

Based on the state-deterministic propagation method, the error will be determined between propagated and actual trajectories. This will give a quantification of the deviation of the aircraft from its initial state set out against the look-ahead time. A large amount of historical flight data will be used to generate a data distribution of trajectory deviations at specific look-ahead times and perform analysis of its characteristics at these moments. In order to assess the differences between an original and a projected trajectory, the cross-track error (CTE) , along-track error (ATE) and resulting total-track error (TTE) are used. Because for both trajectories, the points have the same timestamps, only the distances between them have to be considered. These measures are defined as shown in figure 4.3. This will provide a location independent measures which makes it useful for comparing various flight paths from different aircraft.

Each trajectory part between two points is approached as a straight line segment. The problem of determining the track-errors is then defined as shown in figure 4.3. Here, the trajectory points annotated with an 'a' belong to the original trajectory. Those with a 'b' belong to the projected trajectory. The TTE is then defined as the absolute distance between a projected point and an original point. The CTE is the perpendicular distance between the projected point and the original trajectory. The ATE results from this as the CTE and ATE are perpendicular. Therefore, the ATE is the distance between the original trajectory point and the projection of the propagated trajectory onto the original.

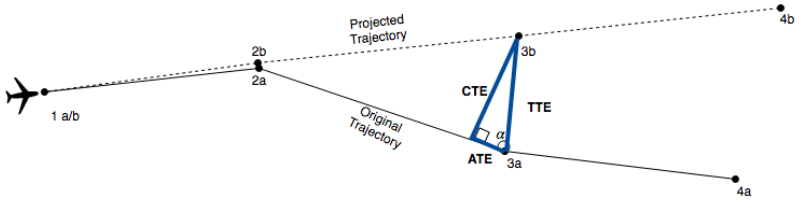


Figure 4.3: Definitions of Cross-Track Error and Along-Track Error

The calculations for this are based on the angle (α) between the TTE line and the original trajectory line at point 3b. The CTE is then defined as $CTE = sin(\alpha) * TTE$. When the angle α becomes larger than 90 degrees, the angle becomes $180^o - \alpha$.

The CTE is defined positive if the angle between the projected trajectory and the original trajectory is clockwise positive. The ATE is defined positive if the projected trajectory is longer than the original trajectory. The combination of both the CTE and ATE for many flight samples aggregated at a certain look-ahead time will yield a location distribution of the aircraft.

Initial analysis of this location distribution is aimed at identifying an initial empirical upper bound on the look-ahead time for which any predictions are still useful considering air traffic management procedures. At a certain look-ahead time the uncertainty will become larger than the IPZ which indicates a bound for the prediction being unreliable for conflict detection. Furthermore, it will be assessed how the error distribution characteristics change for different look-ahead times.

All trajectory related signals are expected to have a time dependency between different entries. Therefore, expected characteristics are the presence of local trends and heteroskedastic behaviour, meaning that the variance in subsections of the signal is significantly different. One also has to take into account the autocorrelation of the flight path deviations. Once an aircraft is flying at a deviation from its planned trajectory at a certain point, this will induce that the second measurement also includes a deviation approximately similar to the initial one which implies an autocorrelation between measurements. It can also be assumed that the variance present in a trajectory is not constant throughout the entire flight, but that the magnitude of the variance will change itself as well implying heteroskedasticity of the signal. Tests need to be performed on the data in order to identify these characteristics.

4.3. Development of a state-probabilistic conflict detection method

This step aims at developing a representative method for a probabilistic approach to state propagation where instead of a single future point, a distribution of location estimates is used. A conflict detection methodology would then be based on the intersection of future state distributions for two aircraft, where a detection is based on a certain probability threshold for intruding the IPZ.

When considering the distribution of the distance between two aircraft over time, one has to take into account that the respective error distributions as derived from a state-deterministic propagation approach are likely to be correlated. This correlation can be due to shared weather circumstances like wind or responses to other surrounding traffic. Therefore, using the distribution derived from the CTE and ATE over time for both aircraft is not valid.

Use is therefore made of the distribution of the error between the expected distance between two aircraft and the actual distance, as stated in equation 4.6. Figure 4.4 provides a visual overview of the metrics and scenario considered. In order to correct for the absolute differences in initial position, the distribution of the variable $D_{ac1/ac2_{error}}$ is used as given in equation 4.9.

$$D_{ac1/ac2_{error}}(D_{ac1/ac2_{actual}}, t_{la}) = \frac{D_{ac1/ac2_{actual}} - D_{ac1/ac2_{est}}}{D_{ac1/ac2_{actual}} \cdot t_{la}} \tag{4.9}$$

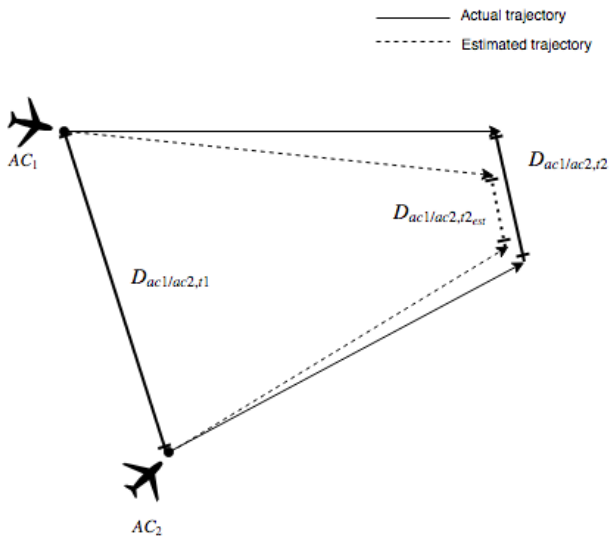


Figure 4.4: Scenario of two aircraft and the considered distances used in the state-probabilistic approach

This gives a scaled equivalent of the estimation difference. The look-ahead time is used as a scaling factor as the distributions grow approximately linear in time. It can be empirically shown that the resulting error term has a approximately the same distribution for different look-ahead times as visualized in figure 4.5.

The term $D_{ac1/ac2_{error}}$ is thereby a term for the estimation error between the actual distance between two aircraft and the estimated distance scaled by the actual distance. This measure is thereby independent of the absolute distance but provides a relative metric. The reasons to choose for this metric are the following:

- The initial convergence speed is not linearly related to the convergence distance over time in a sense that $V_{conv}(t) \neq t \cdot V_{conv}(1)$.
- It is assumed that the distance between two aircraft is correlated with the estimation error.
- It is assumed that the error distributions grows linearly with the look-ahead time

For the detection method, use is made of an approximation of the probability density function (pdf) of $D_{ac1/ac2_{error}}$ using a Kernel Density Estimate (KDE). This will provide a non-parametric distribution estimate for which no assumptions of the actual distribution need to be made. The KDE is based on the flight scenarios filtered from the ADS-B data as described before.

The choice for non-parametric estimation is made as the distribution of the track errors is not normally distributed. The heavy-tails character of the distributions need to be preserved in the distribution estimates as well in order to make the state-probabilistic propagation method reliable. Thereby staying as close as possible to the empirical data distribution. However, overfitting of the distribution needs to be avoided. Artifacts resulting from the data set used in this research which do not generalize shall be smoothed out. This can be controlled by adjusting the kernel bandwidth in the estimation as described below. This choice for a non-parametric approach to the density estimation deems further appropriate as the sample size of the data set is large.

A Kernel Density Estimate is based on the summation of a set of kernel functions fixed at the locations of each data point as in eq. 4.10:

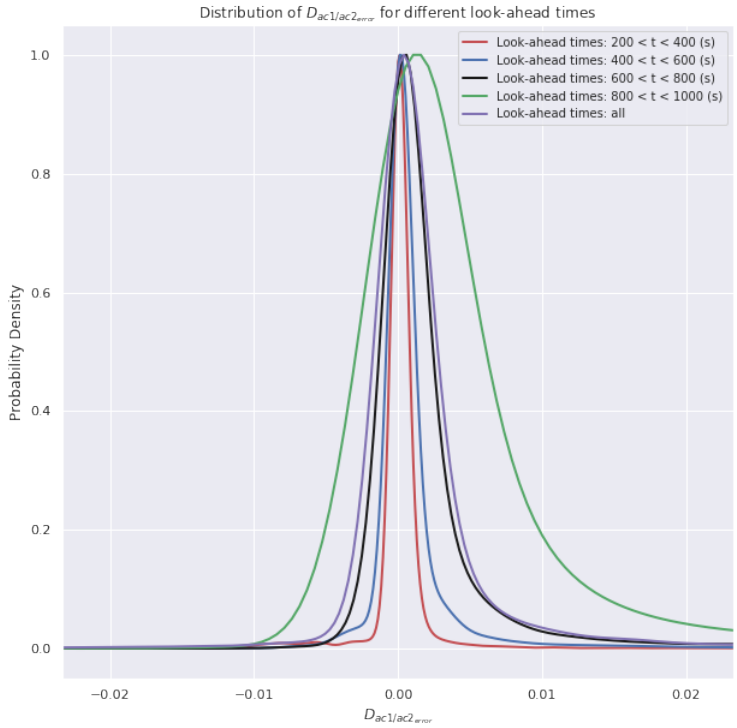


Figure 4.5: Estimated probability density plots (KDE) of variable $D_{ac1/ac2_{error}}$ for different look-ahead times

$$p_b(x) = \frac{1}{n \cdot b} \sum_{i=1}^{n} K\left(\frac{x - x_i}{b}\right) \tag{4.10}$$

Here, parameter b is the bandwith of the kernel which affects the kernel radius, thereby smoothing the estimation for higher values of b. Lower values of b result in a more accurate description of the data, but also cause more variance in the estimate. The bandwidth selection is done using 'Scott's rule of thumb', which makes use of equation 4.11 and can be used as the data closely resembles a gaussian distribution.

$$b = \frac{3.5 \cdot \sigma}{n^{\frac{1}{3}}} \tag{4.11}$$

The resulting pdf created from historic fligth scenarios is thereafter applied in the detection method. Effectively, the same approach as in the state-deterministic case is applied while now taking into account an upper and a lower boundary derived from a chosen quantile in the estimated pdf. These bounds are directly derived from the cumulative distribution function (cdf) which is the integral of the pdf as in $P(x) = \int_{-\inf}^{\inf} p(x)dx$. In the discrete version this becomes $P(X) = \sum_{i=1}^{n} p(x_i)$ for $X = x_0, ..., x_n$.

The respective value x_q belonging to the quantile value $q = \sum_{i=1}^{q} p(x_i)$ is then used as a boundary in further calculations. The estimated time to intrusion is again calculated by applying equation 4.6 for a range of look-ahead times where an intrusion would imply that $D_{ac1/ac2} - d_{lb} < d_{IPZ}$ or $D_{ac1/ac2} + d_{ub} < d_{IPZ}$ with d_{lb} and d_{ub} the respective lower and upper bound as given by the quantiles derived from the estimated cdf of the variable $D_{ac1/ac2_{error}}$.

4.4. Development of an intent-deterministic propagation method using flight plan data

The intent-deterministic conflict detection approach makes use of the next trajectory change point (TCP) as an approximation to the future flight path. Also in the ADS-B system, functionality is included in the newer

system versions to broadcast the next TCP of an aircraft so intent can be shared within the traffic network. For this research, use is made of the Current Tactical Flight Model (CTFM) data which is the same as the ETFMS Model 3 data which contains the last filed flight plan before departure, enriched with correlated position report (CPR) information to describe as close as possible the actual flown route of an aircraft. This will be used as the source for the TCP information. Only the next trajectory change point will be inferred at a certain location in order to keep the approach consistent with the current implementation state of ADS-B technology.

In order to perform conflict detection, a state based projection from the current aircraft position will be made using the compass bearing between the current and next waypoint as the expected heading. Thereby accounting for any translational bias with respect to the actual flight path. The aircraft is thereafter assumed to follow the line between the next waypoint and the following waypoint as soon as its current path crosses the line between this next waypoint and the following one. Only one TCP is considered in the projection. After one heading change the aircraft is assumed to follow its respective heading.

The conflict detection will be performed by simulating the flight path assuming a constant airspeed from the point of projection. The ADS-B flight segments and CTFM routes are matched based on callsign and respective start and end times of the flights. Thereby it is assumed that an ADS-B flight segment always fully falls within the CTFM timespan because the ADS-B flight segments are limited based on the radar range.

In order to perform the heading estimation, the CTFM waypoints and the ADS-B flight segments need to be aligned. Thereby assigning to each ADS-B data point a corresponding current and next waypoint from the CTFM data set as shown in figure 4.6. This assignment is based on the respective angles within the triangle formed by two consecutive waypoints and the aircraft coordinate. If both the angles $\alpha_1 < 90^o$ and $\alpha_2 < 90^o$ the aircraft is assumed to be in the current segment. However, when the waypoint trajectory is curved, this condition might be true for more than one segment. In this case, the segment which comes later in the sequence is chosen. The flight plan data is cropped to the first and last waypoint pairs that can be matched using the ADS-B flight segments. The matched trajectories are stored in a database.

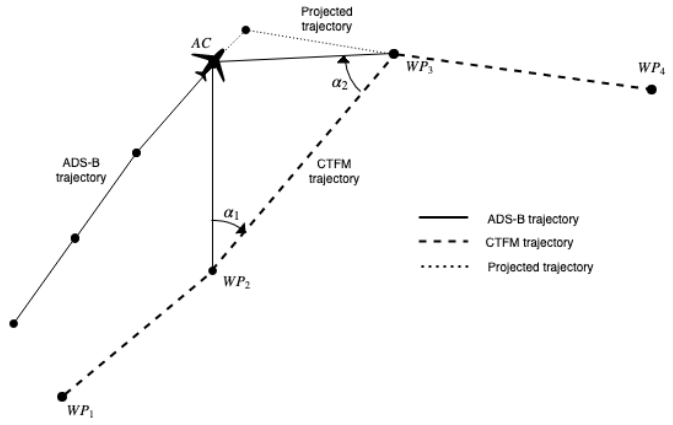


Figure 4.6: Aligning of ADS-B data points with CTFM waypoints

5

Extraction of Conflict Scenarios from ADS-B and ETFMS data

In this research use is made of two different data sets; ADS-B messages obtained from a receiver located at the TU Delft faculty of Aerospace Engineering and sets of historic traffic information provided by Eurocontrol. More information on the underlying systems and characteristics of both data sources is provided in chapter 5.4. In order to execute the experiments on the data, preprocessing of conflict scenarios is required.

5.1. Flight segment extraction from ADS-B and ETFMS data

The available ADS-B data is a set of CSV files containing raw encoded ADS-B messages ordered by time stamp. These separate messages eventually have to be transformed into respective flight segments which is chosen as the basic entity for further analysis work. The raw ADS-B messages have to be decoded and processed in order to extract the required state information on airspeed, altitude, GNSS location and aircraft callsign. The logic in these decoding and preprocessing steps is based on the work of J. Sun et al. [37]. To avoid unforeseen errors in data parsing and to ensure correct and efficient data handling a data pipeline is designed and developed. This pipeline will facilitate the various steps involved in processing the raw data files. It will support the data extraction and transformation from raw ADS-B messages into segregated flight segments. In each of the processing steps, some filtering is applied to ensure sufficient data quality for further analysis steps.

Figure 5.1 shows the different processing steps for both ADS-B data as well as DDR2 data sets. The input files contain ADS-B messages from a single date collected within a range of approximately 300 kilometers around the TU Delft faculty of Aerospace Engineering where the antenna is located. One has to take into account that the data quality degrades if messages originate further from the antenna. Also weather and traffic density play an important role here.

The first processing step transforms these raw messages into data points with position, velocity, altitude and identifier information. The second step clusters the flights based on callsign and splits data points into separate flights.

5.1.1. Decoding and merging of raw ADS-B messages

When the separate ADS-B messages are received, they are encoded in a hexadecimal format and all share the same structure, having five different parts:

1. Downlink Format
2. Capability
3. ICAO address
4. Data string. Depending on the Type Code, this Data string holds specific information.

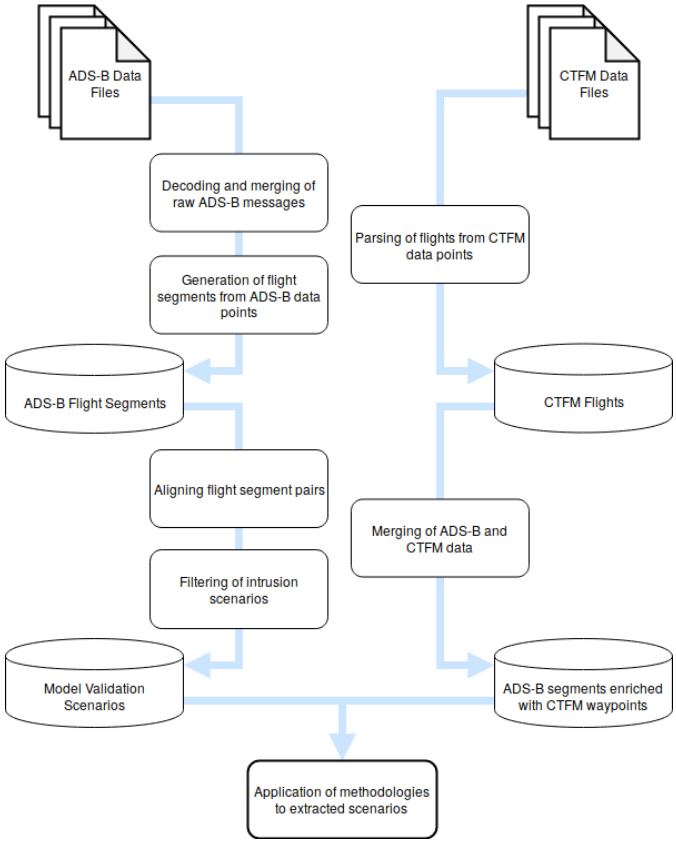


Figure 5.1: Schematic of the proposed data processing steps for ADS-B and CTFM files

5. Type Code
6. Parity ID

The incoming messages are first filtered based on their downlink format, indicating if they are an ADS-B or TIS-B message with downlink id 17 or 18 respectively. For use in this research, three different types of messages, each having different type codes, are considered:

1. **Airborne Position Reports with Barometric Altitude.** ADS-B makes use of Compact Position Reporting to encode the latitudes and longitudes of the aircraft position. In order to do this in a mondial unambiguous way, two separate messages are required. One having a flag with an odd identifier, and one with an even identifier. One could also decode the position based on a single messages using a reference position, for instance from the receiver. However, this way of decoding is only valid for relatively short distances from this reference point. A downside to using the unambiguous way with two separate messages is that not all incoming messages can be decoded correctly if either one of the two required position messages is either corrupted or not received. In the data parser, a pair of odd/even messages is only considered if they are received less than five seconds apart. Otherwise merging of separate position reports might occur resulting in inaccurate location readings.
2. **Airborne Velocity, Heading and Vertical Climb Rate** Airborne Velocity messages are sent in two different types: Ground Velocity and Airborne Velocity. Both indicated by their typecode. Both these message types also contain the encoded heading (if available) and a vertical climb rate. As airborne velocity messages are far more common, only these are used for further processing. These airborne velocity messages are further segregated into True Airspeed and Indicated Airspeed. For this research, the True Airspeed is used. Together with the velocity, this message type also contains data on the compass heading and vertical climb rate. The velocity message is used even if the heading or vertical climb rate information is not available.
3. **Aircraft Identification (Callsign)** The aircraft callsign information comes in a separate message and will

be used to merge the flight plan data from Eurocontrol with the ADS-B data.

All messages contain the unique ICAO address from the sending aircraft, which enables grouping and merging of the right set of raw messages. After grouping, the messages are decoded based on their respective type after which the three sets of message types are merged into one dataset. This merging is done based on the timestamps on which the individual messages are sent. The nearest message in time is used for merging where a tolerance of two seconds is applied. This merge results in only the time stamps being used for which there is both a position report, a velocity report and an identifier report available.

The resulting filtered set of messages contains rows with the following variables:

- A timestamp (in epoch)
- The ICAO address of the aircraft
- The aircraft callsign
- GNSS position in latitude and longitude
- The barometric altitude
- The airspeed
- The heading in degrees (if available)
- The rate of clime (if available)

5.1.2. Extraction of flight segments from ADS-B data points

To generate flight segments out of the aggregate set of resulting messages, the data is split based on the ICAO address. After this, filtering is applied to the difference in timestamps of successive rows. If the time between consecutive message rows is greater then 600 seconds, they are considered to belong to different flight segments. These time gaps are assumed to occur between landings and take-offs at airports or at leave or re-entry of the reception range of the antenna. This results in a set of flight segments belonging to respective ICAO addresses where no time gaps of larger than 600 seconds have occurred between consecutive messages. These flight segments are stored in a central database for use in later analysis stages and can easily be queried and filtered on numerous parameters.

The following list summarizes the different assumptions and filters applied in this processing step:

- Only messages with a downlink format id of 17 or 18 are used (ADS-B or TIS-B).
- If the number of seconds between consecutive messages for the same ICAO address is greater than **600 seconds**, the messages are considered to belong to different flights.
- A pair of odd/even position report messages is only considered if they are received less than **5 seconds** apart.
- For merging the three types of messages (position, velocity and callsign) on timestamp, a tolerance of **2 seconds** is applied.
- Only the merged data points are being used for which there is both a position report and a velocity report available.

5.2. Aligning flight segment pairs

The analysis in this research is focused on the interaction between two aircraft. Therefore, the basic input element is a pair of flight trajectories. This processing step regards combining two flight segments, as described in the previous step, together. Important is to ensure sufficient data quality and filtering on the characteristics of the respective pair of flight segments.

First, the database of flight segments is scanned on segments that overlap in the span of one hour as most of the flight segment are not longer than this timespan. Within this set of flight segments, all possible pairs of segments are checked if they overlap in time by comparing their start and end timestamps. It is ensured

that the timespan of the overlapping segment is equal or greater than the look-ahead time of 1200 seconds (20 minutes) considered in this research. All flight segment points falling outside this window are discarded. In a consecutive filtering step, filtering on altitude is added, leaving only the points with altitudes higher than FL300. This is done to ensure that all segments in further analysis hold mostly en-route flight behavior. After this filter, it is again ensured that the timespan of both segments exceeds the look-ahead time horizon. Otherwise, the set of flight segments is discarded. Finally, it is ensured that the starting points of both flights are not too far apart. Based on the current airspeeds, the flight time between the starting points must not exceed 2400 seconds. This will largely ensure that the flight segments are in approximately the same region of the airspace.

The following list summarizes the set of assumptions:

- Only data points with altitudes higher than **30.000 feet (FL300)** are considered.
- The timespan of two overlapping segments should be equal or greater than **1200 seconds**. Otherwise the pair of segments is discarded. This limit is maintained throughout further processing steps.
- Based on the current airspeeds, the flight time between the starting points of both flight segments must not exceed **2400 seconds**.

5.3. Scenarios for method validation

It is important to create a balanced set of scenarios for validating the performance of the different detection methods. This set of scenarios must both hold intrusions, as well as flight pairs with no significant events. If only scenarios with intrusions would be considered, detection methods can only be benchmarked based on their True Positive and False Negative ratios. However, especially in this context, the False Positive ratio of conflict detection methods is a very relevant metric as they provide insight in the false alarm ratio of a model when applied in practice.

Furthermore, the goal is to extract two different types of flight segment pairs. One type where both flight trajectories cross at some point, which might indicate an intrusion. A second type where the two flight paths are converging towards eachother. The first type makes the search for intrusions more efficient. The second type serves as a compound for a data set for conflict detection model validation.

5.3.1. Preprocessing conflict scenarios based on ADS-B flight segments

Historical flight paths from ADS-B messages include traffic avoidance maneuvers by default based on the current CD&R systems in place. Therefore, a method needs to be developed for filtering out these behaviors to create representative scenarios for the evaluation of different conflict detection methodologies. Important is that the prevalent uncertainty factors and their corresponding distributions are preserved while filtering CD&R related deviations and virtually create conflict scenarios. The goal is to maintain the characteristics of the original flight data while a sufficient number of conflicts needs to be obtained from the data. Real airborne conflicts are existing, but rare, which makes it difficult to obtain a sufficiently large set of conflicts for statistical analysis while avoiding to use a very large set of flight data.

The approach chosen to artificially increase the number of conflicts in the data is to combine aircraft trajectories from different flight levels. The altitude difference cannot be too large however, in order to avoid the influence of large differences in wind direction and speed. Also, the altitude influences the ratio between ground speed and air speed, which should not be significantly different in two compared trajectories. For analysis, only conflicts occuring at heights above 30.000 feet (FL300) are considered in order to limit the influence of behaviour in terminal approach areas. The problem to be solved is to compare each flight in the database with one another and find the times at which both trajectories intrude eachothers IPZ range. A naive approach to compare both flight trajectories is to calculate the distance between each set of points in both trajectories and find the points for which the distance are smaller than the IPZ radius. However, this solution has a time complexity of $O(n^2)$ which makes it computationally expensive to apply in this context.

In order to identify intrusions in a pair of flight segments, the aim is to find the intersection of a pair of trajectories. Both trajectories are not yet aligned in time as the goal is to find an efficient indication if two trajectories cross at some point to identify pairs of flight segments which might have had an intrusion. It is however ineffi-

cient to evaluate every pair of points from both segments and both fights are not yet aligned in time for which the search would entail an operation for each timestamp.

Therefore, the following algorithm is used to check for crossings in the pair of flight segments. A bounding box is defined for each segment stretching from the minimum latitude tot the maximum latitude, and from the minimum longitude to the maximum longitude as found in the coordinate pairs of the segment. The corner points of this box would be *(min(lat),min(lon)), (min(lat), max(lon)), (max(lat), min(lon)), (max(lat), max(lon))*. Then, it is checked if there is any overlap between the two bounding boxes. Because if the flight paths cross at some point, the bounding boxes have to overlap as well, around the point where the paths crossed. If there is an overlap, the trajectories are resampled within the overlapping area and the procedure is repeated. Based on this, the closest point of approach is found between the trajectories. This approach is significantly faster than the naive approach of comparing every point in the flight. However, the limiting factor is that only conflicts are found when trajectories intersect, which does not have to be the case. The approach is shown schematically in figure 5.2

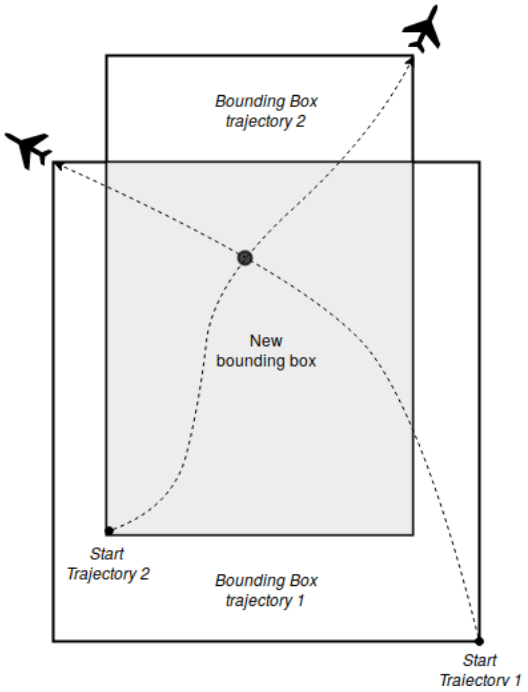


Figure 5.2: Conflict processing method using bounding box overlap

Now, only the distance between trajectories is considered for finding conflicts in the data. However, the time difference between the points of both trajectories needs to be included as well in order to find out if the intrusion actually took place. Therefore, postprocessing of all the conflicts is performed where the trajectories up to the closest point of approach are evaluated by calculating the distance between points matching on timestamp.

Besides the altitude difference being discarded in order to obtain a larger set of intrusions, a buffer in time difference is also used to enlarge the set of resulting intrusion scenarios. If an intrusion using the actual time stamps does not occur, but it does occur if the timestamps of the flight pair are shifted up to 60 seconds, the intrusion is artificially created by shifting the timestamps of one flight by the calculated time difference. By doing this, it is assumed that the scenario is still representative with the time shifting included. This step is performed in order to increase the otherwise very limited set of resulting intrusions.

5.3.2. Check if two flight paths are converging

The filter for converging flight paths is based on the relative approaching speed of the two aircraft at the start of the flight segments. The calculation is based on the initial distance between the initial coordinates of the flight segments, the airspeed of both aircraft and their respective heading. Equation 4.6 is applied for this where the

distance one second ahead is compared to the current distance. If the distance between the aircraft decreased, the flight paths are assumed to converge.

5.3.3. Further filtering of scenarios

After a classification is given to a pair of flight segments, the segments are aligned further. One of the problems is that both flight segments do not have the same amount of data points and that these points are also not aligned well in time. For further analysis it is required that both flight segments have an equal number of data points with the same timestamps. This requires merging both segments and aligning them within a certain tolerance. An inner join is made between both sets of trajectory points based on the nearest timestamp. The timestamps are allowed to have a difference with a maximum of 5 seconds. All data points that can not be matched are discarded during the merge.

It might occur that the flight segments hold gaps of several seconds, for instance because of missing ADS-B messages which were not received. After merging of the flight segments, gaps between consecutive timestamps are used to obtain the longest subsegment in the trajectory pair without any significant gaps. This filtering is based on the running mean of three consecutive time deltas over al timestamps in the trajectory pair. Based on this running mean, a value of 20 seconds is chosen as a splitting point. The subsegment with the highest number of data points between the splitting points is used for further processing. The rest of the data is discarded.

After filtering gaps within the timestamps, the timespan of the remaining segment is checked again if it still exceed the look-ahead time horizon of 1200 seconds. If not, the pair of flights is not used for further analysis.

The flight segment pairs belonging to one of both scenarios are accordingly saved to a database. Here, only the converging flight segments are saved if they do not hold any intrusions in them, otherwise they belong to the set of intrusion scenarios.

The following list of filters summarizes the above processing step:

- Only intrusions are considered in which two flight trajectories cross.
- Altitude is not taken into account for determining intrusion scenarios.
- On the time axis, if an intrusion using the actual time stamps does not occur, but it does occur if the timestamps of the flight pair are shifted up to 60 seconds, the intrusion is used.
- Resulting sets of flight trajectories are combined and matched based on their timestamps, using a tolerance of **five seconds**. All unmatched data points are discarded.
- If, using a running mean with a span of three data points, a gap of > **20 seconds** is identified, the segment before or after this gap which has the shortest time span is discarded.
- Only flight segment pairs that are converging towards eachother or which have intrusions in them are considered
- An ADS-B flight segment is assumed to always fully fit within the respective CTFM flight timespan.

5.4. Data set processing

The data sets to be used are historical data sets provided by the Faculty of Aerospace Engineering of the Delft University of Technology (TU Delft) and Eurocontrol. One data source contains historical ADS-B messages received by the ADS-B SIL antenna which is property of the TU Delft. The data holds raw ADS-B messages which will be decoded according to the structure as described earlier.[37] The second data set contains Enhanced Traffic Flow Management System (ETFMS) flight plan data and is made available by Eurocontrol. The data sets hold the last filed flight plans before departure of all aircraft flying in or moving through the airspace controlled by Eurocontrol.

For data processing use will be made of Python and as the main development language. The size of the available data is significant. There are between 12-16 million raw ADS-B messages received each day which require processing. Considering that there are at least several months of data available, this will require a scalable

data processing setup which includes parallel or distributed processing of datasets and corresponding scalable storage with sufficient throughput. Relying on data from only a select amount of days will induce bias into the results because of several contextual factors like weather, air traffic demand and special circumstances. Proof of concepts and prototypes for different research steps can be developed using smaller data samples and accordingly scaled up for generating final results.

There are 952.347 flight segments derived from the combined data set of ADS-B and flight plan data. The total amount of conflicts derived from this combined data set is 4196.

6

Experiment setup and assumptions

This research step aims at quantifying how different conflict detection methodologies compare by assessing their performance on historic ADS-B data. The simulation environment described in the previous step is needed to do this in a consistent way. This environment shall take care of the consistent processing of ADS-B data and feeding this to the model inside a well defined data pipeline to assure that each method is evaluated in an equal way. Each of the three methods will be applied to the same set of scenarios.

6.1. Evaluation of conflict detection performance

In order to make a valid comparison between the to be assessed methodologies, the evaluation metrics need to be well defined. For conflict detection, an intuitive approach is to look at the conflict detection accuracy based on the binary classification if a conflict is detected, or not, and if an intrusion actually occurred. This classification (C) is based on the differences between the estimated time to intrusion (TTI_{est}) and the actual time to intrusion (TTI_{act}). The conditions for a true positive detection (TP), false positive (FP), true negative (TN) and false negative (FN) classification are as follows:

$$\text{C} = \begin{cases} TP & \text{if } |TTI_{est} - TTI_{act}| < 20(s) \text{ and both } TTI_{est} \text{ and } TTI_{est} \text{ are defined} \\ FP & \text{if } TTI_{est} \text{ is } \textbf{defined} \text{ and } TTI_{act} \text{ is } \textbf{undefined} \\ TN & \text{if both } TTI_{est} \text{ and } TTI_{est} \text{ are } \textbf{undefined} \\ FN & \text{if } TTI_{est} \text{ is } \textbf{undefined} \text{ and } TTI_{act} \text{ is } \textbf{defined} \text{ or } |TTI_{est} - TTI_{act}| \geq 20(s) \end{cases}$$

The summarizing metric for model performance will be the F-measure, which is the weighted harmonic mean of the precision and recall of the detection model as defined in equation 6.1.

$$F_w = (1 + w^2) \cdot \frac{P \cdot R}{w^2 \cdot P + R} \tag{6.1}$$

Where P and R are the precision and recall of the model performance respectively, defined as:

$$P = \frac{TP}{TP + FP} \tag{6.2}$$

$$R = \frac{TP}{TP + FN} \tag{6.3}$$

in the F_w metric.

Recall refers to the amount of correctly classified intrusions compared to the total number of intrusions. Therefore, a perfect recall score equals one, where there would be no false negatives and all intrusions would be correctly classified. Recall thereby provides a measure for how many relevant instances the model is able to select from the total number of relevant instances.

Precision refers to the ratio between true positive classifications over all positive classifications. This relates to the ratio between correctly classified intrusions over all positive classifications. A model which would always output a positive classification would have a low precision score, even though it correctly identified all true positives but also results in a large number of false positives.

The parameter w is used to adjust the score to the respective importance of precision and recall in the evaluation. Higher values of w place more emphasis on the occurence of false negatives compared to false positives. In the context of conflict detection, the ratio of false negatives occurring should be as low as possible, as it implies the occurence of dangerous situations where a model did not detect a conflict while there is a conflict present in the data.

In the evaluation of the methods the value of w is therefore chosen to be $w = 2$ giving the final score as shown in equation 6.4.

$$F_2 = 5 \cdot \frac{P \cdot R}{4 \cdot P + R} = \frac{5 \cdot TP}{5 \cdot TP + 4 \cdot FN + FP} \tag{6.4}$$

True negatives are not taken into account in the F-measure, but they are also less relevant in the evaluation of a conflict detection model.

Furthermore, the positive and negative classes in the evaluation dataset are likely to be imbalanced. In general, there will be a much larger number of data points without an intrusion than data points with an intrusion. In order to overcome this, the evaluation dataset will be constructed from an equal number of flight segment combinations which have an intrusion in them as the number of segment combinations without an intrusion. Thereby balancing the positive and negative classes.

Another measure for model performance is the accuracy A, which is the ratio of correctly classified instances and the total number of instances. The measure for model accuracy is given in equation 6.5. This measure is valid if the positive and negative classes are approximately balanced.

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{6.5}$$

The main interest in this research are the changes in performance of all three models compared to the considered look-ahead time. Therefore, data points from the validation set of flight segment pairs are divided into bins of 20 seconds based on their look-ahead time. For these bins, the F_2 scores calculated for all the evaluated data points in the bin is taken as the performance measure for that look-ahead time for the respective model. This enables the comparison of the models for different look-ahead times.

The following list of filters summarizes the above processing steps:

- False Negatives are emphasized more in the evaluation F-measure than False Positives
- If the absolute time difference between the actual intrusion and the expected intrusion is greater than **20 seconds**, the detection is assumed to be a false negative.

6.2. Dependent and independent measures of the experiments

This section shows a summary of the dependent and independent measures used in the performed experiments.

The independent variables in the experiment are the following:

- **Type of state propagation method** The type of state propagation method; state-deterministic, state-probabilistic or intent-deterministic is a variable which is changed throughout the analysis. The outcomes of these three methods are compared with eachother.
- **Look-ahead time** The look-ahead time determines the time horizon how far into the future the trajectories will be projected. In this research, the maximum is set at 1200 seconds, or 20 minutes.

- **Altitude (flight level)** Only intrusion scenarios above FL300 have initially been selected to assure that mostly en-route flight movements are considered. In order to assess the remaining effect of altitude the scenarios are divided into two categories. The first category is above flight level 360. The second category is between FL300 and FL360. For both categories, the method performances will be calculated.
- **Heading difference** For the intrusion scenarios, the angle between the respective flight paths is chosen as a filter as well as this likely affects the conflict detection performance. For this variable the effects are assessed as well. The boundary between shallow angle conflicts and other conflicts is set at 30 degrees, creating two categories of conflict scenarios. An angle of 30 degrees is chosen to create a category with scenarios having especially shallow angles between the flight paths.
- **Time boundaries for false detections** For the conflict detection methodologies, the detection boundaries as described in 6.1 are used. This implies that when the absolute difference between the estimated and actual time to conflict is smaller than 20 seconds, the conflict is successfully detected. When the time difference falls outside this range, there is either a false negative or a false positive detection.

The resulting dependent measures are:

- **Detection Classification** The main resulting metric is a detection classification into true positives, false positives, true negatives and false negatives.
- **Classification Accuracy scores** From the detection classification counts, accuracy scores are used as output in order to compare the different methodologies. These are the F_2-Measure, Accuracy, precision and recall.

6.3. Overview of Assumptions

The following assumptions are being made throughout the analysis:

- **Constant airspeed for en-route flights** It is assumed for the projections that the airspeed remains constant from the point of projection onwards. Based on this assumption, only flight segments from en-route flights are used for the analysis.
- **Constant airspeed/groundspeed ratio for different altitudes** In the ADS-B data, the measured airspeed is used as the indicated velocity of the aircraft. However, the actual ground speed of the aircraft changes for equal air speeds at different altitudes. This introduces inaccuracies for extrapolated aircraft locations when trajectories from different altitudes are being used.

Considering the different processing steps the following assumptions and filters are used:

Decoding and merging of raw ADS-B messages and extraction of flight segments from ADS-B data points

- Only messages with a downlink format id of 17 or 18 (ADS-B or TIS-B) are considered.
- If the number of seconds between consecutive messages for the same ICAO address is greater than **600 seconds**, the messages are considered to belong to different flights.
- A pair of odd/even position report messages is only considered if they are received less than **5 seconds** apart.
- For merging the three types of messages (position, velocity and callsign) on timestamp, a tolerance of **2 seconds** is applied.
- Only the merged data points are being used for which there is both a position report and a velocity report available.

Alignment of flight segment pairs

- Only data points with altitudes higher than **30.000 feet (FL300)** are considered.
- The timespan of two overlapping segments should be equal or greater than **1200 seconds**. Otherwise the pair of segments is discarded. This limit is maintained throughout further processing steps.

- Based on the current airspeeds, the flight time between the starting points of both flight segments must not exceed **2400 seconds**.

Creating scenarios for method validation

- Only intrusions are considered in which two flight trajectories cross.
- Altitude is not taken into account for determining intrusion scenarios.
- On the time axis, if an intrusion using the actual time stamps does not occur, but it does occur if the timestamps of the flight pair are shifted up to 60 seconds, the intrusion is used.
- Resulting sets of flight trajectories are combined and matched based on their timestamps, using a tolerance of **5 seconds**. All unmatched data points are discarded.
- If, using a running mean with a span of 3 data points, a gap of > **20 seconds** is identified, the segment before or after this gap which has the shortest time span is discarded.
- Only flight segment pairs that are converging towards eachother or which have intrusions in them are considered.
- An ADS-B flight segment is assumed to always fully fit within the respective CTFM flight timespan.

Evaluation of conflict detection performance

- False Negatives are emphasized more in the evaluation F-measure than False Positives
- If the absolute time difference between the actual intrusion and the expected intrusion is greater than **20 seconds**, the detection is assumed to be a false negative.

7

Results

This chapter covers results obtained from the experiments as outlined in chapter 6. Three different sets of results are obtained and discussed: First the statistical properties of the distributions of altitude, heading and airspeed in the ADS-B data set are examined in order to validate the assumptions made with respect to data filtering and data properties. Second, the results from state-deterministic projection applied to a set of 10.000 flight segments derived from ADS-B data are evaluated. This will provide a benchmark for the error distribution when applying a basic projection method on a set of flight segments without regarding any conflict scenarios. Third, the results from the state-deterministic, state-probabilistic and intent-deterministic conflict detection methods are discussed when applied to a set of historic conflict scenarios derived from ADS-B data and flight plan data.

7.1. Statistical properties of altitude, heading and airspeed distribution in historic ADS-B flight segments

Initial statistical properties have been calculated from the set of historic ADS-B flight segments. The distributions of airspeed, heading and altitude deviations are compared for different look-ahead horizons to evaluate how these distributions change through time. A comparison is made between the distribution from flight segments above 30.000 feet and those below 30.000 feet. This way an evaluation is made of the choice for all experiments in this research to incorporate only flights above flight level 300. The aim is to only use en-route flights and filtering the effect of manouvering in the TMA area and during climb and descent procedures on the experiment outcomes.

Derived from the data of 10.000 different flight segments, figures 7.1, 7.4 and 7.8 show the empirical probability density distributions of heading, true airspeed and altitude for different ranges of look-ahead times. The distributions are generated based on the deviations between an initial aircraft state and the state for the same aircraft after a respective number of seconds. This is shown for four different ranges of look-ahead times; 0-300 seconds, 300-600 seconds, 600-900 seconds and 900-1200 seconds.

It can be observed that for all considered variables, the respective distributions remain approximately constant, independent of the look-ahead time. This stability of the different distributions with respect to time is important as it rules out that changes in the underlying data distributions affect the validation of the conflict detection methods. If distributions would contain asymmetries or artifacts this will reflect in the conflict detection results but could not be contributed to the models itself.

Tables 7.1, 7.5 and 7.2 provide the mean, standard deviation, skewness and kurtosis of all distributions.

The effect of filtering the flight segments above 30.000 feet can be clearly visualised as well. The distributions of the same variables from flight segments below this altitude are shown in figures 7.2, 7.6 and 7.9. Tables 7.3, 7.7 and 7.10 show the mean, standard deviation, skewness and kurtosis of the distributions calculated from the flight segments below 30.000 feet. From this it can be observed that there is a significant influence from maneuvers related to descend and climb operations. Therefore, limiting to incorporating flight segments

above 30.000 feet proves to have a positive effect to the stability of the distributions with respect to the look-ahead time.

7.1.1. Distribution of heading changes

Regarding the distributions of heading changes there is a clear difference visible in especially the variance of the samples taken from high and lower altitudes. For samples above FL300, the standard deviation grows from 25.4 degrees to 57.1 degrees at larger look-ahead times. This would be expected behavior as the likelihood of trajectory changes increases with look-ahead time. The mean of all distributions is close to zero but is likely affected by outliers in the sample set especially for look-ahead times above 600 seconds. Looking at the kurtosis of the distribution it can be seen that all distributions show 'fat tail' behavior, indicating a high number of outliers. From this it can be concluded that most of the flights have consistent flight paths but a small subset is showing large deviations. These outliers could likely be responsible for increasing the standard deviation values. From the moderate values for skewness it can be concluded that the distributions are symmetrical.

On the right side of the graph it is visible that some flights include heading changes of about 180 degrees. This can be explained as flights which are descending and make these manouvers at lower altitudes. This artifact is only present in look ahead times above 300 seconds. At these times it is reasonable that aircraft have descended a significant amount to a flight level where these 180-degree turns can be expected.

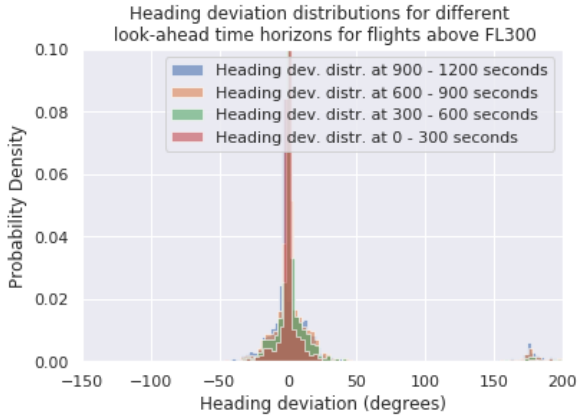


Figure 7.1: Distributions of the heading deviations for different look-ahead times for flight segments above 30.000 feet

	Look-ahead time			
Metric	**0-300 seconds**	**300-600 seconds**	**600-900 seconds**	**900-1200 seconds**
Mean	0.8	4.6	8.0	9.2
Standard Deviation	25.4	36.1	47.2	57.1
Kurtosis	61.2	21.2	11.7	7.2
Skewness	0.8	2.5	1.6	0.9

Table 7.1: Overview of distribution metrics of the heading deviations for different look-ahead times for flight segments above 30.000 feet

The distribution of samples filtered on lower altitude show much higher variance which implies a larger rate of trajectory changes in the sample set. This would be expected as at lower altitudes more manouvering takes place in general as most of the aircraft are in climb or descend operations. Interesting to see is that the kurtosis of the distributions is lower compared to the higher altitude sample set. This indicates that the distribution characteristics are less driven by outliers. Therefore it can be concluded that there is more homogeneity in the sample set regarding changes in flight paths. A larger part of the flights is showing changes in their trajectories compared to the high altitude set. Low levels of skewness indicate more symmetry in the distributions than observed in the high altitude sample set.

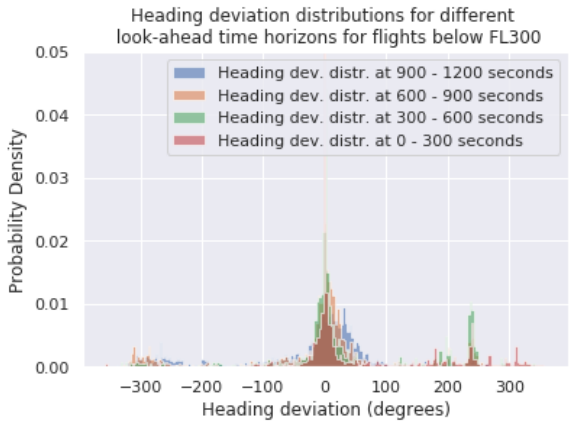


Figure 7.2: Distributions of the heading deviations for different look-ahead times for flight segments below 30.000 feet

Metric	Look-ahead time			
	0-300 seconds	**300-600 seconds**	**600-900 seconds**	**900-1200 seconds**
Mean	43.1	42.2	1.2	-21.2
Standard Deviation	109.9	127.1	134.8	129.0
Kurtosis	1.4	0.6	0.8	0.4
Skewness	0.9	0.0	-0.3	-0.5

Figure 7.3: Overview of distribution metrics of the heading deviations for different look-ahead times for flight segments below 30.000 feet

7.1.2. Distribution of true airspeed changes

In the high altitude sample set the variance in airspeed changes is low. This implies that the assumption for constant airspeed in further analysis seems correct. Moderate values of kurtosis and skewness in the distributions indicate that the distributions do not have many outliers and are fairly symmetrical.

An explanation for the variance in the TAS values could for a part be ascribed to changes in altitude due to which the true airspeed values change. The Mach number then stays constant.

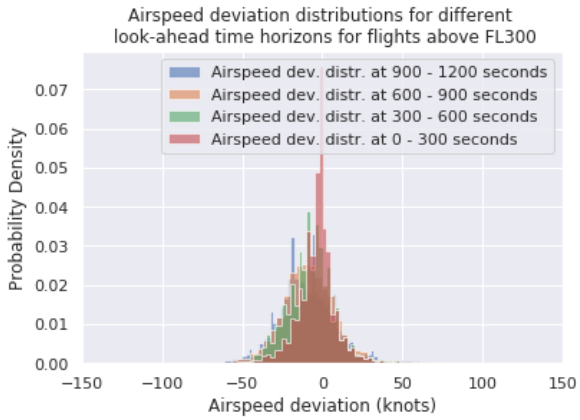


Figure 7.4: Distributions of the true airspeed deviations for different look-ahead times for flight segments above 30.000 feet

Metric	Look-ahead time			
	0-300 seconds	**300-600 seconds**	**600-900 seconds**	**900-1200 seconds**
Mean	-3.5	-7.1	-8.1	-9.2
Standard Deviation	12.3	16.1	18.0	20.4
Kurtosis	7.7	4.8	4.2	3.3
Skewness	0.5	1.0	0.9	0.9

Figure 7.5: Overview of distribution metrics of the airspeed deviations for different look-ahead times for flight segments above 30.000 feet

In the airspeed distributions from the sample set from lower altitudes, bimodal distributions emerge which explains high standard deviation values. The bimodal distributions likely arise from the fact that most flights in this sample set are in climb or descend operations. This transition behavior causes the airspeed to change largely. The skewness values indicate that the distributions are however fairly symmetrical and therefore include a balanced number of climbing and descending flights. Filtering this behavior from the sample set for further analytics will largely improve the stability of the validation of the conflict detection methods as mainly the assumption for constant airspeed is violated for this particular subset of samples.

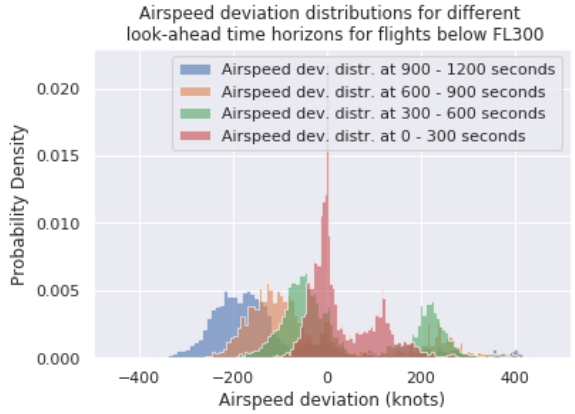


Figure 7.6: Distributions of the airspeed deviations for different look-ahead times for flight segments below 30.000 feet

Metric	Look-ahead time			
	0-300 seconds	**300-600 seconds**	**600-900 seconds**	**900-1200 seconds**
Mean	31.0	41.8	-22.8	-102.7
Standard Deviation	71.7	135.3	159.7	170.8
Kurtosis	0.2	-1.2	-0.2	1.4
Skewness	0.9	0.5	1.0	1.5

Figure 7.7: Overview of distribution metrics of the airspeed deviations for different look-ahead times for flight segments below 30.000 feet

7.1.3. Distribution of altitude changes

Variability in altitude mainly indicates the presence of climb and descend behavior in the data sets. In the sample set fitered above flight level 300 still some variance in altitude can be observed, even though it is much less than in the set of flight segments below FL300. However, most altitude changes are close to zero. For larger look-ahead times more stepwise changes can be observed.

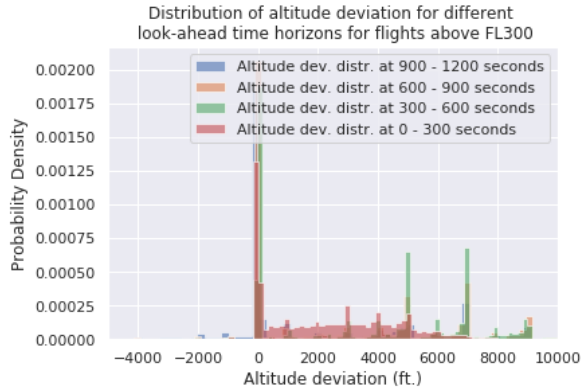


Figure 7.8: Distributions of the altitude deviations for different look-ahead times for flight segments above 30.000 feet

Metric	Look-ahead time			
	0-300 seconds	**300-600 seconds**	**600-900 seconds**	**900-1200 seconds**
Mean	2342.2	3467.6	2840.3	1908.3
Standard Deviation	2152.0	3274.1	3356.5	3263.8
Kurtosis	-0.7	-1.4	-1.1	-0.1
Skewness	0.5	0.2	0.6	0.7

Table 7.2: Overview of distribution metrics of the altitude deviations for different look-ahead times for flight segments above 30.000 feet

In the altitude distributions for lower flight levels a bimodal distribution emerges as well for likely the same reasons as provided at the airspeed distributions as these are related. In the altitude and airspeed similar levels of skewness can be observed indicating symmetry.

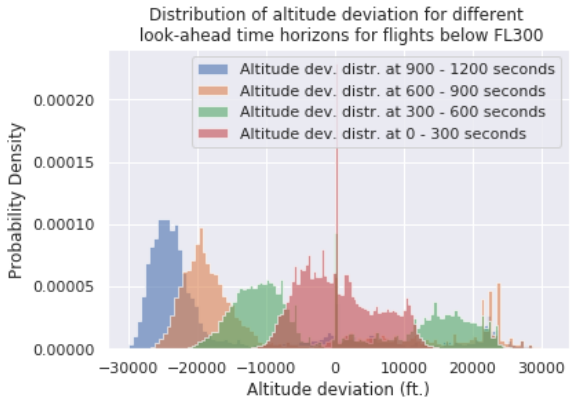


Figure 7.9: Distributions of the altitude deviations for different look-ahead times for flight segments below 30.000 feet

Metric	Look-ahead time			
	0-300 seconds	**300-600 seconds**	**600-900 seconds**	**900-1200 seconds**
Mean	462.9	-421.6	-7647.2	-14592.2
Standard Deviation	5585.2	13193.6	16348.0	16017.3
Kurtosis	-0.4	-1.4	-0.7	0.4
Skewness	0.4	0.4	0.9	1.3

Figure 7.10: Overview of distribution metrics of the altitude deviations for different look-ahead times for flight segments below 30.000 feet

As mostly airspeed and heading affect the quality of state propagation, it can be assumed that the filtered data set is sufficient to compare the state propagation performance with respect to the look-ahead time.

7.2. Analysis of state-deterministic flight projection

This result section will discuss the error distributions derived from state-deterministic projection applied to a set of 16.000 flight segments. The error distributions will serve as a benchmark for the three conflict detection approaches. Furthermore it will provide an empirical upper bound for the look-ahead time above which the standard deviation of the error distribution exceeds the size of the intruder protected zone. This look-ahead time boundary will accordingly be compared to the results from the performance of the three conflict detection approaches of which the results will be discussed in 7.3.

A randomized set of 16.000 flights is used derived from 43 days of ADS-B data in the months of May, September, October, November and December of 2018. As concluded from the previous section, only flight trajectories above 30.000 feet are considered here for the reason of decreasing the effect of TMA manoevering on the results, unless specified otherwise.

As error metrics, the cross-track error (CTE), along-track error (ATE) and total-track error (TTE) are used as described in section 4.2.

The propagation of cross-track errors, along-track errors and total-track errors through time is shown in figures 7.11, 7.14 and 7.16. These plots show the box plots for errors within bin ranges of 20 seconds for look-ahead times of 300, 600, 900 and 1200 seconds for respective flight segments above FL300 and below FL300. The box range is set at the 25% and 75% quantiles of the distribution. The whiskers of the boxplots extend up to 5% and 95% respectively. The outliers outside of the range of the boxplots are not shown in the graphs as these are very frequent and in a much larger range than the distributions shown in the boxplots. Furthermore, the size of the IPZ area is indicated for the TTE results by the dashed line in order to visualize the evaluation of the error distribution compared to the conflict indentification range.

In order to visualise the effect of the flight level on the error levels, the distributions of all metrics are shown for flights above and below the 30.000 feet level. For all distributions, the mean, standard deviation, kurtosis and skweness are provided in tables 7.3, 7.5 and 7.6. Figures 7.12a, 7.15a and 7.17a show the change of the distribution characteristics for the CTE, ATE and TTE samples through time in more detail.

7.2.1. Distribution of Cross-Track Errors (CTE)

In the distribution of cross-track errors for different look-ahead times, symmetric distributions can be observed where the deviation continuously increases.

The data distributions for the CTE appear not visibly skewed and the mean of the CTE error distribution is very close to zero. The low skewness values indicate symmetry in the distributions. The kurtosis of the distributions

is decreasing for larger look-ahead times which indicates that the distribution becomes flatter. This behavior is also observed in the distributions of the heading differences in 7.1 and might be a direct result from this as well. Compared with the distributions obtained from lower flight levels the kurtosis of the distributions are greater for higher flight levels, similar to the results observed before.

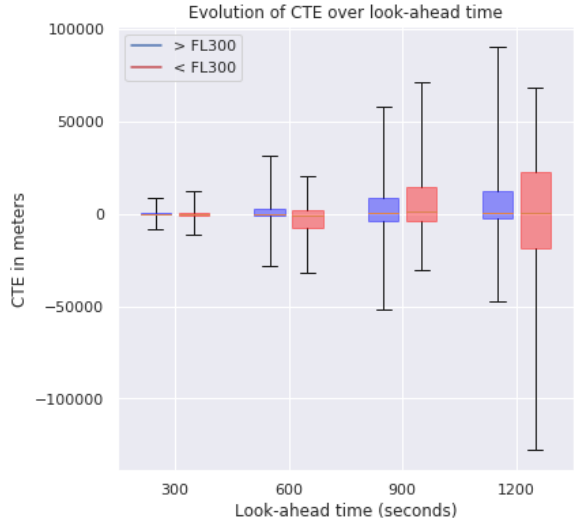


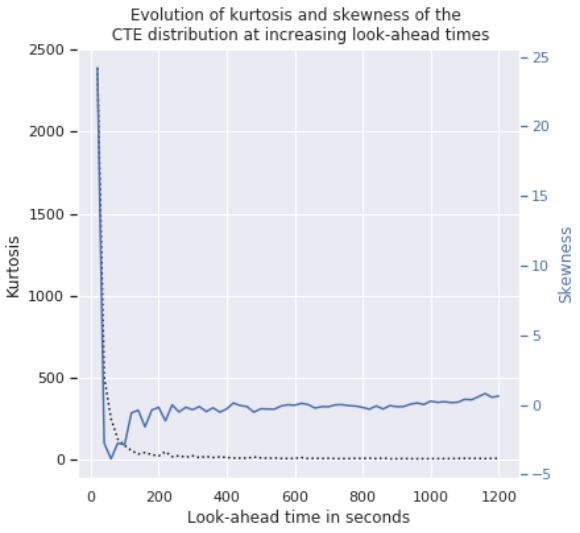
Figure 7.11: Evolution of the CTE for increasing look-ahead times as box-plots, divided in 20-second bins around different look-ahead times separated by flight level

	Look-ahead time			
	Above FL300			
Metric	**300 seconds**	**600 seconds**	**900 seconds**	**1200 seconds**
Mean	-25.7	496.5	1623.6	7192.4
Median	-27.2	3.5	172.0	324.1
Mean Absolute Deviation	2520.6	9662.9	18878.7	25965.6
Median Absolute Deviation	228.8	1898.9	6298.8	6278.0
Kurtosis	24.6	7.7	6.2	8.3
Skewness	-0.3	-0.0	-0.1	0.6
	Below FL300			
Mean	330.9	-2082.0	7427.5	-1685.7
Median	0.6	-780.7	995.9	362.1
Mean Absolute Deviation	3799.6	10738.7	18890.4	40454.2
Median Absolute Deviation	854.4	5287.9	8382.2	20862.2
Kurtosis	7.3	8.0	11.2	4.1
Skewness	0.1	1.0	2.1	0.4

Table 7.3: Overview of distribution metrics of the CTE for different look-ahead times for flight segments above and below 30.000 feet

The influence of outliers on the distributions becomes clear from the difference between the Median Absolute Deviation and the Mean Absolute Deviation boundaries. It can also be observed directly from the mean and median as shown in table 7.3. The median metrics are more robust to outliers. One would expect that in a distribution where no significant outliers are present, the difference between the mean and median is very small. The same effect is visible for the ATE and TTE distributions.

(a) Mean, median, absolute deviations and 99th Percentile



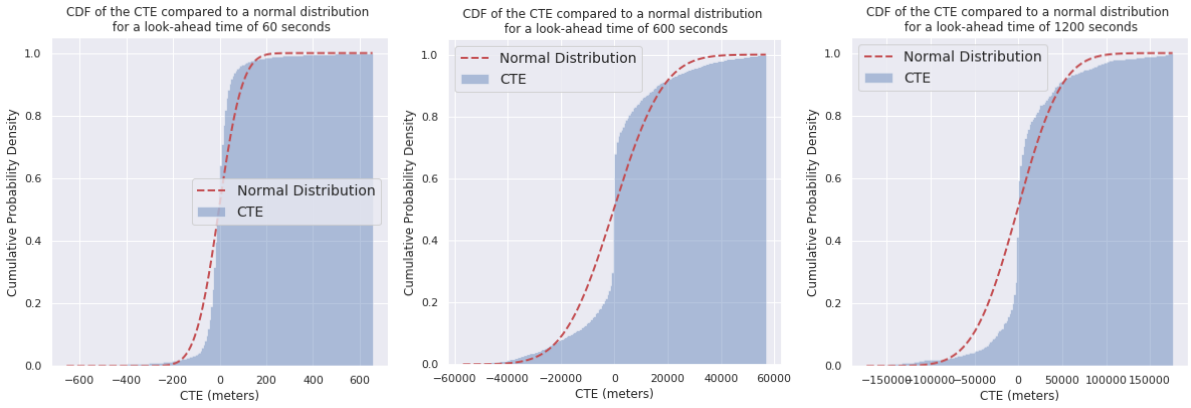
(b) Kurtosis and Skewness

Figure 7.12: Evolution of different metrics for the CTE distribution for increasing look-ahead times

One important aspect to verify is the normality of any of the distributions for especially the CTE metric. Often, normality is assumed as it simplifies further analysis as is reviewed in chapter 2. However, it is worhtwhile to

check if these assumptions are valid based on the empirical distributions obtained here. Especially as a non-parametric estimator will be used as a basis for the state-probabilistic trajectory prediction method, where the distributions are not assumed to be normal.

The comparison of normal distributions and the obtained distributions for the CTE is shown in figures 7.13a, 7.13b and 7.13c. Here, the empirical cumulative distribution functions are set out against a cdf of a normal distribution. These plots show clearly that throughout the look-ahead time horizon, the distributions are significantly fat-tailed. This is observable from the flatter curve in the center of the graph compared to the normal distributions indicated by the red dashed line. However, as observed in table 7.3, the fat-tailedness is declining slightly for larger look-ahead times as the mean absolute deviation of the entire distribution increases accordingly.



(a) Look-ahead time: 60 seconds

(b) Look-ahead time: 600 seconds

(c) Look-ahead time: 1200 seconds

Figure 7.13: Cumulative distribution function of the CTE set out against a normal distribution at different look-ahead times

As a further validation on normality, the Kolmogorov-Smirnov test shows a significant rejection of the null-hypothesis that the empirical distributions are normally distributed. The results of the test are shown in table 7.4. This can be visually verified using the figures mentioned before. This conclusion is in line with the assumptions made for further analysis as a non-parametric distribution estimation is used in the state-probabilistic projection method.

Look-ahead time	**KS Statistic**	**p-value**
20 seconds	0.3257	0.0
300 seconds	0.5050	0.0
600 seconds	0.5232	0.0
1200 seconds	0.5168	0.0

Table 7.4: Results of a Kolmogorov-Smirnov test on normality for the empirical distributions of the CTE for different look-ahead times

7.2.2. Distribution of Along-Track Errors (ATE)

Regarding the along-track errors, the distributions become visibly skewed into the negative side at larger look-ahead times. This means that the propagated trajectories tend to lag behind the original trajectories increasingly. This is especially expressed at the distributions belonging to high flight levels. The kurtosis of the distributions for high flight levels is also very large compared to these obtained from the low flight level sample sets. This might indicate that the skewness is driven by large outliers instead of a recurring pattern. However, for both sample sets, the mean of all distributions shifts down continuously for higher look-ahead times.

The mean absolute deviation values show opposite behaviour from the cross-track error distributions as the higher flight level sample sets show larger deviations in this case.

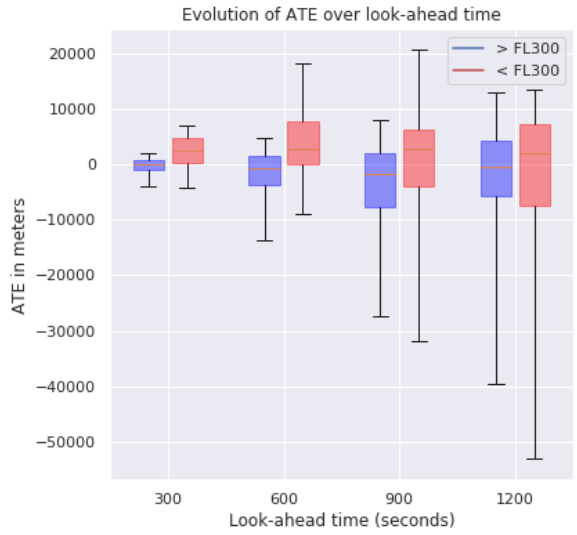
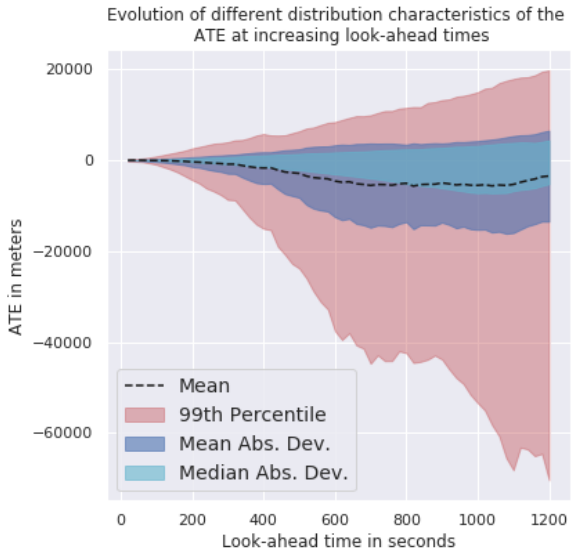


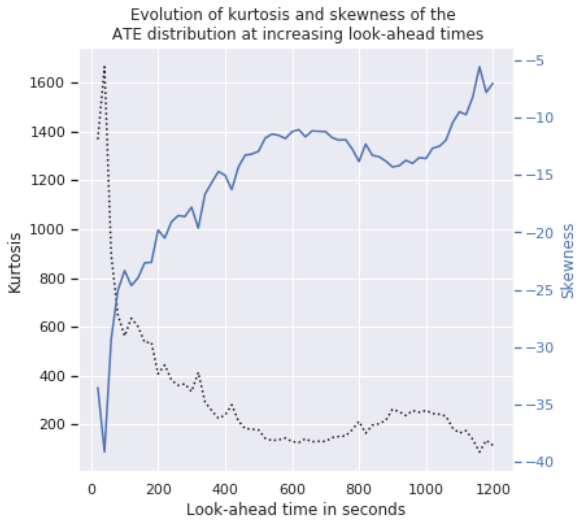
Figure 7.14: Evolution of the ATE for increasing look-ahead times as box-plots, divided in 20-second bins around different look-ahead times separated by flight level

Metric	Look-ahead time			
	Above FL300			
	300 seconds	**600 seconds**	**900 seconds**	**1200 seconds**
Mean	-872.3	-4580.1	-11056.8	-16673.0
Median	-56.3	-657.3	-1760.8	-598.1
Mean Absolute Deviation	2053.6	7898.4	17774.8	31144.5
Median Absolute Deviation	769.8	2396.0	4498.9	4934.9
Kurtosis	335.1	130.4	74.4	53.4
Skewness	-17.8	-11.2	-8.5	-7.3
	Below FL300			
Mean	2154.4	3553.9	-8984.2	-15608.1
Median	2551.8	2745.5	2713.6	1965.7
Mean Absolute Deviation	2881.3	5599.3	24577.6	31430.8
Median Absolute Deviation	2334.1	3631.0	4720.3	6599.6
Kurtosis	5.9	12.4	34.1	45.0
Skewness	-1.2	-1.9	-5.8	-6.6

Table 7.5: Overview of distribution metrics of the ATE for different look-ahead times for flight segments above and below 30.000 feet



(a) Mean, median, absolute deviations and 99th Percentile



(b) Kurtosis and Skewness

Figure 7.15: Evolution of different metrics for the ATE distribution for increasing look-ahead times

7.2.3. Distribution of Total Track Errors TTE and estimation of the look-ahead time boundary

The TTE distributions show a one-sided distribution starting at zero as the TTE metric is only defined for positive values. The effect of the flight level on the distributions is also observable for the TTE distributions where the samples from lower flight levels show larger deviation scores. A likely explanation is higher vectoring activity from air traffic control which affects the distributions as also concluded from previous results in section 7.1. Also climb and descend behavior affects the results as the extrapolations are based on the true airspeed (TAS). The TAS however changes during altitude changes which affects the track error distributions.

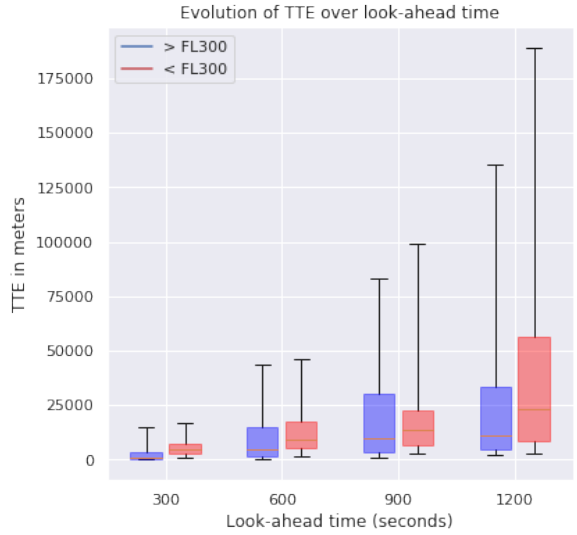
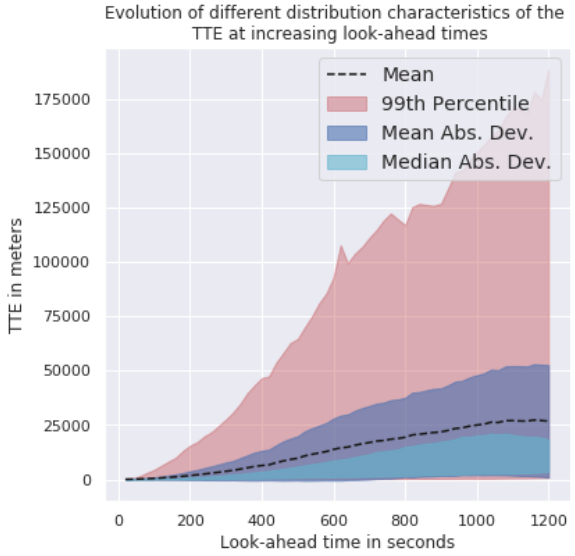


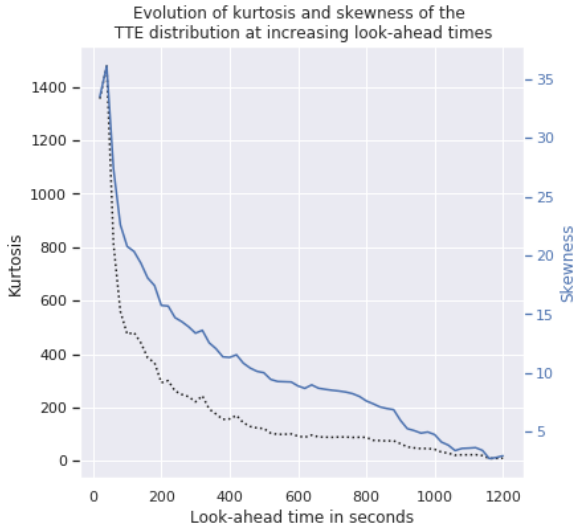
Figure 7.16: Evolution of the TTE for increasing look-ahead times as box-plots, divided in 20-second bins around different look-ahead times separated by flight level

Metric	Look-ahead time			
	Above FL300			
	300 seconds	**600 seconds**	**900 seconds**	**1200 seconds**
Mean	3725.4	13778.0	27906.0	40008.1
Median	1143.0	4830.3	10070.1	11216.5
Mean Absolute Deviation	4099.7	14267.4	28492.7	45207.7
Median Absolute Deviation	834.1	3828.1	8017.7	7809.0
Kurtosis	221.3	93.0	56.3	41.5
Skewness	13.4	8.9	7.1	6.2
	Below FL300			
Mean	6059.4	14089.2	31560.7	53627.0
Median	4964.8	9272.9	13536.9	23042.4
Mean Absolute Deviation	3675.0	9954.5	32622.9	51811.3
Median Absolute Deviation	2212.7	5818.3	7361.6	16200.5
Kurtosis	9.1	12.8	25.9	24.8
Skewness	2.5	3.0	5.0	4.6

Table 7.6: Overview of distribution metrics of the TTE for different look-ahead times for flight segments above and below 30.000 feet



(a) Mean, median, absolute deviations and 99th Percentile



(b) Kurtosis and Skewness

Figure 7.17: Evolution of different metrics for the ATE distribution for increasing look-ahead times

Figure 7.18 shows the evolution of the total track error for look-ahead times up to 840 seconds. The red dashed line in the graph indicates the IPZ radius. The total track error provides a better view on the total amount of error and uncertainty in the state-deterministic projection.

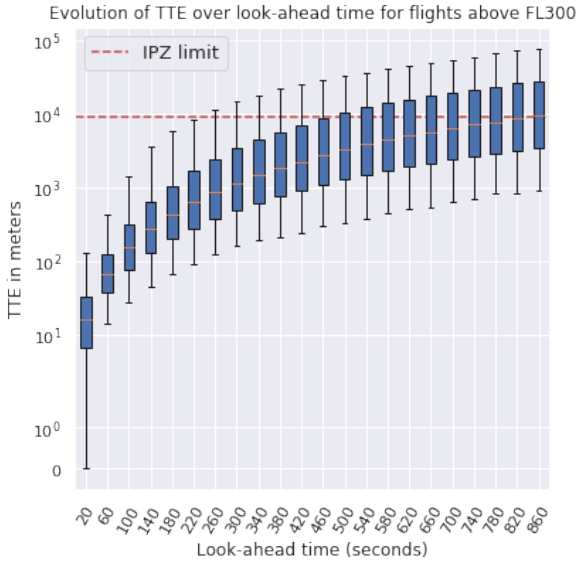


Figure 7.18: Evolution of the TTE, divided in 20-second bins around different look-ahead times

Percentile	Empirical approximate look-ahead time limit (in seconds)
70th	540 (9 minutes)
90th	320 (5 minutes, 20 seconds)
99th	160 (2 minutes, 30 sec.)

Figure 7.19: Empirical limits on the look-ahead time for different percentiles of the distribution based on the total track error (TTE)

Based on the TTE data, which is a combination of ATE and CTE, we can make the following conclusions about when the uncertainty in the distribution exceeds the range of the IPZ, which provides an empirical limit on the look-ahead time for conflict detection based on state-deterministic projection. The respective limits are provided in table 7.19.

This data indicates that when using a state-deterministic projection, in a two dimensional scenario, the look-ahead time horizon is limited to 9 minutes for maintaining reasonable amounts of certainty that the projected aircraft locations does not deviate more than one IPZ radius from the eventual actual location of the aircraft, which could be the difference of detecting a conflict or not.

7.3. Comparison of state-deterministic, state-probabilistic and intent-deterministic conflict detection

This section reviews the results of the comparison between state-deterministic, state-probabilistic and intent-deterministic conflict detection approaches on a preprocessed set of conflict scenarios as described in chapter 5.

7.3.1. Validation data set preparation

For the validation of the detection methods it is important that the data set used for this is well balanced. This implies that the number of positive instances, data points where a conflict is present, should be approximately equal to the number of negative instances, where no conflict is present. If the validation data set is unbalanced in this respect, metrics like model accuracy become unreliable. The F2-measure however is robust for these unbalances as it takes the positive versus negative class ratio into account implicitly.

In order to create a validation data set with well balanced positive and negative classes, a mixture of two types of scenarios are used; One where an intrusion is present and one where no intrusion is present. Using an iterative approach, scenarios from both types are combined into one data set which results in an approximately constant ratio of positive and negative samples of around one. The flight segments itself are not changed in any way in order to achieve this.

7.3.2. Analysis results

A comparison of state-deterministic, state-probabilistic and intent-deterministic conflict detection based on the F2-measure for different look-ahead times is provided in figure 7.20. A total set of 4000 conflict scenarios augmented with 300 scenarios without conflicts is used to create a comparison between the three detection methods.

Figure 7.21 shows the number of positive and negative instances for different look-ahead times. The balance between positive and negative classes is quite constant for the considered look-ahead times with an average ratio of around 1.2 where the number of negative instances is slightly larger than the number of positive ones.

Figure 7.20: F-measure scores from state-deterministic, state-probabilistic and intent-deterministic conflict detection for different look-ahead times

Figure 7.21: Occurrences of the positive and negative classes for different look-ahead times

It can be observed that the F2-measure is decreasing continuously for larger look-ahead times and that the performance of all three methods is fairly similar.

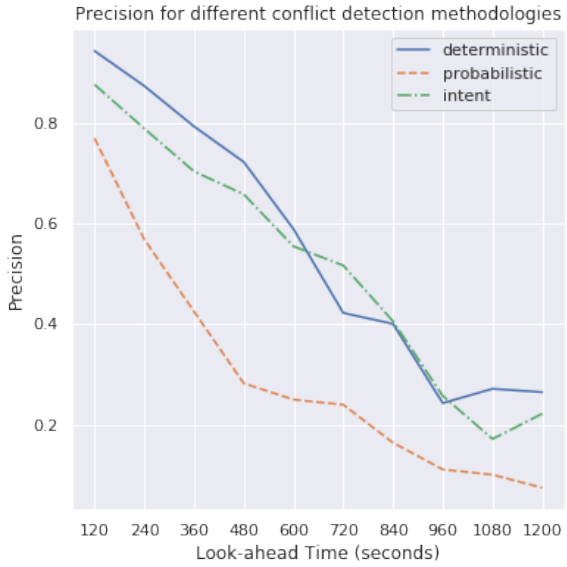
The precision and recall of all three detection methods used to calculate the F2-measure scores are shown in figures 7.22a and 7.22b respectively. The accuracy of the methods is provided in figure 7.22c.

A model which randomly selects between a detection or no detection would have an accuracy of 0.5. It can be seen that the state-probabilistic approach drops below this accuracy level around a look ahead time of 400 seconds. Followed by state-deterministic detection around 700 seconds. The intent-deterministic detection method drops below an 0.5 accuracy level around 840 seconds. This implies that beyond these look-ahead times the models respectuvely loose any predictive power for conflict detection. This corresponds roughly with the look-ahead time limits in table 7.19 as derived from the track-error distributions.

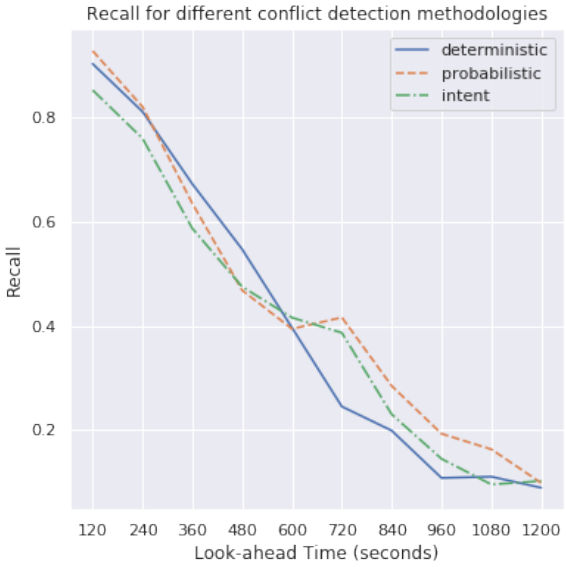
The state-deterministic approach shows the best performance up to approximately 600 seconds after which the state-probabilistic and intent-deterministic approaches perform slightly better. In general the results are in line with the error distributions shown in the previous section. From these it can be expected that no conflict detection method yields reliable performance for look-ahead times above 540 seconds. Except that the intent-deterministic approach shows higher than random accuracies up to 840 seconds ahead.

Recall scores are similar for all three methods. For higher look-ahead times the state-probabilistic approach has a slightly higher recall. However, from the precision metric it can be observed that the state-probabilistic method constantly has a significantly lower precision. As this method can be seen as a deterministic approach with an additional uncertainty margin it would be expected that a larger number of false positives reduces the precision metric. This has a lower number of false negatives as a result as can be seen in figure 7.23.

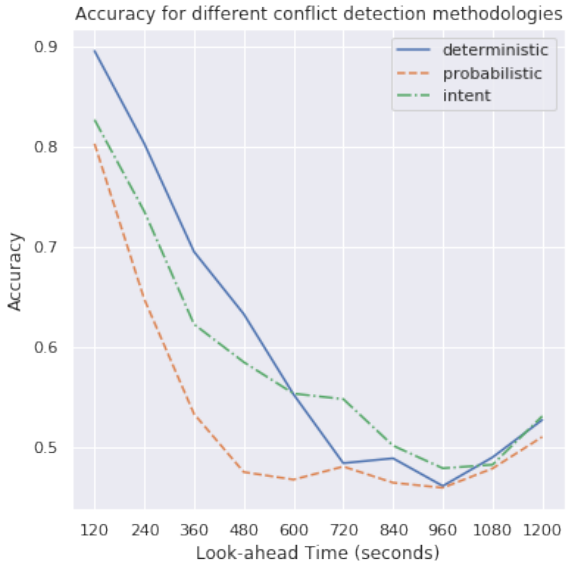
The visible increases in accuracy at a look-ahead time of 1200 seconds is unexpected and a result from a decrease in false positives as visible in figure 7.23b. A possible explanation could be an unbalance in the validation data set for these look-ahead times.



(a) Precision



(b) Recall



(c) Accuracy

Figure 7.22: Comparison of state-deterministic, state-probabilistic and intent-deterministic conflict detection for different look-ahead times based on different metrics

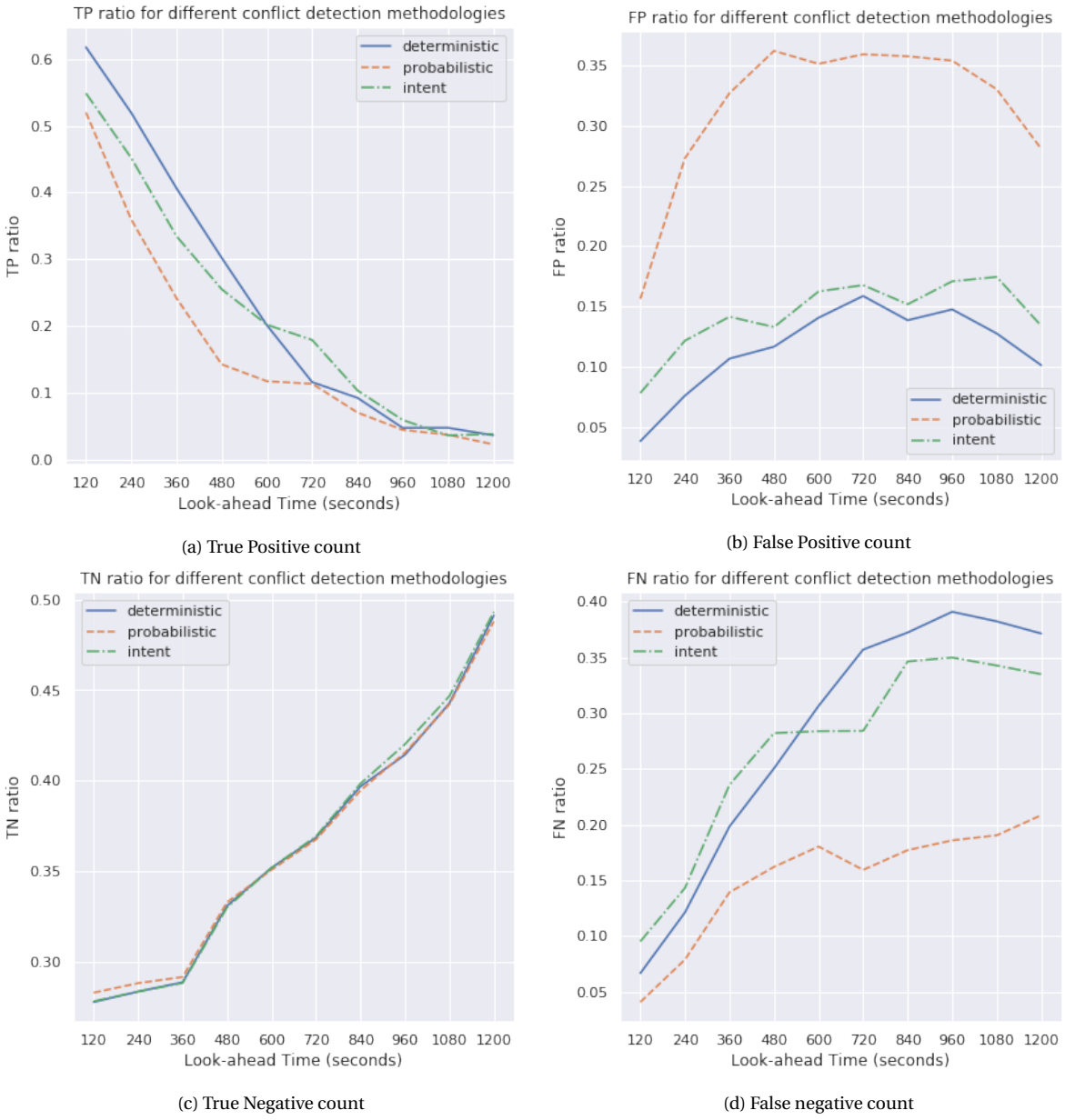
The underlying distribution of true positive, false positive, true negative and false negative detection classes for the full sample set is shown in figures 7.23a, 7.23b, 7.23c and 7.23d. The counts are shown as a fraction the total number of classification points which means that for every look-ahead time $TP + FP + TN + FN = 1$.

The figures show that the true positives counts are continuously decreasing and the false negative rates continuously increasing which shows a natural decrease in detection performance for all detection methods. At a look-ahead time of 1200 seconds all methods show a very poor performance with a true positive count close to zero.

In figure 7.23b, showing the False Positives, it is visible that the state-probabilistic method results in a signifi-

cantly higher number of false positives. However, the amount of false negatives is lower for this method. As the state-probabilistic method is essentially the state-deterministic method with an additional uncertainty margin which scales with the look-ahead time, this method results in fewer missed detections but a higher number of false detections. This trade-off can be changed by choosing a different margin.

For the intent-deterministic approach, the detection performance is similar to the state-deterministic approach although its performance increases slightly for higher look-ahead times. This might indicate that the effect of information from the flight plan starts influencing the method performance for look-ahead times beyond a point of approximately 600 seconds. This can be observed from the model accuracy but also from the true positive and false negative counts.



(a) True Positive count

(b) False Positive count

(c) True Negative count

(d) False negative count

Figure 7.23: Comparison of state-deterministic, state-probabilistic and intent-deterministic conflict detection for different look-ahead times

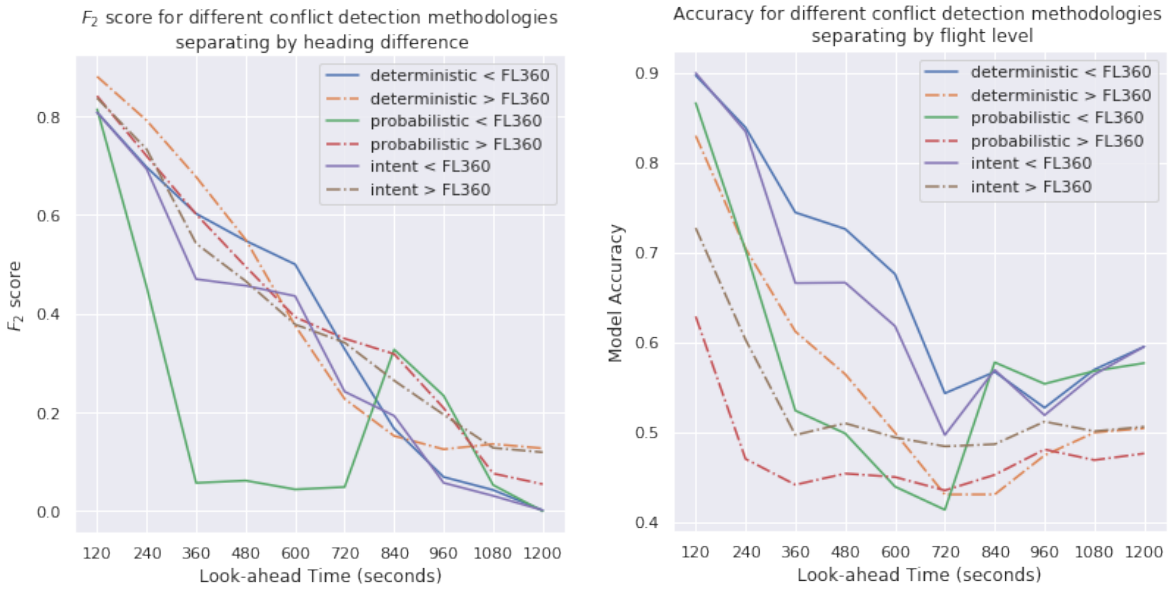
From these results it can be concluded, when balancing accuracy with the number of false detections, that a state-deterministic conflict detection approach yields the most reliable results up to a look-ahead time of approximately 600 seconds. For higher look-ahead times, an intent-deterministic approach proves to be slightly

more accurate. When false negatives form a significant risk, a probabilistic approach would be the most suitable option.

7.3.3. The effect of flight level on conflict detection performance

Although an initial filtering step on the data set is performed in order to only use conflict scenarios above flight level 300, an assessment is made of the effect of the altitude of the conflict on the detection performance. In order to assess the effects of this on the results, the F2-measures for the different methods are also calculated while separating the points in the scenarios by flight level using a split between scenarios above and below flight level 360. The results are shown in figures 7.24a and 7.24b which show the F2-measure and accuracy results respectively.

The F2-measure does not show a very significant difference between the performance of the methods on both datasets except for a poor performance of the state-probabilistic method on the scenarios filtered below FL360 between a look ahead time of 360 and 720 seconds. However, the model accuracy does show a difference where the detection performance on lower flight levels is slightly higher. From earlier results the expected outcome of the filtering would be opposite. The cause of the higher accuracy arises from a larges number of true negative samples in the lower flight level results which result in higher accuracy as can be seen in figure 7.26c. The number of true positives however show the result that would be expected from the flight level segregation. In figure 7.26a it can be observed that all three detection methods yield a larger number of true positives for intrusions at altitudes above FL360 compared to intrusion scenarios below FL360.

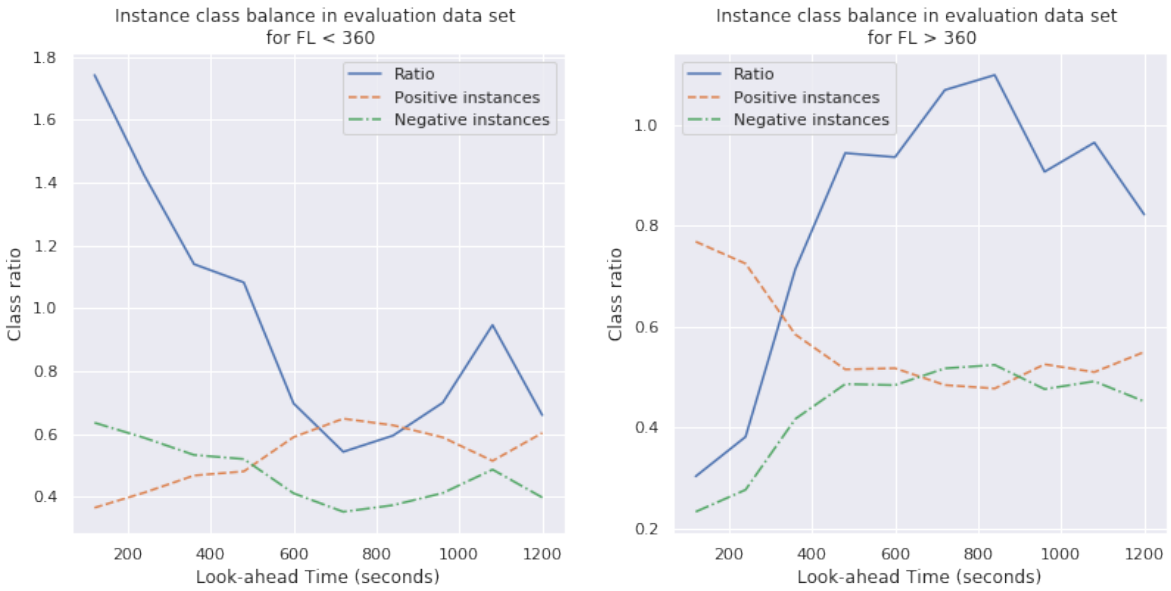


(a) F2-measure

(b) Model Accuracy

Figure 7.24: F-measure and accuracy scores from state-deterministic, state-probabilistic and intent-deterministic conflict detection for different look-ahead times separated by flight level

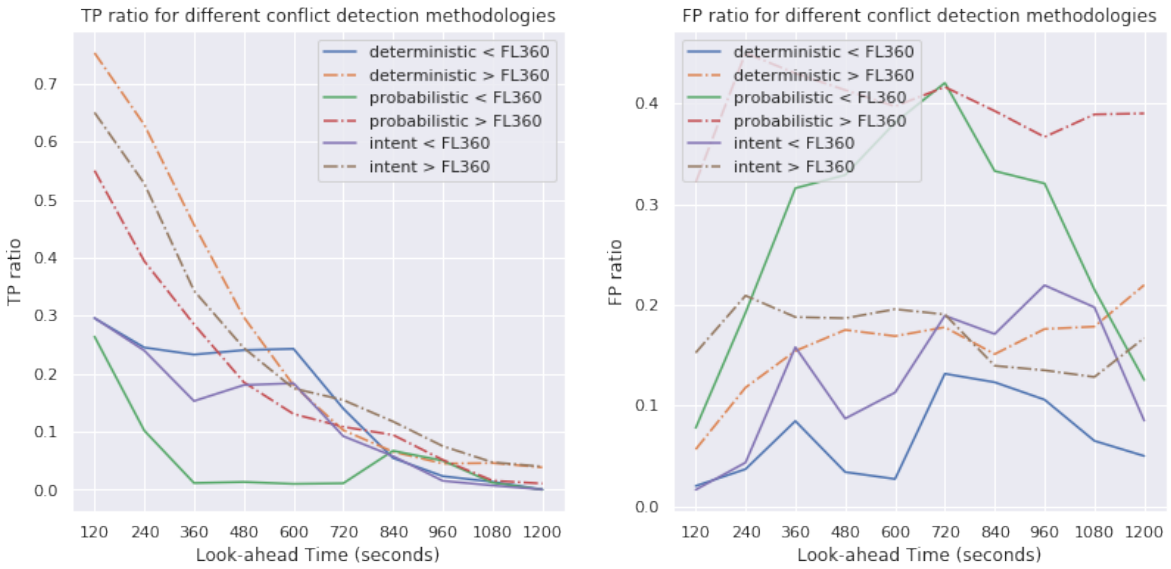
From figures 7.25a and 7.25b it can be observed that the balance between the positive and negative samples is still close to one indicating an acceptable balance between both classes in the validation data set. However, for look-ahead times below 400 seconds, there is a larger unbalance in both validation sets, especially for the scenarios above FL360. Here, significantly more positive instances are present.



(a) Low altitude

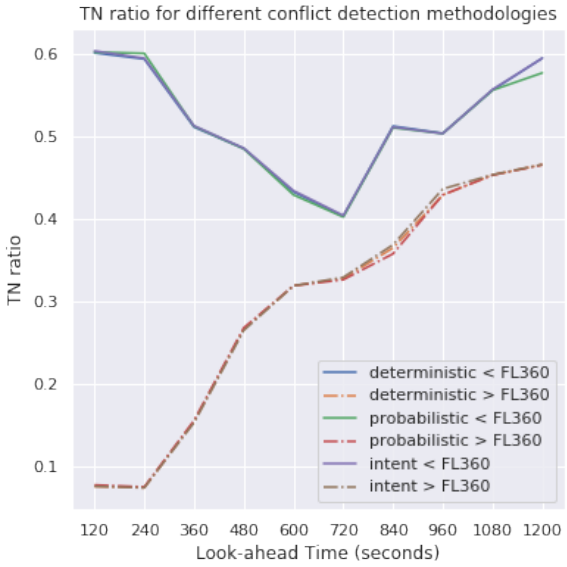
(b) High altitude

Figure 7.25: Occurrences of the positive and negative classes for different look-ahead times

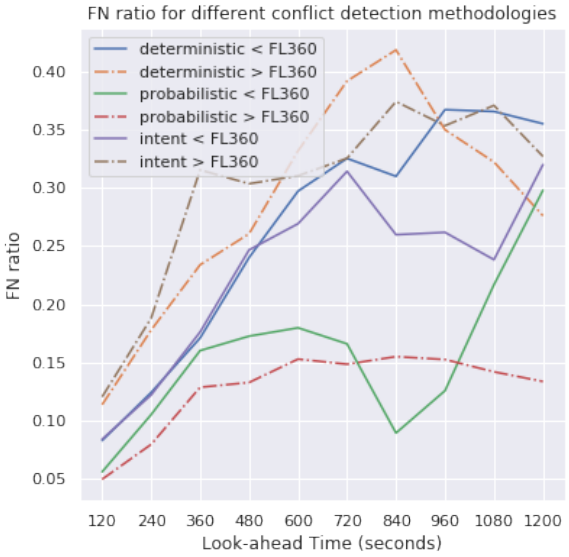


(a) True Positive count

(b) False Positive count



(c) True Negative count



(d) False negative count

Figure 7.26: Comparison of state-deterministic, state-probabilistic and intent-deterministic conflict detection for different look-ahead times separated by flight level

Figures 7.26a, 7.26b, 7.26c and 7.26d respectively show the classification errors by type for the segmentation in altitude. From this it can be seen that mainly a drop in the number of true positives at lower look-ahead times accounts for the difference in performance.

For the intent-deterministic and state-deterministic approaches the false positive and false negative ratios are larger for higher flight levels indicating lower model performance for this validation set. However, figures 7.25a and 7.25b show that there is an unbalance in the positive and negative classes for look ahead times below 400 seconds which might cause the discussed differences in especially the true positive and true negative counts.

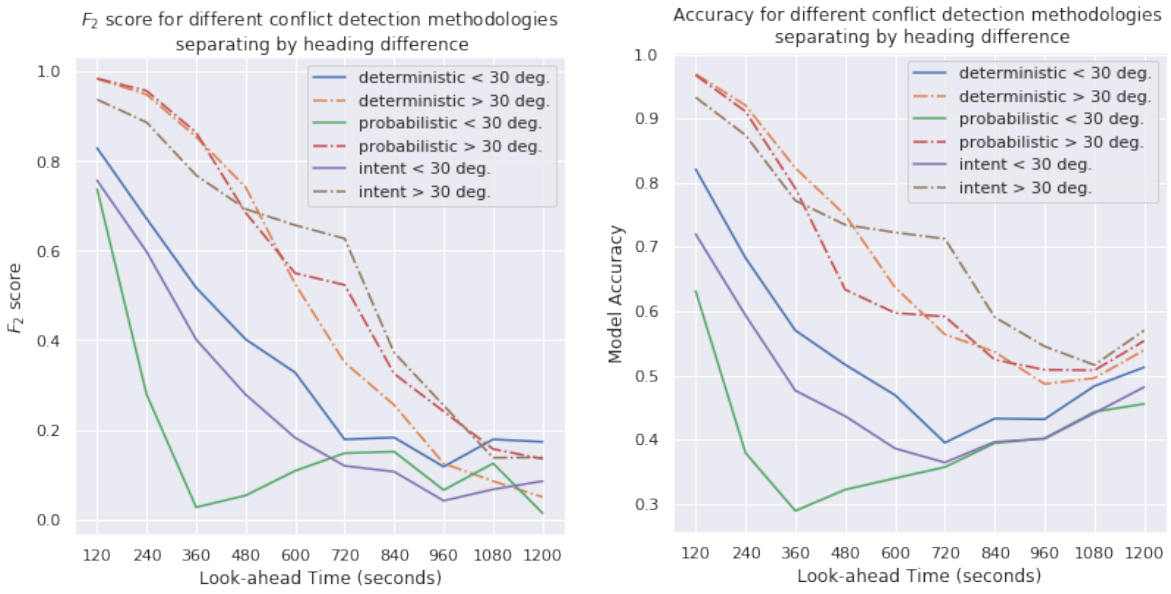
Concluding, it can be stated that the effect of flight level on the conflict detection results is visible when splitting the validation data between intrusion scenarios below FL360 and above FL360. However, the differences are mainly visible in the model accuracy and true positive counts. The F2-measure which takes into account the balance between the positive and negative classes does not reflect these differences significantly. Therefore, it is likely that the observed differences are due to class unbalance.

7.3.4. The effect of heading difference on conflict detection performance

The performance of the three methods is also compared when segregating the validation scenarios by the differences in heading. Angles between the flight paths of both aircraft below 30 degrees and above 30 degrees are considered. The results of this comparison are shown in figures 7.27a and 7.27b which show the F2-measure scores and accuracies.

When both flight segments from a scenario have a larger difference between their headings, which mean they approach eachother at a larger angle, the conflict detection performance increases. This distinction becomes less significant for higher look-ahead times as the detection performance drops in general. Smaller angles, lower than 30 degrees, result in significantly worse conflict detection performance. This effect is also consistent for all of the three methods.

This effect might be caused as intrusion scenarios where two aircraft approach eachother at a shallow angle with an accordingly low approach speed, changes in the flight paths have a larger impact on the time to conflict estimation compared to when the angle between both flight paths is larger.



(a) F2-measure

(b) Model Accuracy

Figure 7.27: F-measure and accuracy scores from state-deterministic, state-probabilistic and intent-deterministic conflict detection for different look-ahead times separated by heading difference

Regarding the accuracy scores, it can be seen that for shallow angle conflict scenarios the three methods only show better than random performance for look-ahead times lower than 400-500 seconds. With the intent-deterministic approach even yielding better than random performance only below a look-ahead time of about 200 seconds. Even when taking into account possible variations between the data sets of the wide angle and shallow angle conflicts, the effect is significant.

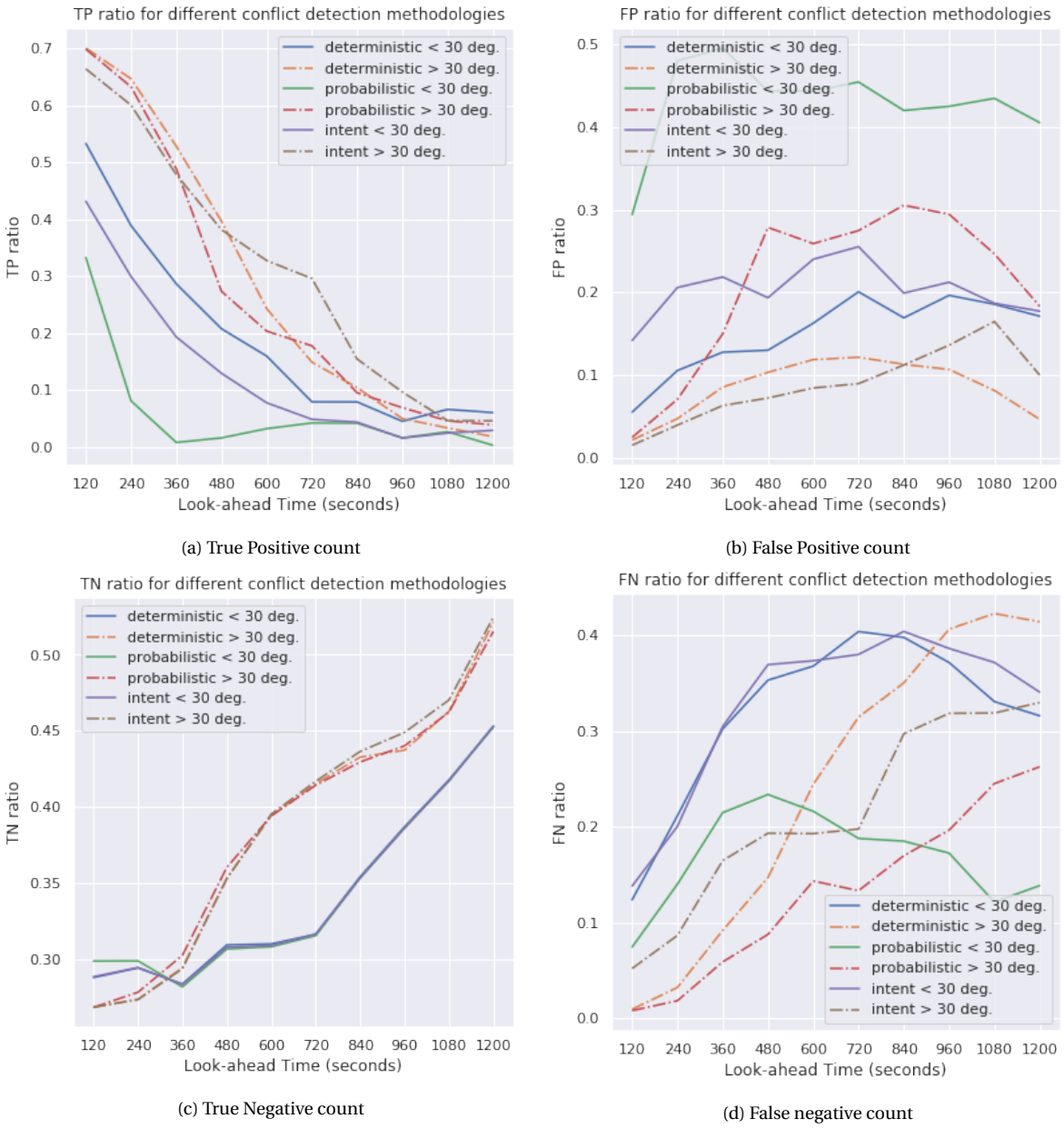
Fow wide angle conflicts, the accuracy decreases below 0.5 only after about 1000 seconds, which is higher than would be expected from the track-error distributions provided in earlier sections.

Figures 7.28a and 7.28b show that in both filtered sets the class balance is better than when segregating by flight level. This means the results are likely to be less affected by the positive and negative class ratio.

(a) Low heading differences

(b) High heading differences

Figure 7.28: Occurrences of the positive and negative classes for different look-ahead times



(a) True Positive count

(b) False Positive count

(c) True Negative count

(d) False negative count

Figure 7.29: Comparison of state-deterministic, state-probabilistic and intent-deterministic conflict detection for different look-ahead times separated by heading difference

Especially the true positive counts in figure 7.29a show a clear distinction between wide and shallow angle scenarios where wide approach angles result in better performance. From figures 7.29b and 7.29d it can be seen that smaller heading differences impacts the number of false positives and false negatives as well.

From these results it can be concluded that the effect of the angle between the aircraft headings has a significant impact on the conflict detection performance. For all three methods this effect is the same. When the angle between two flight paths is more shallow, the detection performance decreases. Taking the model performance accuracy into account the three methods seem to be the most useful for wide angle conflict scenarios.

7.4. Summary of results

From the general comparison of the three conflict detection methodologies the performance shows to be very similar for the observed look-ahead time range to 1200 seconds ahead. For lower look-ahead times a state-deterministic projection approach yields the best results based on model accuracy up to a look-ahead time of approximately 600 seconds. Thereafter, intent-deterministic conflict detection performs better for look-ahead times beyond 600 seconds. Although a probabilistic approach has lower accuracy, the number of false negatives is significantly lower, which in this context might be preferred as false negatives might result in unwanted situations when applied in practice. This lower number of false negatives is however balanced by a larger number of false positives. The probabilistic approach is mainly useful as a risk-averse methodology where a lower number of false negatives is traded for a larger number of false positives but performs worse in general regarding accuracy.

The compared methods prove to have higer than random predictive performance up to approximately 700 seconds ahead. This corresponds to the findings of section 7.2.3 where the track-error distributions were analyzed. In more detail, the state-probabilistic method falls below an accuracy level of 0.5 around 400 seconds, the state-probabilistic method around 700 seconds and the intent-deterministic method around 840 seconds. Fom the F2-measure it can be observed that for higher look-ahead times, the model performance becomes increasingly worse. Around the chosen look-ahead time horizon of 1200 seconds all methods show very poor results which is as expected from initial analyses of the track error distributions.

The results show as well that for the comparison of conflict detection methods, the conflict altitude and heading difference conditions influence the model performance. Effects are especially significant when segregating shallow angle and wide angle conflict scenarios. For scenarios where flight paths have a shallow angle with eachother, the detection performance of all three methods becomes worse. Regarding the model accuracy in this case, the scores drop below a level of 0.5 about 2-3 times earlier for shallow angle conflicts compared to wide angle conflict scenarios.

Considering segragation by altitude, the effect is less significant. The F2-measure shows similar results for both conflict scenarios above FL360 as well as below FL360. However, the accuracy for all methods is higher for scenarios at lower flight levels, which is an opposite effect as would be expected from other analysis results regarding the track errors. Parts of this effect might also result from an unbalance in the positive and negative classes for lower look-ahead times which was present in this experiment.

8

Conclusion

This report presents the research results assessing the predictability of flight trajectories by comparing three conflict detection methodologies using ADS-B data and ETFMS flight plan data from Eurocontrol.

From literature an increasing inclination towards probabilistic and data driven trajectory prediction can be observed. Current approaches to the trajectory prediction problem usually include a dynamic physical model of the aircraft. These models however are often resource intensive to compute and not always as accurate as desired because of the large uncertainty in parameter values and operational context related to for instance aircraft type, state of the airspace and weather. Therefore, data driven models for trajectory prediction are expected to be more accurate and less computationally intensive in operation.

ADS-B transponder data seems a proper data source for development of data driven trajectory prediction models. Studies on ADS-B data quality and accuracy show that this data type proves useful for CD&R applications. The ADS-B transponder system also allows for exchange of the next TCP of aircraft which enables the use of intent information which can be an addition to deterministic or probabilistic state propagation approaches.

Only limited studies to the underlying empirical statistical characteristics of flight paths and intrusion scenarios have been performed. These however are likely to provide useful insights inapplicability of data driven approaches on trajectory propagation. Many times, assumptions on these data distributions are used in research but not often tested against historical data. This for instance regards assumptions on normality of flight path deviations and error distributions.

From a large historic set of ADS-B transponder messages and ETFMS flight plan data from Eurocontrol, a data set of flight trajectories is created. The approach for the processing of ADS-B data into separate flight trajectories is presented as well as the calculation of cross-track, along-track and total track errors. Methods for the processing of conflict scenarios are presented as well to provide efficient ways of deriving conflicts from large amounts of trajectory data. From this data set a large set of intrusion scenarios are used on which the three mentioned conflict detection methods are benchmarked and compared up to a look-ahead time of 20 minutes. This is done using an F-measure score based on a binary conflict detection classification. Before this, a statistical analysis of state-deterministic trajectory extrapolation has been performed with the same look-ahead time boundary.

This statistical analysis on projection errors from a state-deterministic propagation shows that the error distribution of the cross-track, along-track and total track error all show the presence of fat tails. This peakedness of the distributions diminishes slightly for larger look-ahead times as the chance of trajectory changes grows with time. However, in none of the scenarios the error distributions are normally distributed. This has to be taken in mind for any future work based on these data sources.

Altitude largely affects the error distritutions. As flights at lower altitudes are more prone to vectoring directed by ATC, the presence of abrupt changes in the flight path is significantly higher. This reflects in larger variance in the error distributions and a higher presence of outliers. Therefore, the predictability of trajectories is much lower at altitudes below 30.000 feet compared to flights above this flight level with mean absolute deviations of up to a factor three larger.

Based on the combination of total track errors an empirical upper limit for the look-ahead time can be found in the context of state-deterministic projection. This is based on the look-ahead times where the TTE distribution percentiles grow larger than the IPZ. A 70th percentile confidence bound grows beyond the IPZ range around 9 minutes (540 seconds). When using a 90th percentile bound, this look-ahead time reduces to 5 minutes and 20 seconds (320 seconds).

From the general comparison of the three conflict detection methodologies the performance shows to be very similar for the observed look-ahead time range to 1200 seconds ahead. For lower look-ahead times a state-deterministic projection approach yields the best results based on model accuracy up to a look-ahead time of approximately 600 seconds. Thereafter, intent-deterministic conflict detection performs better for look-ahead times beyond 600 seconds. Although a probabilistic approach has lower accuracy, the number of false negatives is significantly lower, which in this context might be preferred as false negatives might result in unwanted situations when applied in practice. This lower number of false negatives is however balanced by a larger number of false positives. The state-probabilistic approach is mainly useful as a risk-averse methodology where a lower number of false negatives is traded for a larger number of false positives but performs worse in general regarding accuracy.

Initially performance was expected to increase significantly for larger look-ahead times using an intent-deterministic approach as the TCP information is expected to account for planned trajectory changes in the future. The effect of this is likely mitigated due to limited adherence to the flight plan and the presence of vectoring procedures resulting in flight plan deviations. The effect of incorporating intent information is not tested for look-ahead times larger than 20 minutes. Therefore it could be that the positive effect of this information becomes more expressive beyond this limit compared to state-probabilistic and state-deterministic approaches.

The results show as well that for the comparison of conflict detection methods, the conflict altitude and heading difference conditions influence the model performance. Effects are especially significant when segregating shallow angle and wide angle conflict scenarios. For scenarios where flight paths have a shallow angle with eachother, the detection performance of all three methods becomes worse. Regarding the model accuracy in this case, the scores drop below a level of 0.5 about 2-3 times earlier for shallow angle conflicts compared to wide angle conflict scenarios.

The compared methods prove to have higer than random predictive performance up to approximately 700 seconds ahead. This corresponds to the findings of section 7.2.3 where the track-error distributions were analyzed and a conservative look-ahead time boundary was set around 540 seconds.

Regarding upper boundaries on look-ahead times for conflict detection, the state-probabilistic method falls below an accuracy level of 0.5 around 400 seconds, the state-probabilistic method around 700 seconds and the intent-deterministic method around 840 seconds. Fom the F2-measure it can be observed that for higher look-ahead times, the model performance becomes increasingly worse. Around the chosen look-ahead time horizon of 1200 seconds all methods show very poor results which is as expected from initial analyses of the track error distributions.

9

Recommendations

This chapter discusses research recommendations based on the results from this research. They are out of the scope of the proposed research in this report, but might be still of interest for future research.

- **Correlation analysis between airspace state and trajectory changes**
  It is known that ATC interventions are likely the source of disturbance resulting in the largest aircraft trajectories deviations. These interventions, based on choices made by the control operator, are expected to be strongly related to the dynamics of the airspace at that moment around the respective aircraft. When for instance a bad weather cell is present or when there is overcapacity in a certain area, it is likely that traffic will be rerouted away from these locations.

  In order to better estimate the chance on future trajectory changes, a relationship between the airspace situation and aircraft intent needs to be further evaluated. The difficulty in this is that not all required information is present at all agents in the system. Deviations from the planned flight path result from unexpected events related to weather, airspace capacity, special use airspace or potential conflicts with other air traffic.

  Performing a study into the relationship between the airspace state, trajectory deviations of surrounding air traffic and the own aircraft enables the modeling of these effects in more detail. This can be of help in improving conflict detection models and their reliability.

- **Analysis of TMA traffic movements in future research**
  In this research, filters based on flight level have been put in place in order to exclude the effects from TMA traffic and manouvering from the results. However, this is an interesting study topic by itself as the effect of vectoring and ATC influence has not been empirically quantified in much detail.

- **Analysis of correlations between trajectories of different aircraft**
  There likely is a relationship between the movements of different aircraft when flying at a relatively close distance. The question is if trajectories in the same airspace area can be viewed as independent movements, in the space and/or time domain. This will have an impact on the way trajectory prediction and conflict detection methods can be approached.

- **Effect of autocorrelation and time dependency in aircraft trajectories**
  In this research, autocorrelation in trajectory errors from different projection methods has not been taken into account. Autocorrelation here refers to consequtive data points from a flight trajectory not being independent. When a certain error from a reference track is established at some point in time, it is likely that following points include this error as well. In this research these effects are assumed to be mitigated by the size of the data set. However, especially when evaluating smaller sample sets of flights, this effect might prove significant for the outcomes of conflict detection methods.

- **Research into the medium -to long term conflict detection**
  Because of the limitations in the available ADS-B data set, the look-ahead time boundary is set at 20 minutes for this research. However, it is recommended to evaluate the differences in the applied method-

ologies for larger look-ahead times. Especially as the effect of intent information is expected to have a more significant effect within this range.

- **Integration with BlueSky**
  BlueSky is an open source air traffic control simulation program developed in Python.[17] It is useful to test different experimental scenarios. Integrating the different trajectory prediction models enables the testing of their applicability in predefined and filtered sets of scenarios.

Bibliography

[1] Elias Alevizos, Alexander Artikis, and George Paliouras. Event Forecasting with Pattern Markov Chains. 12(17), 2017. doi: 10.1145/3093742.3093920. URL http://dx.doi.org/10.1145/3093742.3093920.

[2] Kwang-Yul Baek and Hyo-Choong Bang. ADS-B based Trajectory Prediction and Conflict Detection for Air Traffic Management. *International Journal of Aeronautical and Space Sciences*, 13(3):377–385, 2012. ISSN 2093-274X. doi: 10.5139/IJASS.2012.13.3.377.

[3] G.J. Bakker, H.J. Kremer, and H.A.P. Blom. Geometric and probabilistic approaches towards conflict prediction, 12 2001.

[4] M.G. Ballin and H. Erzberger. An Analysis of Landing Rates and Separations at the Dallas Fort Worth International Airport. *Distributed Control and Stochastic Analysis of Hybrid Systems Supporting Safety Critical Real-Time Systems Design (HYBRIDGE)*, 1996.

[5] K. Blin. Stochastic conflict detection for air traffic management. April 2000.

[6] C. Borst and M. Mulder. Air traffic control & air traffic management. Lecture for course AE4302 - Avionics and Operations, 2015.

[7] EUROCONTROL Experimental Centre. Mtcd concept of operation eatchip iii evaluation and demonstration phase 3a bis. https://www.eurocontrol.int/sites/default/files/library/020_MTCD_Concept_of_Operation.pdf, 9 1999. Accessed: 02-10-2018.

[8] T. Champougny. DDR2 Reference Manual For General Users 2.9.5. https://ext.eurocontrol.int/ddr/files/documentation/ddr2_userguide_generic.pdf, 2018.

[9] DART Consortium. Data Driven Aircraft Trajectory Prediction Exploratory Research, 2017.

[10] Arjen de Leege, Marinus van Paassen, and Max Mulder. A Machine Learning Approach to Trajectory Prediction. *AIAA Guidance, Navigation, and Control (GNC) Conference*, pages 1–14, 2013. doi: 10.2514/6.2013-4782. URL http://arc.aiaa.org/doi/10.2514/6.2013-4782.

[11] Adric Eckstein. Automated flight track taxonomy for measuring benefits from performance based navigation. *Proceedings of the 2009 Integrated Communications, Navigation and Surveillance Conference, ICNS 2009*, 2009. doi: 10.1109/ICNSURV.2009.5172835.

[12] Eurocontrol. Single european sky. http://www.eurocontrol.int/dossiers/single-european-sky. Accessed: 12-03-2018.

[13] EUROCONTROL. Eurocontrol specification for short term conflict alert. https://www.eurocontrol.int/sites/default/files/publication/files/20071122-stca-spe-v1.0.pdf, 11 2007. Accessed: 03-10-2018.

[14] FAA - Federal Aviation Administration. Introduction to TCAS II - Version 7.1. pages 1–50, 2011. ISSN 1098-6596. doi: 10.1017/CBO9781107415324.004.

[15] Tharindu Fernando, Simon Denman, Aaron McFadyen, Sridha Sridharan, and Clinton Fookes. Tree memory networks for modelling long-term temporal dependencies. 2017.

[16] Maxime Gariel, Fabrice Kunzi, and R. John Hansman. An algorithm for conflict detection in dense traffic using ADS-B. *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, pages 1–12, 2011. ISSN 2155-7195. doi: 10.1109/DASC.2011.6095916.

[17] J.M. Hoekstra and J. Ellerbroek. Bluesky atc simulator project: An open data and open source approach. *7th International Conference on Research in Air Transportation*, 2016.

[18] Inseok Hwang and Chze Eng Seah. Intent-Based Probabilistic Conflict Detection for the Next Generation Air Transportation System. *Proceedings of the IEEE*, 96(12):2040–2059, 2008. ISSN 0018-9219. doi: 10.1109/JPROC.2008.2006138. URL http://ieeexplore.ieee.org/document/4745649/.

[19] M.R. Jackson, Y.J. Zhao, and R.A. Slattery. Sensitivity of trajectory prediction in air traffic management. *Journal of Guidance, Control and Dynamics*, 22(2), 1999.

[20] M.J Kochenderfer, J.K. Kuchar, L.P. Espindle, and J.D. Griffith. Uncorrelated Encounter Model of the National Airspace System, Version 1.0. 2008.

[21] Mykel J. Kochenderfer, Jessica E. Holland, and James P. Chryssanthacopoulos. Next-Generation Airborne Collision Avoidance System. *Lincoln Laboratory Journal*, 19(1):17–33, 2013.

[22] Jimmy Krozel and Dominick Andrisani. Intent Inference with Path Prediction. *Journal of Guidance, Control, and Dynamics*, 29(2):225–236, 2006. ISSN 0731-5090. doi: 10.2514/1.14348.

[23] J K Kuchar and L C Yang. A review of conflict detection and resolution modeling methods. *IEEE Transactions on Intelligent Transportation Systems*, 1(4):179–189, 2000. ISSN 15249050. doi: 10.1109/6979.898217. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=898217.

[24] Thom Langejan, Emmanuel Sunil, Joost Ellerbroek, and Jacco Hoekstra. Effect of ADS-B Limitations and Inaccuracies on CD&R Performance. 2016.

[25] T.A. Lauderdale, A.C. Cone, and A.R. Bowe. Relative Significance of Trajectory Prediction Errors on an Automated Separation Assurance Algorithm. *Ninth USA/Europe Air Traffic Management Research and Development Seminar (ATM2011)*, 2011.

[26] Yu Liu and X Rong Li. Intent Based Trajectory Prediction by Multiple Model Prediction and Smoothing. (January):1–15, 2015. doi: 10.2514/6.2015-1324.

[27] Anna Monreale, Fabio Pinelli, and Roberto Trasarti. WhereNext : a Location Predictor on Trajectory Pattern Mining. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*, pages 637–645, 2009. ISSN 00221031. doi: 10.1145/1557019.1557091.

[28] Russell A Paielli. Empirical Test of Conflict Probability Estimation. *2nd USA/Europe Air Traffic Management R&D Seminar (ATM-98)*, 1998.

[29] Russell A Paielli and Heinz Erzberger. Conflict probability estimation for free flight Conflict Probability Estimation For Free Flight. (January), 1997. doi: 10.2514/6.1997-1.

[30] M. Prandini and O.J. Watkins. Probabilistic Aircraft Conflict Detection. *Distributed Control and Stochastic Analysis of Hybrid Systems Supporting Safety Critical Real-Time Systems Design (HYBRIDGE)*, 2005.

[31] Maria Prandini, Jianghai Hu, John Lygeros, and Shankar Sastry. A Probabilistic Approach to Aircraft Conflict Detection. *IEEE Transactions on Intelligent Transportation Systems*, 1(4):199–219, 2000. ISSN 15249050. doi: 10.1109/6979.898224.

[32] T.G. Reynolds and R.J. Hansman. Conformance monitoring approaches in current and future air traffic control environments. *Proceedings. The 21st Digital Avionics Systems Conference*, 2(October):7C1–1–7C1–12, 2002. doi: 10.1109/DASC.2002.1052922.

[33] Ayhan S. and S. Hanan. Aircraft trajectory prediction made easy with predictive analytics. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, pages 21–30, 2016.

[34] Sahawneh Laith R., Duffield Matthew O., Beard Randal W., and Timothy W McLain. Detect and Avoid for Small Unmanned Aircraft Systems using ADS-B. *Air Traffic Control Quartely: An International Journal of Engineering and Operations*, 23(2/3)(1-38):203–240, 2015.

[35] Steven Schelling and Jimmy A. Krozel. Machine Learning Approach for Finding Similar Weather-Impacted Situations in En Route Airspace. *AIAA Guidance, Navigation, and Control Conference*, (January):1–16, 2017. doi: 10.2514/6.2017-1920.

[36] SESAR Joint Undertaking. Digitalising Europe's Aviation Infrastructure. (November), 2017.

[37] J. Sun. The 1090MHz Riddle, An open-access book about decoding Mode-S and ADS-B data, 2018. URL http://mode-s.org/decode/index.html. Accessed: 03-07-2018.

[38] J. Sun, J. Ellerbroek, and J.M. Hoekstra. Large-scale flight phase identification from ads-b data using machine learning methods. *7th International Conference on Research in Air Transportation*, 2016.

[39] Junzi Sun, Joost Ellerbroek, and Jacco Hoekstra. Modeling Aircraft Performance Parameters with Open ADS-B Data. *12th USA/Europe ATM R&D Seminar, Seattle, Washington*, 2017.

[40] T L Verbraak, J Ellerbroek, J Sun, and J M Hoekstra. Large-Scale ADS-B Data and Signal Quality Analysis. *Twelfth USA/Europe Air Traffic Management Research and Development Seminar*, (June), 2017.

[41] George Vouros. Big Data heralds new era for time-critical mobility forecasting and situation awareness Adding significant value to real-time tracking and forecasting mobility. 2017.

[42] Lee C Yang. AIRCRAFT CONFLICT ANALYSIS AND REAL-TIME CONFLICT PROBING USING PROBABILISTIC TRAJECTORY MODELING. 2000.

[43] Yang Yang, Jun Zhang, and Kai Quan Cai. Terminal-area aircraft intent inference approach based on online trajectory clustering. *Scientific World Journal*, 2015, 2015. ISSN 1537744X. doi: 10.1155/2015/671360.

[44] Javier Lovera Yepes, Inseok Hwang, and Mario Rotea. New Algorithms for Aircraft Intent Inference and Trajectory Prediction. *Journal of Guidance, Control, and Dynamics*, 30(2):370–382, 2007. ISSN 0731-5090. doi: 10.2514/1.26750. URL http://arc.aiaa.org/doi/10.2514/1.26750.

[45] N. Yokoyama. Inference of aircraft intent via inverse optimal control including second-order optimality condition. *AIAA Guidance, Navigation, and Control Conference*, 1 2017.

[46] N. Yokoyama. Decentralized conflict detection and resolution using intent-based probabilistic trajectory prediction. *AIAA Guidance, Navigation, and Control Conference*, 2 2018.

[47] Jingbo Zhou, Anthony K.H. Tung, Wei Wu, and Wee Siong Ng. A "semi-lazy" approach to probabilistic path prediction. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13*, page 748, 2013. ISSN 21508097. doi: 10.1145/2487575.2487609.