D4.1 Safety Performance Functions Methodology

DeliverableID	D4.1
Project Acronym	FARO
Grant:	892542
Call:	H2020-SESAR-2019-2
Topic:	ER4-06-2019 "Safety and Resilience"
Consortium coordinator:	CRIDA
Edition date:	17th November 2021
Edition:	00.01.01
Template Edition:	02.00.00

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Fernando GÓMEZ / UPM	WP4 Lead	28 June 2021
Rosa ARNALDO/UPM	Project Contributor	28 June 2021
Raquel DELGADO-AGUILERA / UPM	Project Contributor	28 June 2021
María SÁNCHEZ-VALDEPEÑAS / UPM	Project Contributor	28 June 2021
Mikel LARRAÑAGA / UPM	Project Contributor	28 June 2021

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Fedja NETJASOV/ UB-FTTE	WP6 Lead	7 July 2021
Irene Buselli/ ZENABYTE	Project Contributor	9 July 2021
Anthony SMOKER/ LU	WP5 Lead	12 July 2021
Icíar GARCÍA-OVIES/ CRIDA	WP2 Lead	12 July 2021
Christian VERDONK/ CRIDA	Project Coordinator	13 July 2021
Tamara PEJOVIC/ ECTL	Project Coordinator	13 July 2021

Approved for submission to the SJU By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
Christian VERDONK/ CRIDA	Project Coordinator	19 July 2021
Icíar GARCÍA-OVIES/ CRIDA	WP2 Lead	19 July 2021
Carlo DAMBRA/ ZENABYTE	WP3 Lead	19 July 2021
Fernando GÓMEZ / UPM	WP4 Lead	19 July 2021
Anthony SMOKER/ LU	WP5 Lead	19 July 2021
Fedja NETJASOV/ UB-FTTE	WP6 Lead	19 July 2021
Tamara PEJOVIC/ ECTL	Project Contributor	19 July 2021
Patricia RUIZ/ ENAIRE	Project Contributor	19 July 2021

Rejected By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date

Document History

Edition	Date	Status	Author	Justification
00.00.01	28 June 2021	First Draft	UPM	Initial Draft
00.01.00	31 July 2021	Final Version	UPM	Final Version
00.01.01	17 November 2021	Reviewed Version	UPM	Editorial modifications

© – 2021– FARO Consortium. All rights reserved. Licensed to the SESAR Joint Undertaking under conditions.

FARO

SAFETY AND RESILIENCE GUIDELINES FOR AVIATION

This document is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 892542 under European Union's Horizon 2020 research and innovation programme.

Abstract

FARO WP4 aims at generating predictive models of safety events (Safety Performance Functions - SPFs) by using organisational, technical, human, and procedural precursors to characterise and predict airspace Separation Minima Infringement (SMI), as a function of those precursors.

To accomplish this objective, the work has been organised into 3 tasks:

- T4.1 Safety Performance Functions (SPFs) Development. The development of the SPFs starts with a characterisation of the safety events in terms of the safety dimensions (precursors) and their aggregation.
- T4.2 Safety Performance Functions (SPFs) Calibration, Adjustment and Sensitivity Analysis.
- T4.3 Safety Performance Functions (SPFs) Influence Factors and Applicability Thresholds.

Deliverable 4.1 is tasked with developing a baseline model of the SPF for the characterization and prediction of airspace Separation Minima Infringement (SMI) in particular ATC sectors. This report covers the research conducted in T4.1 and T4.2.

Table of Contents

List of Tables

List of Figures

1 Introduction

Work Package 4 (WP4) in FARO proposes the development of ATM Safety Performance Functions (SPF) as effective tools to measure safety performance in ATM. The concept of Safety Performance Functions refers to explanatory and predictive mathematical models able to explain and predict the occurrence of safety events. FARO project is concerned with a particular type of ATM safety event, the Separation Minima Infringement or SMI. A SMI is a situation in which prescribed separation minima were not maintained between aircraft.

In general, the two project level objectives of FARO project related to safety area are the following:

O1 Capitalisation on the existent knowledge of safety – This objective pursues to systematically extract existent safety knowledge by applying data-driven techniques combined with a knowledge-based approach, levering the knowledge of experts within the consortium and exploiting experience from other transport modes.

This objective pursues the identification of safety events' dimensions in terms of technological, organisational and human aspects associated to specific automation solutions.

O2 Quantification of the impact of increasing the level of automation on ATM safety levels - This objective aims at generating predictive models of safety events as a function of the technological, organisational, human and procedural dimensions and automation solutions defined in the scenarios considered in WP2.

Therefore, WP4 aims at generating predictive models of safety events by using organisational, technical, human, and procedural precursors to characterise and predict airspace Separation Minima Infringement (SMI), as a function of those precursors.

This objective pursues to systematically extract existent safety knowledge by applying data-driven techniques combined with a knowledge-based approach, levering the knowledge of experts within the consortium and exploiting experience from other transport modes.

To accomplish this objective, the work has been organised into 3 tasks:

- T4.1 Safety Performance Functions (SPFs) Development. The development of the SPFs starts with a characterisation of the safety events in terms of the safety dimensions (precursors) and their aggregation. The outcome of this descriptive analysis, together with prior statistical knowledge, serves to select potential models that could provide statistical representations of the frequency and severity of safety events. A data-driven approach complements the SPF development, allowing the identification of the SPFs themselves after the descriptive analysis and the selection of potential statistical models.

- T4.2 Safety Performance Functions (SPFs) Calibration, Adjustment and Sensitivity Analysis. The models proposed in the previous task are adjusted and calibrated from real data. The explanatory power of each model and /or independent variable and the mixed effects are quantified. Sensitivity analysis considering mixed effects enables the characterisation of safety performance in terms of not only the independent dimensions, but also their combinations, identifying prior thresholds of those dimensions that would reduce the frequency of a safety event.

- T4.3 Safety Performance Functions (SPFs) Influence Factors and Applicability Thresholds. Application of the model to the study cases or scenarios defined in the project, to quantify the influence factors of each study case and determine the criteria and thresholds for its applicability.

1.1 Purpose of the document

Deliverable 4.1 is tasked with developing a baseline model of the SPF for the characterization and prediction of airspace Separation Minima Infringement (SMI) in particular ATC sectors. This report covers the research conducted in T4.1 and T4.2.

This document describes the mathematical background and the methodological approach followed, describes the process followed to build up the model, and serves as the underlying framework for subsequent deliverables (D4.2).

The report covers selected SPFs as a function of the safety dimensions (precursors) and their aggregation, the results from the adjustment and calibration processes, and the sensitivity analysis with respect to the independent dimensions.

1.2 Document and content

This document is structured as follows:

- Section 1 introduces the purpose of the document, its contents and the terminology and acronyms used.
- Section 2 provides an introduction to Safety Performance Functions (SPF) and explains that they refer to mathematical models with the ability to predict the occurrence of safety events.
- Section 3 provides a first outlook on the main concepts involved in the construction of a Bayesian Network (BN) and how the BN outcomes can be exploited.
- Section 4 analysis the optimum desirable set of information, from a knowledge-based perspective, to generate the selection of required data, as well as the data finally available within FARO project for being used in the SPF models. It also discusses the data transformation process that has been necessary to exploit, as much as possible, available data to populate variables in the model.
- Section 5 presents the overall methodology and conceptual framework followed to develop the structure of the SPF model.
- Section 6 describes the resulting model in all its details.
- Section 7 presents a list of all input, training and output variables of the network. An explanation of the variable and an example of its discretisation is included.
- Section 8 shows the integration of the subnetworks defining each of the safety barriers into a single compact model.
- Section 9 discusses the conclusions drawn from the methodology developed during this document as well as a proposal for next steps.

1.3 Terminology and acronyms

Table 1: Acronyms list

Term	Definition
3D	3 Dimensional
AC	Aircraft
ANSP	Air Navigation Service Provider
ATC	Air Traffic Control
ATCo	Air Traffic Controller
AFTCM	Air Traffic Flow Capacity Management
ATM	Air Traffic Management
BBN	Bayesian Belief Network
BN	Bayesian Network
CNS	Communications, Navigation and Surveillance
CPA	Closest Point of Approach
dCPA	Distance Closest Point of Approach
EB	Empirical Bayes
ECTS	European Train Control System
ERTMS	European Railway Traffic Management System
ETA	Event Tree Analysis
FARO	safety And Resilience guidelines for aviatiOn
FL	Flight Level
ft	Feet
GSMR	Global System for Mobile communications-Railway
IE	Initial Event
LECBCCC	Barcelona Central Sector
LECBCCU	Barcelona Upper Sector

LECMSAN	Santiago Sector
LoS	Loss of Separation
MAC	Mid Air Collisions
MTS	Maritime Transport System
NM	Nautical Miles
RBC	Radio Block Center
RTM	Regression to the Mean
SMI	Separation Minima Infringement
SPF	Safety Performance Function
STCA	Short Term Conflict Alert
TLC	Time of Last Clearance
Vx	X-axis speed
Vy	Y-axis speed
Vz	Z-axis speed
WP	Work Package

2 Introduction to Safety Performance Functions –SPF.

The term Safety Performance Function (SPF) is used in many industries to refer, in a general way, to mathematical models that have the capacity to explain, but above all, predict the occurrence of safety events.

The expression has been coined in the field of road transport, with the development of models to predict the occurrence of traffic accidents, and over the years its applications have expanded into new fields such as railway or aviation.

Most of the current research on road safety is based on the analysis of crash data that are challenged by well-recognized quality and availability issues. The use of surrogate safety measures, such as traffic conflicts, has been gaining acceptance as an alternative or complementary approach to analyse traffic safety from a broader perspective than collision data alone. Road crashes can occur as a consequence of several factors such as: human behaviour, environment, vehicle, and road characteristics. Khair (Khair S. Jadaan, June 4-5, 2014), using Multivariate Analysis Model, established a mathematical relationship between explanatory variables such as weather, road geometry, traffic volume or human factors and collision frequency in roadways with different sections.

However, there is a need to develop appropriate statistical techniques to analyse conflict data to support various complex safety applications. Frequentist statistical inference, where conclusions are drawn from sample data by emphasizing the frequency or proportion of the data, hypothesis testing, and confidence intervals, is not useful when the number of safety events is limited. The latest works in this area focus on the framework of Bayesian statistics, which is considered the most advanced technique in statistical analysis of collisions (R Arnaldo, 2019) (Rosa María Arnaldo Valdés, 2018 Dec. 14).

At (Sacchi, 2015) SPFs based on Bayesian Networks (BN) were developed to predict the number of rear-end conflicts at different intersection approaches and the functions were validated using posterior predictive checking indicators. Data for traffic conflict observations were automatically extracted with computer vision techniques at several urban and suburban intersections in British Columbia (Canada). The work at (Villaz´an, 2017) focuses on the application of BNs for traffic accident causality analysis as the most adequate statistical model, due to its power to reproduce multidimensional random variables (E. Castillo, 1997), and its capacity to integrate all relevant items of the road in the same model.

Some applications of the Safety Performance Functions have been used to estimate train driver errors and conduct safety assessments of the whole ERTMS (European Railway Traffic Management System) (F. Flammini, 2006). In this application, two complex Bayesian Networks were developed taking into account variables such as tiredness, fatigue, training, policies of organization and so on. The difference between these two networks lies in the equipment need for ERTMS/ECTS operation. They consider two levels. Level 1 fitted with balises, loops, lineside electronic units, lineside signals and track circuits and Level 2 fitted with balises and radio track circuits and radio block. The results of this analysis show that ERTMS Level 2 is safer and less prone to driver errors than ERTMS Level 1; however, it also contains more critical elements (such as GSMR system and RBC) that have a significant impact on the continuity of ERTMS functioning, such that any failure in one of these components will stop the whole ERTMS system. Accordingly, it can be concluded that new systems with advanced technologies will improve safety only if their subsystems and components are reliable and interact with each other reliably.

It is also possible to find some applications of SPF, built upon the technology of Bayesian Networks in the maritime industry. At (P. Truccoa, 2008) a Bayesian Belief Network (BBN) has been developed to model the Maritime Transport System (MTS), by taking into account its different actors (i.e., ship-owners, shipyards, port and regulators) and their mutual influences. The latter have been modelled by means of a set of dependent variables whose combinations express the relevant functions performed by each actor. The study has focused on a collision in open sea hazards. The approach has allowed the identification of probabilistic correlations between the basic events of a collision accident and the BBN model of operational and organisational conditions.

The Bayesian Network –Safety Performance Functions approach has been developed and applied to several case studies in the road, train, and maritime industry, but it can also be utilised in other sectors such aviation and Air Traffic Management. Considering different characteristics which were analysed in depth in WP2 D2.2 (FARO project, 2021), these models can be conceptually extrapolated to the airspace, considering different characteristics: the geometry of the routes followed by the aircraft, the volume of traffic, the mix of traffic and its dynamic variables, the geometry of the encounters between aircraft, the severity or magnitude of Separation Minima Infringement between aircraft, the complexity of the airspace structures, the size and characteristics of the sectors where the aircraft are flying, the complexity of the organization and management of the airspace, etc. One of its main applications focuses on the analysis of Airspace Separation Minima Infringement. Separation Minima Infringement (SMI) is a situation in which prescribed separation minima were not maintained between aircraft. The occurrence of SMIs that could lead to Mid Air Collisions (MAC) is of major concern to Air Traffic Management.

According to this approach, some projects have extended the concept of Safety Performance Functions to ATM to develop models capable of explaining and predicting the occurrence of SMIs considering different precursors. At (R Arnaldo, 2019) a frequentist statistical approach is used to characterise the SMIs between aircraft as count data with an excess of zeros and over dispersion. Subsequently, the relationships between the number of aircraft conflicts in a particular route segment and the airspace design and traffic flow characteristics are modelled using Zero-inflated models. Based on the characteristics of the route segment, the distribution that most closely matches observations of the number of conflicts in airspace segments is a Zero-inflated negative binomial probability distribution. It also takes into account of the large amount of null values that characterise safety occurrences in aviation. However, this first attempt did not exploit some of the potential of the Bayesian Network technology to utilise causality inference and prediction in ATM.

A more complete alternative will be to use empirical Bayesian models (Empirical Bayes, EB). These models allow addressing two common problems associated with predictive safety models (Hauer, 2002):

(1) on the one hand, the consideration of regression to the mean (RTM); and, on the other,

(2) the lack of data when there is an insufficient historical period or with a very low number of occurrences.

Regression to the mean is a common bias when evaluating a network in terms of accident rate or safety, since a point or element in the network can have high occurrence frequencies in a year and, nevertheless, can present a frequency of occurrences smaller and more characteristic the following year. The EB method will allow a better estimation of the safety of a part of the air transport system, taking into account not only the number of safety occurrences at that location, but also the occurrences observed in similar environments, naturally incorporating the knowledge of the experts on the causes that could have produced them (Hauer, 2002).

However, although there are good applications of BN in ATM (Neil, 2003) (Gomez Comendador, Arnaldo Valdés, Villegas Diaz, Puntero Parla, & Zheng, 2019) (Bujor, 2016) (Chen F., 2012), its potential to explain and predict the occurrence of safety events as SMIs has not yet been assessed.

To better understand the potential of Bayesian Networks and to sustain the complex model proposed in this document, the following section provides a first outlook on the main concepts involved in the construction of a BN and how the BN outcomes can be exploited.

3 Principles of Bayesian Network Analysis

3.1 Definition of Bayesian Networks

Bayesian networks are graphical representations that constitute directed acyclic graphs. A graph is a set of nodes and edges (or arcs); acyclic means that this set is linear or open, not circular; and directed that it has a unique direction, which marks the arcs. In the network, nodes are random variables; and arcs represent the direct dependency relationship between variables.

The structure of the network gives information on the relationships between variables, which can be cause-effect relationships. If there is an arc from node X to node Y, X is said to be the parent of Y. The network also represents the conditional independence between variables; in this case, given the parents of a variable, the child is independent of the rest of the nodes in the network.

These networks are based on Bayes' theorem and Bayesian inference. Bayes' theorem calculates the probability of an event A under the condition of another event B, so that the probability of A varies according to whether the event B occurs or not. The a priori probability of A is belief, and the event B is evidence. Bayesian inference makes use of Bayes' theorem and is the process of updating beliefs when evidence is known. Evidence can come from the data obtained or from the knowledge of an expert. This modifies the initial assumptions and results in posterior probabilities.

The graph gives a lot of information about the structure of the network, but not about its numerical properties. Therefore, it is necessary to construct conditional probability tables associated with nodes. The information needed to build the network, once the parameters and the connections between them have been identified, is as follows:

- The a priori probability of nodes without parents
- The conditional probability of nodes having parents

To define the probabilities, one must follow the ancestral order of the graph, knowing which values the parents take from a given parameter, and then the value of the child given his parents. With these data, the following information can be obtained:

- The a priori probability of a child node
- The posterior probability of any node given the observed evidences

When a piece of evidence is introduced into the network, the information travels both upwards and downwards and the probabilities of the other nodes are updated. The model can be fed with both data and expert knowledge.

3.2 Mathematical Foundations of Bayesian Networks: Bayes Theorem

Bayesian networks are based on the conditional probability, in particular the Bayesian theorem and the Bayesian inference. Conditional probability is one of the main ideas of probability theory. The concept of probability considers that the only information is the sample space, however, in the concept of conditional probability there is additional information, with which the probabilities change.

The conditional probability is the probability that an event A occurs knowing that an event B occurs. It is given by the following expression:

$$P(A|\mathrm{B}) = \frac{P(A \cap B)}{P(B)} \tag{1}$$

From the above expression one deduces the joint probability, that is, the probability of the intersection:

$$P(A \cap B) = \mathrm{P(A|B) \cdot P(B)} = \mathrm{P(B|A) \cdot P(A)} \tag{2}$$

For any set of random variables, this expression can be generalized to calculate the joint probability from the conditional probabilities, using the chain rule.

Given n events, A_1, ..., A_n, it is verified:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1|A_2 \cap \dots \cap A_n) \cdot P(A_2|A_3 \cap \dots \cap A_n) \cdot \dots \cdot \mathrm{P}(A_n) \tag{3}$$

Equation (3) is the rule of multiplication, which is to be rewritten as a production:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = \prod_{i=1}^{n} P(A_i|A_{i+1} \cap \dots \cap A_n) \tag{4}$$

Each joint probability distribution of n random variables can be factored into n! different forms, and is the product of the probability distributions of each variable conditioned on other variables.

For example, in equation (2) there are two variables so that the two forms that appear can be factored. Three variables could be factored into 3! different shapes.

Continuing with the concept of conditional probability, the total probability theorem and Bayes' theorem are defined: Let n events be A1, ..., An, which are disjoint two by two, i.e., the intersection of these is the empty set, and the union of n events is the sample space; and given an event B of which the conditioned probabilities are known $P(B|A_i)$, the probability of the event B is given by:

$$P(B) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i) \tag{5}$$

And Bayes' theorem is given by the following expression:

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{P(B)} = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)} \quad \textbf{(6)}$$

Where $P(A_i)$ is the a priori probability, $P(B|A_i)$ is the conditional probability, P(B) is the probability of observing B, the marginal probability, and $P(A_i|B)$ is the posterior probability.

Having made these definitions and explained the Bayes' theorem, a Bayesian network is defined:

Given an acyclic graph directed G and a set of variables A, A is a Bayesian network of G if it meets that the joint probability is the product of the individual probability of a variable given by its parents. This is written as follows:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = \prod_{i=1}^{n} P(A_i|Pa(A_i)) \quad \textbf{(7)}$$

The dissimilarity between the expressions (7) and (4) is the conditional independence of the variables of any node other than its parent, given its parents. That is, it is the chain rule simplified given a series of conditional independence relationships. (Sucar, 2015)

To better understand the concept, a Bayesian network constituted by two events will be seen. These are:

- Event A: To be born in Winter
- Event B: To be born in December

Figure 1 represents the relationships considered in this case between the two events.

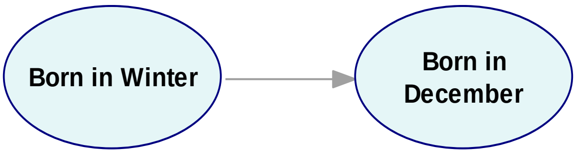


Figure 1: Example of a simple Bayesian network

As it has been explained before, it is necessary to fill in the a priori probability of the parent nodes and the conditional probability of the children.

- To be born in winter: Considering that the winter lasts 90 days and that the year has 365 days, one obtains:
  - A priori probability that a person was born in winter is:

P (to be born in winter) = 90/365 = 0.2466

  - A priori probability of not being born in winter: this is the complementary case:

P (not to be born in winter) = 1- 0.24657 = 0.75343

Next, the conditional probability of the child node:

- To be born in December: Considering that winter has 90 days, 11 of them are from December, and the remaining 20 days of December belong to autumn, the following probabilities are obtained:
    - The probability that a person was born in December knowing he was born in winter:

      P (to be born in December|born in winter): 11/90 = 0.1222

    - The probability that a person was not born in December knowing he was born in winter: this is complementary case to the previous one:

      P (not to be born in December|born in winter): 1 – 0.1222 = 0.8777

    - The probability that a person was born in December knowing that he was not born in winter:

      P (to be born in December|not born in winter): 20/275 = 0.072727

    - The probability that a person was not born in December, knowing that he was not born in winter, is complementary to the previous one:

      P (not to be born in December|not born in winter): 1 - 0.072727 = 0.9272727

With the results obtained from these calculations, the conditional probability table for the child node can be completed, as can be seen in Figure 2:

Born in Winter	Yes	No
Yes	0.1222	0.07273
No	0.8778	0.92727

Figure 2: Conditional Probability Table for the node Born in December

The a priori probability of being born in December is calculated with the equation (5)

P (to be born in December) = 0.1222 · 0.2465 + 0.07272 · 0.7534246 =0.0849

It shows what intuition would have said. If the year has 365 days and December 31, the a priori probability of being born in December is 31/365 = 0.0849. This results is also obtained by the software GeNIe, as can be seen in Figure 3:

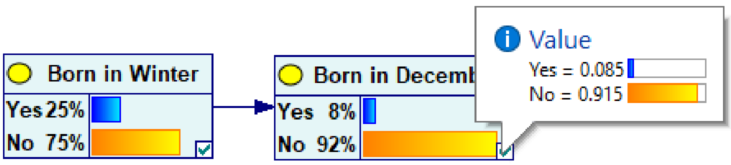


Figure 3: A priori probability for the nodes

Given the evidence, and making use of the Bayes' theorem, one can know the subsequent probability of the rest of variables. Suppose the evidence given is that one was born in December. The posterior probability of an event A (to be born in winter) is calculated with equation (6) and the following is obtained:

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)} = \frac{0.2466 \cdot 0.1222}{0.0849} = 0.3545$$

In Figure 4 it can be seen the same result graphically:

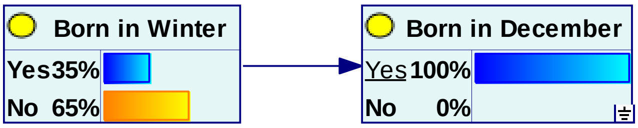


Figure 4: Conditional probability that a person was born in winter having evidence

This example with two nodes can be extended to larger networks. However, the number of values that must be completed in the probability table increases exponentially with the number of parents of a node. For this reason, to work efficiently with Bayesian networks in more complex cases, such as this safety study, it is necessary to use computational tools.

3.3 Methods of constructing Bayesian networks

As mentioned in the previous sections, a Bayesian network can be built from expert knowledge, from real data, or as a mixture of both.

This section will explain how a network should be created depending on the method used.

- **Experts' knowledge:** The experts have to decide the variables to include in the model and establish the causal relationships between them. They also have to complete the conditional probability tables for each child node.

- **Real Data:** In the case of the construction of Bayesian networks directly from data, the causal relationships will be extracted from these data, as well as the conditional probability tables.

- **Mixed case:** In this case, causal relationships created directly from the data can be modified by adding, removing, or changing the directions of the arcs.

In the case that concerns in this document, for the study to be carried out, the method chosen is the mixed one.

3.4 Bayesian Network Construction

The creation of a Bayesian network from a database and using a software must be done in three steps shown in Figure 5:

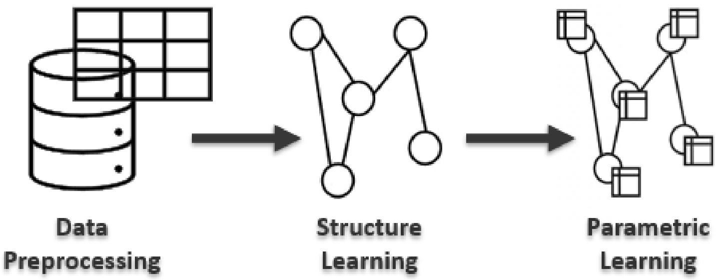


Figure 5: Steps for the creation of a Bayesian Network

- **Data preprocessing**: In this step, the variables to be used to model the problem are selected. A data cleaning is carried out to detect possible failures and correct them, if necessary. This step is essential, because if data are not correct or have a fault, the learning of the network will not be optimal, and therefore the results obtained from them will not be valid.
- **Structure Learning:** In this step, the network structure is determined, that is, the dependency and independence relationships between the variables are established. This learning can be done directly from the data provided, or as in this study, prior knowledge can be introduced into the model.
- **Parametric Learning:** The last step consists of obtaining the required a priori and conditional probabilities, given the structure previously defined. These probabilities are obtained from the observed frequency of the data.

3.5 Example of construction of a Bayesian Network

In this section, a Bayesian network will be built using the mixed method, that is, by combining data and expert's knowledge. This network is about the severity of lung cancer. It is not related to the study case of this document, but it has been considered a useful example to better understand what it is expected to be done. This network will be created from a database, combined with the experts' knowledge. The database has been obtained from a platform called Kaggle (Kaggle, s.f.). Kaggle is an online platform for conducting Data Mining contests, and it provides a repository for companies to publish their data. The dataset includes personal information on symptoms and risk factors of the disease for 1000 patients. The structure of this database can be seen in the Figure 6:

Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	OccuPational Hazards	Genetic Risk	chronic Lung Disease	Balanced Diet	Obesity	Smoking	Passive Smoker
P1	33	1	2	4	5	4	3	2	2	4	3	2
P10	17	1	3	1	5	3	4	2	2	2	2	4
P100	35	1	4	5	6	5	5	4	6	7	2	3
P1000	37	1	7	7	7	7	6	7	7	7	7	7
P101	46	1	6	8	7	7	7	6	7	7	8	7
P102	35	1	4	5	6	5	5	4	6	7	2	3
P103	52	2	2	4	5	4	3	2	2	4	3	2
P104	28	2	3	1	4	3	2	3	4	3	1	4
P105	35	2	4	5	6	5	6	5	5	5	6	6
P106	46	1	2	3	4	2	4	3	3	3	2	3
P107	44	1	6	7	7	7	7	6	7	7	7	8
P108	64	2	6	8	7	7	7	6	7	7	7	8
P109	39	2	4	5	6	6	5	4	6	6	6	6
P11	34	1	6	7	7	7	6	7	7	7	7	7

Figure 6: Structure of the database from Kaggle

The first step for building the network is the pre-processing of the data. In this case, it is not necessary to clean the data since no errors were found in the database. Next, the variables that are believed to be the most representative are selected to model the level of severity of lung cancer. The selected variables are: age, alcohol use, air pollution, smoking, obesity, genetic risk, coughing of blood, fatigue, and the output node, severity level. Finally, it will be necessary to make a discretisation of these variables. This consists of converting the continuous variables into variables grouped by intervals. This step is necessary since most algorithms are optimized for discrete variables. The discretisation of variables can be based either on statistical characterization or on expert knowledge. Discretisation should ensure that no information is lost or considered as an excess of states.

In Table 2, it can be seen the discretisation that has been done for the selected variables.

Table 2: Variable discretisation

Variable	States
Severity level	• Low: 1 • Medium:2 • High: 3
Age	• State 1: Below 28 • State 2: 28-45 • State 3: 45-60 • State 4: 60 up
Alcohol use **Air pollution** **Smoking**	• Low: 1, 2 and 3 • Medium: 4, 5 and 6 • High: 7 and 8
Obesity **Genetic Risk**	• Low: 1 and 2 • Medium: 3, 4 and 5 • High: 6 and 7
Coughing of blood **Fatigue**	• Low: 1, 2 and 3 • Medium: 4, 5 and 6 • High: 7, 8 and 9

For example, for the variable age, it has been discretised in 4 states that are: under 28 years old, between 28 and 45 years old, between 45 and 60 years old, and over 60 years old.

The next step is to obtain the structure for the network. This structure is learned in the first instance directly from the data, using GeNIe software, with an algorithm called Bayesian Search. The Bayesian Search structure learning algorithm is one of the earliest and the most popular algorithms used. It was introduced by (Cooper, 1992) and was refined by (Heckerman, 1995). It essentially follows a hill climbing procedure (guided by a scoring heuristic, which in GeNIe is the log-likelihood function) with random restarts.

This structure is shown in Figure 7:

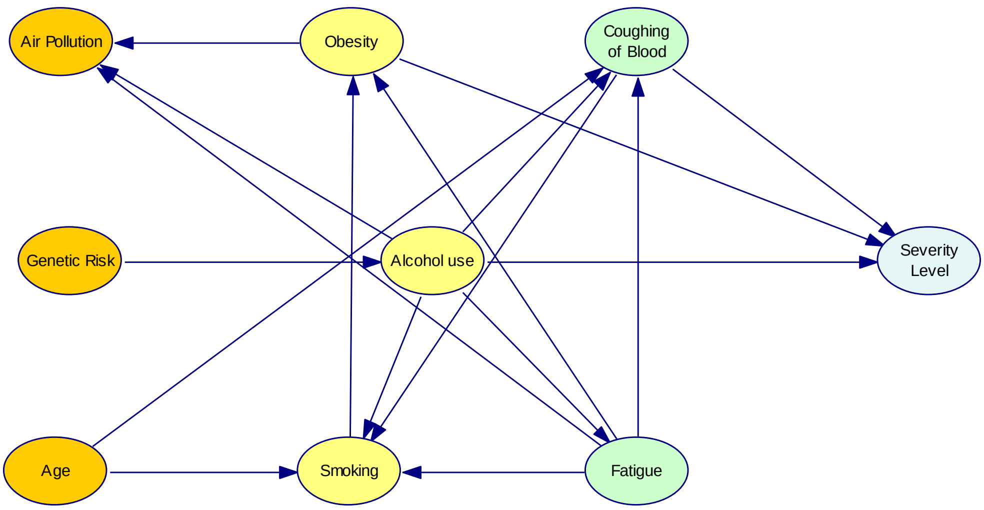


Figure 7: Structure of the Bayesian Network directly obtained from data

When analysing the relationships created by the algorithm, some of them that do not make sense are detected. For example, the variables shown in orange are external factors. Therefore, it is not logical that these factors are children of other nodes.

It is at this time when the knowledge of the experts plays an essential role. They must decide whether to add, remove, or change the directions of the causal relationships created directly from data.

Links that have been deemed meaningless, and that should be removed or changed direction are shown in Figure 8:

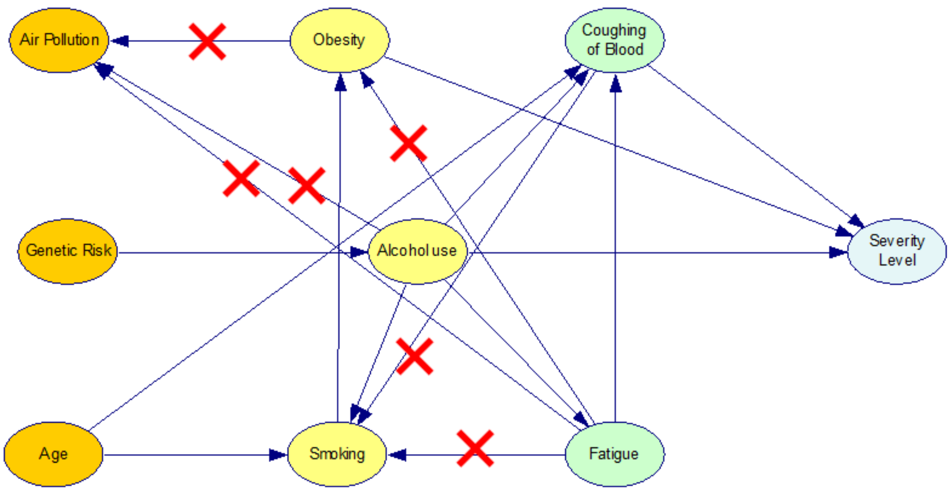


Figure 8: Combination of expert knowledge and the data provided

Once these decisions have been made, the final structure of the network is obtained, which is the result of a combination of expert knowledge and the data provided. This final structure is shown in Figure 9:

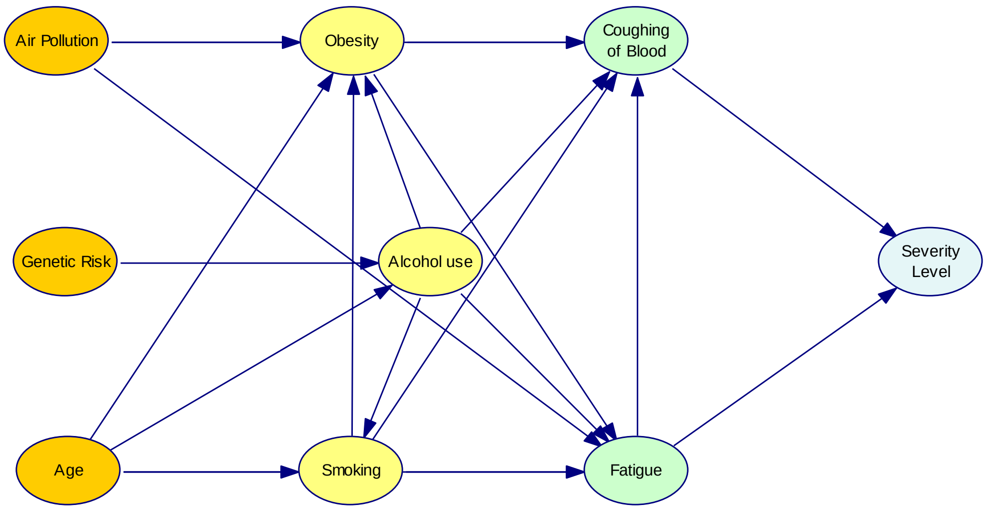


Figure 9: Network structure created as a combination of data and expert knowledge

The next step is parametric learning. As already mentioned, it consists of obtaining the a-priori probabilities of the parent nodes and the conditional probabilities of the child nodes.

In Figure 10, it can be seen that these probabilities are obtained directly from the frequency of the data for each of the states of the variables.

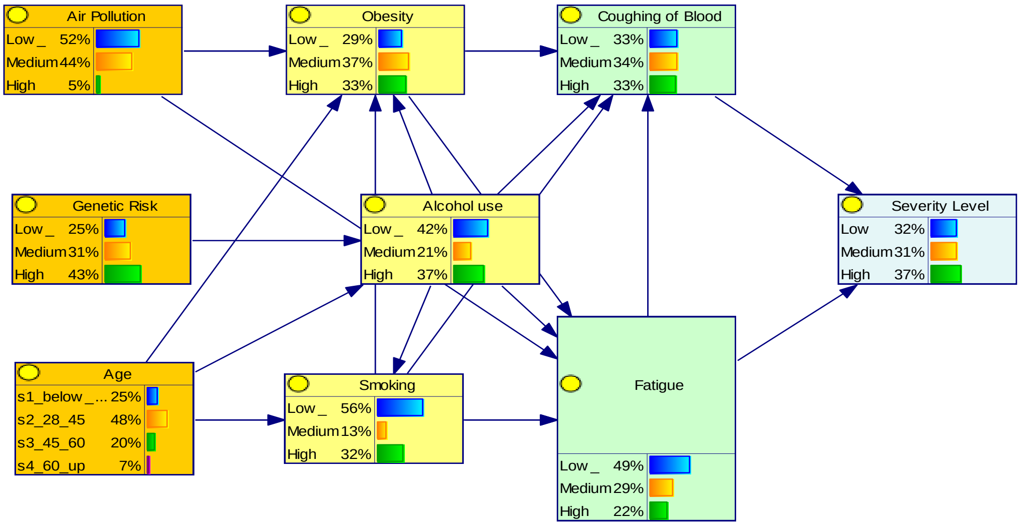


Figure 10: Parametric learning for the Bayesian network

In turn, in this step, the conditional probability tables for the child nodes are also obtained. The Figure 11 shows the conditional probability table for the output node of this network (the severity level).

Coughing of Blood	Low_			Medium			High		
Fatigue	Low_	Medium	High	Low_	Medium	High	Low_	Medium	High
Low	0.62229807	0.16666667	0.03030303	0.81547619	0.0625	0.33333333	0.076923077	0.0033670034	0.0023980815
Medium	0.37656428	0.82795699	0.03030303	0.18154762	0.75625	0.33333333	0.23076923	0.0033670034	0.0023980815
High	0.0011376564	0.0053763441	0.93939394	0.0029761905	0.18125	0.33333333	0.69230769	0.99326599	0.99520384

Figure 11: Conditional Probability Table for the node Severity Level

The conditional probability tables are the results that it is wanted to obtain from the training of the network, since they will help to carry out the subsequent analysis.

Conditional probability tables grow exponentially with the number of parents of a node. For this reason, if the table is to be populated with knowledge of a domain expert then the magnitude of the task constitutes a considerable cognitive barrier.

Once the network is created, the corresponding analysis can be carried out on it.

3.6 Analysis carried out on the network

The analyses that will be carried out with the BN are: sensitivity analysis, backward analysis and forward analysis.

- **Sensitivity analysis:** Sensitivity analysis is used to investigate the effect of small changes in numerical parameters (i.e., prior probability) on the output parameters (e.g., posterior probabilities). Highly sensitive parameters affect the reasoning results more significantly. Identifying them allows for a directed allocation of effort to obtain accurate results of a Bayesian network model. GeNIe implements an algorithm proposed by Kjaerulff and van der Gaag (Kjaerulff, 2000) that performs simple sensitivity analyses in Bayesian networks. Roughly speaking, given a set of target nodes, the algorithm calculates efficiently a complete set of derivatives of the posterior probability distributions over the target nodes over each of the numerical parameters of the Bayesian network. These derivatives indicate importance of precision of network numerical parameters for calculating the posterior probabilities of the targets. If the derivative is large for a parameter p, then a small deviation in p may lead to a large difference in the posterior of the targets. If the derivative is small, then even large deviations in the parameter make little difference in the posterior. The results of the sensitivity analysis are presented graphically as a scale of red tones. The colouring of the individual elements of the definition shows those individual parameters that are important.

- **Backward analysis:** The model is used to deliver a particular configuration of the parent nodes by setting the outcome node (uncertainty level of the severity level) to a target value. In this analysis, the severity level is settled to a high, medium, or low value. Then, the network provides understanding about the main contributors to severity level uncertainty, or what configuration of uncertainty might be admitted in the various input variables to provide the target outcome uncertainty. This case study is useful to answer the following questions: (1) how much will it be necessary to improve uncertainty in the input nodes to achieve a certain uncertainty level in the outcome node?; or (2) what will be the probability of any fault (uncertainty level of the input nodes) given a set of symptoms or results (uncertainty level of the outcome)? This is a typical fault diagnosis scenario.

- **Forward analysis:** The model is used to predict the effects, that is, the uncertainty level in the severity level (output-child node) by setting the probability distribution of the parent-input nodes. This case study is useful to answer the following research question: Given the probability distribution of the uncertainty of the various input nodes, how these uncertainties propagate through the network causes a probability distribution for the uncertainty (% of high uncertainty, % of medium uncertainty or % of low uncertainty) in the outcome of the network, "the severity level"? This is a typical prediction scenario.

The results of these three analyses are presented below for the lung cancer example.

In the sensitivity analysis, the node "severity level" has been marked as the target node. Sensitivity analysis is used to detect highly sensitive parameters for the target node. Identifying these parameters allows to focus our efforts on these variables.

The software shows, on a scale of red tones, the most sensitive parameters to the target node as can be seen in Figure 12. For this particular example, they would be Air pollution and Genetic Risk.

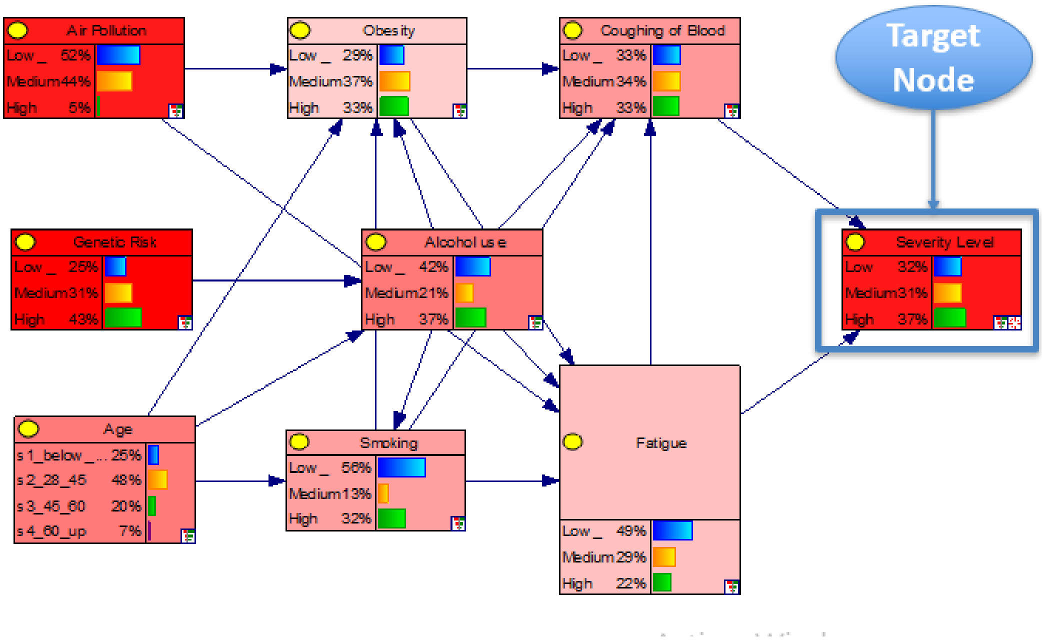


Figure 12: Sensitivity analysis

The Backward analysis is used to deliver a particular configuration of the parent nodes by setting the output node (Severity Level) to a target value. For this case, a severity level of low in a 100% is settled and a configuration of parent nodes to achieve this target is obtained.

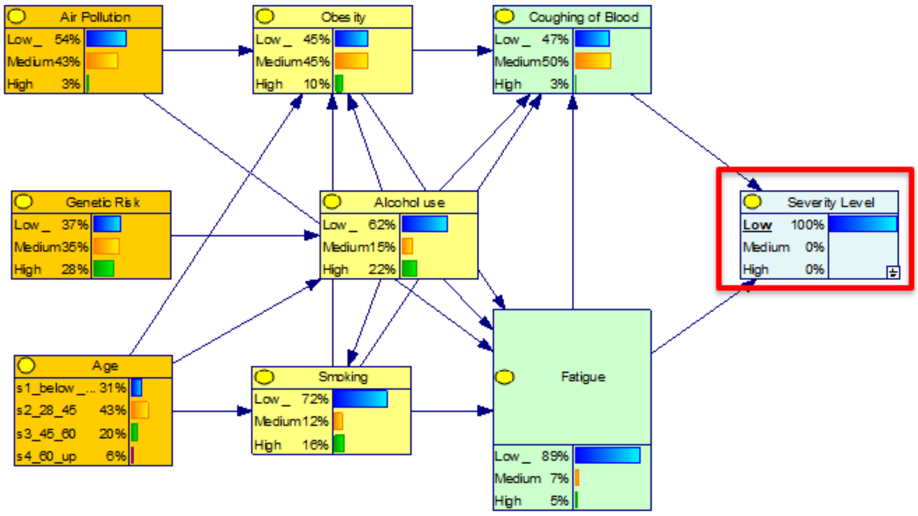


Figure 13: Backward Analysis

In Figure 13, it can be seen how the probabilities of the states have changed. For example: the low state of fatigue has gone from 49% to 89%.

The Forward Analysis shown in Figure 14 is used to obtain the probabilities of the output node. Imagine that there is evidence about some factors of a particular patient, for example, it is known that he lives in a zone with high air pollution, his genetic risk is low, he is 65 years old, he is obese, he drinks a lot of alcohol and smokes a lot. With these evidences is possible to predict the severity level of the patient.

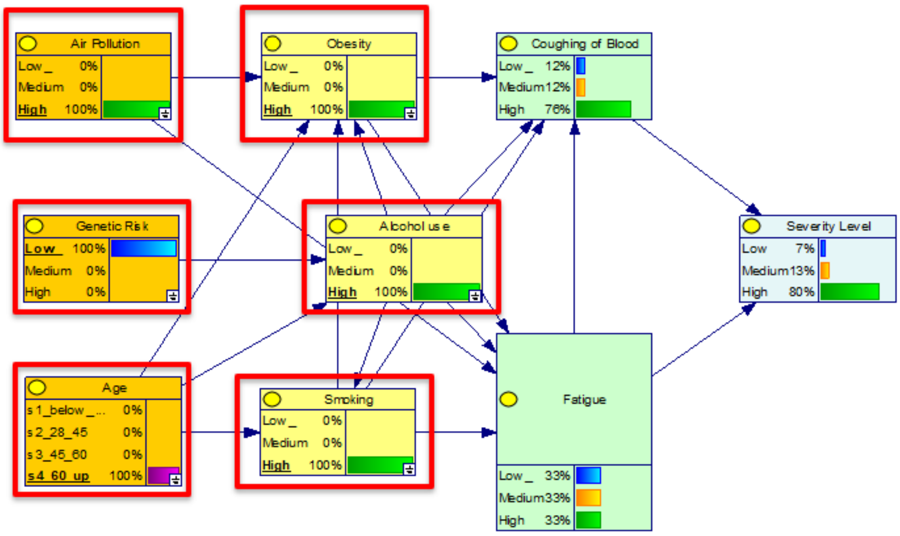


Figure 14: Forward Analysis

3.7 Main reasons for selecting Bayesian network methods

Once it has been seen with the practical example how Bayesian Networks work and the various studies that can be conducted on them, the reasons why it was decided to use Bayesian networks in for fulfilling FARO O2. This is because any change in the system (operations, procedures, etc.), can reflect a change in the prior / posterior distributions, and the cause - effects can be analysed. The main reasons why Bayesian Networks have been selected are explained below:

- Bayesian Networks are very useful for capturing and analysing causality and influence relationships. They are very effective at diffusing uncertainty and updating systems with new data. They are also applicable when the structure of the system is too complex. They provide an intuitive and efficient way to represent a considerable field, making complex systems modelling feasible.

- Bayesian Networks are mainly used to update the probability distribution of the states of hypothetical variables (variables that cannot be observed directly). This probability distribution helps decision makers to determine the appropriate course of action.

- Bayesian Networks provide a convenient and consistent way to express uncertainty in uncertainty models and are increasingly being used to express knowledge of uncertainty. They are used for qualitative and quantitative modelling of uncertainty and its causes.

- Due to the conditional dependence of variables in the network, BN provides the ability to predict or diagnose (i.e., they can determine impact and causes). Bayesian Networks are used to model multidirectional uncertainties forward and backward.

Bayesian Networks can perform qualitative *cause* and *effect* evaluations and can quantitatively update the probability distribution of unobservable variables.

Qualitative analysis: Given a scene, the Bayesian Network graphically represents the causal relationship between the various elements of the scene.

Quantitative analysis: update the probability distribution. Given the hypothetical variables representing possible actions and the prior probability distribution, the Bayesian Network provides the function of updating this probability distribution when new data and information are acquired.

4 Data Management

This section discusses the main data management tasks performed. The process of identifying the key information to perform the model is presented firstly, followed by the data exploratory analysis conducted to examine the characteristics of the available data, and finalising with data filtering and data transformation.

As specified in the previous sections, the aim of this work is to define the ATM Safety Performance Functions by using organisational, technical, human, and procedural precursors to characterise and predict airspace separation minima infringement as a function of those precursors.

The identification of those precursors, as well as the data and variables to support its characterization and quantification, will determine the optimal set of information that it would be desirable for the model.

In this section, this optimal set of information is analysed from a knowledge-based perspective. Unfortunately, not all the necessary data will be available for use in the model.

First of all, a brainstorming exercise was carried out between all partners with the aim of identifying all possible variables that could affect the occurrence of a violation of the minimum separation distance between en-route aircraft (FARO project, 2021; FARO project, 2020). These variables were then grouped into different causal factors, and these causal factors were in turn grouped into different analysis areas, to establish a general and more visual framework of the information required. The following 12 analysis areas were obtained in a first attempt:

- Traffic demand
- Airspace
- Organization and Management of Human Resources
- Human Resources
- ATFCM Regulations
- Planning Compliance
- Operations
- Potential Conflict
- Safety Management
- Economic Management and Results
- Aeronautical Information
- CNS / ATM Systems

In parallel, indicators were identified for each of these variables, with the objective of converting them into quantifiable variables. Once all this information was available, a selection process had to be carried out to decide which of them were the most realistic and achievable. All of this process was set out in the Progress Meeting of 30th November 2020 and is compiled in the FARO data description document.

In table 3, it can be viewed a fragment of the aforementioned document, which includes the traffic demand analysis area, the causal factors into which it is divided, and the parameters that make it up, as an example.

Table 3: Fragment of the FARO data description document.

Analysis area	Potential Precursors or causal factors	Parameter to characterise any event
Traffic demand	Temporal distribution of demand	Instantaneous demand of a sector / Instantenous workload measures or traffic loading
		Total volume hourly demand
		Variations in peaks of traffic
		A/c in sector above the declared capacity
		Seasonality of traffic
	Traffic density	Hourly entry counts
		Traffic complexity
		Distribution of flight time per aircraft under ATCO responsibility in the given timeframe
		Percentage of flights out of standard flows
		Number of AC passing through a node
		Traffic Flow density
		Distribution/dispersion of traffic in volume
		En-route flows
		Ascent flows
		Descent flows
		Ascent-descent flows
	Traffic mix	Altitude AC changes
		Altitude AC distribution
		Speed AC distribution
	Traffic Interaction	Number of interactions
		Time difference at crossing points
		ASMT or SMF output for all interactions recorded for the sector. Height/level to be determined
		Vertical and horizontal convergence (diverging, constant or converging)

However, data limitations were also encountered that prevented access to the information that would ideally be desired in this first approximation. These limitations were due to the impossibility of obtaining part of the information, or because it was felt that the effort involved in obtaining the variable was not worth the information it would provide. Other limitations were the difficulty of quantifying some variables or their loss of relevance because they could not be compared with data from other Air Navigation Service Providers (ANSP) than ENAIRE. For these reasons, the following analysis areas and their related variables were finally discarded:

- CNS / ATM Systems
- Aeronautical Information
- Safety Management
- Economic Management and Results

4.1 Exploratory analysis of available data

After being provided with the available data by the ENAIRE-CRIDA data warehouse (DWH), the next step was to pre-visualize the data and run an exploratory analysis.

In this context, the first aim was to understand the structure that was followed in each type of file, so that was possible to know what information was included in each one.

The content and structure of each type of file provided by the ENAIRE-CRIDA data warehouse (DWH) is detailed below (GARCÍA, GARCÍA-OVIES, VERDONK, & GARCÍA, 2020).

- **Flights:** This file gathers the invariant information related to the dimension Flight such as the flight key or the origin and destination airport. It also includes information about the aircraft model and the flight rules, among others, despite being all of them parameters not used in the analyses performed. It includes two relevant variables for each flight key, the cruise speed and the cruise flight level. Both were taken as reference values when comparing them to the speed and flight level of each flight at any point in its trajectory, being very useful parameters in several stages of the analyses conducted**.**

- **Tracks:** This kind of file is one of the most relevant, as it was very useful in all stages of the developed study. The file contains information about the aircraft state vector. It includes trajectory information such as latitude, longitude, and flight level of all the flights analysed, at intervals of time of 5 seconds. This means that it has been possible to locate each flight at any moment in the Spanish airspace.

  The heading of each aircraft, its speed in the three axes of space (X, Y, Z) and its speed module each 5 seconds are also available. Therefore, apart from the location, information about the motion of each aircraft during its flight could be also used.

  Studying the location and motion of each aircraft in the precise moments such as the instance of CPA (Closest Point of Approach), the instance an aircraft enters or exits a sector, or the instance when an ATC instruction was issued has been possible thanks to the information in this file. Knowing the conditions of each aircraft was crucial in all stages of the analyses performed.

- **Sector Entry:** This file was expected to provide the geographical transition of flights between sectors, that is, the times and points of entry and exit from ATC sectors for each aircraft. The purpose of using them was initially to access the exact time each flight enters the ATC sectors, as well as the time it leaves the sector. In addition, location and motion variables were included associated to each register, allowing to know all these aircrafts characteristics at the exact time they enter/leave each sector.

  However, data in the Sector Entry files were not finally fed into the model because several records were duplicated or included inaccurate information, some flights were not registered in the file, and some difficulties were encountered to locate flights coming from outside the Spanish airspace.

- **LoS:** This file provides information for each aircraft pair whose separation was registered to be below the separation minima during at least an instance of time. Loss of Separation Minima is defined as less than 5NM and 1000 ft. It includes the vertical, horizontal and 3D separation during the duration of the SMI, the sector where it happened and the coordinates of the point where the Closest Point of Approach took place, (longitude, latitude and flight level).

- **ATC Event:** It consists of the register of Sector Tactical Controller actions or clearances on the ATCo working position. Actions are identified as events associated with a flight and to the ATC working position involved. ATCo actions are manually entered by the controller in the system. The file records the date and time when the controller records each action.

  This file has been very useful to recognise which clearances have been issued to each aircraft. However, it was not possible to associate clearances to an aircraft pair, just to the flight that received the clearance. This made analysis of conflict solving strategies difficult.

  Apart from this, the information in this file has also been used to derive sector characteristics, ATCo workload and the association between ATCo working positions and ATC sectors based on counting techniques.

- **Separation:** This file filters all pairs of aircraft that, at any given time, have been at a distance equal to or less than 20NM. This limit of 20NM was determined by CRIDA based on previous work, considering technical needs for the development of the Automatic Safety and Monitoring Tool of ENAIRE.

  For every one of these pairs of aircraft, vertical (ft), horizontal (NM), and 3D (NM) separations are provided, as well as the latitude, longitude, and flight level of each aircraft. The file includes a record every 5 seconds.

  This file has been the key to knowing, at any time, the separation between the aircraft involved in a possible conflict. It gives the opportunity to assess how a conflict evolves and provide evidence of its resolution. This file has been used repeatedly.

- **STCA Alert**: A short-term conflict alert is a ground-based safety net intended to assist the controller in preventing SMI between aircraft by generating an alert of a potential infringement of separation minima**.**

  This file provides the log of STCA alerts detected in the system. Each STCA is associated with an individual aircraft, so some data processing is necessary to identify the aircraft pairs affected by each STCA.

  Apart from the flight key to which the STCA alert is related, its location information (latitude, longitude, flight level and sector), and the exact time in which the STCA alert has started and finished are also provided.

- **Route:** The information provided in this file consists of the definition of the route followed by the aircraft in each sector of the Spanish airspace. The time of entry and exit in each sector is also included. However, these files did not include all flights. These discrepancies discouraged the use of this file.

4.2 Data filtering

A prefiltering of all original data files has been developed to reduce the size of the files to work with. As only a few sectors relative to the study cases selected are analysed, those flights that have been found at some point in the sectors of interest are filtered, discarding all additional information of no interest. This process has been necessary due to the large size of the files available initially. Many of these original files include irrelevant information that has therefore been removed.

The filter was not a straightforward process and has involved the development of an elaborate heuristic. Despite not being useful for the processes followed later, the Sector Entry files were used to identify all flights whose trajectories have crossed the sector of interest at some point. Their flight's keys were stored and then used to obtain the information of their complete trajectory, contained in all other available files.

For Tracks and Separations files, due to their large size, it has been necessary to create a code to automatically import long lists of files, filter them by flight code, and store them in separate files, one for each day of collected data.

4.3 Data processing

The models developed in the project are sector specific. It has been developed one model for LECMSAN, one for LECBCCC and one for LECBCCU. Each model considers only the data of flights in its corresponding sector.

The elementary piece of analysis in the data is every aircraft pair closer than 20NM. The different variables considered in the model are assessed for each aircraft pair.

We have processed around 40.000 pairs of aircraft per sector, 120.000 in total. The time frame considered in the analysis is one and a half years. Data correspond to 80 days spread between one and a half years. Those days were selected by CRIDA as on these days there were SMIs.

It has been necessary to generate a specific ad hoc file for the training of each subnetwork included in the model, as well as for each scenario and type of analyses. Therefore, the objective of the data processing has been to generate each single file needed in each stage of the model training, with the required combination of variables for all flight keys considered in each case, and in the needed formats.

In this section, some of the tasks conducted to process the data and generate all files needed for the different stages of the modelling are summarised. It is focused on the most complex tasks that have been developed through data processing, showing some of the methodologies followed and the difficulties that have been faced in each stage of the project.

- **Definition of entries and exits in sectors:** The information on the Sector Entry files was not entirely reliable, so it had to be recalculated from the basic radar flight trajectories.

Sector entry time, exit time and its corresponding coordinates were recalculated by crossing every aircraft trajectory with the geographical boundary of each sector. Every trajectory was correlated with the actual entry point at the sector boundary. As the aircraft positions were provided every 5 seconds, their trajectories were drawn together with the coordinates of the geographical limits of the sectors of interest, obtaining the entry and exit instances for each aircraft, those associated with the closest points to the limits of the sectors in the drawn plan.

- **Aircraft conditions in a given instance.** Another data transformation that required a complex process was assigning specific characteristics of each aircraft in terms of position and motion at specific times. This exercise has been necessary in multiple parts of the analyses to generate specific training files. For a given time instance and flight key it has been necessary to obtain the flight and position characteristics of each aircraft at that moment, merging data from different files has been of great importance

Aircraft trajectories and flight parameters were provided each 5 seconds. However, the time stamps in the different files did not always match. Time differences between 1 and 4 seconds sometimes make time interpolation necessary.

With the aim of dealing with this difficulty, the merging of data from different files was made using time intervals instead of single time stamps. That has added a certain complexity to the calculations.

For this purpose, the data of the Tracks files has been sorted by flight key and time, generating time intervals between consecutive data rows of the same flight key. In this way, specific instances associated with a flight key in the defined time intervals have been inscribed, treating those specific instances as time intervals that start and finish at the same time instance.

- **ATC working position for each sector.** An additional calculation required was the identification of the Sector Tactical Controller working position (or air traffic controller) responsible for each sector in each hour of operation. This information was not available in the files and it has been necessary to calculate it indirectly to correctly attribute the ATC events.

Additionally, the time of entry or exit of the aircraft from the sector will not necessarily coincide with the time of the first and last action of the controller over the aircraft. The first ATC action over an aircraft may occur before or after it has crossed the geographical limits of the sector. It may be said that the geographical and operative boundaries of the sector are not necessarily the same. Transfer of control may be given before the aircraft has reached the defined transfer of control point. And also transfer of communication may take before the transfer of Communication or control points. Some of these are defined in the letter of agreement or standing agreements.

The ATC event files have been used for this task. A preliminary filtering saves the ATC actions on the flights of interest throughout their entire trajectory. For each ATC sector, a list is generated with all possible operational ATC working positions every hour. Possible operational positions are identified by filtering the ATC events file for each flight by its time of entry and exit in the ATC sector. The most probable ATC working operational position is selected as the one from which most ATC clearances are recorded each hour. An ATC sector can be associated with different ATC working positions depending on the configuration of the ATC control room.

The accuracy of this information determines the quality of certain derived parameters, such as the hourly count of ATCo clearances in each sector, or the time of the first and last ATCo clearance before the CPA.

- **Identification of STCAs.** The last data processing task was the identification of the STCAs and the association of this type of alerts to pairs of flights.

The system is supposed to register an STCA two minutes before the SMI. Although there will be at least two aircraft involved in an STCA, the alert is registered individually by the flight key. An STCA has been defined only if there are at least two flights for which an STCA is registered in the same time instance.

5 Model development methodology. Conceptual framework

The proposed methodology aims to derive a model able to characterise and predict the occurrence of separation minima infringements (SMIs) between en-route aircraft (level flight and climbing and descending as well). Traditional statistical approaches are not useful in this case due to the scarce number of SMIs. With a much reduced number of separation minima infringements in the total sample of flights, conventional approaches will not have statistical relevance. It has been opted for a technique with high predictive capacity, that allows integrating knowledge modelling with data inference, and have proven to be useful to estimate low probability events: **Bayesian Networks** (BN).

To develop a Bayesian Network model for such a complex problem as SMI prediction is not straightforward. It has been necessary to set up a conceptual framework that integrates the current available knowledge about SMIs causality and precursors with the hindsight derived from the analysis of the type of data available in the project, particularly those that reflect the ATCo interventions.

This section describes the overall methodology followed to develop this conceptual framework and the structure of the SPF model.

5.1 Initial approach to SPF: focus on the Closest Point of Approach-CPA

The conceptual framework that backs up the proposed BN model considers the general scenario where aircraft trajectories evolve and focuses on the analysis of the Closest Point of Approach (CPA), for any possible aircraft pair in an air traffic sector, and on the understanding and quantification of the process that leads to such CPA.

The three main elements in the conceptual framework will be considered. Figure 15 illustrates the interaction between two aircraft pairs in a sector and their respective CPAs, which are represented by a red circle. The actual final CPA between an aircraft pair can be interpreted as the outcome of a process where the expected aircraft trajectories become modified as the results of the ATCo clearance. Then the CPA between an aircraft pair may be considered as an aircraft pair as the actual shortest distance between those two aircraft, expressed as vertical separation and horizontal separation. This magnitude is called "final CPA". It can be also calculated what the CPA would have been between this aircraft pair if both had followed their planned trajectories without any modification or ATCo intervention. This magnitude is named "prior CPA". The difference between both magnitudes, final CPA and prior CPA, is attributed to alterations of the expected trajectory that are induced by ATCo intervention.

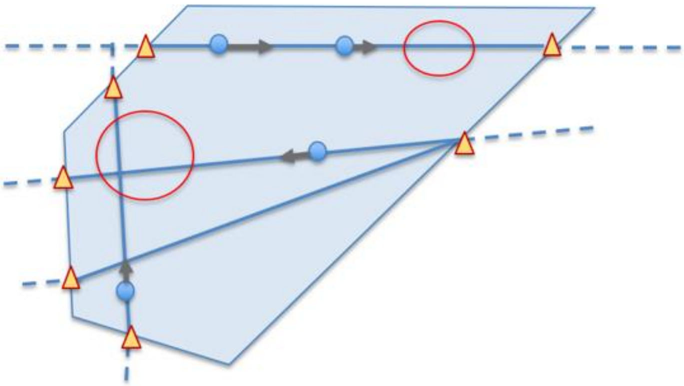


Figure 15: Representation of a general scenario and the CPAs of two interactions.

Based on this approach, the framework considers three main elements in the interaction of each aircraft pair within the area of responsibility of a specific air traffic controller, represented in Figure 16.

- **Distance Closest Point of Approach prior (dCPA prior):** This concept corresponds to the distance at which two aircraft would cross considering their planned trajectories. This distance will be measured horizontally in Nautical Miles (NM) and vertically in feet (ft).
- **Time of Last Clearance (TLC):** This term refers to the time elapsed since the last ATCo clearance to any of the aircraft in the pair and the instance in which the CPA occurs. This time will be measured in seconds.
- **Distance Closest Point of Approach final (dCPA final):** This last concept corresponds to the final shortest distance the aircraft pair cross after receiving a clearance from the controller. Units of measurement are the same as for dCPA prior.

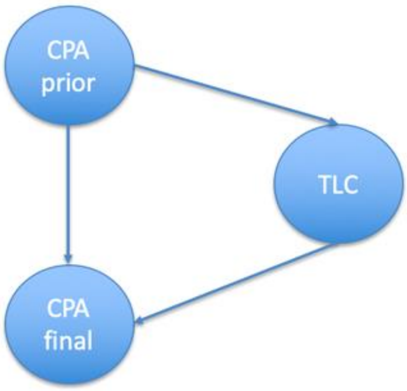


Figure 16: Relationship between the three subnets

Figure 17 shows a diagram with the concepts explained above. The two blue dots in the diagram correspond to an aircraft pair, Ai and Aj. In this figure, *the dCPA prior* is symbolised in red. *dCPA prior* corresponds to the distance at which the two aircraft would cross considering only their planned trajectories. The controller's task is to detect whether this *dCPA prior* could constitute a possible SMI between this aircraft pair and acting on them to prevent it. The distance between the aircraft at the time the controller issues the last clearance on them is shown in the diagram as *d at TLC*. The aircraft will eventually cross each other at a distance called *dCPA final*, which is represented in the Figure 17.

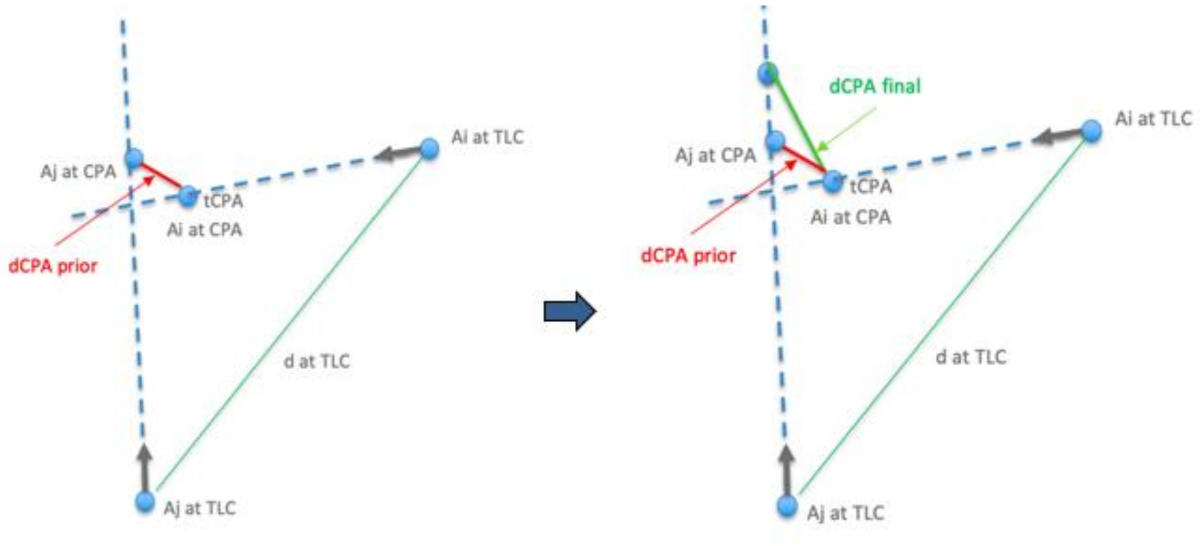


Figure 17: Comparison of the CPA without ATCo action and with ATCo action

Figure 18 illustrates the concept of the 3 subnetworks (CPA prior, TLC and CPA final), their precursors, and the causal interrelationships. In the diagram it can be seen that some precursors may have an influence on more than one network. In turn, it is seen that three networks are connected to each other, forming a single final model.

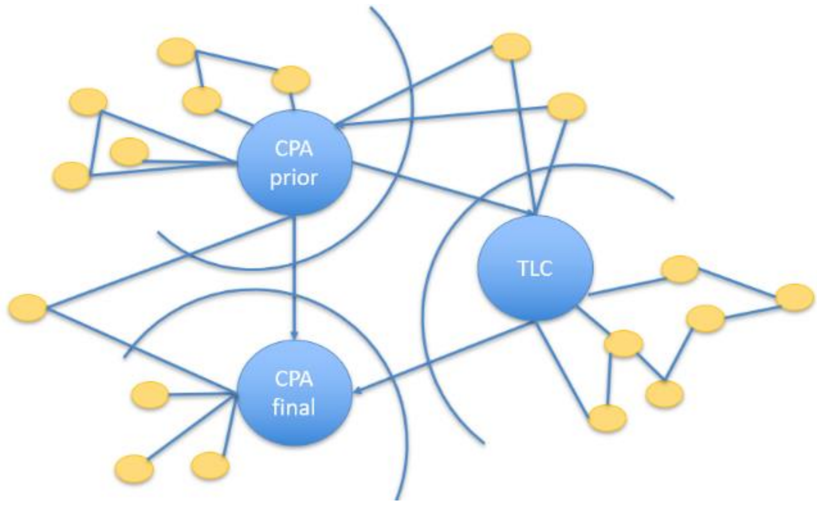
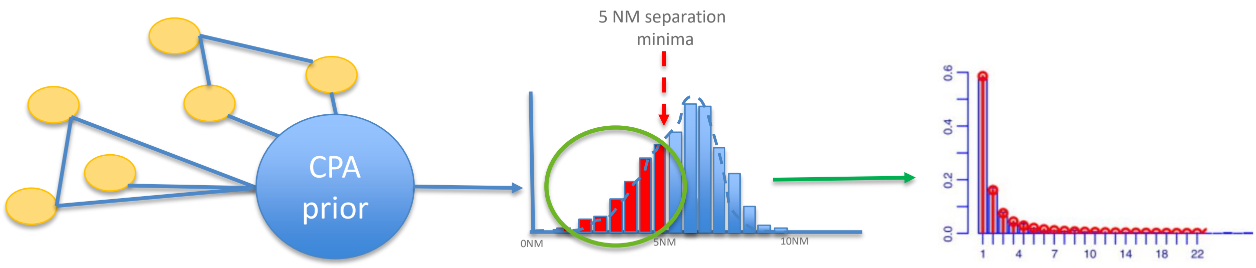


Figure 18: 3 subnetworks integrated into a single final model

- **CPA Prior.** The subnetwork indicated by the CPA prior bubble will estimate, based on a set of selected precursors, the vertical and horizontal separation probability distribution between any aircraft pair at their CPA prior, which is the predicted CPA between the two aircraft if they only follow their plan trajectories without any outside or ATCo intervention.

Possible precursors to be considered in this subnetwork might be those related to temporal distribution of demand, traffic density, flows and airspace structure, ATFCM Measures, changes in the Flight Plan, adherence to the trajectory, among others.

Figure 19 represents the outcome of the subnetwork. It shows the diagram of the separation probability distribution that could be obtained for all pairs of aircraft in an ATC sector considering their previous CPA. By comparing this separation distribution with the applicable separation minima, the probability of potential conflicts can be derived. In Figure 19, the red bars represent the aircraft pairs whose CPA prior distance is expected to be below the applicable en-route separation minima. By drawing a frequency diagram for just those cases, the graph on the right side of the figure represents the distribution of the expected number of "potential conflicts".



CPA prior subnet and its precursors

Probability distribution of distance between every pair of aircraft at the CPA prior.

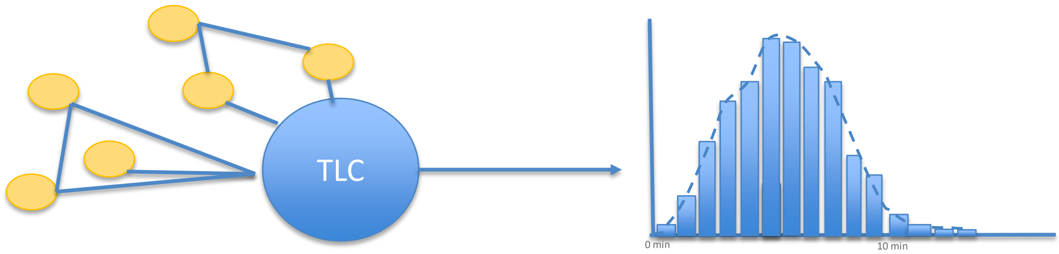
Number of potential conflicts

Figure 19: CPA Prior - Separation probability distribution - Potential conflict frequency diagram

- **TLC:** The subnetwork indicated by the TLC bubble accounts for the ATCo clearance for each aircraft pair in the sector and its precursors. The role of ATCos is to ensure a safe and efficient flow of air traffic in the airspace for which they have responsibility.

Possible precursors to be considered in this subnetwork might be those related to the performance of the controller at his workplace, specifically everything that influences the controllers' workload. These parameters will be those related to organization and management of human resources, human resources information, automation, complexity, operations, precursors related work shifts, among others.

Figure 20 represents the outcome of the subnetwork, indicated as the time distribution of the controller's clearances on the aircraft pairs.



TLC subnet and its precursors

Probability distribution of the ATCo's clearance on every pair of aircraft in the sector

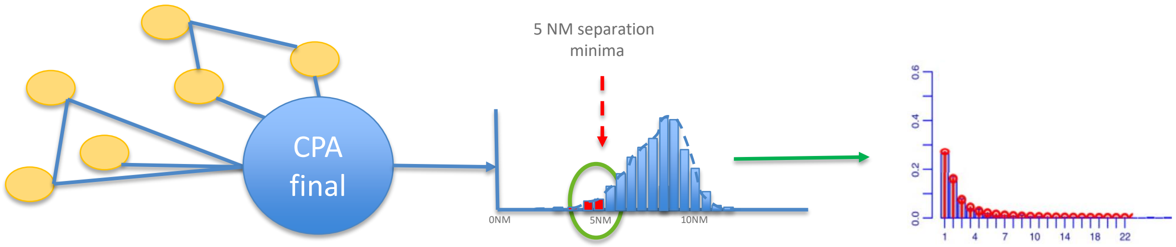
Figure 20: TLC – ATCo's clearance probability distribution

- **CPA Final.** The same concept applies to the subnetwork indicated by the CPA final bubble. The Closest Point of Approach Final refers to the final shorter distance at which the aircraft pairs cross after receiving the last clearance from the air traffic controller.

Possible precursors to be considered in this subnetwork might be those that could impair the application of operational solutions or the effectiveness of the ATCo clearances, as well as the reaction of the aircraft pilot.

In **Figure 21**, it can be seen how the distribution will shift to the right compared to the CPA prior figure, as there will be fewer aircraft violating the minimum separation.

The following figure represents the outcome of the subnetwork. It shows the diagram of the separation probability distribution that could be obtained for all pairs of aircraft in an ATC sector considering their final actual CPA. By comparing this separation with the applicable separation minima, the number of Separation Minima Infringements –SMIs- can be derived. In the figure, the red bars represent the aircraft pairs whose actual CPA turned to be below the applicable en-route separation minima. As controllers are the most effective ATC barrier, it is expected that CPA distances are safer after ATCo's clearance than before, and the number of resulting true SMIs are lower than the number of potential conflicts identified in the CPA prior.



CPA final subnet and its precursors

Probability distribution of distance between every pair of aircraft at the final actual CPA

Number of Airspace SMI

Figure 21: CPA Final: Actual Separation probability distribution – Airspace SMI frequency diagram

Finally, Figure 22 illustrates the integration of the three subnets and the interrelationship between their outcomes.

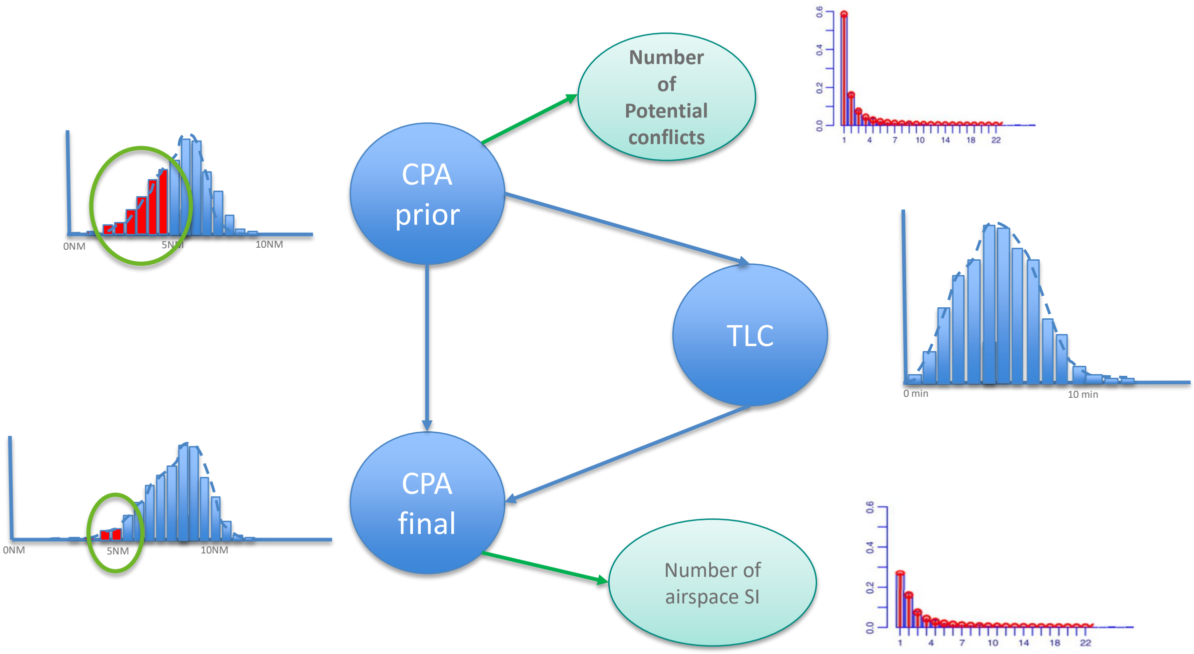


Figure 22: Model integration

To translate the conceptual framework into a set of causal subnetworks, the concepts of **ATM barrier model** and **event trees** have been incorporated. The following sections describe these elements of the model in detail.

5.2 ATM Barrier Model: An abstraction of the Aircraft Separation Provision function

The barrier model in Figure 23 is an adaptation of the one proposed by Eurocontro (Perrin & Kirwan, 2007) l, taken into account the knowledge and data driven approach followed in the project. It can be seen the main barriers identified, the areas of analysis to which they belong, the order in which they occur and the direction of the flow.

Based upon the application of the *Swiss cheese* model, the **ATM barrier model** explicitly presents the progression of a safety incident and can be used as a "live" model to prevent future breaches of separation or to intervene in an incident to stop its development. The ATM Barrier Model presented is an abstraction of the ATC separation provision function.

It divides the aircraft separation provision process into different stages where safety barriers are identified. In particular, it reflects stages and barriers in the progress of an SMI that could be studied and analysed with the data available in the project. To build this model for a specific incident, the analyst needs to identify the **barriers**, and then their failures.

Although it will be covered it in detail later, it outlines the main ATM barriers that can be quantified from the available data, the areas of analysis to which they belong, the order in which they occur and the direction of the flow.

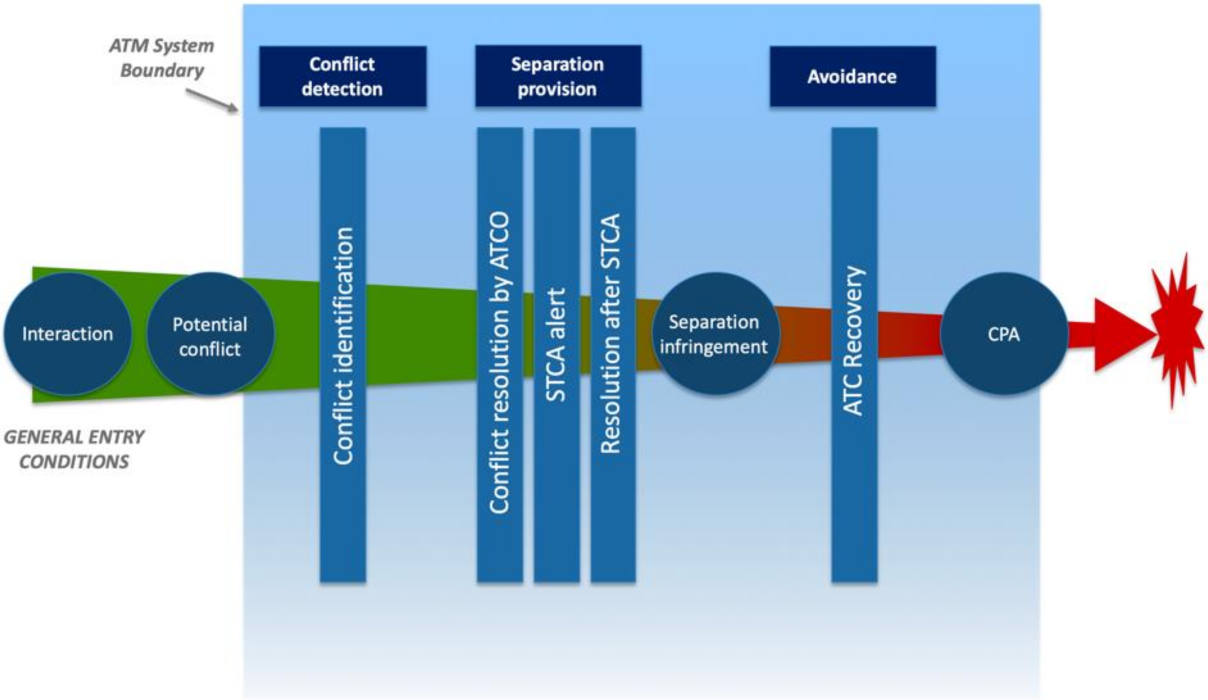


Figure 23: Barrier model

The process of identifying the sequence of barriers that occur before a loss of separation occurs is carried out (Figure 24):

- **Interaction identification:** The first step is to find out whether the two aircraft constitute an interaction. An *Interaction* has been defined as two aircraft within 20 NM of each other. When two aircraft constitute an interaction, they are considered to be a pair. Due to the data nature, the scope of FARO safety model is aircraft pairs.

- **Assessment of potential conflict:** If two aircraft constitute a pair, the probability that the pair constitutes a potential conflict under the conditions of the situation existing at that moment must be evaluated, that is, without the action of the controller.

- **Conflict identification:** Again, in the case of a potential conflict, it will have to be assessed whether it is detected by the ATCo.

- **Conflict resolution in identified conflicts:** The probability of conflict resolution by the ATCo of those cases that ARE potential conflicts and are detected by the ATCo.

- **STCA alert:** Finally, an assessment is made of the probability of triggering the STCA alert for conflicts not identified by the ATCo.

- **Conflict resolution:** For detected conflicts but that are not resolved and the probability of conflict resolution after the activation of STCA alert.

Figure 24: Sequence of barriers identified

5.3 Event tree analysis (ETA)

We use the principles of **event tree analysis** (Figure 25) to effectively translate this barrier model into a causal network representation. Event tree provides a top-down logic modeling technique for success and failure that explores responses through a single initiating event. It establishes a pathway to **evaluate the probabilities** of outcomes and overall system analysis. The output of each node represents a Boolean logic.

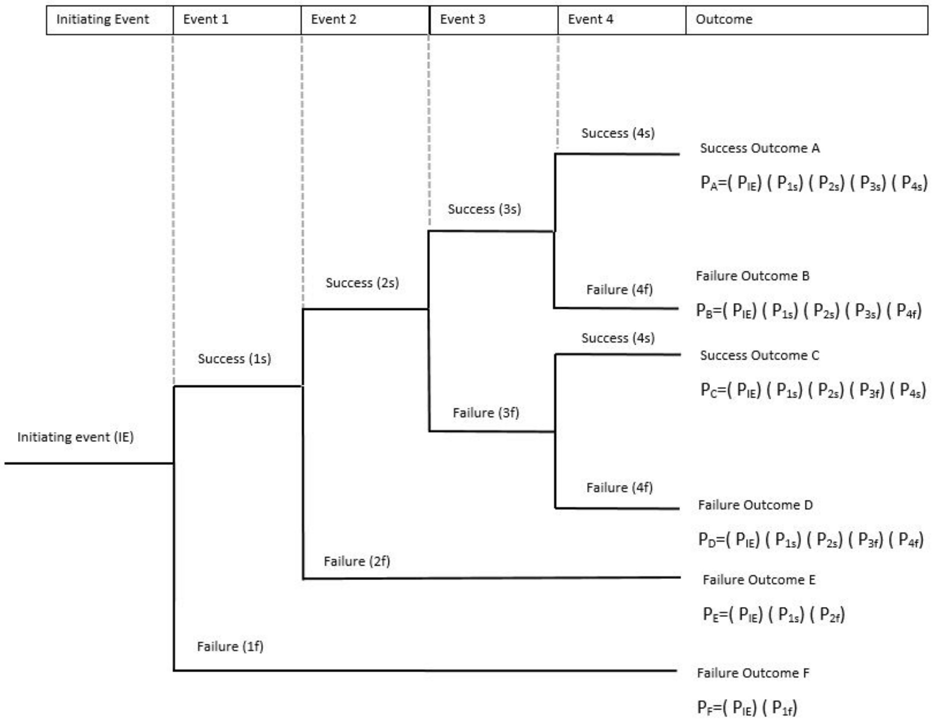


Figure 25: Event tree analysis example

This model provides a very visual approach to cause-and-effect relationships as well as exploring all possibilities. In addition, it allows complex models to be simplified and approached in a more understandable way. The initiating challenge must be identified by the analyst and success, or failure probabilities are usually difficult to find.

Figure 26 shows how the ATM barriers and the event tree are combined in our conceptual framework. The sequence of barriers can be identified in the upper bar. The probabilities of each event occurring or not, the bifurcation lines of each decision and, finally, the consequences of each of the branches can also be observed.

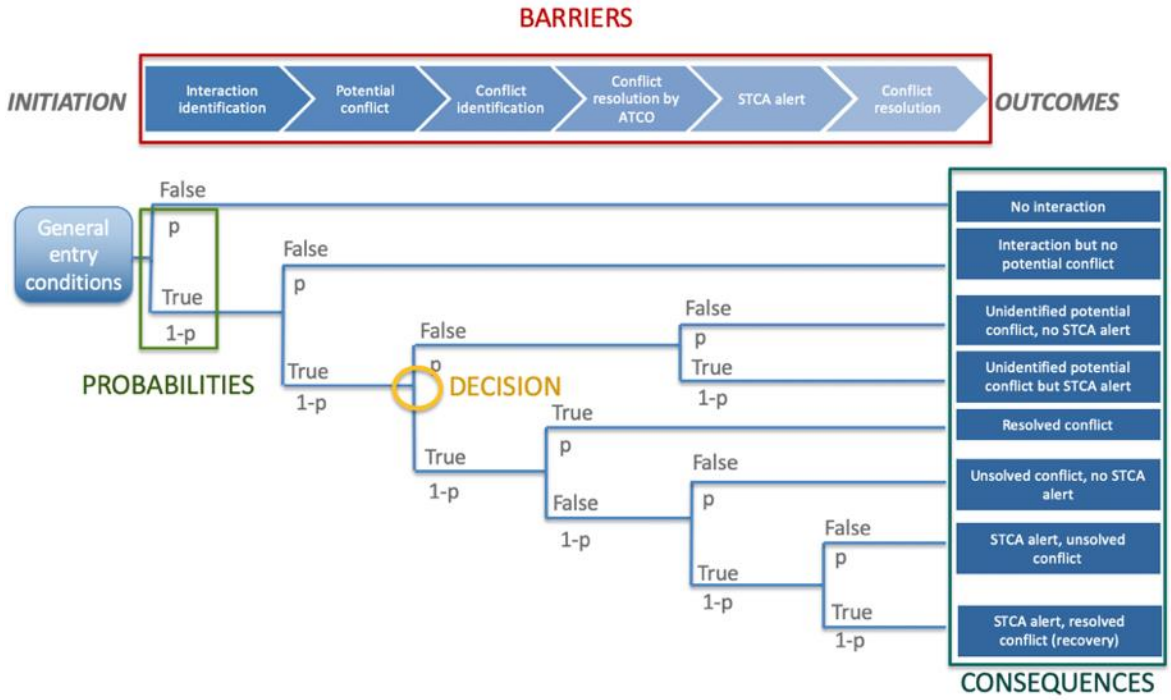


Figure 26: Integrated model

In addition, the probability of each of the exits can be expressed as a conditional probability of each of the branches of the tree, as indicated in Figure 27.

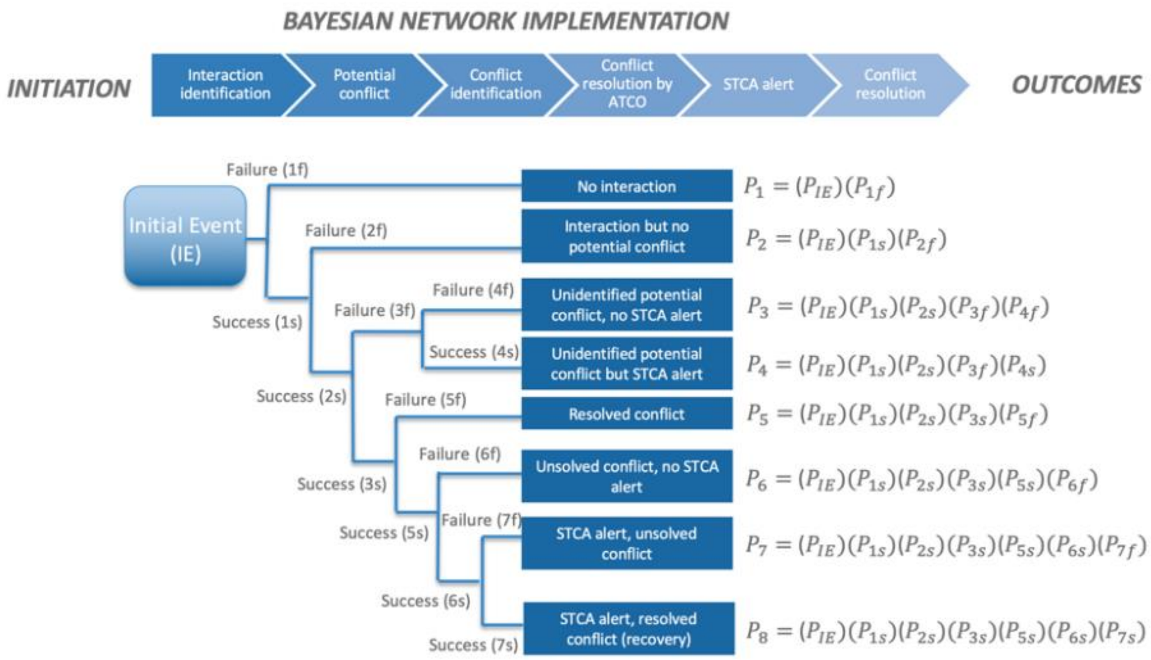


Figure 27: Conditional probability

5.4 Bayesian Network definition

Now that the conceptual framework and its causal model have been developed at a more detailed level of granularity, it can be implemented using **Bayesian networks**. Based on the barrier model and the event tree described above, each of the ATM barriers and the outcomes of the event tree will be modelled using a Bayesian subnetwork. The complete large-scale scheme of the proposed network can be seen in the following Figure 28.

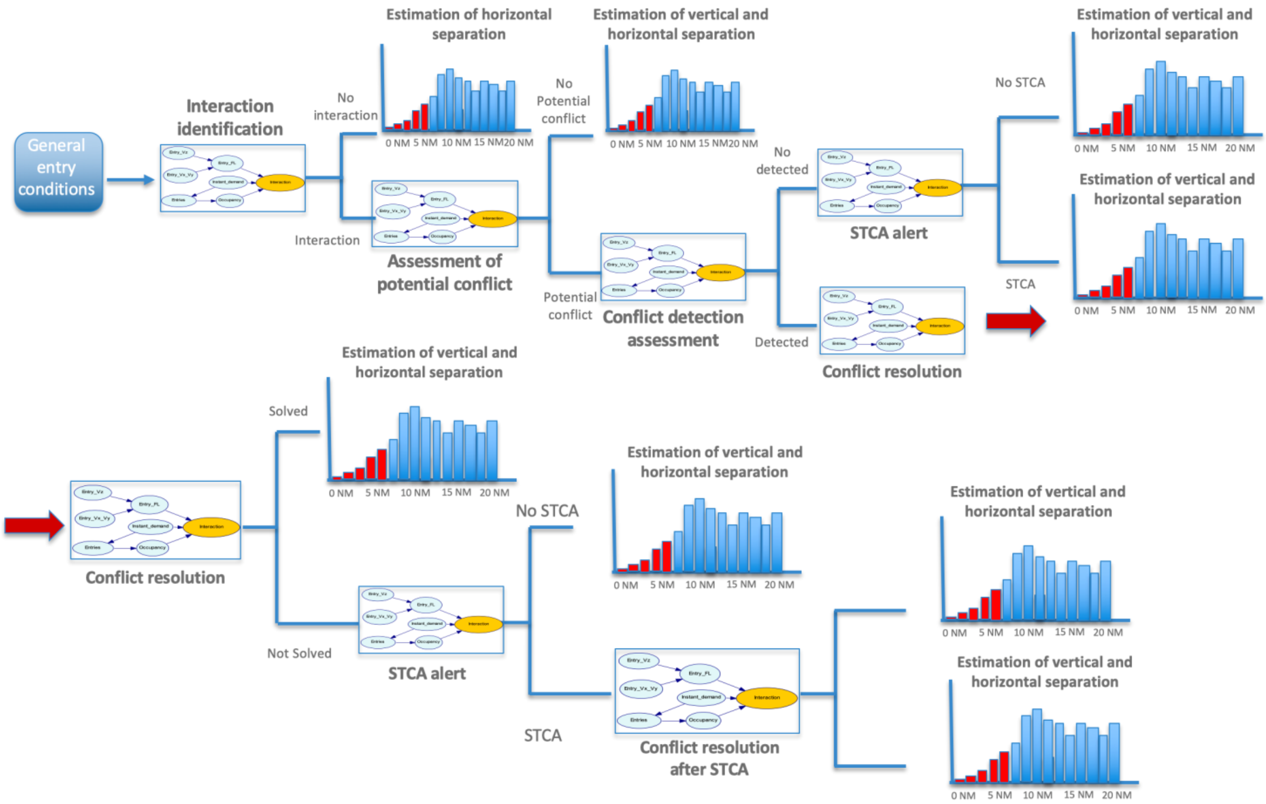


Figure 28: BN large-scale scheme

Two types of subnetworks can be identified. In the first place, those destined to estimate the probability of an event occurring and, on the other hand, those destined to estimate the vertical and horizontal distance at the CPA between aircraft. The first type of subnetwork explains the modelling of the ATM barriers and its effectiveness. These are represented in Figure 28 by a box with blue and yellow bubbles inside.

The second type explains the modelling of the outcomes of the event tree and the probability distribution of the vertical and horizontal separation between the pairs of aircraft included in each outcome. These subnetworks are represented in Figure 28 by a red and blue histogram.

It has to be noted that the whole model is composed by a set of 24 subnetworks integrated into a large BN.

As final results, the weighted sum of the vertical and horizontal distributions obtained as the final node of each branch will be carried out (Figure 29). However, this is an aspect that will be dealt with in greater depth in due course.

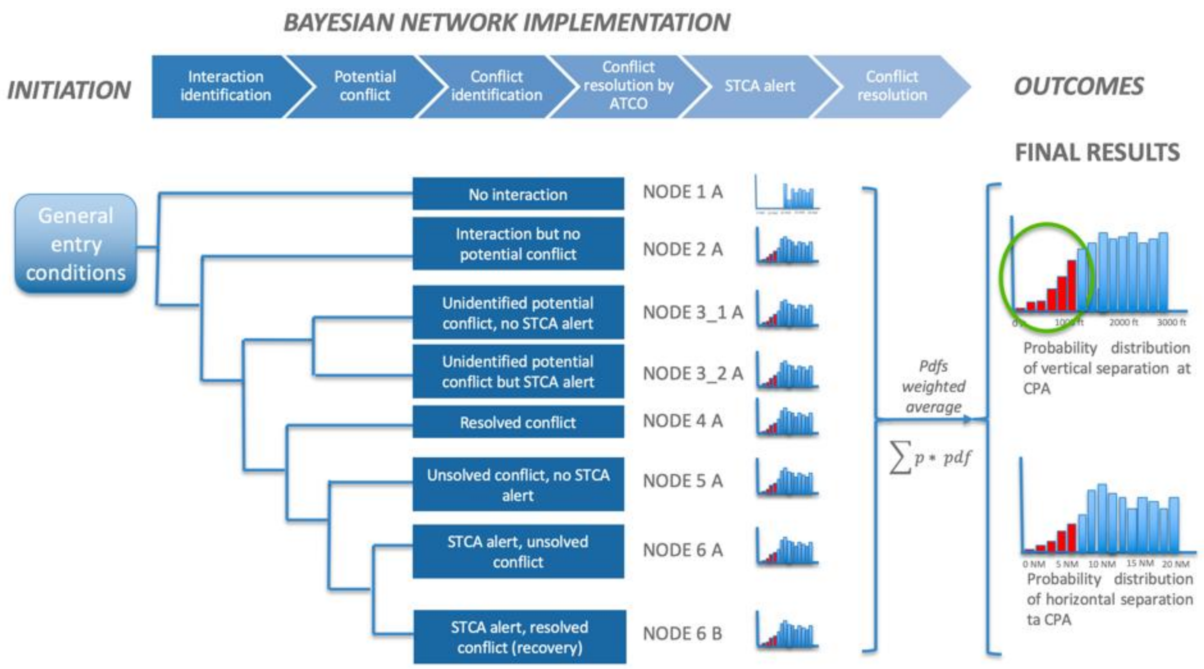


Figure 29: Data driven Event tree + Bayesian network

To complete the presentation of the methodological framework, the principles of every type of subnetwork as well as the main analysis can be carried out to tune and adjust each of those networks will be explained in the following sections.

5.4.1 Type 1 subnetwork: ATM safety barrier efficiency estimation

Part of the subnetworks in the model are built to capture the performance and effectiveness of ATM safety barriers. The main characteristics in this type of networks is that the outcome is always a binary variable/node that represents the effectiveness of an ATM safety barrier. To illustrate how these subnetworks work, the example of the conflict detection barrier will be examined (Figure 30).

Figure 30: Conflict detection assessment

The purpose of this subnetwork is to assess the probability of potential conflicts being detected by the ATCo. Figure 30 schematically indicates the main inputs and outcomes of this subnetwork. It may be seen that the probability of conflict detection will be influenced by ATCo workload and alertness as well as by the traffic complexity and overall scenario conditions. The outcome of the subnetwork will be either the “detection” or “not detection” of a conflict between an aircraft pair, represented as a binary variable or node.

Factors influencing the probability of conflict detection are identified based on experts’ knowledge, and variables to characterise them are proposed based upon the data available. Each of the variable is represented by a node in the BN, and the interrelationship between these variables are defined.

Each node is a discrete element and imply that the variables considered as nodes in the BN need to go through a discretisation process. This step is necessary since most algorithms are optimized for discrete variables. This consists of converting the continuous variables into variables grouped by intervals. The discretisation of variables can be based either on statistical characterization or in expert knowledge. Discretisation should ensure that no information is lost or considered as an excess of states. Table 4 shows some examples of node discretisation.

Table 4: Input variables and states example

Parameter	Description	Discretisation criteria	Discretisation example
Occupancy	Occupation Sector data at the AC entry hour	It is divided into 4 intervals that represent the percentage of occupation in the sector at the AC entry hour respect to the maximum value detected in the sector. The values of the intervals depend on the sector data.	Value <40% 40% <value <65% 65% <value <85% 85% <value
Entry	Entries Sector data at the AC entry hour	It is divided into 4 intervals that represent the percentage of entries in the sector respect to the sector entry declared value at the AC entry hour. The values of the intervals depend on the sector data.	Value <55% 55% <value <65% 65% <value <80% 80% <value
Instant demand	AC entry data in the sector in the 5 minutes period	It is divided into 4 intervals that represent the percentage respect to the maximum instantaneous demand value. The values of the intervals depend on the sector data.	Value <60% 60% <value <90% 90% <value <120% 120% <value

After that, the a priori probability of nodes without parent and the conditional probability dependences for nodes with parents are calculated and learnt from the data following the parametric learning process detailed in chapter 6. These tables are obtained directly from the frequencies observed in the data.

Figure 31 illustrates the resulting network obtained after structure learning and parametric learning. Each node shows the probabilities directly observed from data. It can be seen that the outcome node, indicted by the box “Detected conflict” on the right-hand side of the figure is a binary node, with two states. State 0 represents the probability of a failure in the ATM barrier, while state 1 represents the probability of the barrier success.

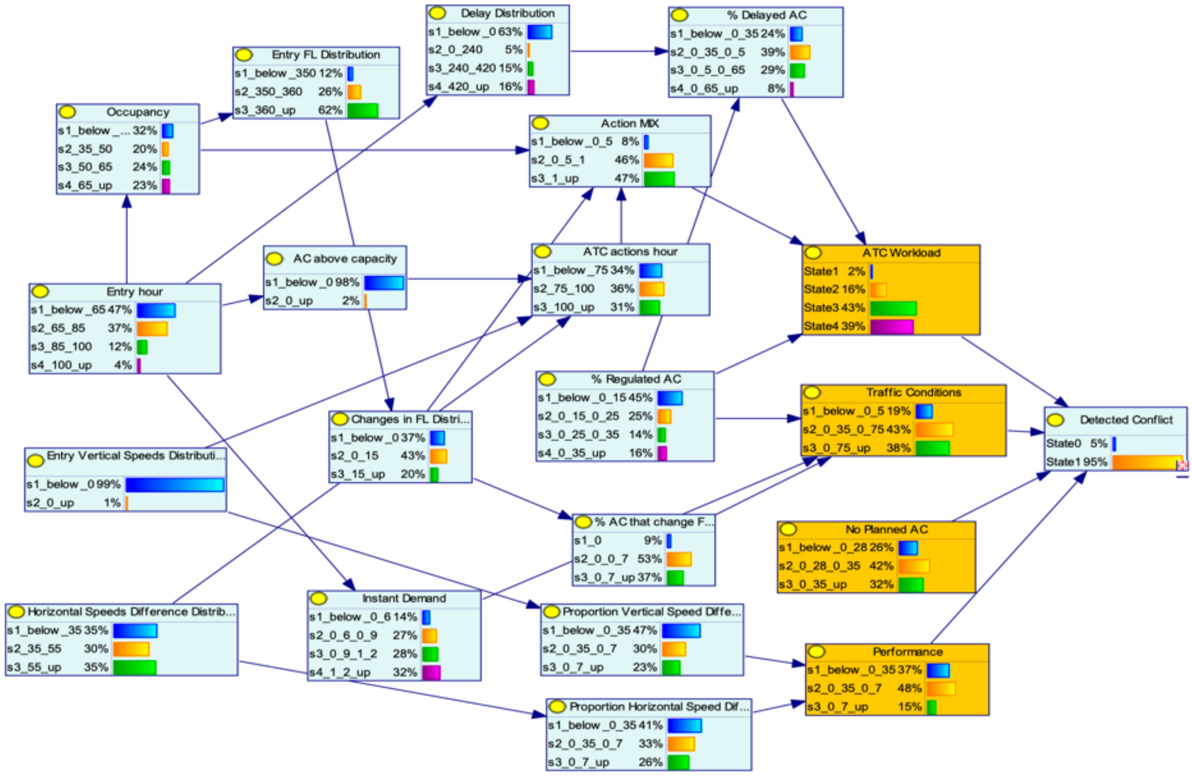


Figure 31: Scheme of a type 1 Bayesian Network

5.4.2 Type 2 subnetwork: Vertical and horizontal separation distribution estimation.

Part of the subnetworks in the model are built to capture the resulting vertical and horizontal separation distribution for each branch of the probabilistic event tree. The main characteristics in this type of networks is that the outcomes are always a discrete probability distribution which represent either the vertical or the horizontal separation at the CPA for those pairs of aircraft pertaining to each branch of the event tree.

The concept is illustrated in Figure 32. The vertical and horizontal dimension of the movement have been decoupled in two different networks. The network that estimates the vertical separation distribution considers mainly variables related with the vertical dimension of the aircraft movement. Whereas the network that estimates the horizontal separation distribution considers principally variables that accounts for the horizontal dimension of aircraft movements.

The horizontal separation distribution is estimated only for those aircraft pairs whose vertical separation at CPA is less than 1000’. Horizontal separation follows a uniform distribution for those aircraft pairs whose vertical separation at CPA is 1000’ or higher.

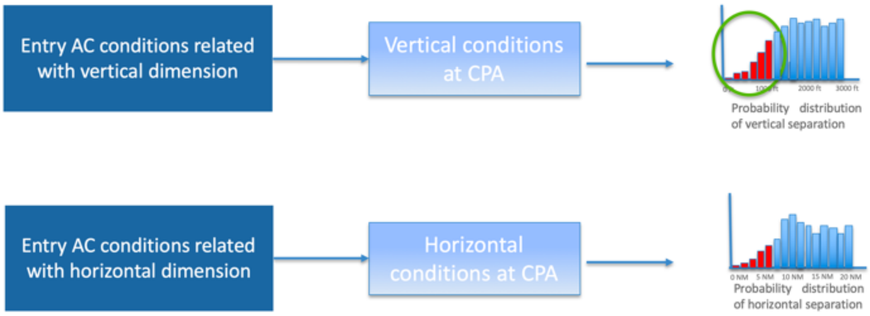


Figure 32: Vertical and horizontal distance estimation process

Figure 33 provides more detail on each of the subnetworks. The whole subnets can be observed with the input variables and the discretisation of the output variable.

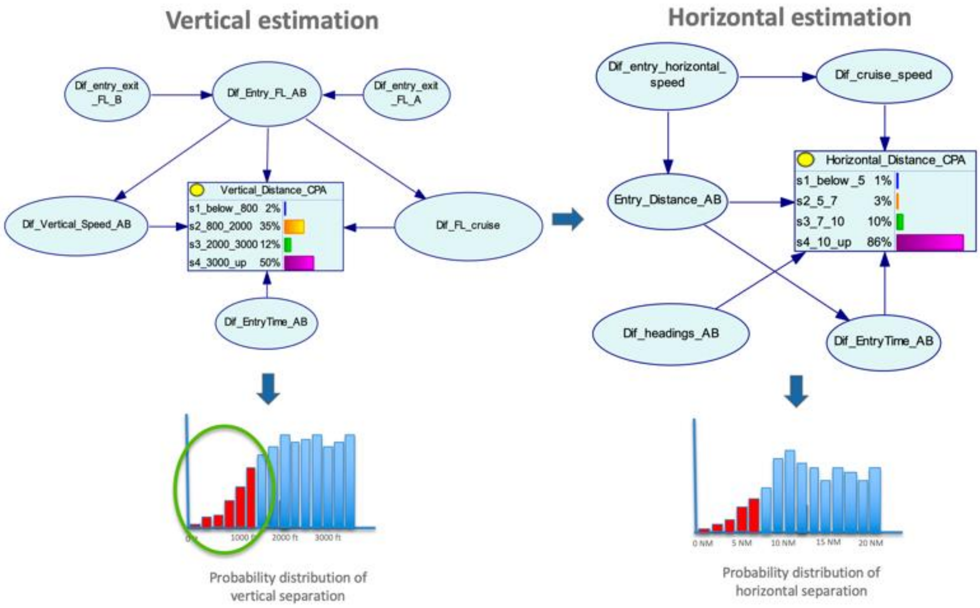


Figure 33: Vertical and horizontal distance estimation networks

In the case of vertical estimation, input variables such as the difference in entry flight levels, the difference in vertical speed of the aircraft pairs, the difference in cruise speeds and the distribution of entry times of each aircraft to the sector are considered.

As mentioned above, the distribution of vertical distances will be obtained and those that are less than a thousand feet apart will feed the horizontal separation network to obtain this distribution as well.

In the case of the horizontal subnetwork, the difference in horizontal speed between aircraft entering the sector, the difference in cruise speed, the distance and time at which the aircraft enter the sector and, finally, the difference in headings are considered.

6 Network development and training

During the development of this section, the description of the subnetworks that make up the general model will be covered. The same scheme will be followed for all of them to cover all information that may be relevant.

Thus, first of all, the description of the subnetwork and its objective will be presented. Next, an overview of the subnetwork and its high-level structure will be given.

The next step will consist of defining in detail the structure of the subnetwork. It consists of finding the causal relationships between the variables, that is, the topology of the Bayesian Network. This structure can be defined directly from the database, or through the knowledge of experts. In this case, it has been defined by a mixture of both, since once the subnetwork has been obtained directly from the data, it has been complemented by adding or removing arcs and changing the direction of the causal relationships.

Later, the input variables, their possible states and the discretisation criteria will be discussed in more detail. The discretisation of variables consists of converting the variables that are continuous into variables grouped by intervals. This step is necessary, since Bayesian networks consider discrete or continuous variables, but it is not possible to obtain a hybrid model from the data. The discretisation of variables can be based either on statistical characterization or in expert knowledge. Discretisation should ensure that no information is lost or considered as an excess of states.

Once the variables have been transformed into nodes in the subnetwork, it will be proceeded to obtain the conditional probability tables for each node. These tables are obtained directly from the frequencies observed in the data. This process is included in the parametric learning section.

Once the learning process of the network has been completed, its evaluation and verification of its usefulness is necessary. For this purpose, a sensitivity analysis will be carried out. Sensitivity analysis is a technique that can help validate the probability parameters of a Bayesian network. This is done by investigating the effect of small changes in numerical parameters on the output parameters. High-sensitivity parameters affect the reasoning results more significantly. Identifying them allows for a directed allocation of effort to obtain accurate results of a Bayesian network model. The implemented algorithm was proposed by (Kjaerulff, 2000). Given a set of target nodes, the algorithm calculates efficiently a complete set of derivatives of the posterior probability distributions over the target nodes over each of the numerical parameters of the Bayesian network. These derivatives indicate the importance of the precision of network numerical parameters for calculating the posterior probabilities of the targets. If the derivative is large for a parameter p, then a small change in p may lead to a large change in the posterior of the target. If the derivative is small, then even large changes in the parameter make little difference in the posterior.

The objective of this deliverable D4.1 is to explain the construction of the network as well as to demonstrate the feasibility of using Bayesian analysis. In this deliverable (D4.1), the development of the generic BN model is described using data from the Santiago sector (LECMSAN), as an example for the sake of illustration.

However, it has to be noticed that the SPF model is sector dependant, what means that the generic BN model needs to be adapted for each ATC sector. The model needs to be fully informed from the data of a particular ATC sector. That the conditional probability tables in the model are specific to each

ATC sector and have to be learned from data of the specific sector under analysis. Adaptation of the generic model to the characteristics of the sector analysed in each Use Case implies:

- Gathering and processing all data from the sector/traffic under analysis in the Use Case. Processing involves the discretisation of the data for each model variable. Discretisation scheme is, at this stage, a knowledge-based process, it cannot be automated, and it might be different for each sector because it depends on sector features and traffic profile.
- Parametric learning, i.e., obtaining the required a priori and conditional probability tables from the frequency observed directly from data.
- Sensitivity analysis to tune the network and identify the most influential variables for each particular sector.
- Backward and forward analysis to define thresholds for the variables that might impact safety performance in a scenario, if applicable.

Upon completion of the above work, the BN will be useable for a particular use case and validation activities can be performed. The adaptation of the model to the characteristics of the sectors analysed in Use Cases 1 and 2 will be detailed in Deliverable 4.2 (expected in October 2021).

In this section, an example of each of the two types of networks discussed in chapter 5 will be explained in detail. This has been done with the intention of simplifying the document and making it easier to read.

Specifically, subnets 2 and 2A will be shown. Subnet 2 aims to identify whether a pair of aircraft is a potential conflict or not. On the other hand, the objective of subnetwork 2A is to estimate both the vertical and horizontal distance of those pairs of aircraft that have not been identified as a potential conflict in subnetwork 2.

For a correct understanding of the model developed, it is recommended to consult Annex I, which includes all the subnets that model the safety barriers, as well as the evaluation of distances at which the pairs of aircraft are in the CPA.

6.1 Subnetwork 2: Assessment of potential conflict

6.1.1 Description and objective

The objective of the subnetwork is to identify the probability of an aircraft pair being a potential conflict before the ATCo acts. For this purpose, the sector entry conditions of the aircraft pair and the relative conditions between them will be considered. According to the ATM barrier model (Figure 23) this event is posterior to the event “Interaction”; therefore this network is concerned with aircraft pairs that were identified as interactions in subnetwork 1.

In other words, the aim is to analyse whether, due to the geometrical and dynamic condition of two aircraft, the pair will constitute a possible potential conflict. This subnetwork will only process a subset of the aircraft considered in the previous subnetwork (subnetwork 1). In particular, only those pairs of aircraft belonging to branch “1-p” in subnetwork 1, i.e., the aircraft pairs that became an interaction in the previous subnetwork.

As mentioned above, the process will be illustrated using the Santiago sector (LECMSAN) as an example. The process will be similar for the other two sectors under study Barcelona Upper (LECBCCU) and Barcelona Central (LECBCCC). Although the input data for each of these scenarios are unique, the construction and layout of the network will be similar as the model and the expert judgement identify the same variables as the most influential in terms of interaction.

6.1.2 Network construction

As mentioned above, the starting point is the data on the sector entry conditions of an aircraft pair and the relative conditions between them.

Subnetwork 2 is a particular type of BN, in which the central node involves a deterministic relationship with two conditions that have to be fulfilled simultaneously.

A potential conflict is defined if an aircraft pair is expected to lose their separation based only on their planned trajectories. Potential conflicts are expected to require higher attention and ATCo intervention to provide and monitor aircraft separation. The data initially provided by CRIDA did not include this information for any aircraft pair. Therefore, a simple heuristic has been defined in this subnetwork. An aircraft pair in the same sector are considered a potential conflict if two conditions are met simultaneously: 1) their headings are convergent, i.e., the difference in headings is less than 180° and, at the same time, 2) the entry and exit flight levels of the two aircraft overlap. Thus, a new binary variable called "potential conflict" will be generated, which will take the value 1 only when the two conditions mentioned above are met. If the conditions for a “potential conflict” change, the subnetwork 2 should be modified accordingly.

This subnetwork will generate two different branches in the event tree, “p” and “1-p”. The aircraft that do not involve any potential conflict, included in the first branch “p” will be further analysed in subnetwork 2A. While aircraft that do represent a potential conflict, indicated by the branch “1-p,” will be further analysed in subnetwork 3.

Thus, the construction of the network shown in Figure 34 is based on:

- **Network entries:** Flight level and heading data for both aircraft in each pair.
- **Training data:** For each aircraft pair, the range of flight levels and headings are calculated, and an assessment is made as to whether the conditions for a potential conflict are met.
- **Network exit:** Probability of conflict. Node with two states “yes/no”.

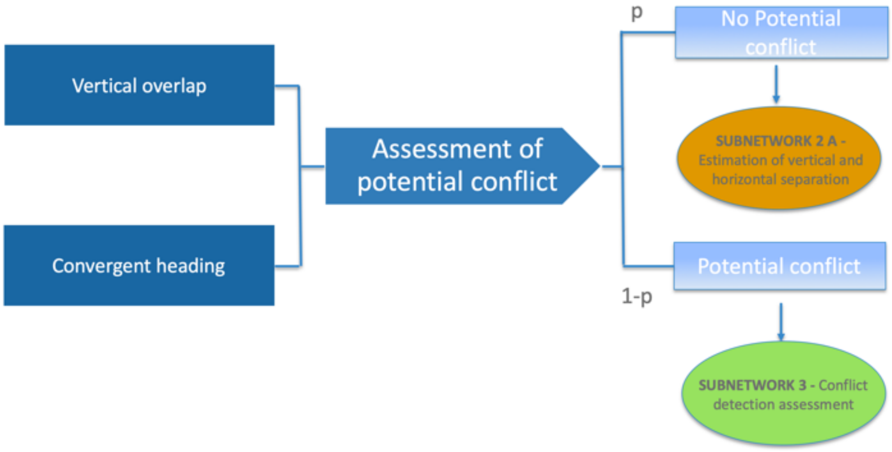


Figure 34: Assessment potential conflict network

Figure 35 shows the proposed Bayesian network scheme. It can be seen how the variables that directly affect the calculation of a potential conflict are in turn determined by other variables.

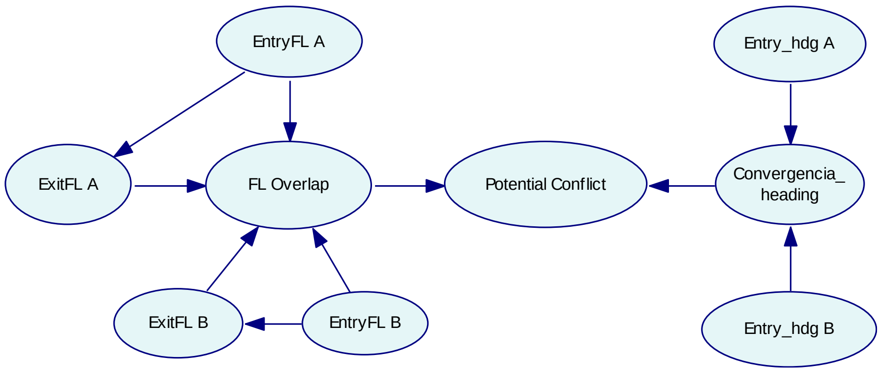


Figure 35: Subnetwork 2 structure

Specifically, the flight level overlap is defined by the entry and exit flight level of each aircraft. In addition, the entry flight level will condition the exit flight level.

On the other hand, the convergence of headings only considers the heading of each aircraft at the sector entry.

6.1.3 Input and output variables and states

In a first approximation, the input variables could be divided into two main areas:

- **Flight Levels Overlap:** This area aims to determine if the flight levels of an aircraft pair overlap. This area considers as variables the sector entry flight level (“Entry_FL”) and the sector exit flight level (“Exit_FL”) for each aircraft in the pair (aircraft A and aircraft B). Two intermediate variables have been created: Flight level interval aircraft A and Flight level interval for aircraft B, which will not appear in the network. These intermediate variables consider the entry and exit flight level in the sector. There is overlap of flight levels if the flight level intervals of both aircraft intersect

- **Convergence of headings:** This area aims to determine the heading convergence between pairs of aircraft. This area considers as variables the heading of each aircraft when entering the sector ("Entry_hdg_A" and "Entry_hdg_B"). One intermediate variable has been defined, which will not appear in the network: Difference of headings. This intermediate variable calculates the difference between the courses of both aircraft when entering the sector. If the difference is between 0 and 180 degrees, they do not converge, and if it is between 181 and 360 degrees, they do converge.

Detailed definition of all variables and its discretisation states is provided in the chapter 7.

6.1.4 Parametric learning

In Figure 36, it can be seen the probabilities directly observed from data for the parameters of subnetwork 2, particularised for Santiago Sector (LECMSAN). It is observed that 24% of all the aircraft pair in this subnetwork are considered potential conflict.

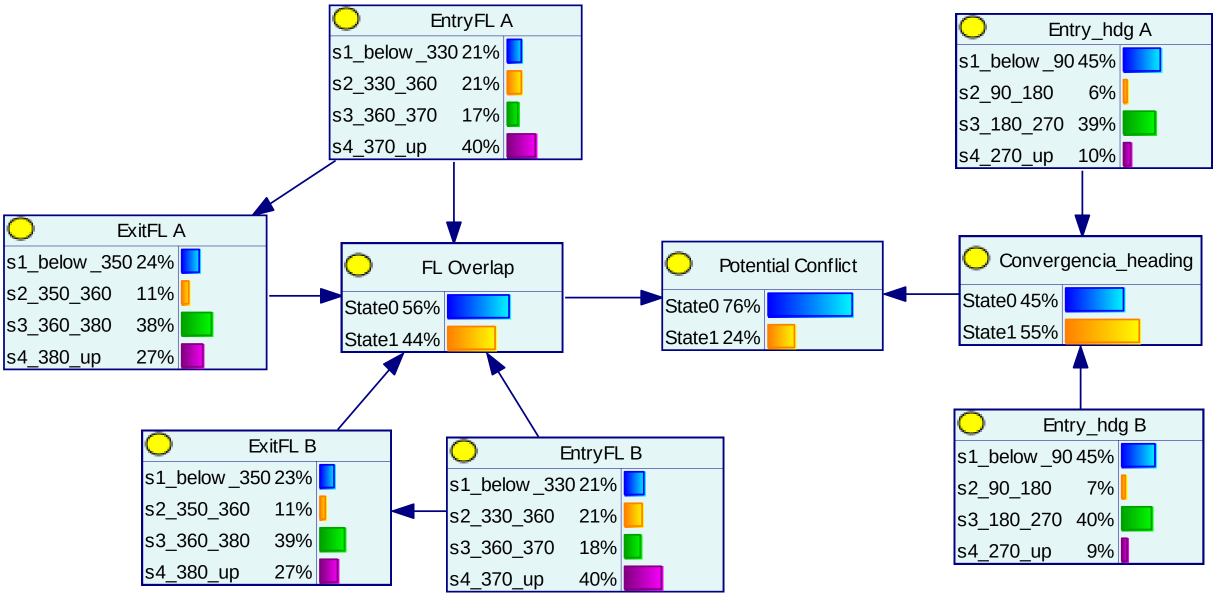


Figure 36: Probabilities directly observed from data for Node 2, Santiago Sector

6.1.5 Sensitivity analysis

To carry out the sensitivity analysis, the output variable, "Potential Conflict", is set as a target. The Kjaerulff algorithm is applied for sensitivity analysis. Figure 37, shows the results of the sensitivity analysis carried out on this subnetwork for Santiago Sector.

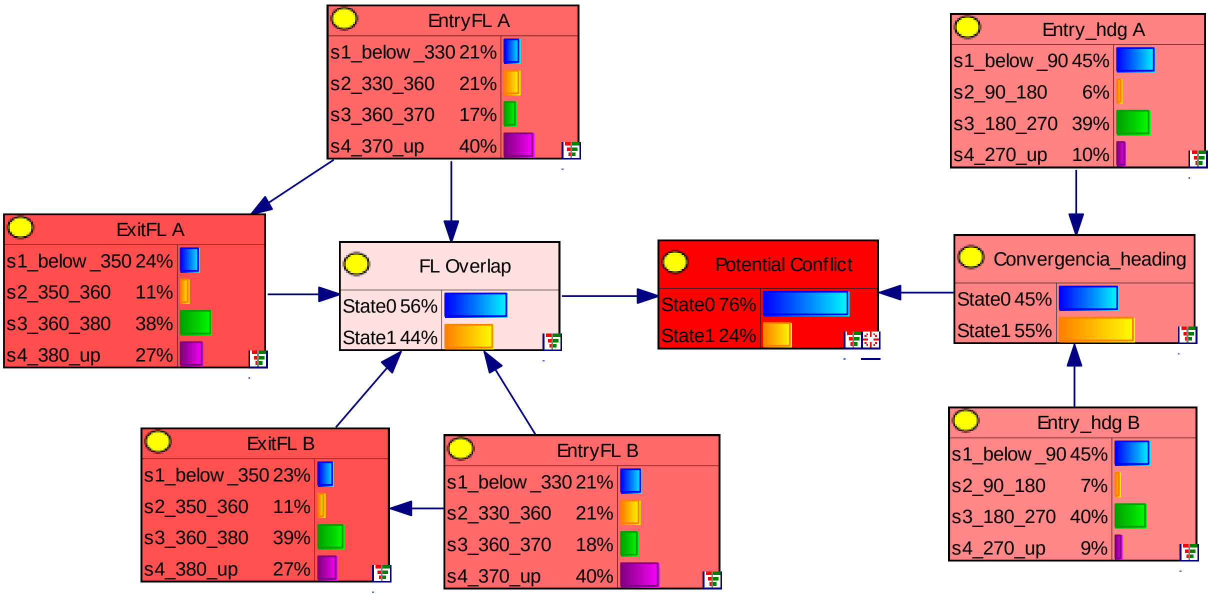


Figure 37: Sensitivity analysis Subnetwork 2, Santiago Sector

As mentioned in previous sections, one of the main objectives of conducting a sensitivity analysis is to detect the variables that have a biggest impact under conditions of uncertainty. It can be observed that all the input variables have a high degree of sensitivity with respect to the output variable. This is reasonable considering that the central node in this subnetwork involves a deterministic relationship with two conditions that have to be fulfilled simultaneously by an aircraft pair to be considered a potential conflict: convergence of headings and overlap of flights levels.

6.2 Subnetwork 2A: Distance evaluation between aircraft. Cases of non potential conflict

6.2.1 Description and objective

The objective of this subnetwork is to assess the actual CPA horizontal and vertical separation for those pairs of aircraft not considered as potential conflicts in subnetwork 2. The separation distribution will depend on the aircraft's trajectories and the aircraft's conditions when entering the sector.

The study of the vertical separation will be carried out separately from that of the horizontal separation, so it will actually be two Bayesian networks.

6.2.2 Network construction

The vertical and horizontal separation distributions are modelled with two decoupled BNs. First, to estimate the vertical separation distribution at which aircraft were found at the CPA, all aircraft pairs that were not identified as potential conflicts in the previous subnetwork are considered. This network considers mainly variables related with the vertical dimension of the aircraft movement.

To proceed with the second network, data are segregated. For the estimation of the CPA horizontal distance distribution, only aircraft pairs with less than 1000 vertical feet separation are retained. It has to be remembered that the horizontal separation follows a uniform distribution for those pairs of aircraft whose vertical separation at CPA is 1000' or higher. This second BN considers principally variables that accounts for the horizontal dimension of aircraft movements.

Thus, the construction of the two networks is showed in Figure 38 and based on:

- **Network entry:** Relative vertical or horizontal condition data of each aircraft pair at the sector entry point.
- **Training data:** The vertical and horizontal separation at the CPA for each aircraft pair.
- **Network exit:** Vertical and horizontal distance between aircraft at the CPA.

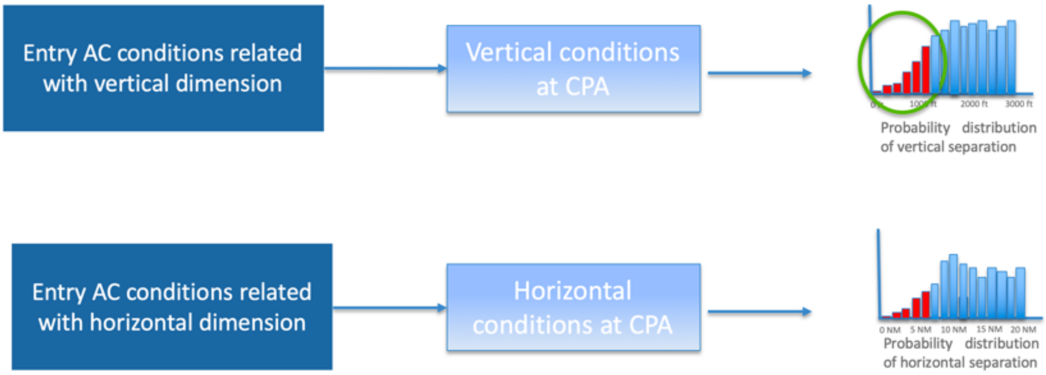


Figure 38: Assessment of distance in "no" potential conflict network

The Bayesian network model for the vertical dimension is presented first. For this network, only data on the relative vertical conditions between aircraft at the sector entry point for each aircraft pair are considered.

The links representing the causal relationships between variables have resulted from a combination of the model proposals and the experience provided by expert judgement.

Figure **39** shows the structure for the vertical Bayesian Network:

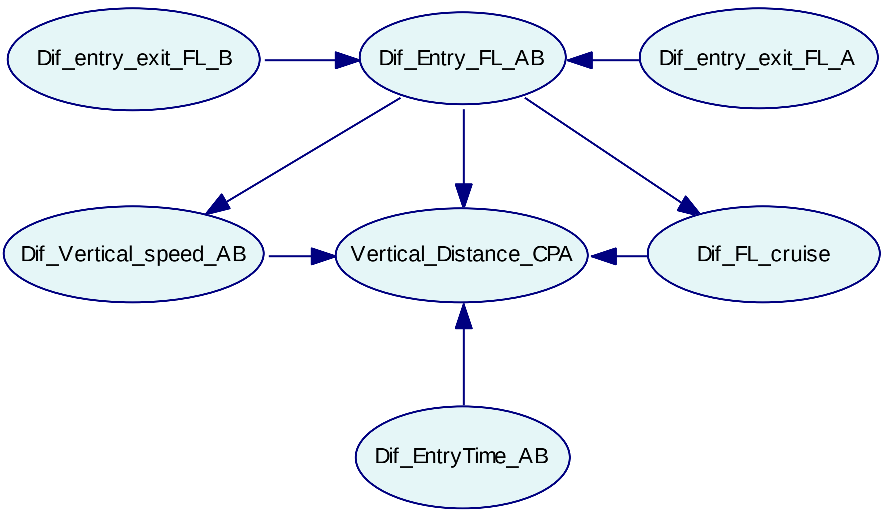


Figure 39**: Subnetwork 2 A. Vertical Bayesian Network**

Figure 40 presents the Bayesian network model for the horizontal dimension. In this case, all network input variables refer to the horizontal sector input conditions of the aircraft pair.

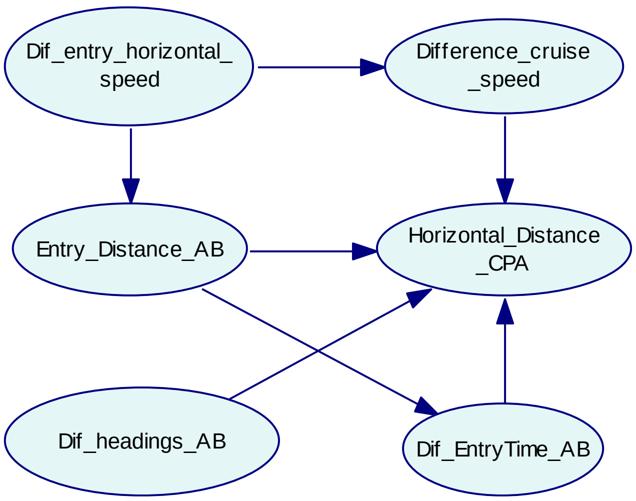


Figure 40: Subnetwork 2 A. Horizontal Bayesian Network

6.2.3 Input and output variables and states

The aspects that will be considered to model these two networks are shown below:

- **Vertical distance:** The variables to be considered for the modelling of the vertical distance are those related to the aircraft entry conditions into the sector. i.e., their flight levels, vertical speed, and entry times.

- **Horizontal distance:** The variable to be considered for the modelling of the horizontal distance are those related to the aircraft entry conditions into the sector. i.e., their headings, horizontal speed, and entry times.

In Figure **39** and Figure 40 it can be seen the network structure that represents the causal relationships between the variables that constitute this node.

6.2.4 Parametric learning

Figure 41 shows the probabilities directly observed from data for the parameters of subnetwork that estimated vertical separation distribution (subnetwork 2A), particularized in Santiago Sector.

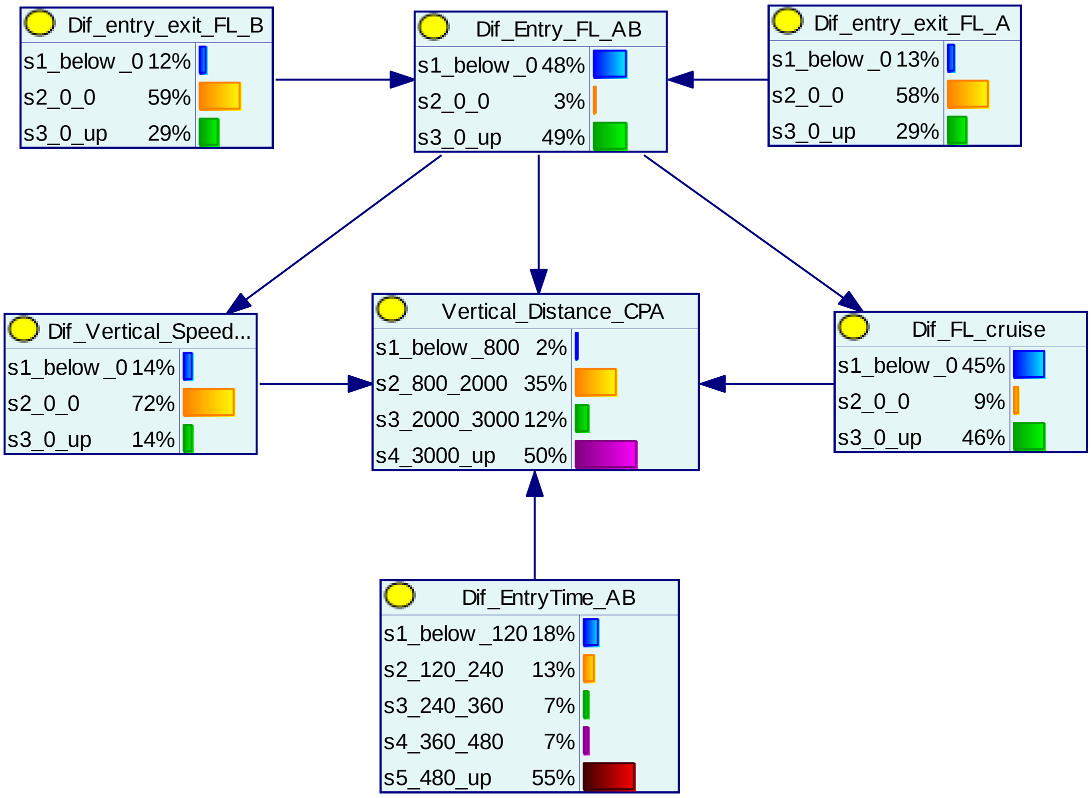


Figure 41: Probabilities directly observed from the data for Subnetwork 2 A for vertical distance, Santiago Sector.

The vertical distance at the CPA have been discretised to account for some anomalies in the data. The aircraft position information available is obtained from the radar tracks. The altitude or flight level information is based upon the barometric information enclosed in the secondary radar messages, and it is therefore affected by barometric altitude inaccuracies, differences between the estimated and true position of the aircraft. This error has a high impact in the model, particularly to identify the aircraft whose vertical separation is below 1000 ft. To avoid false cases, a limit of 800 ft instead 1000 ft has been considered. Therefore, the lower interval of the variable "vertical distance at the CPA" has been set between 0 and 800 ft instead between 0 and 1000 ft.

It is observed that 50% of the aircraft pairs have a vertical separation above 3000 ft. On the other hand, only 2% of the aircraft pairs will have a separation of less than 800 ft. Aircraft with a separation of less than 800 ft will be evaluated in the horizontal separation network.

The Figure 42 shows the probabilities directly observed from data for the parameters of subnetwork that estimated horizontal separation distribution (subnetwork 2B), particularized in Santiago Sector.

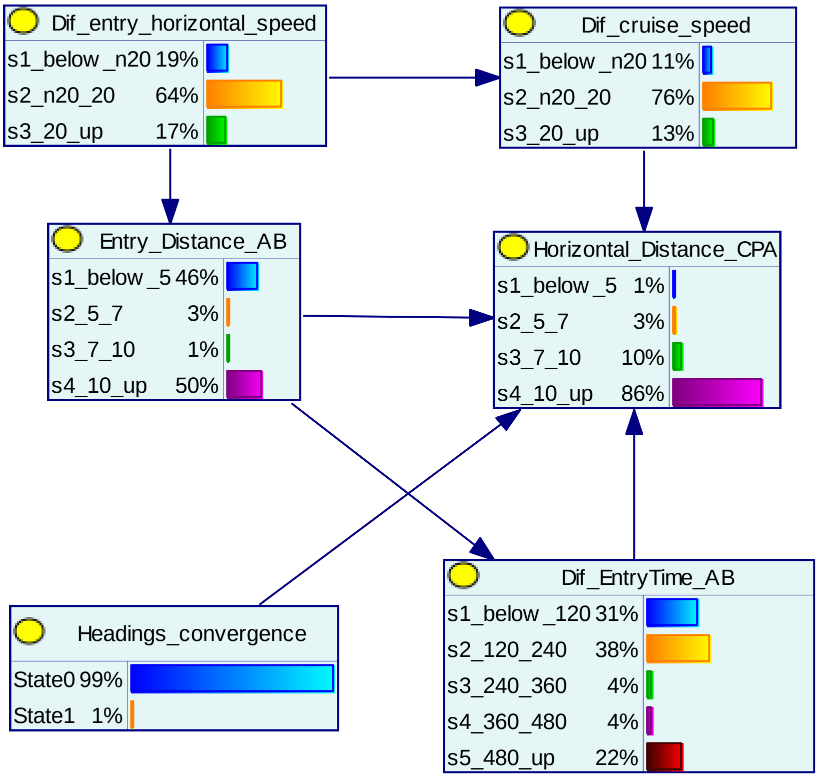


Figure 42: Probabilities directly observed from data for Subnetwork 2A for horizontal distance, Santiago Sector

It is observed that there is 1% probability of an aircraft pair with a horizontal separation below 5 NM and vertical separation below 800ft.

6.2.5 Sensitivity analysis for fine tunning

To carry out the sensitivity analysis, the output variables, "Vertical Distance CPA" and "Horizontal Distance CPA", have been set as target nodes in their respective Bayesian networks.

Figure 43 shows the results of the sensitivity analysis carried out on subnetwork 2A for Santiago Sector. It can be observed that all the parameters are highly sensitive to the output variable. Therefore, it can be concluded that the variables that have been selected following the expert's criteria to model this node are highly representative.

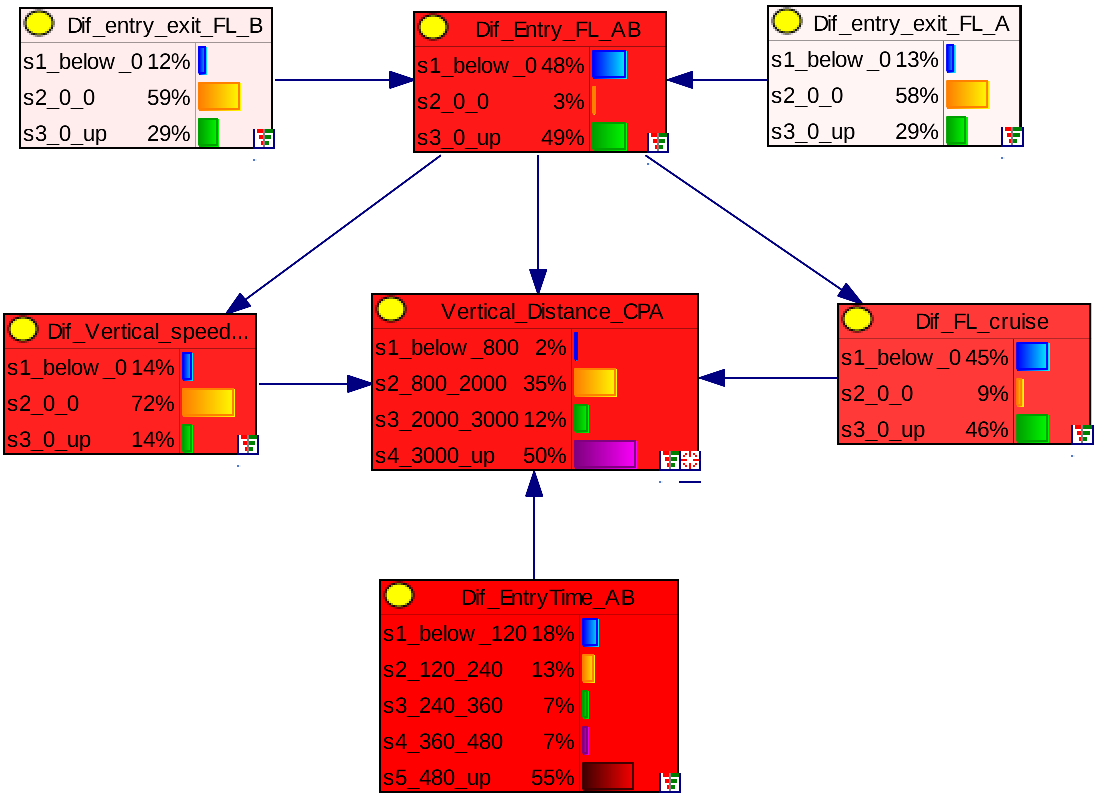


Figure 43: Sensitivity analysis subnetwork 2A for vertical distance, Santiago Sector

Figure 44 shows the results of the sensitivity analysis carried out on subnetwork 2B for Santiago Sector. It can also be observed that all variables are highly sensitive to the target variable, Horizontal Distance at CPA.

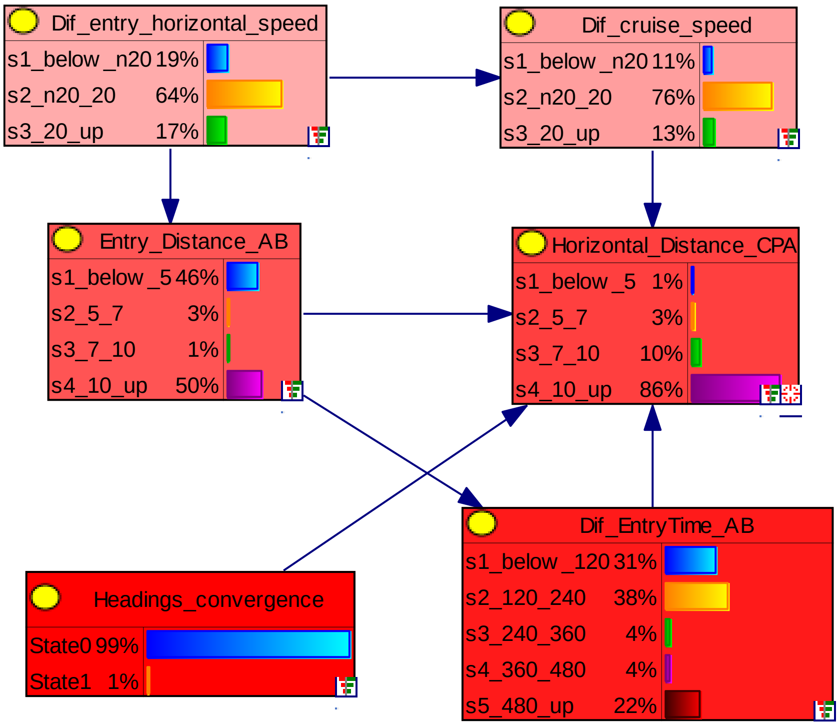


Figure 44: Sensitivity analysis Subnetwork 2 A for horizontal distance, Santiago Sector

Therefore, it can be concluded that the variables chosen for both the vertical and horizontal separation cases are quite representative for modelling the distribution of distances at which the aircraft pairs are separated.

7 Input variables and states

Until this point, the input, training and output variables of each of the developed subnetworks have been mentioned on several occasions. However, this section aims to develop them in greater depth. These variables can be involved in a single subnetwork or, as will be the case for most, be relevant to more than one.

Thus, a description of the variable is included in Table 5 and Table 6, an explanation of the discretisation criterion followed for each one of them. Discretisation consists of converting continuous variables into variables grouped by intervals. This step is necessary since most algorithms are optimized for discrete variables. Variable discretisation can be based on statistical characterization or expert knowledge. The discretisation must ensure that no information is lost or that too many states are considered.

A possible example of discretisation for each variable is also included for information purposes only, as it is considered more visual. However, this discretisation may vary for the same variable depending on which subnetwork and scenario it is in.

Table 5: Training variables

Parameter	Description	Discretisation criteria	Discretisation example
Occupancy	Occupation Sector data at the AC entry hour	It is divided into 4 intervals that represent the percentage of occupation in the sector at the AC entry hour respect to the maximum value detected in the sector. The values of the intervals depend on the sector data.	Value <40% 40% <value <65% 65% <value <85% 85% <value
Entry	Entries Sector data at the AC entry hour	It is divided into 4 intervals that represent the percentage of entries in the sector respect to the sector entry declared value at the AC entry hour. The values of the intervals depend on the sector data.	Value <55% 55% <value <65% 65% <value <80% 80% <value
Instant demand	Ac entry data in the sector in the 5 minutes period	It is divided into 4 intervals that represent the percentage respect to the maximum instantaneous demand value. The values of the intervals depend on the sector data.	Value <60% 60% <value <90% 90% <value <120% 120% <value
AC distribution by FL	Distribution of entry FL	It is divided into 6 intervals with uniform counts to model the vertical distribution in the sector. The values of the intervals depend on the sector data.	FL <250 250 <FL < 330 330 <FL <360 360 <FL <370 370 <FL <380 380 <FL

Parameter	Description	Discretisation criteria	Discretisation example
Horizontal ground speed distribution.	AC horizontal speed in the entry time	It is divided into 3 intervals. An interval will contain the mean value of the sector data. There will be another interval above and another below. The values of the intervals depend on the sector data	Value < 430 430< value < 470 470 < value
Vertical ground speed distribution	AC vertical speed in the entry time	It its divided into three intervals to separate ascending flights (>0), descending flights (<0) and stabilized flights (=0)	Value < 0 Value = 0 Value > 0
Distribution of entry times	10 minutes interval in which the AC enters the sector.	It is divided into 6 states. Each AC is assigned a value from 1 to 6, depending on the 10 minutes period in which the AC enters the sector.	Value= 1 Value= 2 Value= 3 Value= 4 Value= 5 Value= 6
Interaction	Interaction detected between aircraft. Aircraft distance less or equal than 20 NM as defined in the separation file described in section 4.1	It is divided into 2 states, 0 (when it is considered that AC is not an interaction with any other AC in the sector) and 1 (when it is considered that the AC is an interaction with another AC in the sector)	Value= 0 Value= 1
Entry FL A	Distribution of entry FL for the first Aircraft	It is divided into 4 intervals with uniform counts to model the vertical distribution in the entry of the sector. The values of the intervals depend on the sector data.	FL <330 330 <FL < 360 360 <FL <370 370 <FL
Entry FL B	Distribution of entry FL for the second Aircraft	It is divided into 4 intervals with uniform counts to model the vertical distribution in the entry of the sector. The values of the intervals depend on the sector data.	FL <330 330 <FL < 360 360 <FL <370 370 <FL
Exit FL A	Distribution of exit FL for the first Aircraft	It is divided into 4 intervals with uniform counts to model the vertical distribution in the exit of the sector. The values of the intervals depend on the sector data.	FL <350 350 <FL < 360 360 <FL <380 380 <FL
Exit FL B	Distribution of exit FL for the second Aircraft	It is divided into 4 intervals with uniform counts to model the vertical distribution in the exit of the sector. The values of the intervals depend on the sector data.	FL <350 350 <FL < 360 360 <FL <380 380 <FL

Parameter	Description	Discretisation criteria	Discretisation example
Entry heading A	Distribution of entry headings for the first Aircraft	It is divided into 4 intervals with uniform widths to model the entry headings distribution in the entry of the sector.	value <90 90 < value < 180 180 < value <270 270 < value
Entry heading B	Distribution of entry headings for the second Aircraft	It is divided into 4 intervals with uniform widths to model the entry headings distribution in the entry of the sector.	value <90 90 < value < 180 180 < value <270 270 < value
Flight Levels Overlap	Evaluation of FL overlap for the aircraft pairs	It is divided into 2 states, 0 (when there is not FL overlap) and 1 (when there is FL overlap)	Value= 0 Value= 1
Convergence of headings	Assessment of heading convergence at the entry for the aircraft pairs	It is divided into 2 states, 0 (when there is not Convergence of headings) and 1 (when there is Convergence of headings)	Value= 0 Value= 1
Potential Conflict	It is considered that there is a potential conflict when there is FL overlap and convergence of headings of the aircraft pairs	It is divided into 2 states, 0 (when there is not potential conflict) and 1 (when there is potential conflict)	Value= 0 Value= 1
Difference entry and exit FL AC A	Calculation of the difference between the aircraft A enters the sector and its exit flight level	It its divided into three intervals to separate ascending flights (>0), descending flights (<0) and stabilized flights (=0)	Value < 0 Value = 0 Value > 0
Difference entry and exit FL AC B	Calculation of the difference between the aircraft B enters the sector and its exit flight level	It its divided into three intervals to separate ascending flights (>0), descending flights (<0) and stabilized flights (=0)	Value < 0 Value = 0 Value > 0
Difference vertical speed of ACs	Calculation of the difference between vertical speed of aircraft A and B at the CPA instance	It its divided into three intervals to identify three interaction types	Value < 0 Value = 0 Value > 0
Difference FL entry	Calculation of the difference between entry flight level of aircraft A and B into the sector	It its divided into three intervals to identify three interaction types, depending on whether they enter the same flight level, above or below	Value < 0 Value = 0 Value > 0

Parameter	Description	Discretisation criteria	Discretisation example
Difference FL cruise	Calculation of the difference between cruise flight level of aircraft A and B	It its divided into three intervals to identify three interaction types, depending on whether they have the same flight level, above or below	Value < 0 Value = 0 Value > 0
Difference entry time	Calculation of the difference between the entry time (seconds) of aircraft A and B into the sector	It its divided into five intervals of the same widths.	t <120 120 < t< 240 240 <t <360 360 <t <480 480 <t
Difference cruise speed	Calculation of the difference between cruise speed of aircraft A and B	It is divided into three intervals to model the difference of cruise speed between the ACs	Value < -20 -20 < value < 20 Value > 20
Difference entry speed	Calculation of the difference between entry speed of aircraft A and B into the sector	It is divided into four intervals to model the difference of entry speed into the sector between the ACs	Value < -20 -20 < value < 20 Value > 20
Entry distance	Calculation of the entry distance into the sector between the aircraft pairs	It is divided into four intervals to model the distance	0<Value<5 5< Value <7,5 7,5< Value <10 10< Value
Vertical Distance at the CPA	Vertical distance between the aircraft pairs at CPA instance	It is divided into four intervals	D<1000 1000<D<2000 2000<D<2000 3000<D
Horizontal Distance at the CPA	Horizontal distance between the aircraft pairs at CPA instance	It is divided into four intervals	0<Value<5 5< Value <7,5 7,5< Value <10 10< Value
AC above capacity	Percentage of aircraft exceeding declared capacity per hour	It is divided into two states, when the percentage of AC above capacity is equal to 0 and when it is greater than 0	Value = 0 Value > 0
Entry FL Distribution	Calculation of the median flight level with which AC enter the sector at the entry hour	It is divided into three intervals that represent the entry FL distribution in the sector	Value < 350 350 <Value <360 Value > 360

Parameter	Description	Discretisation criteria	Discretisation example
Horizontal Speeds Differences Distribution	Calculation of the median horizontal speeds with which AC enter the sector at the entry hour	It is divided into three intervals that represent the entry Horizontal Speed Difference Distribution in the sector	Value < 35 35 <Value <55 Value > 55
Entry Vertical Speeds Distribution	Calculation of the median vertical speeds with which AC enter the sector at the entry hour. It is calculated in absolute value	It is divided into two states, when the median of entry Vertical speed is 0 and when it is greater than 0. This discretisation is intended to represent the complexity for the controller, modelling the AC that are ascending or descending.	Value = 0 Value > 0
Changes in FL distribution	Calculation of the median percentage of AC that change their FL in the sector at the entry hour	It is divided into three intervals, that depends on the sector data.	Value = 0 0% <Value <15% Value > 15%
Delay Distribution	Calculation of the median time delayed of AC at the entry hour into the sector	It is divided into four intervals that represent the delay time (seconds) in the sector. These intervals depend on the sector data	Value =0 0<Value <240 240 <Value <420 420<Value
% Regulated AC	Calculation of the percentage of AC that have some regulation at the entry hour	It is divided into four intervals. These intervals depend on the sector data	Value <15% 15% <Value <25% 25% <Value <35% 35% <Value
% Delayed AC	Calculation of the percentage of AC that have some delay at the entry hour	It is divided into four intervals. These intervals depend on the sector data	Value <35% 35% <Value <50% 50% <Value <65% 65% <Value
ATCo actions hour	Calculation of the total actions that the controller has given in one hour in the sector	It is divided into three intervals, which are intended to represent the workload of ATCo according to the action they have undertaken. These intervals depend on the sector data	Value < 75 75 <Value <100 Value > 100
Action MIX	Sum of resolution actions of any kind. This variable considers flight level changes, vectors and directs.	Resolution actions are added, and the same three categories are defined	Value < 0.5 0.5 <Value <1 Value > 1

Parameter	Description	Discretisation criteria	Discretisation example
% AC that change FL in 5 min	Distribution of aircraft by number of flight levels that change	It is divided into three intervals	Value =0% 0%< Value <70% 70% <Value
Proportion Vertical Speed Difference	Calculation of the percentage of AC with vertical speed difference other than 0, on a scale of 0-1	It is divided into three intervals that represent the complexity when ACs are ascending or descending with different speeds.	Value <0.35 0.35< Value <0.7 0.7 <Value
Proportion Horizontal Speed Difference	Calculation of the percentage of AC with horizontal speed difference greater than 50 knots: on a scale of 0-1	It is divided into three intervals that represent the complexity when ACs have speed difference	Value <0.35 0.35< Value <0.7 0.7 <Value
ATC Workload	Combination of total actions with resolution actions	This output node is divided into four states, as defined in the first sections	Level 1 Level 2 Level 3 Level 4
Traffic Conditions	This variable has been calculated as explained in the first section	This output node is divided into three states, in order to model the traffic conditions at the conflict moment	Value <0.5 0.5< Value <0.75 0.75 <Value
No Planned AC	This variable has been calculated as explained in the first section	This output node is divided into three states from 0 to 1	Value <0.28 0.28< Value <0.35 0.35 <Value
Performance	This variable has been calculated as explained in the first section	This output node is divided into three states from 0 to 1	Value <0.35 0.35< Value <0.7 0.7 <Value
Conflict identification	It is considered that there is a detected conflict according to the conditions set out in the first section of this node	It is divided into 2 states, 0 (when there is not a detection of the conflict) and 1 (when the conflict is detected)	Value= 0 Value= 1
ATC actions in instance t	Controller actions in the time period	It is divided into 3 states to model the actions that the ACO has to perform in the period of 5 minutes after the first action on the aircraft. The first state represents a low workload, the second	Value < 3 3< value < 5 5 < value

Parameter	Description	Discretisation criteria	Discretisation example
		medium workload and the third high workload	
Dif FL t act AB	Estimated difference in flight level over the time of action	It is divided into 2 states. The firs state corresponds to the aircraft that are at the same FL at the time at which the ATCo acts on them (their difference is equal to 0). The second state corresponds to the pairs of aircraft that are at different FL at the time of the actions (their difference in absolute value is greater than 0)	Value = 0 Value > 0
Dif Vz t act AB	Estimated difference in vertical speeds over the time of action	It is divided into 2 states. The firs state corresponds to the aircraft that are flying with the same vertical speed (Vz) at the time at which the ATCo acts on them (their difference is equal to 0). The second state corresponds to the pairs of aircraft that are flying with different vertical speeds at the time of the actions (their difference in absolute value is greater than 0)	Value = 0 Value > 0
Dist t act AB	Estimation of distance between aircraft in time of action	It is divided into 4 states which represent the horizontal distance at which the aircraft pair is at the moment in which the ATCo acts on them	5< Value 5< Value < 20 20< Value < 60 60 <Value
Dif VxVy t act AB	Estimation of the difference of the horizontal speed of the aircraft in the time of action	It is divided into 3 states to model if the aircraft pairs fly with a very similar horizontal speed, the state two if the speed difference is medium and the state three for a high speed difference	0< Value <10 10< Value < 60 60 <Value
t entry to t act	Time difference between aircraft entering the sector and the first ATCo action	It is divided into three states to model the time (in seconds) that the ATCo has to assess the situation. The first state represents a short time, the second a mean time and the third a long time	t <120 120 < t< 300 300 <t
ATC Workload in instance t	ATCo's workload at the time of action	This output node is divided into four states, as defined in the first sections	Level 1 Level 2 Level 3 Level 4
t entry to t CPA	Time from aircraft entering the sector to reaching the CPA	It is divided in three states which represent whether the ACTO has had a short time (seconds) to assess the situation (state 1), has had a medium time (state 2), or has had a long time (state 3)	t <120 120 < t< 300 300 <t
ATC actions to CPA	Actions taken by the ATCo before the CPA is reached	It is divided into two states that represent whether the conflict was easy to resolve (the ATCo performs one action or less on them) or it was complex to resolve (the ATCo performs more than one action on them)	Value ≤ 1 Value > 1

Parameter	Description	Discretisation criteria	Discretisation example
Conflict Resolved	Outcome of the conflict resolution	It is divided into 2 states, 0 (when the ATCo does not resolve the conflict) and 1 (when the ATCo resolves the conflict)	Value= 0 Value= 1
t last act t CPA	Time elapsed since the ATCo's last action on the aircraft until the CPA is reached	4 states are defined to model the time that elapses from the las ATCo action to the CPA. State 1 corresponds to the negative values, which represent the cases in which the CPA occurs when the aircraft are so far apart that the ATCo does not take any action before the CPA. The next state shows when the ATCo resolves the conflict with a low lead time, state 3 with a medium lead time, and state 4 with a high lead time.	t<0 0< t <120 120 < t< 300 300 <t
STCA	Likelihood of STCA alert being triggered	It is divided into 2 states, 0 (when there is not a STCA alert) and 1 (when there is a STCA alert)	Value= 0 Value= 1
T act ATC to t STCA	Time elapsed since the ATCo's last action on the aircraft until the STCA is triggered	4 states are defined to model the time that elapses from the las ATCo action to the STCA. State 1 corresponds to the negative values, which represent that the ATCo action on the aircraft pair has been carried out after the STCA. The next states represent whether the time(seconds) between the last action and the STCA is low (state 2), medium (state 3) or high (sate 4)	T<0 0<t<60 60<t<120 120<t
Dif Vx Vy STCA AB	Estimation of the difference in horizontal speed at the time of STCA	It is divided into 3 states to model if the aircraft pairs fly with a very similar horizontal speed, the state two if the speed difference is medium and the state three for a high-speed difference when the STCA triggers	Value <10 10 < value < 60 60 < value
Dist STCA AB	Estimation of distance between aircraft at the time of STCA	It is divided into 4 states which represent the horizontal distance at which the aircraft pair is at the moment in which the STCA triggers	Value < 5 5 < value < 7.5 7.5 < value < 10 10 < value
Dif Vz STCA AB	Estimation of the difference in vertical speed at the time of STCA	It is divided into 2 states. The firs state corresponds to the aircraft that are flying with the same vertical speed (Vz) at the time at which the STCA triggers (their difference is equal to 0). The second state corresponds to the pairs of aircraft that are flying with different vertical speeds at the time of the STCA (their difference in absolute value is greater than 0)	Value = 0 Value >0
Dif FL STCA AB	Estimation of the flight level difference at the time of STCA	It is divided into 2 states. The firs state corresponds to the aircraft that are at the same FL at the time at which the STCA triggers (their difference is equal to 0). The second state corresponds to	Value = 0 Value >0

Parameter	Description	Discretisation criteria	Discretisation example
		the pairs of aircraft that are at different FL at the time of the STCA (their difference in absolute value is greater than 0)	
T entry to STCA	Time from aircraft entering the sector until the STCA is triggered	It is divided into three states to see if the alert occurs very close to the entry of the aircraft in the sector (state 1), in an medium time (state 2) or much later (state 3)	t < 180 180 < t < 360 360 > t
ATC actions STCA to CPA	Actions taken by the ATCo on the aircraft pair from the time the STCA is triggered until the CPA is reached	It is divided into two states that represent whether the conflict was easy to resolve (the ATCo performs one action or less on them) or it was complex to resolve (the ATCo performs more than one action on them)	Value ≤ 1 Value > 1
T STCA to CPA	Time from triggering STCA to reaching CPA	It is divided into 3 states that represent if the time(seconds) that elapses between the STCA and the CPA is low (state 1), medium (state 2) or high (state 3)	t < 60 60 < t < 120 120 < t
Resolution	Likelihood of conflict resolution after the STCA	It is divided into 2 states, 0 (when the conflict is not resolved after STCA) and 1 (when the conflict is resolved after STCA)	Value = 0 Value = 1

The table above included all the variables that appear in the network, whether they are input, output or training variables.
However, a summary table of the predictor variables, i.e. input and output variables for each subnetwork, is included below.

Table 6: Prediction variables

Parameter	Description	Discretisation criteria	Discretisation example
Occupancy	Occupation Sector data at the AC entry hour	It is divided into 4 intervals that represent the percentage of occupation in the sector at the AC entry hour respect to the maximum value detected in the sector. The values of the intervals depend on the sector data.	Value <40% 40% <value <65% 65% <value <85% 85% <value
Entry	Entries Sector data at the AC entry hour	It is divided into 4 intervals that represent the percentage of entries in the sector respect to the sector entry declared value at the AC entry hour. The values of the intervals depend on the sector data.	Value <55% 55% <value <65% 65% <value <80% 80% <value
Instant demand	Ac entry data in the sector in the 5 minutes period	It is divided into 4 intervals that represent the percentage respect to the maximum instantaneous demand value. The values of the intervals depend on the sector data.	Value <60% 60% <value <90% 90% <value <120% 120% <value

Parameter	Description	Discretisation criteria	Discretisation example
AC distribution by FL	Distribution of entry FL	It is divided into 6 intervals with uniform counts to model the vertical distribution in the sector. The values of the intervals depend on the sector data.	FL <250 250 <FL < 330 330 <FL <360 360 <FL <370 370 <FL <380 380 <FL
Horizontal speed distribution.	AC horizontal speed in the entry time	It is divided into 3 intervals. An interval will contain the mean value of the sector data. There will be another interval above and another below. The values of the intervals depend on the sector data	Value < 430 430< value < 470 470 < value
Vertical speed distribution	AC vertical speed in the entry time	It its divided into three intervals to separate ascending flights (>0), descending flights (<0) and stabilized flights (=0)	Value < 0 Value = 0 Value > 0
Distribution of entry times	10 minutes interval in which the AC enters the sector.	It is divided into 6 states. Each AC is assigned a value from 1 to 6, depending on the 10 minutes period in which the AC enters the sector.	Value= 1 Value= 2 Value= 3 Value= 4 Value= 5 Value= 6
Interaction	Interaction detected between aircraft	It is divided into 2 states, 0 (when it is considered that AC is not an interaction with any other AC in the sector) and 1 (when it is considered that the AC is an interaction with another AC in the sector)	Value= 0 Value= 1
Entry FL A	Distribution of entry FL for the first Aircraft	It is divided into 4 intervals with uniform counts to model the vertical distribution in the entry of the sector. The values of the intervals depend on the sector data.	FL <330 330 <FL < 360 360 <FL <370 370 <FL
Entry FL B	Distribution of entry FL for the second Aircraft	It is divided into 4 intervals with uniform counts to model the vertical distribution in the entry of the sector. The values of the intervals depend on the sector data.	FL <330 330 <FL < 360 360 <FL <370 370 <FL

Parameter	Description	Discretisation criteria	Discretisation example
Exit FL A	Distribution of exit FL for the first Aircraft	It is divided into 4 intervals with uniform counts to model the vertical distribution in the exit of the sector. The values of the intervals depend on the sector data.	FL <350 350 <FL < 360 360 <FL <380 380 <FL
Exit FL B	Distribution of exit FL for the second Aircraft	It is divided into 4 intervals with uniform counts to model the vertical distribution in the exit of the sector. The values of the intervals depend on the sector data.	FL <350 350 <FL < 360 360 <FL <380 380 <FL
Entry heading A	Distribution of entry headings for the first Aircraft	It is divided into 4 intervals with uniform widths to model the entry headings distribution in the entry of the sector.	value <90 90 < value < 180 180 < value <270 270 < value
Entry heading B	Distribution of entry headings for the second Aircraft	It is divided into 4 intervals with uniform widths to model the entry headings distribution in the entry of the sector.	value <90 90 < value < 180 180 < value <270 270 < value
Flight Levels Overlap	Evaluation of FL overlap for the aircraft pairs	It is divided into 2 states, 0 (when there is not FL overlap) and 1 (when there is FL overlap)	Value= 0 Value= 1
Convergence of headings	Assessment of heading convergence at the entry for the aircraft pairs	It is divided into 2 states, 0 (when there is not Convergence of headings) and 1 (when there is Convergence of headings)	Value= 0 Value= 1
Potential Conflict	It is considered that there is a potential conflict when there is FL overlap and convergence of headings of the aircraft pairs	It is divided into 2 states, 0 (when there is not potential conflict) and 1 (when there is potential conflict)	Value= 0 Value= 1
Vertical Distance at the CPA	Vertical distance between the aircraft pairs at CPA instance	It is divided into four intervals	D<1000 1000<D<2000 2000<D<2000 3000<D
Horizontal Distance at the CPA	Horizontal distance between the aircraft pairs at CPA instance	It is divided into four intervals	0<Value<5 5< Value <7,5 7,5< Value <10 10< Value

Parameter	Description	Discretisation criteria	Discretisation example
Vertical Distance at the CPA	Vertical distance between the aircraft pairs at CPA instance	It is divided into four intervals	D<1000 1000<D<2000 2000<D<2000 3000<D
Horizontal Distance at the CPA	Horizontal distance between the aircraft pairs at CPA instance	It is divided into four intervals	0<Value<5 5< Value <7,5 7,5< Value <10 10< Value
AC above capacity	Percentage of aircraft exceeding declared capacity per hour	It is divided into two states, when the percentage of AC above capacity is equal to 0 and when it is greater than 0	Value = 0 Value > 0
Conflict resolved	It is considered that there is a detected conflict according to the conditions set out in the first section of this node	It is divided into 2 states, 0 (when the ATCo doesn't resolve the conflict) and 1 (when the ATCo resolves the conflict)	Value= 0 Value= 1
STCA	Likelihood of STCA alert being triggered	It is divided into 2 states, 0 (when there is not a STCA alert) and 1 (when there is a STCA alert)	Value= 0 Value= 1
Resolution	Likelihood of conflict resolution after the STCA	It is divided into 2 states, 0 (when the conflict is not resolved after STCA) and 1 (when the conflict is resolved after STCA)	Value= 0 Value= 1

8 Subnetwork integration

So far, a fragmented analysis of each of the subnetworks that make up the safety barriers identified for the model has been carried out. The aim of this section is to integrate each of these subnetworks into a single compact model.

The objective is to obtain a joint network to apply different analyses and carry out studies that provide global results, since until this point, conclusions could only be drawn in isolation for each of the barriers.

As a particularity, despite the fact that interaction identification subnetwork 1 is highly relevant for the model, since it provides its starting point, it has not been included in the integration. This is because in subnetwork 1 aircraft were considered independent and not pairs as in the other subnetworks. Moreover, from this subnetwork it is only possible to draw the conclusion that aircraft are more than 20 NM away at all times, and precisely, the study maintains the criterion of evaluating the evolution of aircraft that are less than this distance away. Therefore, in order to keep the input criteria and boundary conditions consistent across all distance assessments, this subnetwork does not need to be included in the global model.

The networks integration is supported by the event tree and the barrier model approach presented in the conceptual framework in chapter 5 and synthetized in Figure 45. This scheme will be followed, completing the probabilities, so far identified as "p" and "1-p", with the actual values obtained from the outputs of each of the subnetworks.

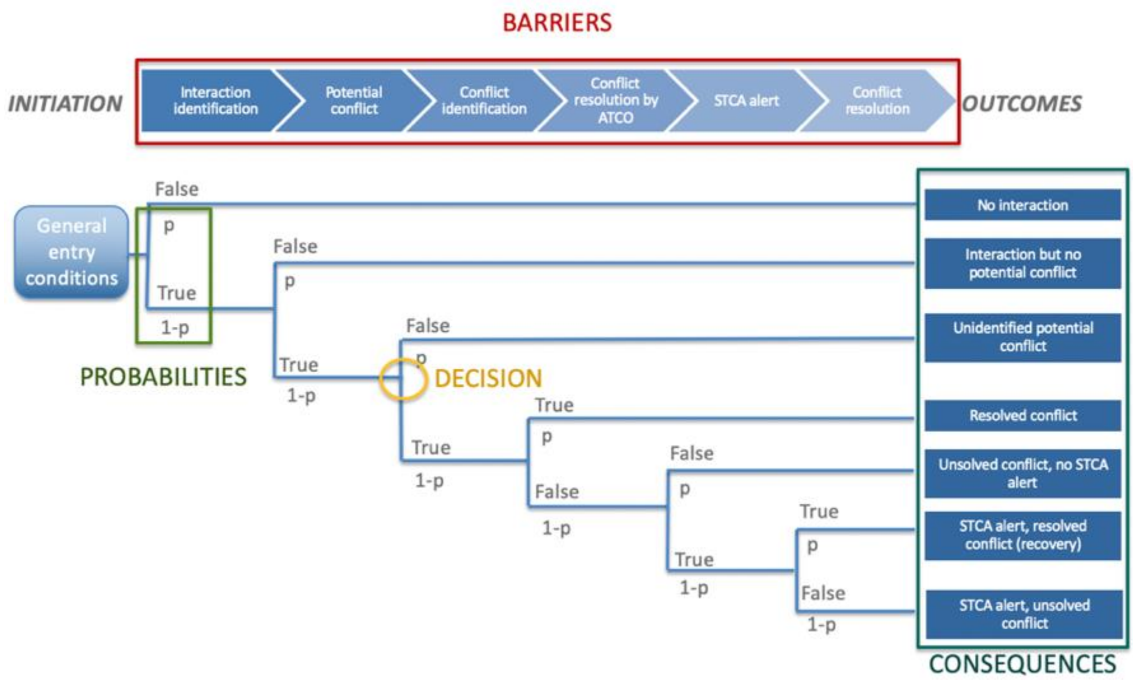


Figure 45: Data driven event tree

The outputs through each of the branches will be calculated, as can be seen in the Figure 46, as a conditional probability.

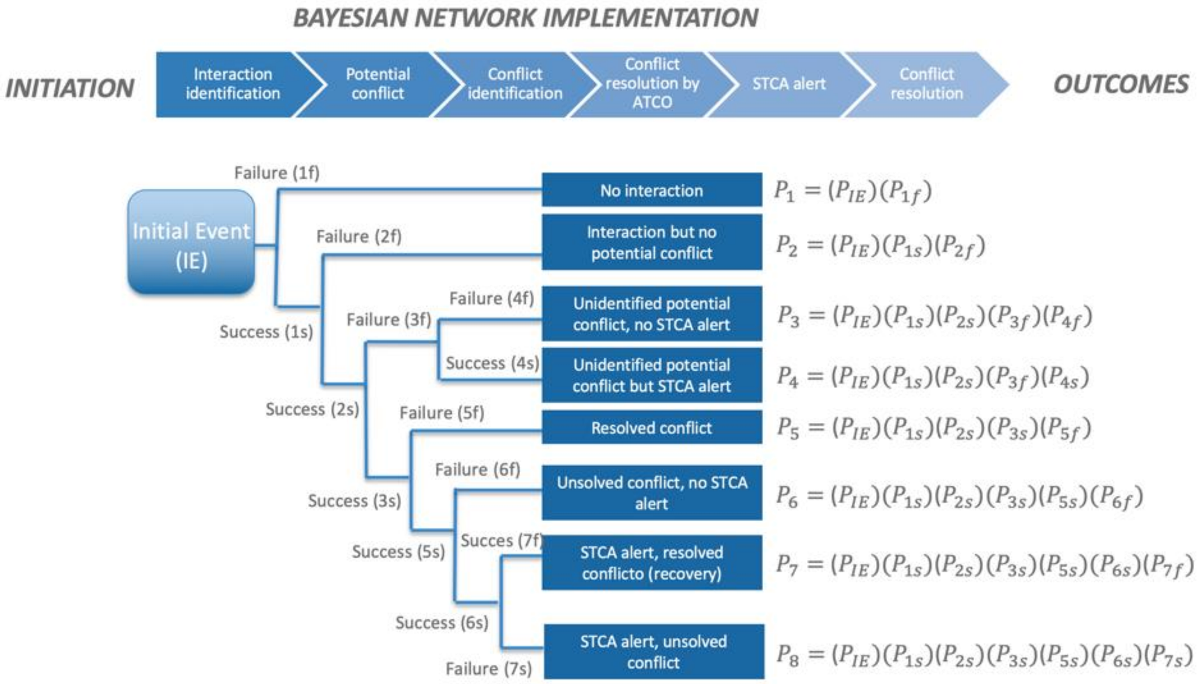


Figure 46: Conditional probabilities

8.1 Methodology and application of the integration model

The integration model is applied to a particular case that serves as an example. To do this, in the first place, the probabilities of the output of each of the barriers have been obtained directly from the subnetworks described in the previous chapters of this document.

For example, the value of the output of subnetwork 2, in which the probability of an aircraft pair constituting a potential conflict was evaluated, as can be seen in figure Figure 47.

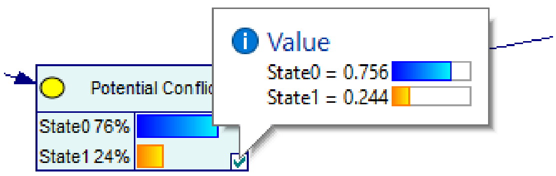


Figure 47: Probability of subnetwork 2. Potential conflict

In Figure 48, probabilities in the event tree model have been populated with the values of “p “ and “1-p” obtained from the subnetworks that calculates the probability of failure of the ATM barriers.

In the right-hand column, the conditional probability outputs of each tree branches can be seen. These outputs are the result of the probability multiplication by traversing the branch of the event tree.

It can be seen that the sum of all the outputs considered is equal to 1 as presented theoretically.

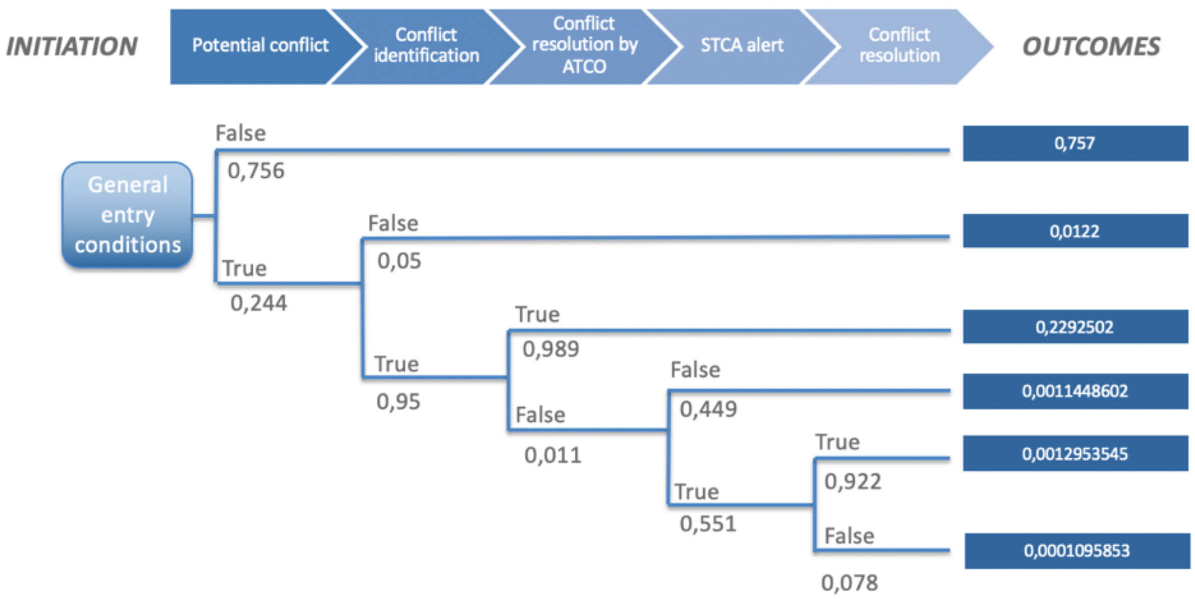


Figure 48: Event tree with sample data

These final outputs represent the weighing that each one of the events has on the complete network integration. Therefore, their corresponding separation distributions, both vertical and horizontal, will be multiplied by these weights.

Figure 49 shows the GENIE software interface directly with the integration equation of the horizontal distance distributions. As can be seen, each node is multiplied by its corresponding weight.

```
Final_Horizontal_Distribution=Horizontal_Distribution2A*0.756
+Horizontal_Distribution3A*0.0122+Horizontal_Distribution4A*
0.22925+Horizontal_Distribution5A*0.00114486
+Horizontal_Distribution6A*0.00129535
+Horizontal_Distribution6B*0.000109585
```

Figure 49: Integration equation

The integration of complex networks in GENIE is a complex process that required some intermediate steps and modelling. The intermediate process carried out by the software to obtain the complete integration of the network is detailed hereafter.

The distance distribution node of each of the discretised subnetworks is the starting point. To carry out the integration, these nodes are transformed into sums of uniform distributions.

Each of these uniform distributions will be bounded by the ends of each of the discretisation intervals. The process followed by the software to carry out the summation of uniform distributions consists of keeping the percentages of the initial states constant, giving random values within the established interval.

Once the sums of uniform distributions have been obtained for each node, they are integrated into the distribution node of distances between aircraft in the CPA, both vertically and horizontally. To do this, the weighted sum of each distribution by its weight is carried out as described above.

This process is shown in Figure 50 and Figure 51 for vertical and horizontal distances, respectively.

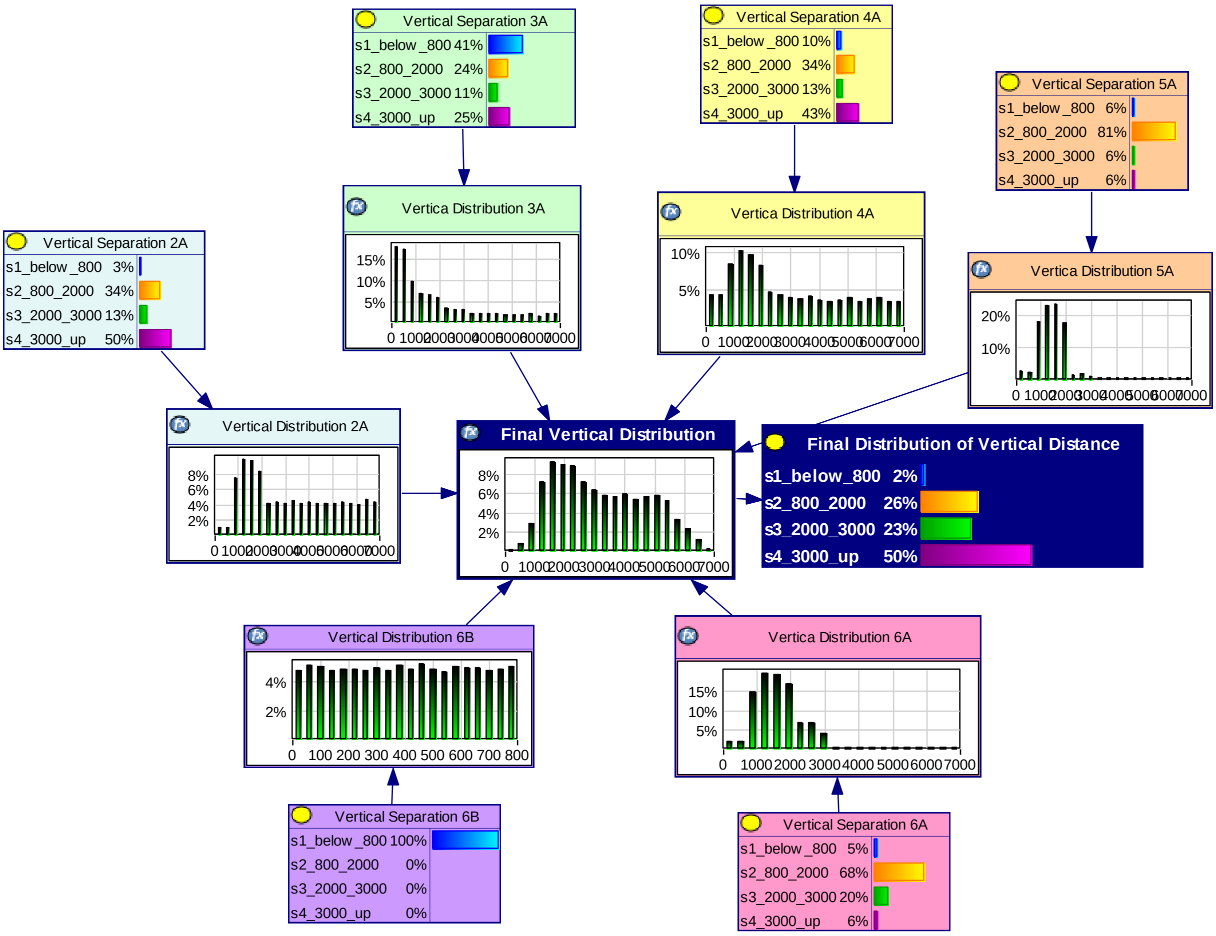


Figure 50: Integration of vertical distances

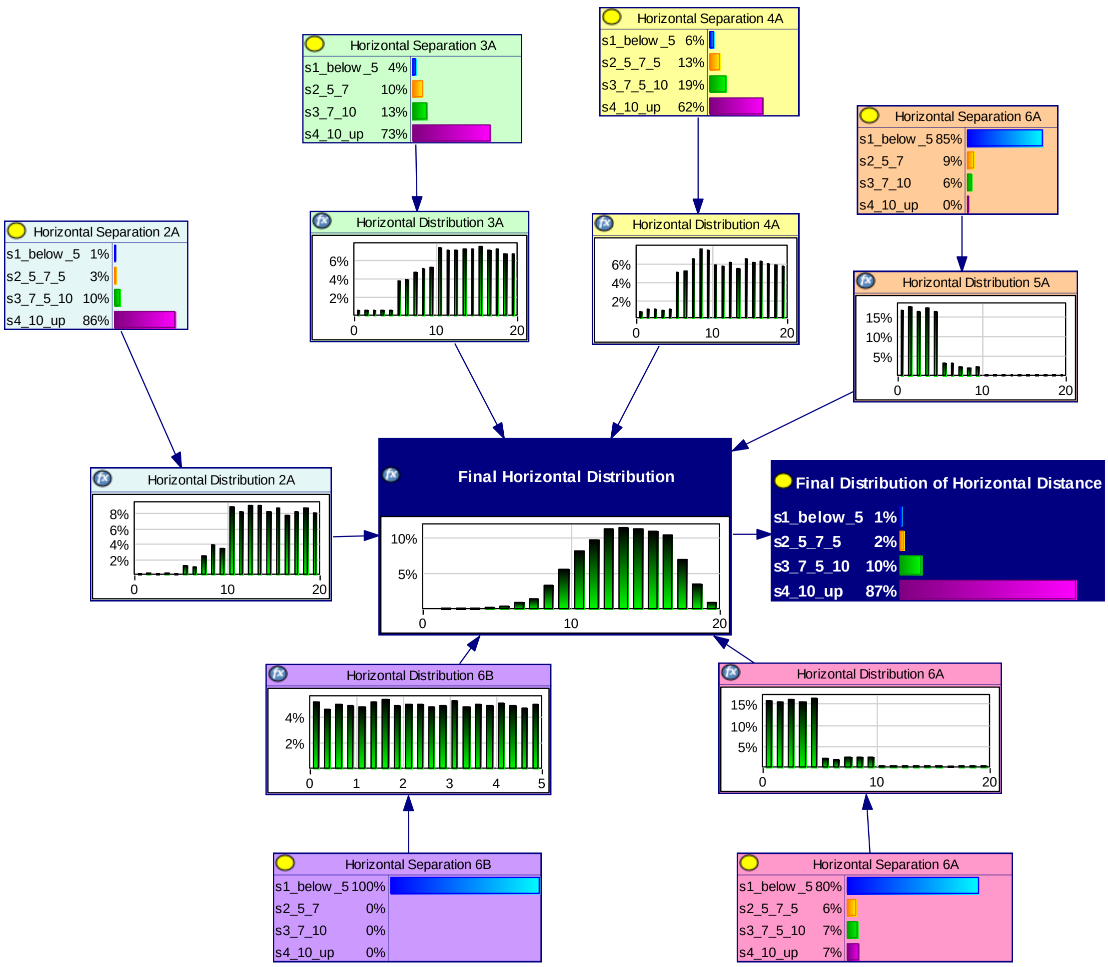


Figure 51: Integration of horizontal distances

Once the final distribution of distances has been obtained, the inverse transformation is carried out to obtain a final discretised node in four states. Both the final distributions and their discretised nodes can be located in the centre and right of the attached graphs in dark blue.

9 Conclusion and next steps

FARO WP4 aims at generating predictive models of safety events, that we call Safety Performance Functions (SPFs), by using organisational, technical, human, and procedural precursors to characterise and predict airspace Separation Minima Infringement (SMI), as a function of those precursors.

This document has described the process followed to develop a baseline model of the SPF for the characterisation and prediction of airspace Separation Minima Infringement (SMI) in particular ATC sectors. The document summarises the mathematical background and the methodological approach followed, the process followed to build up the model, and serves as the underlying framework for subsequent deliverables. The report covers selected SPFs as a function of the safety dimensions (precursors) and their aggregation, the results from the adjustment and calibration processes, and the sensitivity analyses with respect to the independent dimensions.

The term Safety Performance Function (SPF) is used in many industries to refer, in a general way, to mathematical models that have the capacity to explain, but above all, predict the occurrence of safety events. The expression has been coined in the field of road transport, with the development of models to predict the occurrence of traffic accidents, and over the years its applications have expanded into new fields such as railway or aviation.

These models can be conceptually extrapolated to any airspace, considering different characteristics: the geometry of the routes followed by the aircraft, the volume of traffic, the mix of traffic and its dynamic variables, the geometry of the encounters between aircraft, the severity or magnitude of Separation Minima Infringement between aircraft, the complexity of the airspace structures, the size and characteristics of the sectors where the aircraft are flying, the complexity of the organization and management of the airspace, etc. Among the various possible techniques with high predictive capacity **Bayesian Networks** (BN) have been selected as the more appropriate for this purpose.

To develop a Bayesian Network model for such a complex problem as SMI prediction has not been straightforward. It has been necessary to set up a conceptual framework that integrates the current available knowledge about SMIs causality and precursors with the hindsight derived from the analysis of the type of data available in the project, particularly those that reflect the ATCo interventions.

The conceptual framework that backs up the proposed BN model considers the general scenario where aircraft routes evolve and focuses on the analysis of the Closest Point of Approach (CPA), for any possible aircraft pair in an air traffic sector, and on the understanding and quantification of the process that leads to such CPA.

The three main elements in the conceptual framework have been considered. The actual final CPA between an aircraft pair can be interpreted as the outcome of a process where the expected aircraft trajectories become modified as the results of the ATCo intervention. It may be then considered the CPA between an aircraft pair as the actual shortest distance between those two aircraft, expressed as vertical separation and horizontal separation. This magnitude is called "final CPA". It can be also calculated what the CPA would have been between this aircraft pair if both had followed their planned trajectories without any modification or ATCo intervention. This magnitude is named "prior CPA". The difference between both magnitudes, final CPA and prior CPA, is attributed to alterations of the expected trajectory that are induced by ATCo intervention.

To translate the conceptual framework into a set of causal subnetworks, the concepts of **ATM barrier model** and **event trees** have been incorporated in the model. Based upon the available operational data the barriers considered in this model refer to: Assessment of potential conflict, Conflict identification, Conflict resolution by the ATCo, STCA, and Conflict resolution after STCA. Each ATM barrier is analysed in a bespoke BN and all of them are finally integrated following the structure of the proposed event tree.

Two types of subnetworks have been built. In the first place, those destined to estimate the probability of an event occurring and, on the other hand, those destined to estimate the vertical and horizontal distance at the CPA between aircraft. The first type of subnetwork explains the modelling of the ATM barriers and its effectiveness. The second type explains the modelling of the outcomes of the event tree and the probability distribution of the vertical and horizontal separation between the pairs of aircraft included in each outcome. The whole model is composed by a set of 24 subnetworks integrated into a big BN.

The outputs of the network are characterised by probability distributions. For a given state of the input variables, the model will predict the predicted probability of success for the ATM barriers; for example, the probability of interaction between aircraft, probability of potential conflict, probability of detection of the conflict, probability of resolution of the conflict, etc. Also, for a given state of the input variables, the model will predict the probability distribution of the aircraft's vertical and horizontal separation at its CPAs. The inputs to the model are also characterised as probability distributions and refer to the general conditions that characterise the sector and its traffic (for example the distribution of aircraft entries in the sector, the distribution of FL at the point of entry of aircraft in the sector, the distribution of aircraft speed, etc.)

Thus, with this calculation capacity, the network would allow conclusions to be drawn about the impact that a modification in the network entry conditions would have on the effectiveness of the barriers and on the final distance distributions between aircraft in the CPA, thus estimating the probability of SMI.

The global network allows for forward and backward analysis. In a backward analysis, the model will be used to deliver a particular configuration in the input variables by setting the result variables to a target value. In forward analysis, the model is used to predict the outcome variables by establishing the probability distribution of the input nodes.

Once the generic BN model has been developed and described into this Deliverable 4.1 the following activities will be accomplished ins task 4.3 and documented into a future Deliverable 4.2:

1) The adaptation of the generic model to the characteristics of the sectors analysed in Use Cases 1 and 2.

The SPF is sector dependant, what means that the generic BN model needs to be adapted for each ATC sector. That means that the conditional probability tables in the model are specific to each Use Case and have to be learned from data of the specific sector under analysis. Adaptation of the generic model to the characteristics of the sector analysed in each Use Case implies:

- Gathering and processing all the data from the sector/traffic under analysis in the Use Case. Processing involves the discretisation of the data for each model variable. Discretisation scheme is knowledge driven, it cannot be automated and it might be different for each sector because it depends on sector features and traffic profile.
- Parametric learning, i.e. obtaining the required a priori and conditional probabilities tables from the frequency observed directly from data.
- Sensitivity analysis using the tornado diagram to tune the network and identify the most influential variables for each particular sector.

- Backward and forward analysis to define thresholds in the variables that might impact safety performance in a scenario, if applicable.

This approach will be carried out subnetwork by subnetwork, so that particularized information can be acquired for each of the defined barriers and lead to a summary of the network as a whole.

Upon completion of the above work, the BN will be useable for a particular use case and validation activities can be performed.

2) Derivation of the most Influencing Factors and Applicability Thresholds in each of the models.

One of the main task still remaining in WP4 is the application of the model to the study cases or scenarios defined in the project to quantify the influence factors of each study case and determine the criteria and thresholds for its applicability.

For each subnetwork the variables that have the largest influence in the subnetwork outcome will be identified. A systematic +/-10% probability variation will be applied to each state of each variable, so the ones producing a highest change in the outcome variable can be identified and retained. The percentage of variation will be used to set some threshold of variables or states were applicable. This analysis will be accomplished in task 4.3 and properly documented in deliverable 4.2.

3) Validation of the model

The basic concept in the validation of the SPF is the goodness of fit of the model. The proposed model is indeed and hyper-structure with a huge combination of BN sub-networks. Convectional BN goodness of fit techniques have been applied during the development and tuning of each individual sub-network. The validation of the model itself accounts for the goodness of fit of the entire model, that is, the hyper-structure of 19 Bayesian subnetworks. For such a complex structure, a specific validation approach is needed.

Taking inspiration from classical validation techniques, two different scenarios will be defined to test the goodness of fit of the entire BN hyper-structure in each Use Case. The two scenarios will represent extreme conditions of the use case. For each scenario, a representative data set will be separated from the learning data, and subsequently used to test the predictive ability of BN.

Scenario 1 will consist of a set of data with high occupancy rate in the sector and Scenario 2 will consist of a set of data with low occupancy rate in the sector.

The predicted outcomes of the BN (Predicted probability of success of the ATM barriers included in the model and Predicted probability distribution of the vertical and horizontal separation of the aircraft at their CPAs) will be compared with the actual ones, to determine the validity of the statistical approach.

4) Development of a graphical scheme to presents the outcomes of the model.

The SPF developed is a complex tool composed by 19 interconnected BN. AS part of task 4.3 it will be developed a graphical synthetic representation of the outcomes of the model that allows a quick, direct and easy to read, interpretation of the result predicted by the network.

The graphical representation will have the form of a triptych, that presents a schematic graphic summary of the inputs and outcomes of the model developed so far. The left side of the triptych will summarise the input variables. Inputs refer to variables that account for the general conditions that characterise the sector and its traffic, and that will be characterised as probability distributions, for example the distribution of aircraft entries in the sector, the distribution of FL at the point of entry of aircraft in the sector, the distribution of aircraft speed, etc… For a given state of the input variables, the central part of the triptych shows the predicted probability of success for the ATM barriers; for example, the probability of interaction between aircraft, probability of potential conflict, probability of detection of the conflict, probability of resolution of the conflict, etc. For a given state of the input variables, the right side of the triptych will show the predicted probability distribution of the aircraft's vertical and horizontal separation at its CPAs.

With this type of graphics representation all the information about a predictive scenario can be summarised into a single unique graphic, that provides information about the scenario and its traffic and about the impact of safety in terms of vertical and horizontal separation distribution and probability of success /failure of the ATM barriers.

5) Application of the model to different study cases.

Once the model has been tuned for each scenario and it goodness of fit has been validated, it can be used to assess the impact of changes in a scenario on the its safety performance.

Every change to be analysed in a scenario will be feed into the model as a variation in the probability distribution of the parent nodes. The main use of the model is to predict the effects, that is, the probability of the output-child nodes by setting the probability distribution of the parent-input nodes.

Given the new probability distribution of the various input nodes, then the model will propagate these uncertainties through the network and will cause a change in the probability distribution in the outcomes of the network. By varying the probability distribution of the input nodes it will be possible to predict the probability distribution of the outputs as summarised the following table.

Input variables	Outcome variables	
General features that characterise the scenario and its traffic	Predicted probability of success of the ATM barriers included in the model	Predicted probability distribution of the vertical and horizontal separation of the aircraft at CPAs.
Input variables as defined in chapter 7 of this document: e.g. Occupancy, Entries, AC distribution by FL, etc	Probability of aircraft Interaction defined as two aircraft within 20 NM of each other. Probably of Potential conflict, based upon predicted trajectory, without the action of the controller. Probability of conflict identification Probability of conflict resolution Probability of STCA alert Probability of Conflict resolution after STCA	Probability distribution of the vertical separation between aircraft Probability distribution of the horizontal separation between aircraft, for those with vertical separation less than 800 feet.

10 References

Bujor, A. a. (2016). Assessing Air Traffic Management Performance Interdependencies through Bayesian Networks: Preliminary Applications and Results. *Journal of Traffic and Transportation Engineering , 4*.

Chen F., S. Y. (2012). Study on Safety Assessment of Air Traffic Management System Based on Bayesian Networks. *3rd International Conference on Teaching and Computational Science .*

Cooper, G. H. (1992). A Bayesian method for the induction of probabilistic networks from data. Mach Learn 9. *Mach Learn 9*, 309–347.

E. Castillo, J. G. (1997). Sensitivity analysis in discrete Bayesian networks. *IEEE Trans. Syst. Man Cybern. Part A.*

F. Flammini, S. M. (2006). *Modeling system reliability aspects of ERTMS/ETCS by fault trees and Bayesian networks.* London,: Taylor & Francis Group,.

FARO project. (2020). *D3.1 Data Pipeline.*

FARO project. (2021). *D2.2 Case Studies.*

GARCÍA, M., GARCÍA-OVIES, I., VERDONK, C., & GARCÍA, J. (2020). *D3.1 Data Pipeline.*

Gomez Comendador, V., Arnaldo Valdés, R., Villegas Diaz, M., Puntero Parla, E., & Zheng, D. (2019). Bayesian Network Modelling of ATC Complexity Metrics for Future SESAR Demand and Capacity Balance Solutions. *Entropy , 21*, 379.

Hauer, E. H. (2002). Estimating safety by the empirical Bayes method: a tutorial. . *Journal of the Transportation Research Board,* , (1784), 126-131.

Heckerman, D. G. (1995). The Combination of Knowledge and Statistical Data. *Machine Learning,20*, 197-243.

Kaggle. (s.f.). Obtenido de https://www.kaggle.com/christopherwsmith/how-to-predict-lung-cancer-levels-100-accuracy

Khair S. Jadaan, L. F.-M. (June 4-5, 2014). Modeling of Accidents Using Safety Performance Functions. *International Conference on Agriculture, Environment and Biological Sciences (ICFAE'14).* Antalya (Turkey).

Kjaerulff, U. &. (2000). Making Sensitivity Analysis Computatinally Efficient. *Sixteenth Annual Corference on Uncertainty in Artificial Intelligente (UAI)*, (págs. 317-325).

Neil, M. &. (2003). *Modelling an Air Traffic Control Environment Using Bayesian Belief Networks. .*

P. Truccoa, ,. E. (2008). A Bayesian Belief Network modelling of organisational factors in risk analysis: A case study in maritime transportation . *Reliability Engineering and System Safety , 92*, 823–834 .

Pearl, J. (1998). *Probabilistic Reasoning in Intelligent Systems.* San Mateo, CA, USA: Morgan Kauffman.

Perrin, E., & Kirwan, B. (2007). A Systemic Model of ATM Safety: The Integrated Risk Picture.

R Arnaldo, F. G. (2019). Development of safety performance functions (SPFs) to analyse and predict aircrafts loss of separation in accordance with the characteristics of the airspace. *Safety Science*, Vol 186.

Rosa María Arnaldo Valdés, S. Z. (2018 Dec. 14). *Application of Bayesian Networks and Information Theory to Estimate the Occurrence of Mid-Air Collisions Based on Accident Precursors* . Entropy (Basel), 20(12):969. doi: 10.3390/e20120969.

Rosa Maria Arnaldo Valdés, V. F. (s.f.). *Development of safety performance functions (SPFs) to analyse and predict aircraft loss of separation in accordance with the characteristics of the airspace.* Reliability Engineering & System Safety, Volume 186, 2019, Pages 143-161, ISSN 0951-8320.

Sacchi, E. &. (2015). Bayesian Estimation of Conflict-Based Safety Performance Functions. . *Journal of Transportation Safety & Security.* .

Sucar, L. E. (2015). Redes Bayesianas, Capítulo 1. Sta. María Tonantzintla, Puebla, México.

Villaz´an, E. M. (2017). *A Bayesian Network Approach for Probabilistic Safety Analysis of Traffic networks. Phd Thesis.* UNIVERSITY OF CANTABRIA, Department of Applied Mathematics and Computational Sciences. Santande r, Spain.

Annex I: Detailed description of the safety model subnetworks

This annex shows the description of all the subnets that make up the safety model exposed in the document.

As in the example in chapter 6, the construction of the subnet, its parametric learning and the sensitivity analysis to detect the most influential variables are shown for each of the subnetworks.

The *Parametric Learning* and *Sensitivity Analysis* for each subnet have been marked as **confidential** by the data provider.

1. Subnetwork 1: Assessment of aircraft interaction

1.1. Description and objective

The aim of this subnetwork is to identify the probability that an aircraft becomes an interaction with another aircraft before the ATCo acts on any of them. In this study, an aircraft pair is considered an interaction if at any point of their trajectories in the sector they are at a distance of 20NM or lower.

This subnetwork is necessary due to the nature of the data available to build the model, as only data detailed data were available for those pairs of aircraft that, at any given time, have been at a distance equal to or less than 20NM. This limit of 20NM has been determined by CRIDA based on previous work, considering that only the pairs of flights that are less than 20NM. For every one of these pairs of aircraft, vertical (ft), horizontal (NM), and 3D (NM) separations are provided, as well as the latitude, longitude, and flight level of each aircraft, every 5 seconds. This subnetwork is aimed to reproduce the filtering process carried out at the CRIDA data warehouse. It is to be noticed that depending upon the data source, this network might require some fine tuning.

The input data to this subnetwork are either general conditions of the sector, or conditions of the traffic when they enter the sector. Namely, all variables that are actually known at the time the aircraft enters the sector and that are considered to have a direct influence on the probability of two aircraft interacting.

For those aircraft that do not constitute a pair with another, the only available information is that they are at a distance greater than 20NM. It is possible to determine the percentage of aircraft that constitute an interaction or not.

1.2. Network construction

As introduced above, the starting point is the sector entry condition data of an aircraft and the general sector conditions during the aircraft entry time. In addition, this information will include whether two aircraft have had an interaction, so this data will be used as training data for the network.

According to the event tree model this subnetwork will generate two different branches “p” and “1-p”. Thus, it is intended to obtain the probability "p" that a pair of aircraft does not involve an interaction, in which case it would only be possible to know that these aircraft are more than 20 NM horizontally. The subnetwork will provide also the probability “1-p” that a pair of aircraft becomes an interaction. The aircraft on the first branch “p” will be further analysis in subnetwork 1A, while aircraft in the branch “1-p” will be further analysed in the subnetwork 2.

For non-interacting aircraft, the distance assessment will only be able to ensure that all of these ACs will be in the major category of horizontal separation (>20NM).

Thus, the construction of the network is based on:

Network entries: Input data of sector macroscopic conditions.

Training data: For each AC it is checked whether it has formed a pair with any other in the pairs file.

Network exit: Probability of interaction. This node has two states: yes / no.

Figure 52 shows in a schematic way the concept of subnetwork 1.

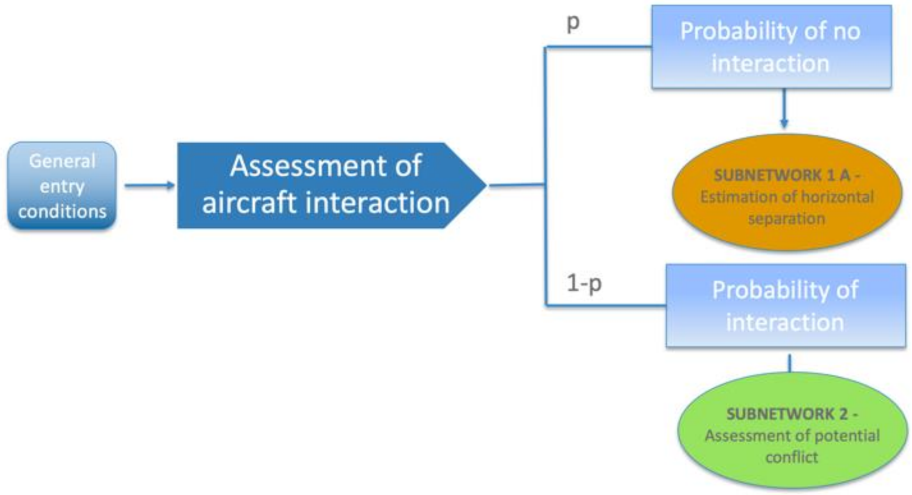


Figure 52: Assessment of aircraft interaction network

Figure 53 shows the proposed Bayesian network scheme. The selected input variables, which will be explained in more detail in later sections, and the causal relationships between them can be observed. These causal relationships are represented by the arrows linking the input variables, i.e. the influence that one variable may have on another. These causal relationships have been obtained by the model directly from the input database and completed by expert judgement.

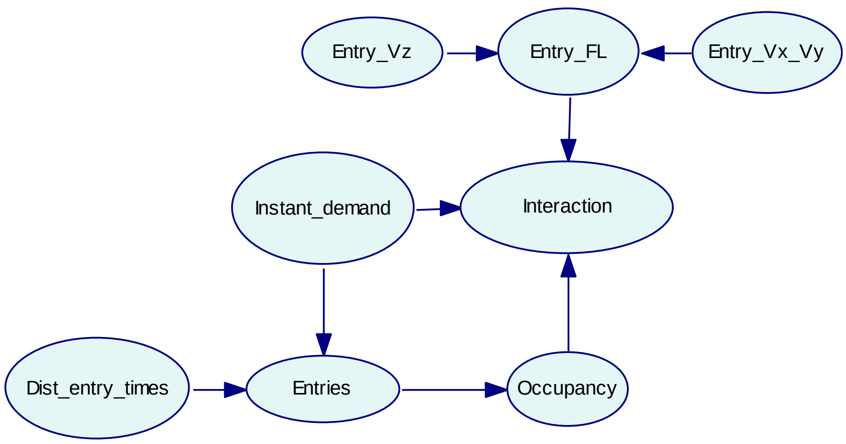


Figure 53: Subnetwork 1 structure

In the causal relationships that can be seen in the network structure, it should be noted that both the horizontal speed (“Entry_Vx_Vy”) and vertical speed (“Entry_Vz”) of the aircraft are linked to the entry flight level (“Entry_FL”). Higher flight levels are generally associated with higher speeds.

On the other hand, the network shows the link between the entries every five minutes (“Instant_Demand”) and the hourly entry count (“Entries”). It also shows the relationship between the hourly entry count (“Entries”) and the occupancy (“Occupancy”). The distribution of entry times in the sector over a period of 10 minutes (“Dist_entry_times”) is linked to the hourly entry count (“Entries”).

Finally, three variables have a direct impact on the exit variable (“Interaction”): entry flight level (“Entry_FL”), entries every five minutes (“Instant_demand”), and occupancy (“Occupancy”).

1.3. Input and output variables and states

The input variables that have been selected for the definition of this first subnetwork are intended to model the general conditions of the sector at the time of entry of the aircraft, as well as the traffic conditions at that time.

In a first approximation, the input variables could be divided into four main areas, which are the ones shown below:

- **Hourly traffic volume:** It aims to represent the volume of traffic in the studied sector, to see its influence on a possible interaction between aircraft. This area will be represented by two variables: the occupancy (“Occupation”), and the hourly entry count, i.e., the number of aircraft that enter the sector in an hour (“Entries”).
- **Instant traffic volume:** It aims to model the volume of the traffic the controller would have to attend simultaneously or in short periods of time of 5 minutes. This area will be represented by two variables: the instance demand (“Instant_demand”), which represents the aircraft that enters the sector in a five-minute period, and the distribution of aircraft sector entry times in a period of ten minutes (“Dist_entry_times”).
- **Speeds**: Both the vertical speed (“Entry_Vz”) and horizontal speed (“Entry_Vx_Vy”) of the aircraft at the entry of the studied sector will be considered.
- **Flight levels**: The network considers the flight level at which the aircraft enters in the studied sector and models the vertical occupation in the sector. This will be represented with the variable Flight Level (“Entry_FL”) that shows the vertical distribution of the aircraft FL when entering in the sector. For this, the flight level at which aircraft enter in the studied sector will be considered, and in this way also model the vertical occupation in the sector.

1.4. Parametric learning

Confidential

1.5. Sensitivity analysis for fine tuning

Confidential

2. Subnetwork 2: Assessment of potential conflict

2.1. Description and objective

The objective of the subnetwork is to identify the probability of an aircraft pair being a potential conflict before the ATCo acts. For this purpose, the sector entry conditions of the aircraft pair and the relative conditions between them will be considered. According to the ATM barrier model (Figure 23) this event is posterior to the event "Interaction", therefore this network is concerned with aircraft pairs that were identified as interactions in subnetwork 1.

In other words, the aim is to analyse whether, due to the geometrical and dynamic condition of two aircraft, the pair will constitute a possible potential conflict. This subnetwork will only process a subset of the aircraft considered in the previous subnetwork (subnetwork 1). In particular, only those pairs of aircraft belonging to branch "1-p" in subnetwork 1, i.e., the aircraft pairs that became an interaction in the previous subnetwork

2.2. Network construction

As mentioned above, the starting point is the data on the sector entry conditions of an aircraft pair and the relative conditions between them.

Subnetwork 2 is a particular type of BN, in which the central node involves a deterministic relationship with two conditions that have to be fulfilled simultaneously.

A potential conflict is defined if an aircraft pair is expected to lose their separation based only on their planned trajectories. Potential conflicts are expected to require higher attention and ATCo intervention to provide and monitor aircraft separation. The data initially provided by CRIDA did not include this information for any aircraft pair. Therefore, a simple heuristic has been defined in this subnetwork. An aircraft pair in the same sector are considered a potential conflict if two conditions are met simultaneously: 1) their headings are convergent, i.e., the difference in headings is less than 180° and, at the same time, 2) the entry and exit flight levels of the two aircraft overlap. Thus, a new binary variable called "potential conflict" will be generated, which will take the value 1 only when the two conditions mentioned above are met. If the conditions for a "potential conflict" change, the subnetwork 2 should be modified accordingly.

This subnetwork will generate two different branches in the event tree, "p" and "1-p". The aircraft that do not involve any potential conflict, included in the first branch "p" will be further analysed in subnetwork 2A. While aircraft that do represent a potential conflict, indicated by the branch "1-p," will be further analysed in subnetwork 3.

Thus, the construction of the network shown in Figure 54 is based on:

- **Network entries:** Flight level and heading data for both aircraft in each pair.
- **Training data:** For each aircraft pair, the range of flight levels and headings are calculated, and an assessment is made as to whether the conditions for a potential conflict are met.
- **Network exit:** Probability of conflict. Node with two states "yes/no".

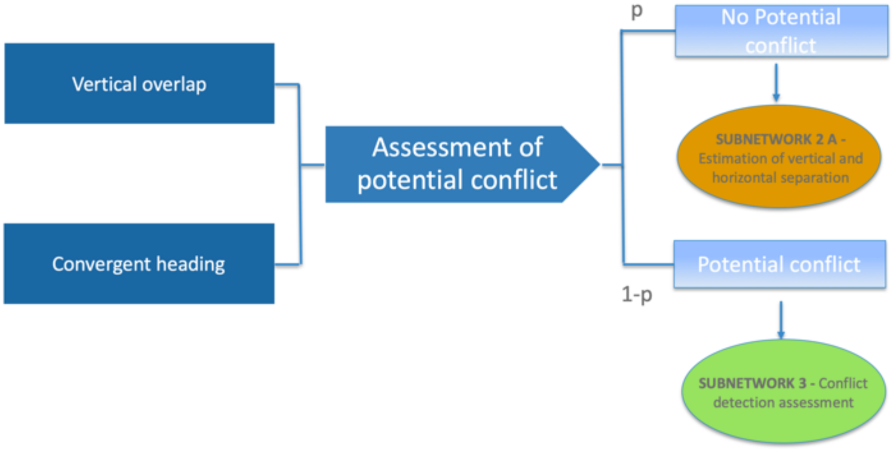


Figure 54: Assessment potential conflict network

Figure 55 shows the proposed Bayesian network scheme. It can be seen how the variables that directly affect the calculation of a potential conflict are in turn determined by other variables.

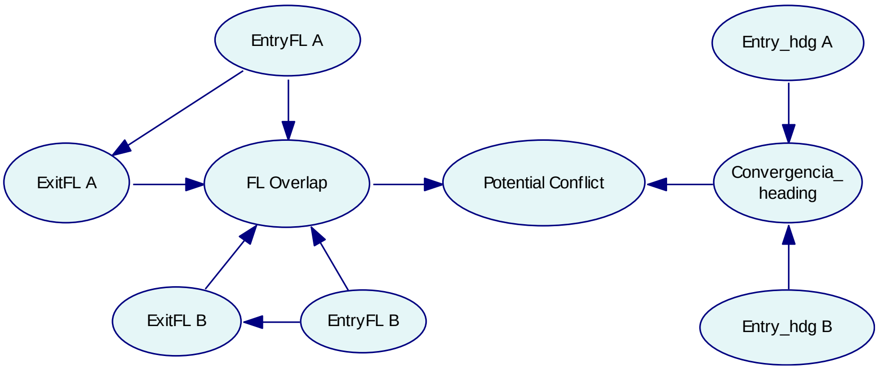


Figure 55: Subnetwork 2 structure

Specifically, the flight level overlap is defined by the entry and exit flight level of each aircraft. In addition, the entry flight level will condition the exit flight level.

On the other hand, the convergence of headings only considers the heading of each aircraft at the sector entry.

2.3. Input and output variables and states

In a first approximation, the input variables could be divided into two main areas:

- **Flight Levels Overlap:** This area aims to determine if the flight levels of an aircraft pair overlap. This area considers as variables the sector entry flight level ("Entry_FL") and the sector exit flight level ("Exit_FL") for each aircraft in the pair (aircraft A and aircraft B). Two intermediate variables have been created: Flight level interval aircraft A and Flight level interval for aircraft B, which will not appear in the network. These intermediate variables consider the entry and exit flight level in the sector. There is overlap of flight levels if the flight level intervals of both aircraft intersect.

- **Convergence of headings:** This area aims to determine the heading convergence between pairs of aircraft. This area considers as variables the heading of each aircraft when entering the sector ("Entry_hdg_A" and "Entry_hdg_B"). One intermediate variable has been defined, which will not appear in the network: Difference of headings. This intermediate variable calculates the difference between the courses of both aircraft when entering the sector. If the difference is between 0 and 180 degrees, they do not converge, and if it is between 181 and 360 degrees, they do converge.

Detailed definition of all variables and its discretisation states is provided in the chapter 7.

2.4. Parametric learning

Confidential

2.5. Sensitivity analysis

Confidential

3. Subnetwork 2A: Distance evaluation between aircraft. Cases of non potential conflict

3.1. Description and objective

The objective of this subnetwork is to assess the actual CPA horizontal and vertical separation for those pairs of aircraft not considered as potential conflicts in subnetwork 2. The separation distribution will depend on the aircraft's trajectories and the aircraft's conditions when entering the sector.

The study of the vertical separation will be carried out separately from that of the horizontal separation, so it will actually be two Bayesian networks.

3.2. Network construction

The vertical and horizontal separation distributions are modelled with two decoupled BNs. First, to estimate the vertical separation distribution at which aircraft were found at the CPA, all aircraft pairs that were not identified as potential conflicts in the previous subnetwork are considered. This network considers mainly variables related with the vertical dimension of the aircraft movement.

To proceed with the second network, data are segregated. For the estimation of the CPA horizontal distance distribution, only aircraft pairs with less than 1000 vertical feet separation are retained. It has to be remembered that the horizontal separation follows a uniform distribution for those pairs of aircraft whose vertical separation at CPA is 1000' or higher. This second BN considers principally variables that accounts for the horizontal dimension of aircraft movements.

Thus, the construction of the two networks is showed in Figure 56 and based on:

- **Network entry:** Relative vertical or horizontal condition data of each aircraft pair at the sector entry point.
- **Training data:** The vertical and horizontal separation at the CPA for each aircraft pair.
- **Network exit:** Vertical and horizontal distance between aircraft at the CPA.

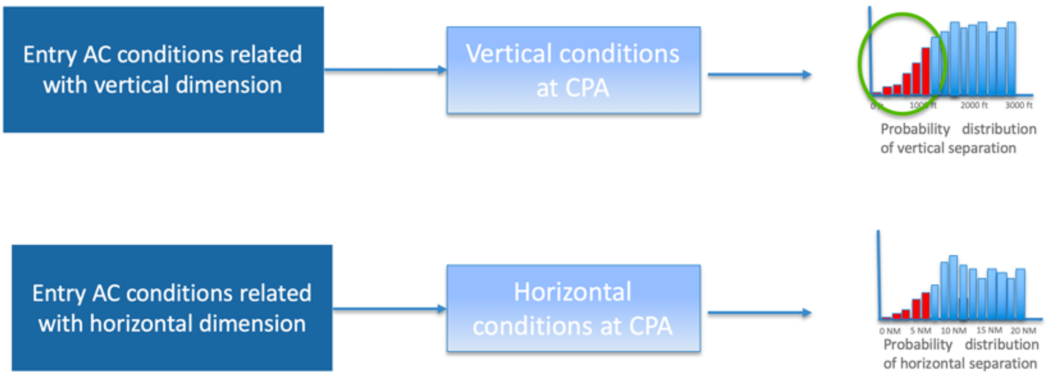


Figure 56: Assessment of distance in "no" potential conflict network

The Bayesian network model for the vertical dimension is presented first. For this network, only data on the relative vertical conditions between aircraft at the sector entry point for each aircraft pair are considered.

The links representing the causal relationships between variables have resulted from a combination of the model proposals and the experience provided by expert judgement.

Figure 57 shows the structure for the vertical Bayesian Network:

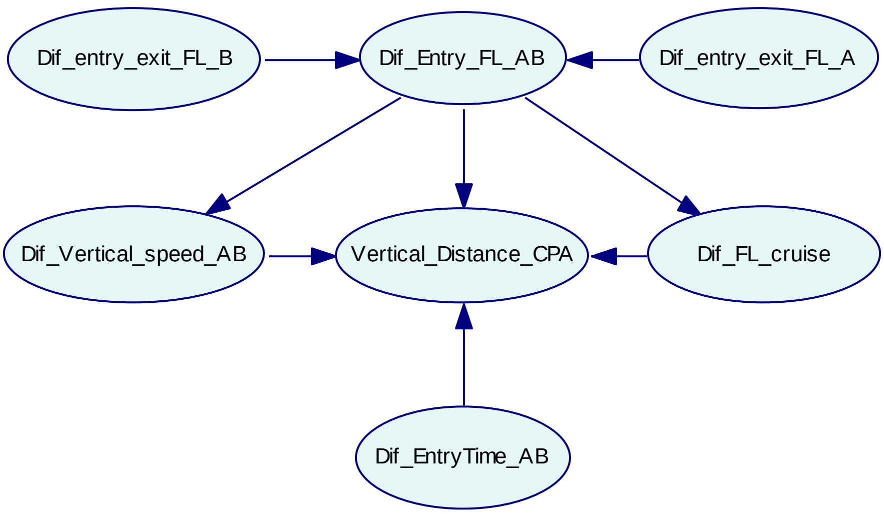


Figure 57: Subnetwork 2 A. Vertical Bayesian Network

Figure 58 presents the Bayesian network model for the horizontal dimension. In this case, all network input variables refer to the horizontal sector input conditions of the aircraft pair.

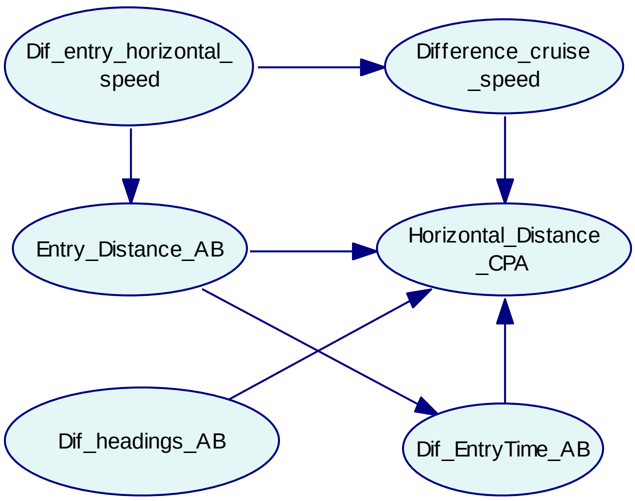


Figure 58: Subnetwork 2 A. Horizontal Bayesian Network

3.3. Input and output variables and states

The aspects that will be considered to model these two networks are shown below:

- **Vertical distance:** The variables to be considered for the modelling of the vertical distance are those related to the aircraft entry conditions into the sector. i.e., their flight levels, vertical speed, and entry times.

- **Horizontal distance:** The variable to be considered for the modelling of the horizontal distance are those related to the aircraft entry conditions into the sector. i.e., their headings, horizontal speed, and entry times.

In Figure 57 and Figure 58 it can be seen the network structure that represents the causal relationships between the variables that constitute this node.

3.4. Parametric learning

Confidential

3.5. Sensitivity analysis for fine tunning

Confidential

4. Subnetwork 3: Conflict detection assessment

4.1. Description and objective

The purpose of this subnetwork is to assess the probability of conflict detection by the ATCo. As indicated in ATM barrier model (Figure 23), this barrier is posterior to the event Potential Conflict, therefore the subnetwork is concerned with the study of the possible potential conflicts that came out of subnetwork 2.

4.2. Network construction

This subnetwork models the ATCo's ability to detect potential conflicts as a result of short- and mid-term effects. The high-level structure of the subnetwork is presented in Figure 59. For this purpose, two concepts are mainly considered:

- **ATCo alert state**. This alertness is assessed according to the air traffic controller workload during the hour at which the CPA of the aircraft pair occurs. Workload will consider the traffic in that hour, the trajectories and conditions of these traffic, as well as the actions performed by the ATCo during that hour. This concept accounts for the mid-term effects influencing the ATCo performance. This idea is represented by the blue box on the top left-hand side of Figure 59 with the label “Hourly ATCo workload & Situational awareness”.

- **Scenario and traffic conditions at the event time.** This concept considers the short time effects influencing the ATCo performance. These derive from the relative situation of each aircraft pair at the moment the ATCo search for potential conflicts. This moment "t" is considered to be the 5-minute period in which the event occurs. To account for short time effects, every hour will be divided into 12 periods of 5 minutes, and the variables representing the situation of aircraft pairs will be calculated for each of the 5-minute periods. This idea is represented by the blue box on the bottom left hand side of Figure 59 with the label “Traffic complexity at time conflict”. The scenario conditions are subdivided into three variables:

- Traffic conditions.
- Aircraft performance.
- Not expected aircraft in the sector.

In this way, the probability of detecting the conflict will depend on the ATCo alertness throughout the hour, and the complexity of the overall situation in short 5-minute periods, when aircraft pair events occur.

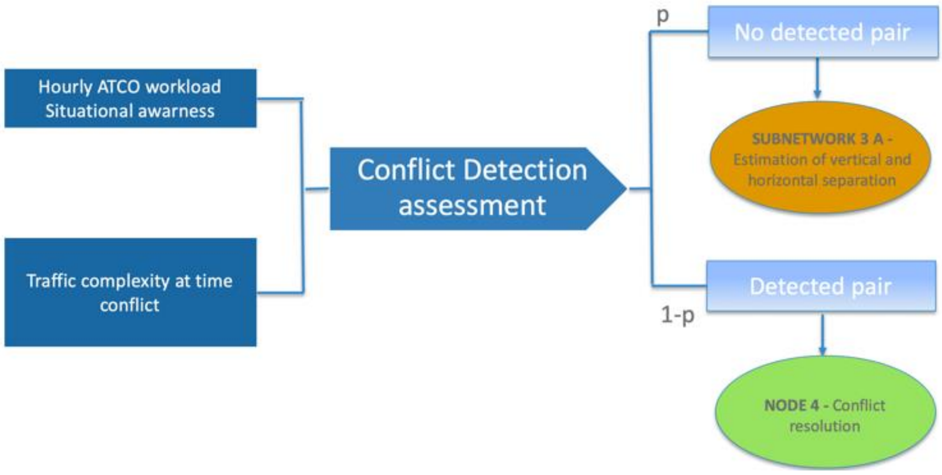


Figure 59: Conflict detection assessment network

Finally, the probability that the conflict will be detected by the ATCo will be evaluated for each aircraft pair and marked with a "YES/NO - 1/0" label in the corresponding file. Two possible outcome branches are envisaged as shown in Figure 59. If the potential conflict is not detected by the controller, the estimation of vertical and horizontal distance at the CPA will be assessed. On the other hand, if the ATCo has been able to detect the conflict, its resolution will be evaluated at subnetwork 4.

Thus, the construction of the network is based on:

- **Network entry:** It will consist of the four inputs already mentioned, composed of the ATCo workload plus the three variables that characterize the scenario (traffic conditions, aircraft performance and unexpected aircraft).
- **Training data:** The network is trained with the conditions of each record.
- **Network exit:** The output will be the probability that the ATCo will detect the potential conflict. It should be noted that the conflict will be considered to have been detected whenever the air traffic controller takes any action on either of the two aircraft involved.

This high-level approach is incorporated into a single network (Figure 60) in which, ultimately, four main variables directly feed the probability of a potential conflict being detected by the ATCo. These variables are represented by the four areas of analysis that are considered to be of greatest relevance in determining whether a controller successfully detects the conflict. Each of the above-mentioned areas are covered separately below:

5. ATCo Alert Status analysis area

The alert state will depend directly on the workload of the controller during the operating hour. In this respect, both "saturation" and "relaxation" can be a problem.

Therefore, it is a matter of having an assessment of the workload during the hour in which the event occurs, estimated by the actions performed by the ATCo during that time.

This area of analysis could be treated as an independent network, whose construction has been based on the following elements:

- **Network entry:** Global parameters of the hour, referring to the traffic density and the ATCo actions during the hour.
- **Training data:** Considering the sector conditions, it is trained to identify the ATCo actions in the hour, and in particular the clearances to solve a conflict (change flight level, direct to, rerouting).
- **Network output:** The workload level, defined in 4 states:
    - Level 1: High number of resolution actions, and high number of total actions.
    - Level 2: No. of medium resolution actions, with a high total number of actions.
    - Level 3: Medium number of resolution actions and medium total number of actions.
    - Level 4: Reduced total number of actions.

6. Scenario Conditions analysis area

This is the assessment of the scenario at the moment when the second aircraft of the pair enters the sector as this would be the moment from which a conflict is assessed and detected.

This is implemented through three independent subnets in which each output variable would constitute a direct input to determine the probability that the ATCo detects the conflict. However, the type of input data for each of them can be described in a general way:

- **Network inputs:** Average of the scenario parameters over the 5-minute period identified in time.
- **Parameters calculation:** Considering the conditions of the scenario, it is trained to evaluate the three parameters that characterize the scenario and that are described below:
    - Traffic conditions: Defined by parameters such as instantaneous demand (5 minutes entry counts), the percentage of regulated aircraft and the distribution of flight levels. All these variables evaluated in 5 minutes time intervals
    - Aircraft performance: Defined by both vertical and horizontal speed distributions.
    - Not expected aircraft in the sector: Both out-of-flow and unexpected aircraft are considered; however, since it was not possible to obtain data related to these flows, only one variable will be used.

The detailed calculation of the scenario parameters is discussed hereafter.

- **Traffic conditions:** It considers the situation of the demand.
    - Instantaneous demand: relation between the instantaneous demand with the average of the maximum instantaneous demand on the trading day (the maximum will be 1).
    - % Aircraft regulated: % on a scale of 0 - 1. The fact that aircraft are regulated means that the flow will be continuous.
    - % Aircraft that change FL in the period: % on a scale of 0 - 1. The FL change makes the scenario more complex.

- **Performances**: Considers the difference in aircraft performances at time *t*.

  Speed Distribution: Complexity occurs when aircraft have differences in speed. Therefore, the speed difference between the two aircraft is compared:

  % Aircraft with a speed difference greater than 50 knots: % on a scale of 0-1.

  Vertical Speed Distribution: Complexity occurs when ACs climb or descend with different speeds. So, the difference in speeds between the two aircraft is compared:

  % AC with non-zero vertical speed difference: % on a scale of 0-1.

- **Unplanned aircraft:** Considers the complexity contributed by unexpected aircraft or aircraft operating outside standard flows in the 5 min period.

  % AC Out of Flows: % on a scale of 0 - 1.

  % Unexpected AC: % aircraft that have changed the EOBT time against the IOBT, or that have changed flow, including those.

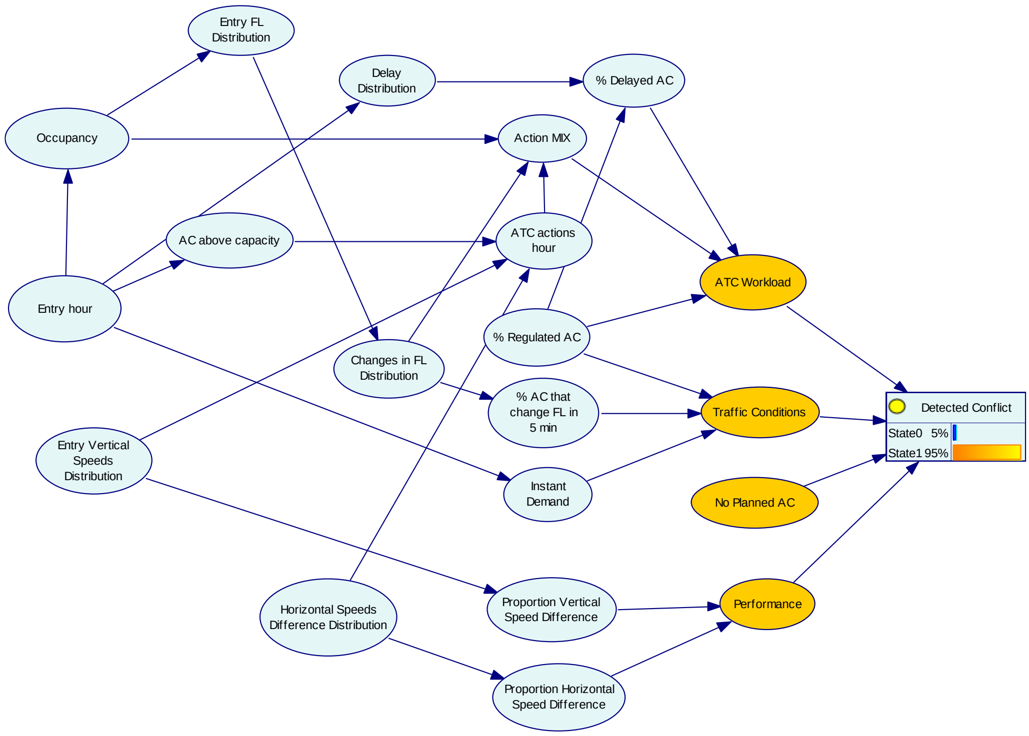


Figure 60: Subnetwork 3. Conflict identification

The Figure 60 shows the detailed proposed Bayesian network scheme. In this case it is a three-level distributed network. It represents the selected input variables and the causal relationships between them. Again, these causal relationships have been obtained by the model directly from the input database and completed by expert judgement.

4.3. Input variables and states

In this section, the selected variables will be explained in order to define the subnetwork and is corresponding purpose. The objective of this subnetwork is to evaluate the detection probability of a conflict between an aircraft pair by the air traffic controller.

The input variables for this node have been selected to model both the controller alert status and the stage conditions in the event at the time of the event, as explained below:

- **Controller alert status:** It is evaluated according to the workload that the controller has at the time the event occurs, considering the general conditions of the time and the controllers' actions.
- **Stage Conditions at the instance of the event:** It is going to be considered as instance "t" the 5-minute period in which the event occurs. The hour divided into 12 periods of 5 minutes will always be considered, and the parameters for each of the periods of the hour are calculate.

4.4. Parametric learning

Confidential

4.5. Sensitivity analysis

Confidential

5. Subnetwork 3A: Distance evaluation between aircraft. Cases of no conflict detection

5.1. Description and objective

As a common objective of the "Type A" subnetwork, this network is intended to evaluate the horizontal and vertical distance of the aircraft pairs in the CPA. This subnetwork is concerned with those pairs of aircraft for which a potential conflict has not been detected in subnetwork 3.

5.2. Network construction

In the particular case of the Santiago sector used to illustrate the model, this set of aircraft pairs did not experience any ATCo clearance nor any SCTA alert. Therefore, the structure of this network, and the process followed to derive it, is the same as subnetwork 2A.

However, it should be remembered that the set of aircraft pair data used for the parametric learning of the network is different. There are actually two Bayesian networks, one for the vertical distance estimation and a second one for the estimation of the horizontal distance of the aircraft pair whose verticals separation is lower than 800 ft. Figure 61 and Figure 62 illustrate the structure of the resulting two networks.

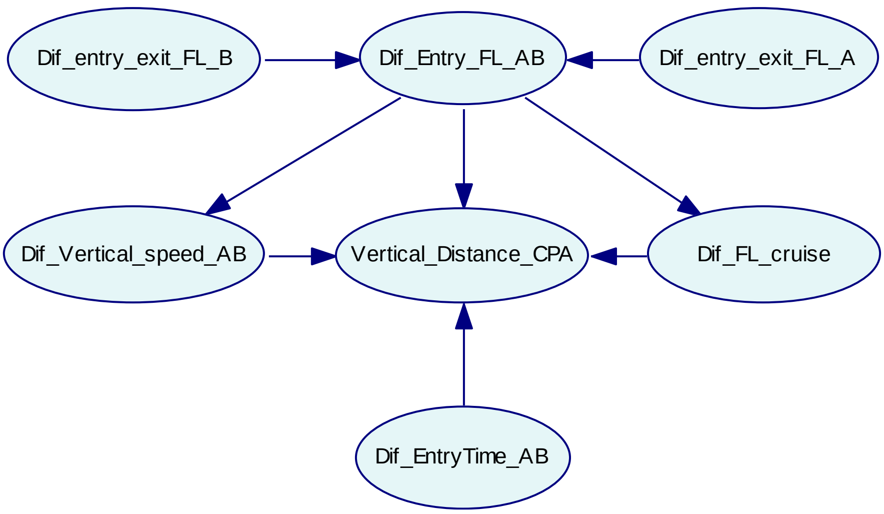


Figure 61: Subnetwork 3A. Vertical distance distribution Bayesian Network

.

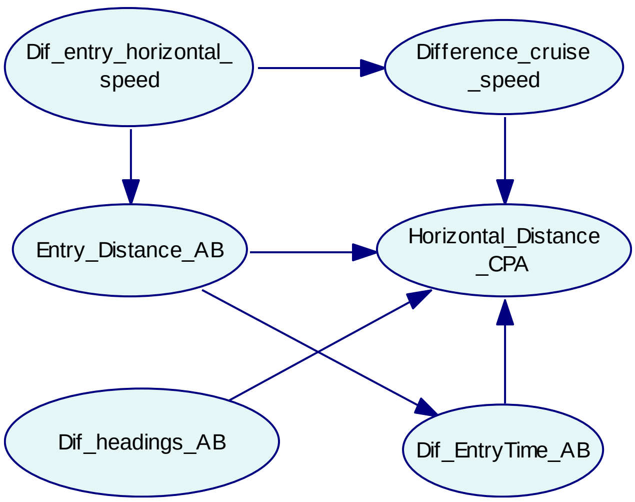


Figure 62: Subnetwork 2A. Horizontal distance distribution Bayesian Network

5.3. Input variables and states

Consequently, the aspects that will be considered to model these two networks, will also be the same as in subnetwork 2A:

- **Vertical distance:** The variables to be considered for the modelling of the vertical distance are those related to the aircraft entry conditions into the sector. i.e., their flight levels, vertical speed, and entry times.

- **Horizontal distance:** The variables to be considered for the modelling of the horizontal distance are those related to the aircraft entry conditions into the sector. i.e., their horizontal speed, and entry times.

5.4. Parametric learning

Confidential

5.5. Sensitivity analysis for fine tuning

Confidential

6. Subnetwork 4: Conflict resolution assessment

6.1. Description and objective

The objective of this subnetwork is to assess the probability of conflict resolution by ATCo, once a potential conflict has been detected. As indicated in ATM barrier model (Figure 23), this barrier is posterior to the Conflict Detection barrier, therefore the subnetwork is concerned with the study of the detected potential conflicts that came out of subnetwork 3.

6.2. Network construction

This subnetwork models the ATCo's ability to resolve a potential conflict once it has been identified. It is mainly concerned with short term effects affecting the controller performance. This short time frame "t" is considered to be the 5-minute period in which the event occurs. Therefore, the hour shall always be considered divided into 12 periods of 5 minutes, and the parameters shall be calculated for each of the periods of the hour. The high-level structure of the subnetwork is presented in figure 59. For this purpose, two concepts are mainly considered:

- **ATCo alert state.** This alertness is assessed according to the workload of the air traffic controller at the time of the event. Workload will consider the general conditions at that time and the controller's actions. This concept is represented by the blue box on the top left-hand side of Figure 63. with the label "Instant ATCo workload (5 minutes) & Situational awareness".

- **Scenario conditions at the time of the event.** These concept accounts for the relative situation of each aircraft pair at the moment the ATCo acts on the potential conflict. This idea is represented by the blue box on the bottom left hand side of Figure 63 with the label "Specific conflict situation".

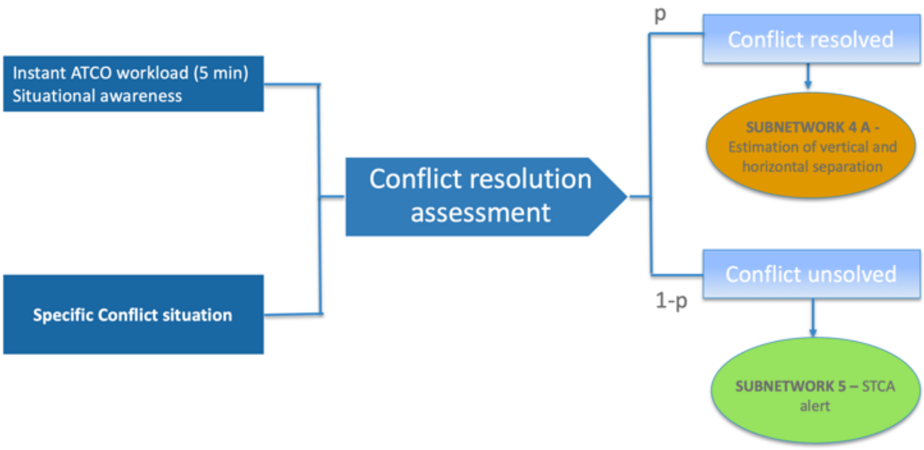


Figure 63: Conflict resolution assessment network

The probability that the conflict will be resolved by the ATCo will be evaluated for each pair f aircraft and marked with s "YES/NO" or "1/0" 1label in the corresponding field. Two possible outcome branches are envisaged as shown in the figure 59. If the potential conflict is not resolved by the controller, the possibility of activating the STCA is evaluated at a consecutive subnetwork. On the other hand, if the ATCo has been able to resolve the conflict, the estimated CPA distances between the aircraft will be evaluated at subnetwork 4A.

Thus, the construction of the network is based on:

- **Network entry:** It will be composed by the ATCo workload, the actions on the traffic carried out by the ATCo and the time since the aircraft pair enters the sector until they reach their CPA.
- **Training data:** The network is trained with the conditions of each record.
- **Network output:** The output will be the probability that the ATCo resolves the conflict. If there is no STCA and no loss of separation, it has been resolved, otherwise it has not been resolved (this may be due to ATCo action or any other cause).

Going into more detail, in Figure 60 it can be seen the variables that make up this subnetwork. These variables correspond to the two areas of analysis mentioned at the beginning of this section. Thus, each of the above-mentioned areas will be covered separately below:

1. **ATCo instantaneous workload (5-minute time period) analysis area**

This network evaluates the ATCo situation at the time of the event, estimated by the actions performed on the traffic during that period of time.

This area of analysis can be developed as an independent network, whose construction is based as follow:

- **Network entry:** Parameters already defined in the conflict detection probability subnetwork such as traffic conditions, performances and unplanned aircraft as these variables affect the traffic condition in the event period.

- **Training data:** Considering the traffic conditions of the period, the network is trained to identify the ATCo actions in the period, and in particular the actions of each type (traffic actions and routine actions).

- **Network output:** The workload level, defined in 4 states:

    - Level 1: High number of resolution actions, and high number of total actions.
    - Level 2: No. of medium resolution actions, with a high total number of actions.
    - Level 3: Medium number of resolution actions and medium total number of actions.
    - Level 4: Reduced total number of actions.

2. Analysis area of the complexity of aircraft conditions at the instance of first action

The difficulty of the situation to be resolved by the ATCo for this aircraft pair is assessed. The difficulty will be impact on the set of actions the ATCo has to take to resolve the potential conflict and the time period available to do so.

This analysis area can be modelled as a subnetwork that follows the following structure:

- **Network entry:** Relative vertical and horizontal condition data between aircraft at sector entry.

- **Training data:** For each pair, the conditions that the aircraft will have at the time the ATCo takes the first action on one of the aircraft are calculated. They are used to train the actions and the time.

- **Network output:** The ATCo's actions and the time from the aircraft pair's entry into the sector to its CPA are obtained as output.

Thus, the probability of the ATCo resolving the conflict will depend on the workload at the time of the action, and the complexity of the overall situation in the period in which the event occurs. This high-level approach is incorporated into a single network (Figure 64).

Figure 64: Subnetwork 4. Conflict resolution

Figure 64 shows the proposed Bayesian network scheme, the selected input variables and the causal relationships between them. In this case it is a five-level distributed network, as can be seen graphically through the proposed colours.

6.3. Input variables and states

This is a unique network in which, ultimately, only three variables directly feed into the probability that an already detected potential conflict will be resolved by the ATCo.

These variables are represented by the two areas of analysis that are considered to be of most relevance in determining whether a controller successfully resolves the conflict.

The most relevant factors are represented by the instantaneous workload of the ATCo at the time of the event, the time he/she has to resolve the conflict, i.e. from the time the aircraft pair is within the sector until they reach their CPA and, finally, the number of traffic actions the air traffic controller has to perform to resolve the conflict.

6.4. Parametric learning

Confidential

6.5. Sensitivity analysis for fine tuning

Confidential

7. Subnetwork 4A: Distance evaluation between aircraft. Cases of conflict resolution

7.1. Description and objective

The objective of this subnetwork is to evaluate the horizontal and vertical distance in the CPA, in the case of aircraft pairs that constituted a potential that were detected and resolved by the ATCo.

7.2. Network construction

For this purpose, the conditions of the aircraft pair when the controller issue the last clearance to each of them before the CPA are estimated. Then the influence of these conditions on the final actual distance at the CPA will be evaluated. As in previous similar subnetworks, two decoupled BN have been built to estimate separately the vertical and the horizontal distance distributions. The resulting networks are shown in Figure 65 and Figure 66, respectively.

Both of them are a four-level network:

- The first level, constituted by the nodes in blue pale, accounts for the relative vertical or horizontal conditions data of each aircraft pair at the sector entry point.
- The second level, composed by the nodes in green, accounts for the relative vertical or horizontal conditions data of each aircraft pair at the point the controller issue the last clearance to them.
- The third level, nodes in orange, accounts for the total number of actions performed by the ATCo and the time he has available to resolve the conflict.
- Finally, the fourth level, representing the outcome of the network in yellow, refers to vertical and horizontal distance distribution at the CPA.

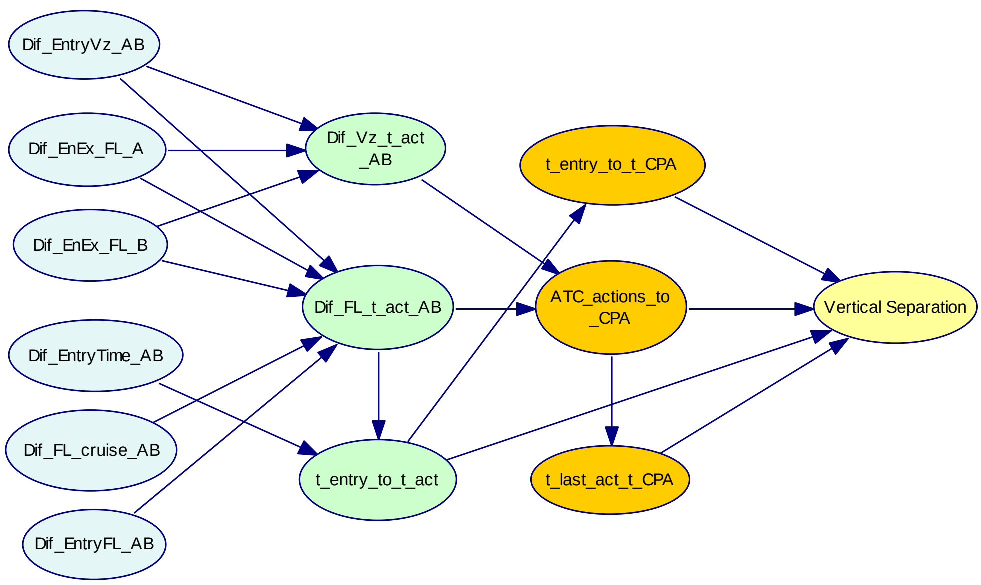


Figure 65: Subnetwork 4 A. Vertical Bayesian Network

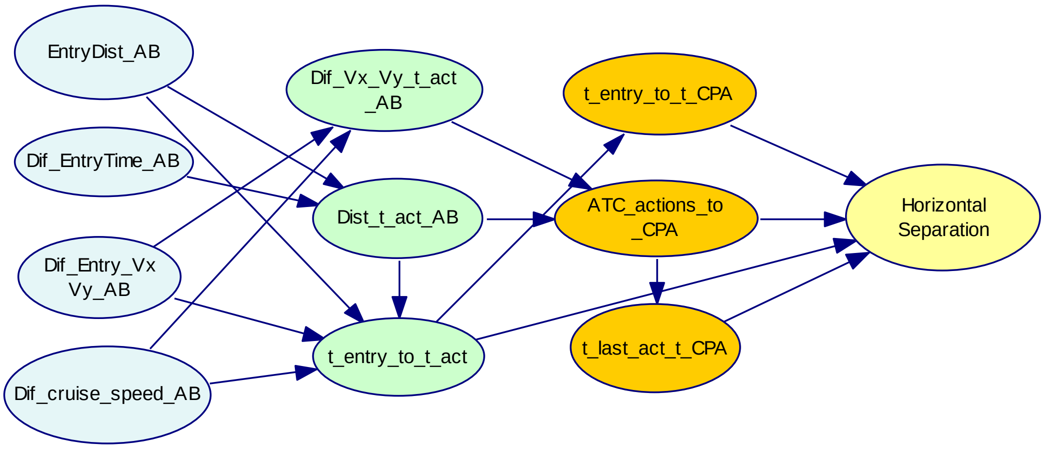


Figure 66: Subnetwork 4A. Horizontal Bayesian Network

7.3. Input variables and states

The aspects that will be considered to model these two networks are shown below:

- **Vertical distance:** The variables to estimate the vertical distance at the CPA are those related to the aircraft pairs vertical conditions at the sector entry point, that is, the differences of flight levels ("Dif_EntryFL_AB", "Dif_FL_cruise_AB", "Dif_EnEx_FL_AA", "Dif_EnEx_FL_B"), differences of cruise and vertical speeds ("Dif_EntryVz_AB"), and the differences of sector entry times ("Dif_Entry_time_AB"). Additionally, this network also considers the vertical conditions of the aircraft pair at the time of the last ATCo clearance ("Dif_Vz_t_act_AB", "Dif_FL_t_act_AB", "t_entry_to_t_act"). As well as the total number of ATCo action up to the moment in which the CPA occurs, and intermediate times ("t_entry_to_t_CPA", "ATC_actions_to_CPA", "t_last_act:t_CPA")

- **Horizontal distance:** The variables to estimate the horizontal distance are those related to the aircraft pairs horizontal conditions at the entry time, that is, the differences in horizontal distance and horizontal speed of the aircraft pair ("EntryDist_AB", "Dif_EntryTime_AB", "Dif_Entry_Vx_Vy_AB", "Dif_cruise_speed_AB"). Additionally, this network must also consider the horizontal conditions that the aircraft will have at the time of the last ATCo action (Dif_Vx_Vy_t_act_AB, Dist_t_act_AB, t_entry_to_t_act). As well as the actions that the ATCo has carried out up to the moment in which the CPA occurs, and intermediate times ("t_entry_to_t_CPA", "ATC_actions_to_CPA", "t_last_act_t_CPA").

7.4. Parametric learning

Confidential

7.5. Sensitivity analysis for fine tuning

Confidential

8. Subnetwork 5: STCA probability

8.1. Description and objective

The objective of this subnetwork is to assess the probability of an STCA activation for aircraft pairs involving a potential conflict not resolved by the ATCo.

The activation of this alert will depend on the geometry of the conflict and the position of the aircraft at the time of the conflict is considered.

Considering the relative conditions of the aircraft pair at the entrance to the sector, the conditions of the aircraft at the time of the controller's action are estimated. From these the percentage of aircraft pairs leading an STCA alert are predicted.

8.2. Network construction

The high-level structure of the network is shown in Figure 67.

As stated above, the activation of this alert will depend on the geometry of the conflict and the actions carried out by the ATCo. Therefore, it is important to estimate the scenario conditions and the geometry of the aircraft at the moment when the controller takes the last action. This is indicated by the blue box on the left side of Figure 67 with the label “Geometric condition for the ACs modified by ATCo actions”.

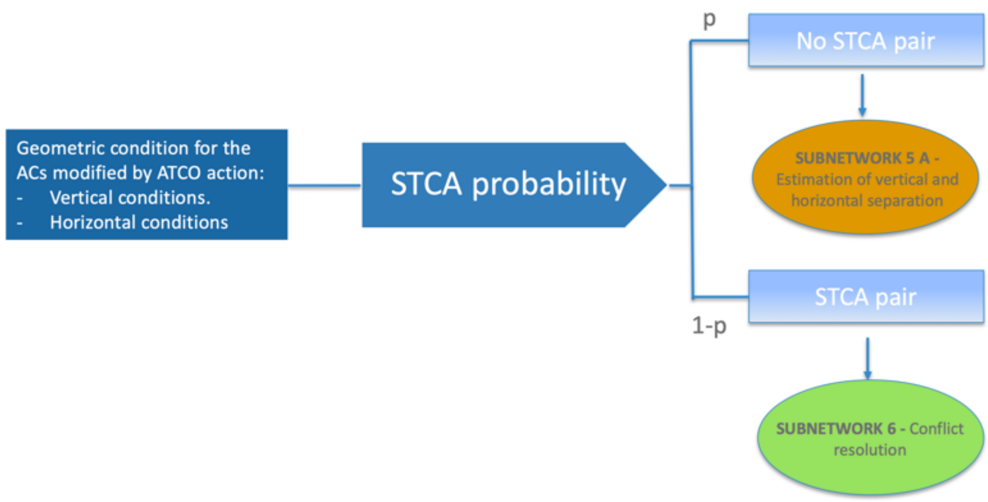


Figure 67: STCA alert probability assessment network

The probability of an STCA triggering will be evaluated for each aircraft pair and marked with “YES/NO” or ”1/0” label in the corresponding field. Two possible outcome branches are envisaged as shown in Figure 67. If an STCA is not generated the vertical and horizontal distance estimation will be carried out in a successive subnetwork (subnetwork 5A). On the other hand, if an STCA is generated, it will be necessary to evaluate whether the conflict will be finally resolved (subnetwork 6).

The detailed structure of the network is presented in Figure 68. The figure shows the proposed Bayesian network scheme, the variables and the causal relationships between them. Thus, the model will be organised as a four-level distributed network, as can be seen graphically through the proposed colours:

- The first level, constituted by the nodes in blue pale, accounts for the relative vertical or horizontal conditions data of each aircraft pair at the sector entry point.
- The second level, composed by the nodes in green, accounts for the relative vertical or horizontal conditions data of each aircraft pair at the point the controller issue the last clearance to them.
- The third level, nodes in orange, accounts for the total number of actions performed by the ATCo and the time since the last clearance until the STCA.
- Finally, the fourth level, representing the outcome of the network in yellow, refers to the activation of the STCA.

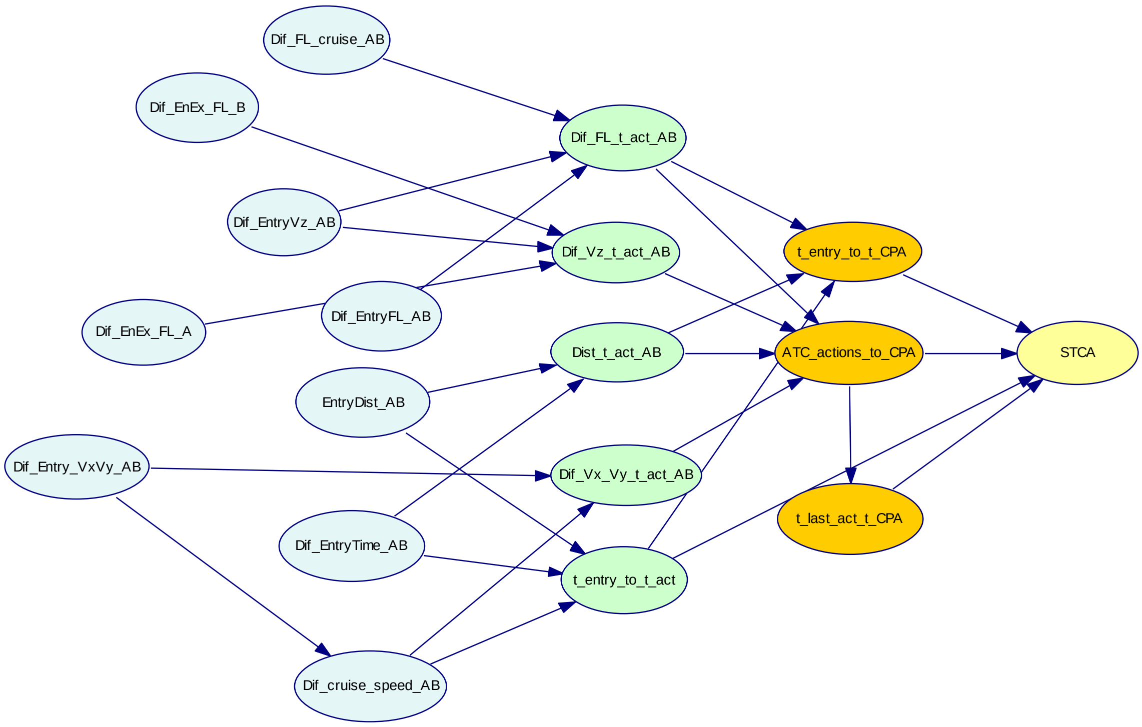


Figure 68: Subnetwork 5. STCA probability

8.3. Input and output variables and states

The variables that have been taken into account to model the objective of this subnetwork can be divided into two large blocks:

- **Scenario conditions at the sector entry time:** As in previous subnets, the aircraft sector entry conditions are related to the flight levels of the aircraft, their speeds, both vertical and horizontal, and the time difference and distance with which they enter into the sector.
- **Scenario conditions at the instance "t" of last ATCo Clearance:** These variables are those that model the scenario at the moment in which the ATCo has carried out the last clearance on an aircraft pair. That variables are related to the time from the sector entrance until the last ATCo clearance, the difference in speeds, both vertical and horizontal, the horizontal distance between both aircraft and if they are flying in the same flight level or not.

8.4. Parametric learning

Confidential

8.5. Sensitivity analysis for fine tunning

Confidential

9. Subnetwork 5A: Distance evaluation between aircraft. Cases where there is a conflict are detected, not resolved and the STCA is not triggered.

9.1. Description and objective

The objective of this subnetwork is to assess the horizontal and vertical distance at the CPA, in the case of conflicts that were not finally solved by the ATCo, and ended in an SMI without triggering an STCA alert.

9.2. Network construction

The structure of this network, and the process followed to derive it, is the same as in subnetwork 4A. The difference is that the aircraft population for training is different.

There are actually two Bayesian networks, one for the vertical distance estimation and a second one for the estimation of the horizontal distance of the aircraft pair whose verticals separation is lower than 1000 feets.

Thus, the first step is to estimate the conditions at the time of the last ATCo clearance, and then evaluate the CPA distance based on these conditions. Both networks are structure into four levels:

- The first level, constituted by the nodes in blue pale, accounts for the relative vertical or horizontal conditions data of each aircraft pair at the sector entry point.
- The second level, composed by the nodes in green, accounts for the relative vertical or horizontal conditions data of each aircraft pair at the point the controller issue the last clearance to them.

- The third level, nodes en orange, accounts for the total number of actions performed by the ATCo and the time he has available to resolve the conflict.
- Finally, the fourth level, representing the outcome of the network in yellow, refers to vertical and horizontal distance distribution at the CPA.

Thus, the Bayesian network model for the vertical dimension is presented first (Figure 69). It will not go into greater detail since the structure is similar to that proposed in the 4A subnetwork.

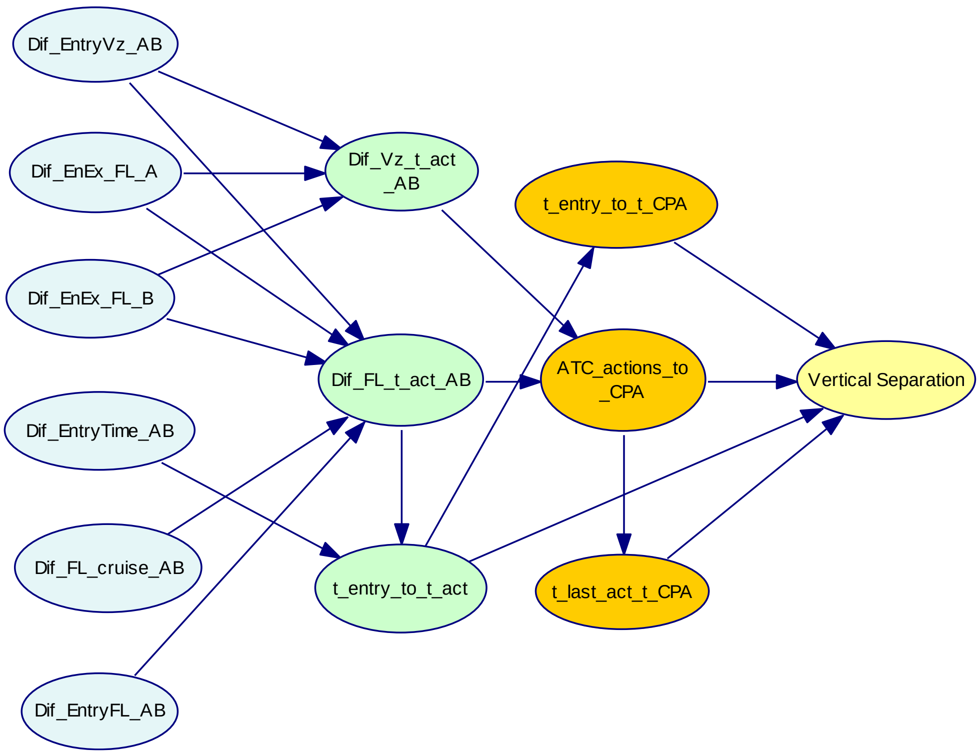


Figure 69: Subnetwork 5A. Vertical Bayesian Network

In the same way, the structure of the horizontal distance estimation network between aircraft pairs in the CPA is included (Figure 70).

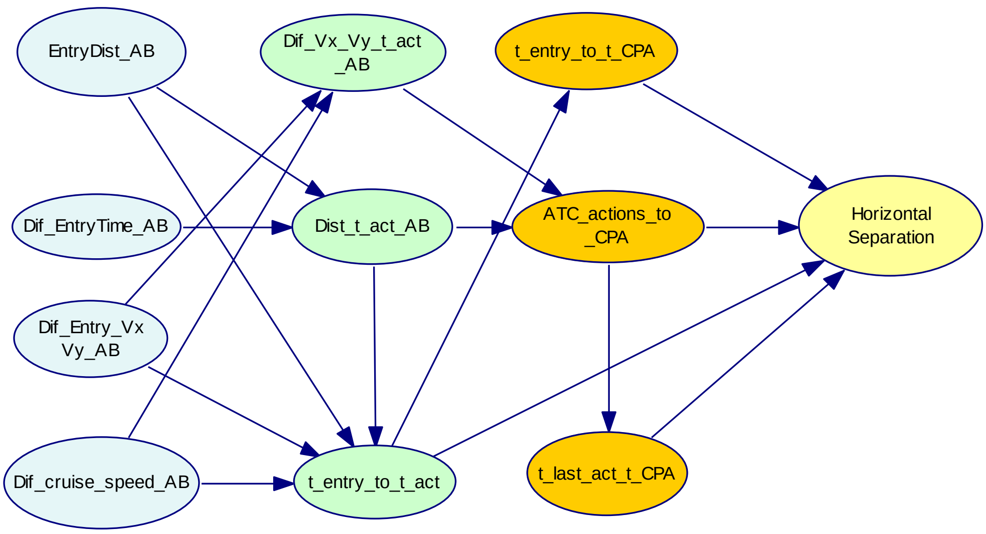


Figure 70: Subnetwork 5A. Horizontal Bayesian Network

9.3. Input and output variables and states

Network variables and its discretisation states are the same than in subnetwork 4A. Detailed definition of all variables and its discretisation states is provided in the chapter 7.

9.4. Parametric learning

Confidential

9.5. Sensitivity analysis for fine tuning

Confidential

10. Subnetwork 6: Conflict resolution after STCA.

10.1. Description and objective

The objective of this subnetwork is to assess the probability of conflict resolution by the ATCo after the Short Term Conflict Alert (STCA) has been triggered.

10.2. Network construction

This subnetwork models the ATCo's ability to resolve a potential conflict once the STCA has been activated. It is mainly concerned with short term effects affecting the controller performance. This short time frame is considered to be the 5-minute period in which the STCA event occurs. Therefore, the hour shall always be considered divided into 12 periods of 5 minutes, and the parameters shall be calculated for each of the periods of the hour. The high-level structure of the subnetwork is presented in Figure 71.

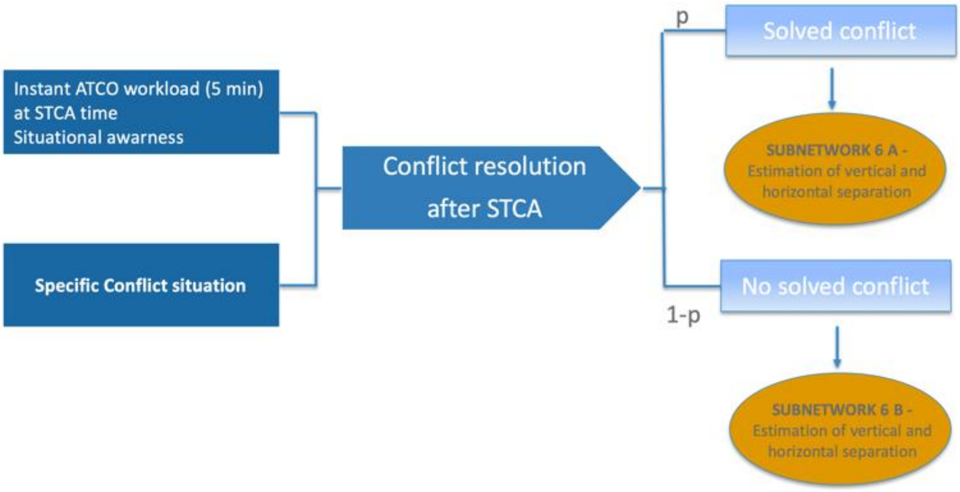


Figure 71: Conflict resolution after STCA network

Therefore, all that remains is to establish the final distances between the aircraft pairs in the CPA for both cases, whether the conflict has been resolved or not. These distances will be evaluated in subnets 6A and 6B.

The probability that the ATCo will resolve the conflict once the STCA alert has been triggered depends on the controller workload at that time, and the complexity and conditions of the scenario in that moment. The first element is represented by the blue box on the top left-hand side of figure 81 with the label “Instant ATCo workload (5 minutes) at STCA & Situational awareness”. The second element is represented by the blue box on the bottom left hand side of figure 81 with the label “Specific conflict situation”.

The probability that the conflict will be resolved by the ATCo after an STCA will be evaluated for each aircraft pair and marked with “YES/NO” or ”1/0” label in the corresponding field. Two possible outcome branches are envisaged as shown in the Figure 71, depending on whether the situation after the SCTA is resolved. In both tree branches a subsequent network will be estimated CPA distances between the aircraft (subnetwork 6A and 6B).

The detailed scheme of the proposed Bayesian network is presented in Figure 71. In this scheme it can be seen all the variables taken into account as well as the links between them.

This BN is integrated by two main parts. The first one models the ATC workload at the SCTA time, and the second part of the network models the complexity of aircraft conditions at the STCA instance. Each of the above-mentioned areas are further elaborated below:

1. **ATCo workload network at the STCA time analysis area**

This part of the network evaluates the ATCo workload considering the actions performed by the air traffic controller on the traffic during the STCA duration. This part of the network can be identified in Figure 72 by all the nodes convergent into the orange node labelled “ART Workload in instance t”.

To evaluate this analysis area, the following structure is followed:

- **Network entry:** Parameters already defined such as traffic conditions, performances and unplanned aircraft as these variables affect the traffic condition in the event period.

- **Training data:** Considering the traffic conditions of the period, the network is trained to identify the ATCo actions in the period, and in particular the actions of each type.
- **Network output:** The workload level, defined in 4 states:
    - Level 1: High number of resolution actions, and high number of total actions.
    - Level 2: Medium number of resolution actions, with a high total number of actions
    - Level 3: Medium number of resolution actions and medium total number of actions
    - Level 4: Reduced total number of actions.

2. **Complexity of aircraft conditions at the STCA instance.**

This part of the network assesses the difficulty of the situation to be resolved by the ATCo for each aircraft pair. The difficulty will depend on the set of actions the ATCo has to take to resolve the potential conflict, the time period available to do so and the geometry of the STCA. Therefore, this second part of the networks consists of 4 level structure that follows the previous steps. It can be identified in Figure 72 by all the nodes convergent into the 3 orange nodes labelled “t_entry_to_t_STCA”, “ATC_actions_STCA_to_CPA”, and “T_STCA_to_CPA”

The defined sequence to estimate traffic conditions at the SCTA time would be the following:

a. Conditions of entry to the sector of the aircraft pair. They are indicated by the pale blue nodes. They include flight levels of the aircraft, the distance at which they enter, and their vertical and horizontal speeds, among others.

b. The ATCo acts on the aircraft pair, therefore the conditions at the time of last clearance before the STCA are estimated from the entry conditions. They are indicated by the pink nodes and includes variables such as the relative position, speed and the time until last ATCo action.

c. After the ATCo acts, the STCA occurs, therefore, the conditions at this time are estimated from the conditions at the time of the last controller action. They are indicated by the purple nodes and accounts for variables such as the relative position, speed and the time until the STCA.

d. After the STCA, the ATCo may act on the aircraft pair, therefore the number of ATC clearances and the time available until the CPA are also calculated.

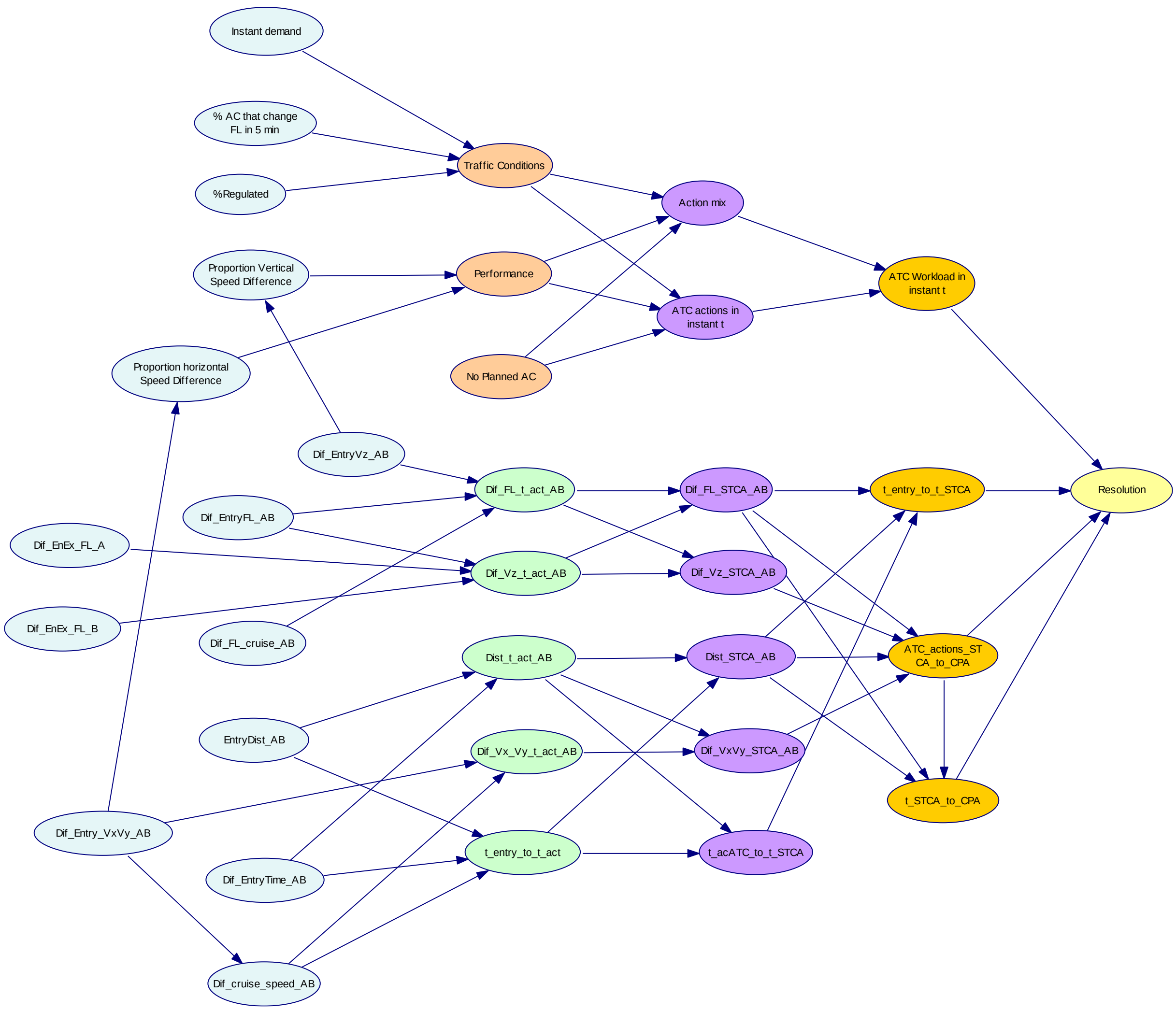


Figure 72: Subnetwork 6. Conflict resolution after STCA

10.3. Input and output variables and states

The variables that have been considered to model the objective of this subnetwork can be divided into two large blocks:

- **ATCo workload at the instance "t":** In this case, the instance "t" is considered the 5-minute period in which STCA occurs. To model the ATCo workload, the variables related to traffic conditions, aircraft performances and not expected aircraft in the sector have been chosen. These variables have been explained in previous sections.

- **Scenario conditions at the instance "t" of STCA:** These variables are those that model the scenario at the instance "t", the 5-minute period in which the STCA occurs. To model these conditions, variables related to the entry conditions into the sector will be considered. These entry conditions are related to the flight levels of the aircraft, the distance at which they enter, their vertical and horizontal speeds, among others. The conditions of the aircraft will also be taken into account at the time the ATCo performs the last action on the aircraft pair and at the time the STCA occurs.

All the chosen variables and the relationships that have been established between them can be seen in more detail in Figure 72 of the network structure set out in the previous section. Detailed definition of all variables and its discretisation states is provided in the chapter 7.

10.4. Parametric learning

Confidential

10.5. Sensitivity analysis

Confidential

11. Subnetwork 6A: Distance evaluation between aircraft. Cases where there is a conflict, the STCA is activated and resolves them.

11.1. Description and objective

The objective of this subnetwork is to assess the vertical and horizontal distance in the CPA. In this case for pairs of aircraft that constituted a potential conflict that were resolved after triggering an STCA alert.

Therefore, as the conflict is considered to be resolved, it can be assured that the final output of the distance distribution will be defined by pairs of aircraft that are above the minimum separation distance. However, the actual distribution of these distances will have to be ascertained.

To assess these distances, the conditions at the STCA will have to be estimated as it is the reference point immediately upstream of the CPA.

11.2. Network construction

To estimate the conditions of the aircraft pair in the STCA, the following sequence must be considered:

- The conditions at the sector entry time.

- Once the aircraft are in the sector, the air traffic controller acts on them. Therefore, the conditions at the action time must be estimated from the conditions at the sector entry time.

- After the ATCo carries out the relevant actions on the traffic, the moment arrives when the STCA is triggered. Therefore, the conditions at the time of the STCA are estimated on the basis of the conditions at the time of action.

Following this sequence, this subnetwork will be organised in five levels. Each level is represented in a different colour, as can be seen in Figure 73 and Figure 74.

- The first level, constituted by the nodes in blue pale, accounts for the relative vertical or horizontal conditions data of each aircraft pair at the sector entry point.

- The second level, composed by the green nodes, accounts for the relative vertical or horizontal conditions data of each aircraft pair at the point the ATCo issues the last clearance to them.
- The third level, nodes in purple, accounts for the relative vertical or horizontal conditions data of each aircraft pair when STCA occurs and the time that elapses from the last ATCo clearance to the STCA.
- The fourth level, constituted by the nodes in orange, accounts the total number of actions performed by the ATCo and the time since the STCA until the CPA.
- Finally, the fifth level, representing the outcome of the network in yellow, refers to the Vertical and horizontal separation of the aircraft pairs assessed in this section.

Again, the structure proposed in the rest of the distance estimation subnetworks developed throughout this document will be maintained. There are actually two Bayesian networks, the vertical distance estimation network and the horizontal distance at which the aircraft were located at the CPA network.

Thus, the Bayesian network model for the vertical dimension is shown first in Figure 73.

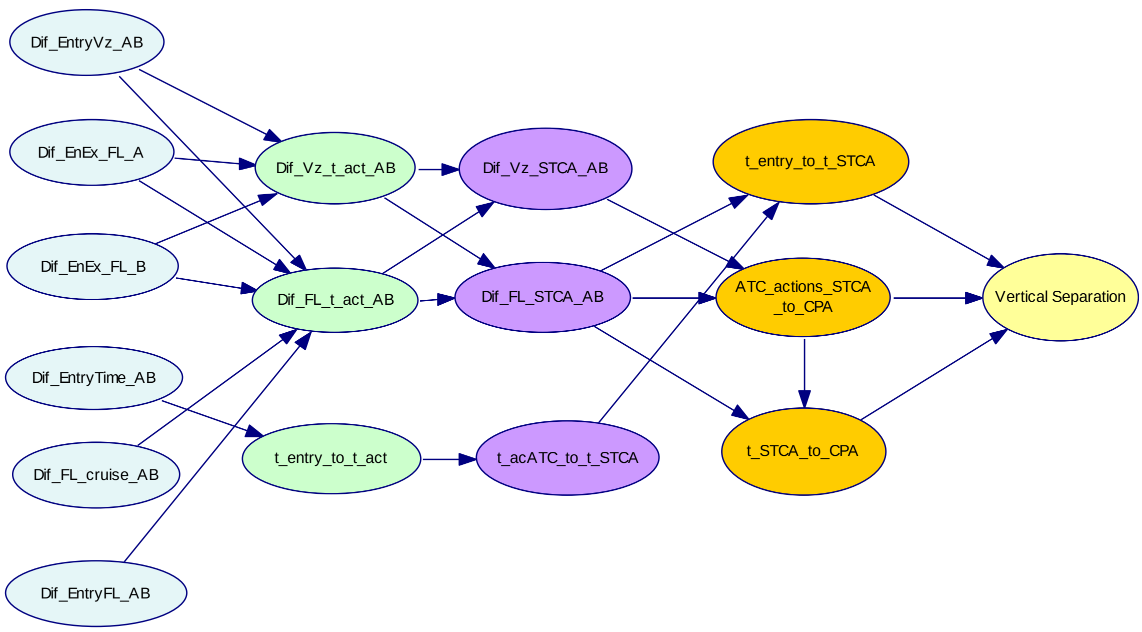


Figure 73: Subnetwork 6A. Vertical Bayesian Network

In the same way, the structure of the horizontal distance estimation network between aircraft pairs in the CPA is included in Figure 74.

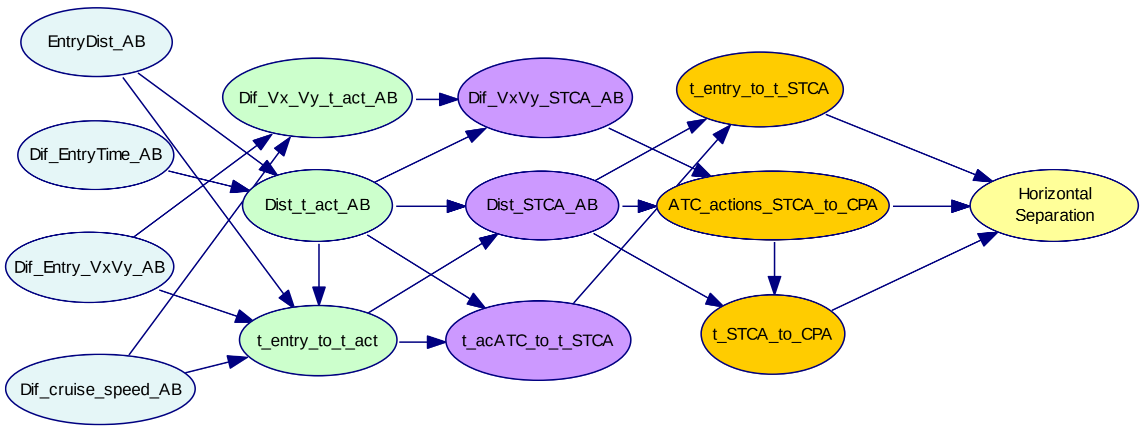


Figure 74: Subnetwork 6A. Horizontal Bayesian Network

11.3. Input and output variables and states

The aspects that will be considered to model these two networks are shown below:

- **Vertical distance:** The variables that have been considered to evaluate the estimation of the vertical distance are those related to the vertical conditions at the entry time, that is, the flight levels and the vertical speeds of the aircraft pair. In addition, this subnetwork will also take into account the vertical conditions of the aircraft at the time the ATCo performs the first action on them, and at the time the STCA trips, as well as the time and the actions that the ATCo carries out on the aircraft from the instance of the STCA to the CPA.

- **Horizontal distance:** The variables that have been considered to evaluate the estimation of the horizontal distance are those related to the horizontal conditions at the entry time, that is, the horizontal distance and the horizontal speed of the aircraft pair. In addition, this subnetwork will also take into account the horizontal conditions of the aircraft at the time the ATCo performs the first action on them, and at the time the STCA trips, as well as the time and the actions that the ATCo carries out on the aircraft from the instance of the STCA to the CPA.

In Figure 73 and Figure 74it can be seen the network structure that represent the causal relationships between the variables that constitute this subnetwork.

11.4. Parametric learning

Confidential

11.5. Sensitivity analysis for fine tuning

Confidential

12. Subnetwork 6B: Distance evaluation between aircraft. Cases where there is a conflict, the STCA is activated but does not resolve them.

The objective of this last subnetwork is the evaluation of the vertical and horizontal distance for those potential conflict which triggered an STCA alert, but were not resolved and ended in an SMI.

The structure of this network and the process to develop it is, in essence, the same as in network 6A. The difference is that the aircraft population for training is different.

The final separation distance between the aircraft pairs in this network will be always below the minimum separation distance since all them will be unresolved conflicts.

In addition, the data file generated for the training of the network has very few cases that meet these conditions. As it is logical, in the airspace it is very unlikely that a true loss of separation will occur. For this reason, it has not been possible to train the network and it has only been developed theoretically. Therefore, it can only be ensured that the separation minima are violated.

The variables that will form this theoretical network are the same that have already been explained in subnetwork 6A.

Figure 75 and Figure 76 show the theoretical structure that both the vertical and horizontal estimation subnetwork will have.

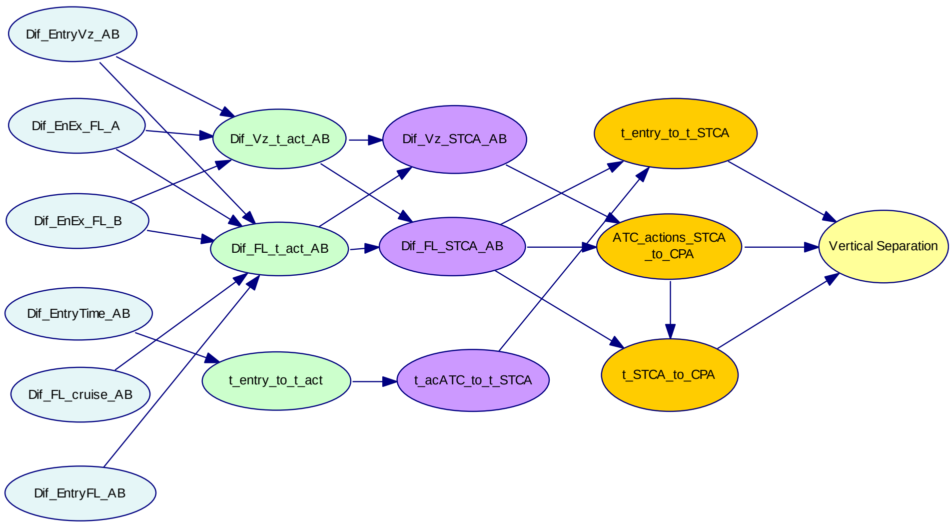


Figure 75: Subnetwork 6B. Vertical Bayesian Network

In the same way, the structure of the horizontal distance estimation network between aircraft pairs in the CPA is included.

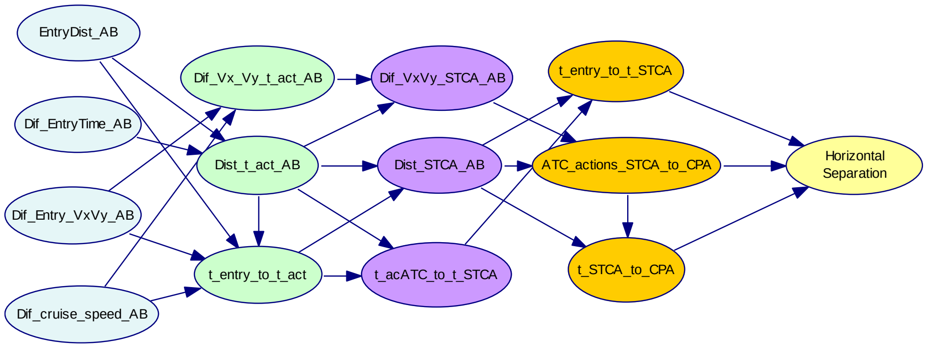


Figure 76: Subnetwork 6B. Horizontal Bayesian Network

Obviously, the parametric learning of the variables and the sensitivity analysis on the network have not been carried out, as there was not enough data available.