

On the Use of Machine Learning and Evidence Theory to Improve Collision Risk Management

Luis Sánchez Fernández-Mellado^{a,*}, Massimiliano Vasile^a

^a*Aerospace Center of Excellence, University of Strathclyde, United Kingdom*

Abstract

This paper presents an element of an Artificial Intelligence (AI) system to assist operators with the management of traffic in orbit.

The system is based on a classification of collision events using Dempster-Shaffer's theory of Evidence (DSt). DSt is proposed to capture and quantify the epistemic uncertainty in the estimation of the Probability of Collision (P_C). By capturing the epistemic uncertainty in the calculation of P_C , we mitigate the paradoxical Dilution of Probability that affects the usual definition of this quantity. This phenomenon gives the counterintuitive idea that the lower the amount of information available to the operators, the smaller the probability of collision. This lack of information corresponds to an uncertainty that is epistemic in nature. Furthermore, when different sources provide contradictory information, the level of epistemic uncertainty in the calculation of P_C can lead to a false confidence in the likelihood of a collision with either an undesirable increase in the number of collision avoidance manoeuvres or an equally undesirable number of false negative.

In order to make the classification automatic and provide continuous decision support to operators, we investigated the use of different Machine Learning techniques and identified the algorithm of choice to deliver the correct classification of collision events. We first considered a classification where the Belief and Plausibility in the correctness of the P_C were used as additional classification criteria. Then we tested a second classification method that avoids the direct computation of Belief and Plausibility but retains the

*8th Floor. James Weir Building. 75 Montrose Street. University of Strathclyde. G1 1XJ, Glasgow. United Kingdom.

Email address: `luis.sanchez-fdez-mellado@strath.ac.uk` (Luis Sánchez Fernández-Mellado)

same added information on the credibility of the P_C . Results suggest that Machine Learning can be effectively used in conjunction with DSt to provide decision support to operators and render the conjunction and collision analysis automatic.

Keywords: Space Traffic Management, Machine Learning, Evidence Theory, Collision Probability, Probability Dilution

1 . Introduction

The growth of the private space sector and the emergence of the so-called New Space movement [1] have brought along new challenges in the management of traffic in orbit. Furthermore, the increase in the size of constellations and the ramping up of the number of launches of satellites of small size, down to pico-sats, are expected to exacerbate the the current proliferation of space debris [2].

The IAA (International Academy of Astronautics) in the Cosmic Study of Space Traffic Management (STM) in 2006, defines STM as "the set of technical and regulatory provisions for promoting safe access into outer space, operations in outer space and returns from outer space to Earth free from physical or radio-frequency interference"[3].

Because of the increase in traffic and size of the overall population of resident objects, an effective STM system will need new technologies, tools and regulations to support space operators [4].

In particular it is envisaged that a higher degree of automation is essential [5], [6], [7]. In this context, Artificial Intelligence (AI) and Machine Learning (ML) can play a fundamental role due to their potential to learn from real and synthetic data and provide decision support with limited computation effort. The goal is to predict events and automatise the allocation of remediation actions.

Applications of AI to decision making can be found in the works of Ramirez-Atencia et al. [8] and Barhydt and Krishnamurthy [9]. In Ramirez-Atencia et al. the authors introduced a decision support tool to support Unmanned Aerial Vehicles (UAV) operators in order to reduce their workload in a increasingly crowded environment, support the decision making and automating tasks. In Barhydt and Krishnamurthy the authors addressed the problem of automatic flight trajectory planning with conflict detection and resolution with separations constrains in air flights with an Autonomous Op-

erator Planner that provides flight guidance to pilots considering the pilot goals in the process. The interesting aspect of these works is the similarities that can be found between the problematic that arises on UAV and Air Traffic Management systems and the issues to be addressed when automating STM: operators workload reduction, the need of automation to cope with a more populated environment or the problems related with collision avoidance maneuvers, time restrictions and the trade-off between false-alerts and missed detections.

However, their application to STM and space safety is still in its infancy. Peng and Bai [10, 11] proposed ML to predict orbit detection and orbit propagation error, Vasile et al. [12] and Sanchez et al. [13] proposed Artificial Neural Networks to predict possible conjunctions and execute optimal correction manoeuvres. Abay et al. [14] also presents a ML-based system for sub-optimal automatic collision avoidance maneuver planning.

In [12] the authors presented a system for supporting the planning and implementation of collision avoidance maneuvers (CAM) using ML (Elastic Nets) and a maneuvers dataset, with special attention to the future consequences on the global space environment. Sanchez et al. introduced an ANN model using exclusively the knowledge of the Keplerian elements of pairs of space objects at a given time. This system was shown to be able to predict the future state of large population of space debris within a given time interval. The prediction of the equinoctial parameters compared with the results from a high fidelity propagator were accurate enough to determine the occurrence of conjunction events and associated impact risk.

This paper extends and refines the work presented in Sanchez et al. [13] and proposes an Intelligent Event Classification system to support the automatic management of traffic in space. The problems addressed in this paper are the assessment of the actual collision probability and the automatic classification of conjunction events. The outcome of the classification provides a recommendation to operators on the possible execution of Collision Avoidance Manoeuvres or the acquisition of further and better observations.

The first part of the paper introduces a methodology to mitigate the dilution of probability associated to the use of the Probability of Collision, P_C . The probability of collision is a widely used metric for risk assessment [15, 16, 17, 18], however, there are concerns about its suitability since it suffers from a dilution of probability that leads to a false confidence in the realisation of collision events [19, 20]. The root of this problem can be found in the fact that epistemic uncertainty is incorrectly treated as aleatory. Here,

we proposed to properly model epistemic uncertainty with Dempster-Shafer Evidence Theory (DSt) [21]. A set of new classification criteria are proposed that accounts for the level of confidence associated to the quality of the observations.

The second part of the paper makes use of the new classification criteria to train a Machine Learning system so that the classification can be rendered automatic. The use of ML is at the core of the proposed Intelligent Event Classification system. The system is intended for the support of space operators under the assumption that multiple sources of information are available. Each source is expected to have a degree of epistemic uncertainty. The existence of multiple sources, potentially in disagreement, introduces a further level of epistemic uncertainty on the reliability of the sources, which is captured with a quantification of the degree of belief in each source.

The paper is structured as follows. In Section 2 , a brief explanation of dilution of probability and its relation with the epistemic uncertainty is presented, as well as an introduction to Evidence Theory and its application on conjunction risk assessment. Section 3 presents a classification of risk events taking into account epistemic uncertainty under DSt and compare it with probabilistic approaches for classifying events. In Section 4 , the AI-based classification model is presented with the Machine Learning techniques considered in this work. Finally, Section 5 summarizes the main conclusions of the paper and suggests future research directions.

2 . Evidence-based Collision Risk Quantification

In this section we will revisit the calculation of the probability of collision between two space objects and will introduce a quantification of the epistemic uncertainty in the value of the probability of collision. More specifically, we will start from the assumption that the uncertainty associated to the position of the two bodies can be well described with two families of uncorrelated Gaussian distributions. However, the parameters of the two families are assumed to be poorly known and we can only assign a degree of belief to the possible sets of values that they can assume.

2 .1. Dilution of Probability of Collision

We start from the usual computation of the Probability of Collision under the assumptions of short-term encounter [15, 16, 17, 18]: i) the relative

motion between objects is assumed to be rectilinear; ii) the uncertainty distributions of the positions of the two bodies are Gaussian and uncorrelated; iii) the velocity vectors are not uncertain; iv) the objects are modeled as hard spheres. Under these assumptions, the Probability of Collision P_C is computed by integrating the combined uncertainty ellipsoid, centered on one of the bodies and projected on the impact plane, over the closed region $\mathcal{B}((0,0), R)$ defined by the sphere of radius R enveloping the two objects (or Hard Body Radius, HBR) [18]:

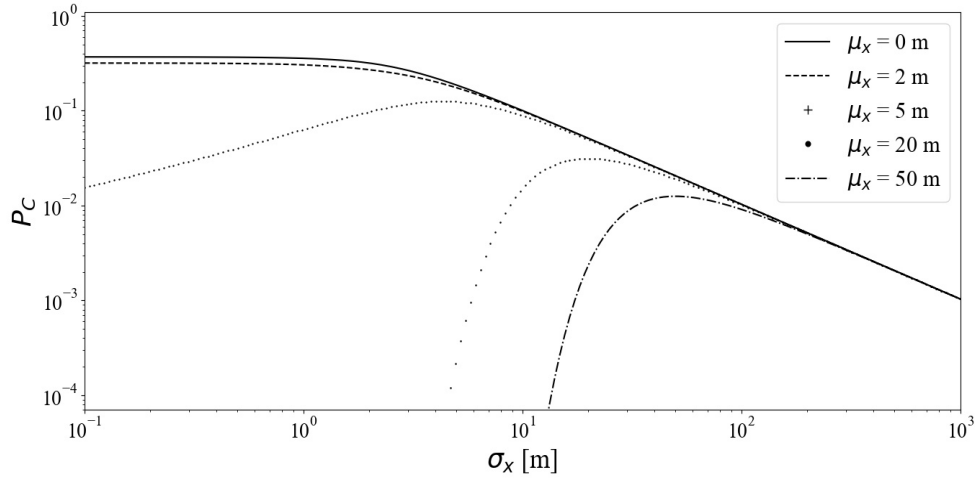
$$P_C = \frac{1}{2\pi\sigma_x\sigma_y} \int_{\mathcal{B}((0,0), R)} e^{-\frac{1}{2}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)} dx dy \quad (1)$$

where x and y are the coordinates on the impact plane of the target object at the time of expected impact, $[\mu_x, \mu_y]$ define the coordinates of the centre of the ellipsoid on the impact plane and $[\sigma_x, \sigma_y]$ its semi-minor and semi-major axes. Note that, although this way to compute the P_C is not the most general, the considerations that follow are generally applicable.

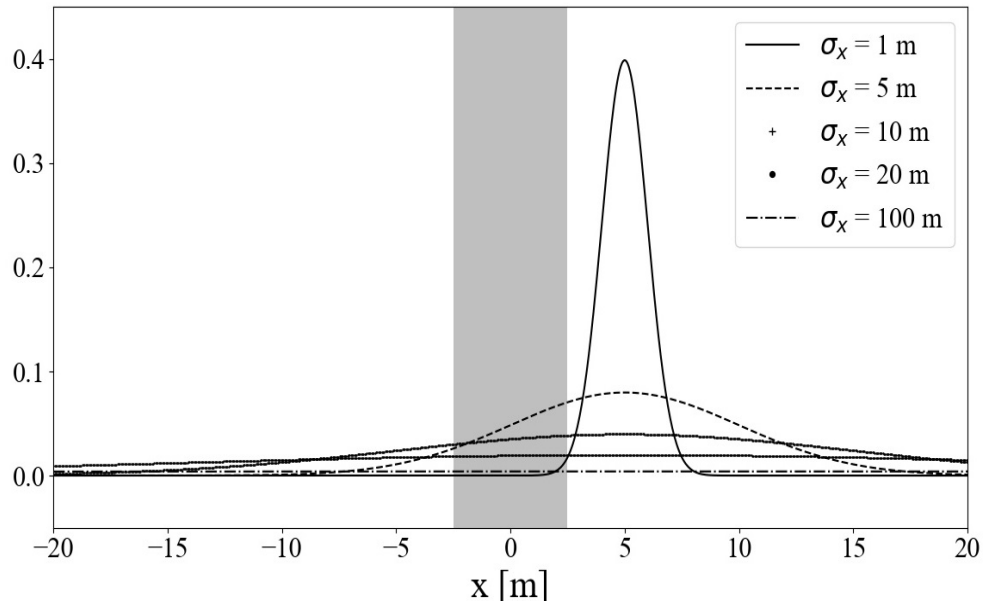
Eq. (1) leads to the known paradoxical phenomenon of dilution of probability [19]. This phenomenon is well represented in Fig. 1a: an increase in the standard deviation yields at first an increase of the probability of collision, up to a maximum value, and then a progressive decrease. Thus, if the uncertainty in the observations grew beyond the point of maximum P_C the obvious conclusion would be a lower risk of collision. In other words, Eq. (1) seems to suggest that the risk of a collision decreases as the amount of information on the position of the two objects decreases.

As explained also in [20], the central problem is the model of uncertainty. If the uncertainty in the position of the two objects was purely aleatory (an irreducible random process), the dilution of probability would describe an actual reduction in the risk of a collision. This can be understood from Fig. 1b where an increase in the standard deviation leads to a reduction of the total probability mass falling within the HBR. However, we can argue that this is true only when the uncertainty on dynamics and sensors is perfectly known. On the contrary, when uncertainty is epistemic, i.e. there is a lack of knowledge on the position or velocity of an object, the correct reasoning should lead to the conclusion that a higher ignorance corresponds to a higher risk of a collision.

In the following we maintain the calculation of the P_C with Eq. (1) but we propose to model the epistemic uncertainty in the observations using



(a)



(b)

Figure 1: (a) Dilution of probability. An increase on the standard deviation (σ_y) beyond certain value yields to a reduction on the computed P_C independently the value of the miss distance (μ_x). For this examples $\mu_y = 6m$, $\sigma_x = 3m$ and $HBR = 5m$. (b) Flatten of Normal Distribution when standard deviation increases. The grey area (HBR) below the curves increases for low values of σ ($\sigma < 5m$) but decrease for higher values ($\sigma > 5m$). $HBR = 5m$ and $\mu = 5m$.

Dempster-Shafer theory of Evidence. In particular we consider the case in which the values of $[\mu_x, \mu_y]$ and $[\sigma_x, \sigma_y]$ are partially known and we can assign a degree of belief (or confidence) to the set to which they belong. This situation can derive, for example, from the fusion of different conflicting observations or from a lack of knowledge on the quality of the sensors.

2.2. Evidence Theory

In this section we provide a brief description of the main features of DSt that will be used in the remainder of the paper. In DSt one can associate a degree of belief in the realisation of an event without an exact knowledge of an associated probability distribution.

Given an event space, the set Θ of all the mutually exclusive and collectively exhaustive elementary events (or hypotheses) $\Theta = \{\theta_1, \theta_2, \dots, \theta_{|\Theta|}\}$ is considered. The collection of all non empty subsets of Θ , including the Empty Space and Θ itself, is the Power Set $2^\Theta = (\Theta, \cup)$. One can now assign a probability mass, called basic probability assignment (*bpa*), to the elements of 2^Θ . Each element of 2^Θ with a non-zero *bpa* is called a *Focal Element* (FE) and is represented with the symbol γ in the following. The *bpa* functions have the following properties:

$$bpa(\emptyset) = 0 \quad (2)$$

$$\forall \gamma_i \in \Gamma, 0 \leq bpa(\gamma_i) \leq 1 \quad (3)$$

$$\sum bpa(\gamma_i) = 1, \forall \gamma_i \in \Gamma \quad (4)$$

In this work, the pair $\langle \Gamma, bpa_\Gamma \rangle$ - where $\Gamma \ni \gamma$ and $bpa_\Gamma \ni bpa_\gamma$ - is called the *Body of Evidence* and the power set $U = 2^\Theta$ the *Uncertainty Space*. Consider now the set:

$$\Omega = \{\mathbf{u} \in U | f(\mathbf{u}) \in \Phi\} \quad (5)$$

where f is a quantity of interest, \mathbf{u} a vector of uncertain parameters and Φ is a desirable target set for f . Given Ω we define the two quantities, Belief (Bel) and Plausibility (Pl):

$$Bel(\Omega) = \sum_{\gamma_i \subset \Omega, \gamma_i \in U} bpa(\gamma_i), \quad (6)$$

$$Pl(\Omega) = \sum_{\gamma_i \cap \Omega \neq \emptyset, \gamma_i \in U} bpa(\gamma_i), \quad (7)$$

which give, respectively, the lower and upper limit on our confidence on Ω given the available evidence supporting statement (5). Belief and Plausibility have the following properties:

$$Bel(\Omega) + Pl(\neg\Omega) = 1 \quad (8)$$

$$Bel(\Omega) + Bel(\neg\Omega) \leq 1 \quad (9)$$

$$Pl(\Omega) + Pl(\neg\Omega) \geq 1 \quad (10)$$

$$Pl(\Omega) \geq Bel(\Omega) \quad (11)$$

The difference between Belief and Plausibility, $Pl(\Omega) - Bel(\Omega)$, associated to Ω , is called *Degree of Confidence* and can be used to determine the degree of epistemic uncertainty associated to an event given the available evidence. This concept will be exploited in this paper to classify conjunction events.

2.3. Evidence Theory for Collision Risk Assessment

The idea is that when $[\mu_x, \mu_y]$ and $[\sigma_x, \sigma_y]$ are affected by epistemic uncertainty, their values are not precisely defined and we can only say that they belong to a given set with a given degree of belief. A higher uncertainty translates into a larger set or in more, possibly disjoint, focal elements.

We then collect all the pieces of evidence in support to the possible values of $[\mu_x, \mu_y]$ and $[\sigma_x, \sigma_y]$ and calculate the corresponding values of the P_C with Eq. (1). The evidence on the values of $[\mu_x, \mu_y]$ and $[\sigma_x, \sigma_y]$ comes from our degree of knowledge or ignorance on the source of the observations. The bpa assignment can derive from a quantitative analysis or simply from a subjective opinion on the credibility of the source of information.

Given a belief assignment to $[\mu_x, \mu_y]$ and $[\sigma_x, \sigma_y]$, i.e. a bpa assigned to a set of focal elements, we can compute the values of Bel and Pl associated to the corresponding P_C . Because $[\mu_x, \mu_y]$ and $[\sigma_x, \sigma_y]$ now belong to a collection of sets (the focal elements), also P_C is not a single value but belongs to a set. With reference to Eq. (5) we can now call $f = P_C$ and Φ the desired set of value for P_C .

For example we can have $\Phi = \{P_C | P_C < P_{C0}\}$ with P_{C0} a given threshold on the values of P_C . We can then define the joint uncertainty space U so that $\mathbf{u} = [\mu_x, \mu_y, \sigma_x, \sigma_y]^T$ and write:

$$\Omega = \{\mathbf{u} \in U | P_C(\mathbf{u}) \in \Phi\} \quad (12)$$

With Eq. (12) we could then apply Eqs. (6) and (7) to compute the corresponding *Bel* and *Pl*. At this point it is important to underline that although we started from the simple calculation of the P_C in Eq. (1), this methodology is applicable to more sophisticated definitions of the P_C as long as one can identify the source of epistemic uncertainty and translate it into a bpa assignment.

2.3.1. Example

As an illustrative example, in this paper, we will assume that there exist two sources of observations but each source provides only one interval per each uncertain variable. Hence, each source provides a single set of values for $[\mu_x, \mu_y]$ and $[\sigma_x, \sigma_y]$, which translates into a single family of ellipses on the impact plane of a given target object. We then data-fuse the different sources to build the uncertainty space U and proceed with the calculation of *Bel* and *Pl*. Among the different possible combination rules available in Evidence Theory, the Mixing Rule has been chosen in this paper [22].

Fig. 2 illustrates this procedure. The figure represents the impact plane of a hypothetical target object and a hypothetical scenario where two sources provide one observation each. In Fig. 2, we consider only uncertainty intervals on μ_x and σ_x , while μ_y and σ_y are supposed to be well defined. One of the sources, named Source 1, provides the following information: $\mu_{x1} \in [4, 7]$, $\mu_{y1} = 6$, $\sigma_{x1} \in [1, 2.5]$ and $\sigma_{y1} = 3$. The set of ellipses induced by these parameters is labeled as Source 1 in Fig. 2. The other source, named Source 2, provides the following parameters: $\mu_{x2} \in [15, 20]$, $\mu_{y2} = 6$, $\sigma_{x2} \in [2, 6]$ and $\sigma_{y2} = 3$, which translate into the set of ellipses labelled as Source 2 in Fig.2. The HBR, with radius $R_{HBR} = 5$ is represented with a yellow solid circle centred in (0,0) in Fig. 2.

If both sources are credible and one cannot discard any of the two, the simplest thing to do, under the assumption of purely aleatory uncertainty, is to sample both families and build a joint distribution that can be used to compute P_C . The method used to combine the two sources is not important at this point. The key assumption is that, no matter the combination rule, we need to define a single distribution for the computation of P_C with Eq. (1). We represent this joint distribution with the red ellipse in Fig. 2 and we call the associated value of the Probability of Collision \hat{P}_C . If we judge one source more reliable than the other, we can assign weights to each individual distribution to represent the relative credibility of the sources. The red ellipses in 2a, represents the case in which both sources are equally reliable,

while the red ellipse in 2b, represents the case in which Source 1 is assigned a weight of 0.9 and Source 2 a weight of 0.1. The uncertain ellipse in Fig. 2a is defined by $\mu_{x0} = 11.4$, $\mu_{y0} = 5.8$, and $\sigma_x = 43.1$, $\sigma_y = 8.6$. The uncertain ellipse in Fig. 2b is defined by $\mu_{x0} = 6.6$, $\mu_{y0} = 5.8$, and $\sigma_x = 18.2$, $\sigma_y = 8.6$. Note that, in the case in which the combined ellipse is titled with respect to the x and y axes, a rotation of the reference frame is applied before the computation of P_C .

Given the conflict between the two sources it is clear that there is a degree of epistemic uncertainty that cannot be neglected in the computation of P_C . Thus we can now try to use DSt to capture this epistemic uncertainty and derive a quantification of the uncertainty in the value of P_C .

We start by associating a degree of credibility to each source based on their trustworthiness. If both sources are reliable, we assign $bpa_1 = bpa_2 = 0.5$. This means that we assign a $bpa = 0.5$ to both the intervals of $[\mu_x, \mu_y]$ and $[\sigma_x, \sigma_y]$ coming from each source. This situation is equivalent to the weights used in Fig. 2a to build the joint probability distribution.

On the other hand, as we did in Fig. 2b, we can assign $bpa_1 = 0.9$ to Source 1, $bpa = 0.1$ to Source 2. This example describes the situation in which an operator receives information from sources who are fully confident in their own prediction but the operator is considering one source more reliable than the other and, therefore, allocates their own bpa , reflecting their opinion on the sources.

Starting from the intervals of each variable and the bpa assigned by the operator it is possible to build the Joint Body of Evidence. The operator considers that the sources are independent and there is no constraint on the values of mean and variance. Thus the Body of Evidence is computed through the Cartesian product of all the intervals along each dimension. In this case two intervals per variable for a total of four sets. Each of the resulting set with a non-zero bpa is a Focal Element. Given that the sources are independent the bpa of each FE is given by the product of the bpa 's of the associated intervals. Table 1 shows the FE in the first columns with the associated combined bpa for both cases: sources equally reliable (second column) and Source 1 more credible than Source 2 (third column). To be noted that different combinations of intervals lead to a different range of possible values for P_C . The minimum and maximum value of P_C for each FE are reported in the last two columns. The corresponding Belief and Plausibility Curves can be found in Fig. 3, for the geometry presented in Fig. 2. Fig. 3a represents the case when the two sources are equally reliable and Fig. 3b the case in

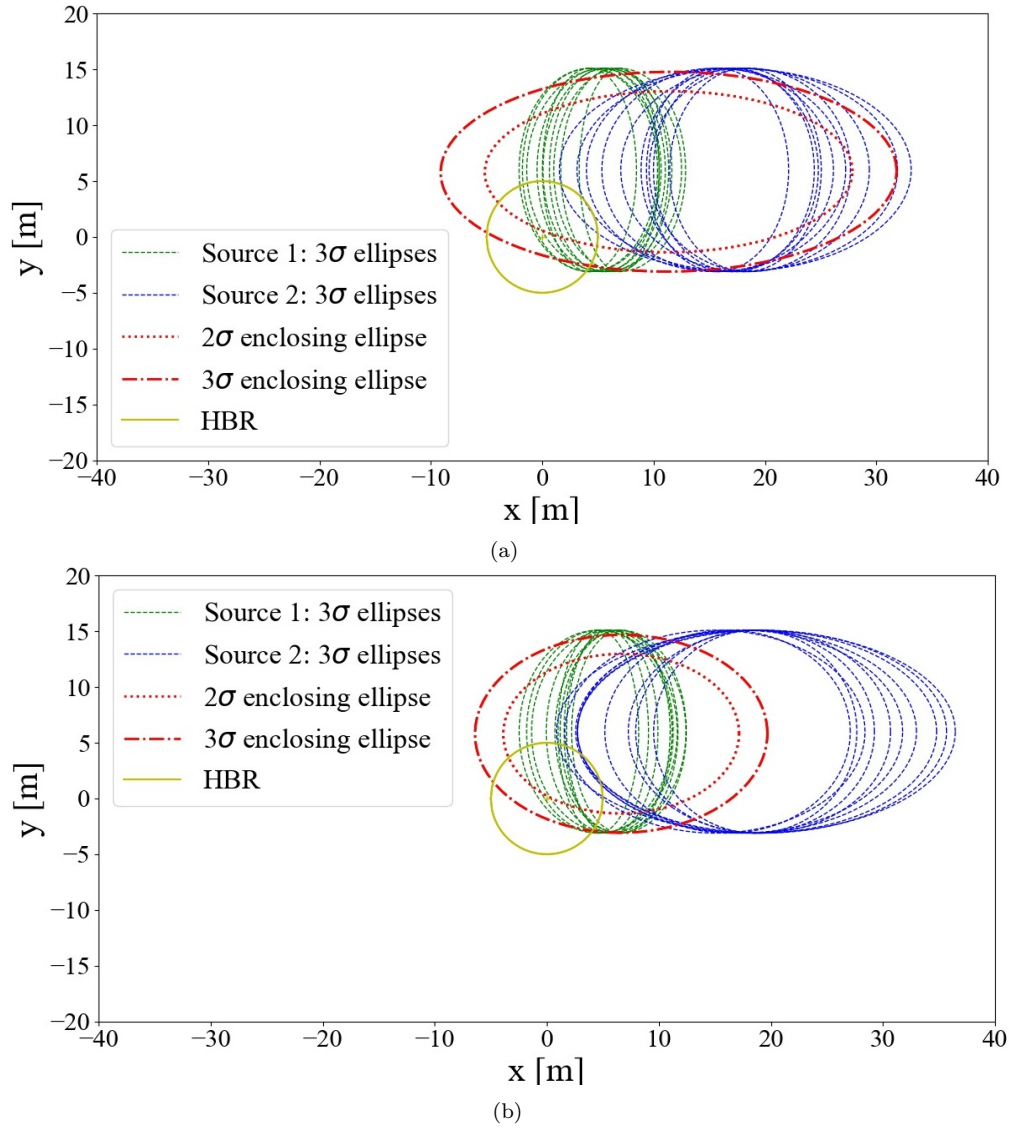


Figure 2: Conjunction geometry on the impact plane, centered on the secondary body. (a) Sources Equally reliable. (b) Sources not equally reliable. Source 1 (dotted green) and Source 2 (dotted blue) represent the family of Normal Distributions provided by intervals. The ellipse enclosing both family of distributions is represented in red: 2σ (solid red) and 3σ (dash red) ellipses. The HBR is represented by the yellow solid line.

which Source 1 is assigned a higher confidence.

As one can see the gap between *Bel* and *Pl* is lower when Source 1 is deemed to be more reliable. This is due to the fact that in this case the operator is less uncertain about the sources.

Table 1: Focal Elements with the associated *bpa* (for sources equally and no equally reliable) and the minimum and maximum P_C on each Focal Element.

Focal Element ([\mathbf{\mu}],[\mathbf{\sigma}])	<i>bpa</i> (equally reliable)	<i>bpa</i> (not equally reliable)	Min P_C	Max P_C
[4, 7],[1, 2.5]	0.25	0.81	1.73×10^{-3}	1.58×10^{-1}
[4, 7],[2, 6]	0.25	0.09	2.37×10^{-2}	1.59×10^{-1}
[15, 20],[1, 2.5]	0.25	0.09	8.88×10^{-37}	3.12×10^{-6}
[15, 20],[2, 6]	0.25	0.01	1.91×10^{-15}	1.06×10^{-2}

3 . Evidence-based Collision Classification Criteria

This section presents how to use *Bel* and *Pl* as additional criteria to classify conjunction events. We will compare a classification purely based on P_C with a classification that includes *Bel* and *Pl*. This classification system will form the learning ground of the machine learning algorithm we propose in this paper.

3 .1. P_C as Classification Criterion

When Probability of Collision in Eq. (1) is used as single classification criterion, it is customary [4] to define a series of thresholds on the value of P_C , each of which activates an alert or a recommended action. Generally speaking, any event with a $P_C > 10^{-7}$ starts a Collision Avoidance Manoeuvre (CAM) assessment process, a $P_C > 10^{-6}$ triggers additional tracking, a $P_C > 10^{-5}$ triggers the planning of a CAM and an event with a $P_C > 10^{-4}$ triggers the execution of a CAM. The details of the process vary from operator to operator, with variations on the specific value of the thresholds, the action to be carried out or the inclusion of other considerations, like the time to Time to Closest Approach (TCA). For examples, the CARA team of NASA uses two thresholds (10^{-7} and 4.4×10^{-4}) [23]. Anything below the

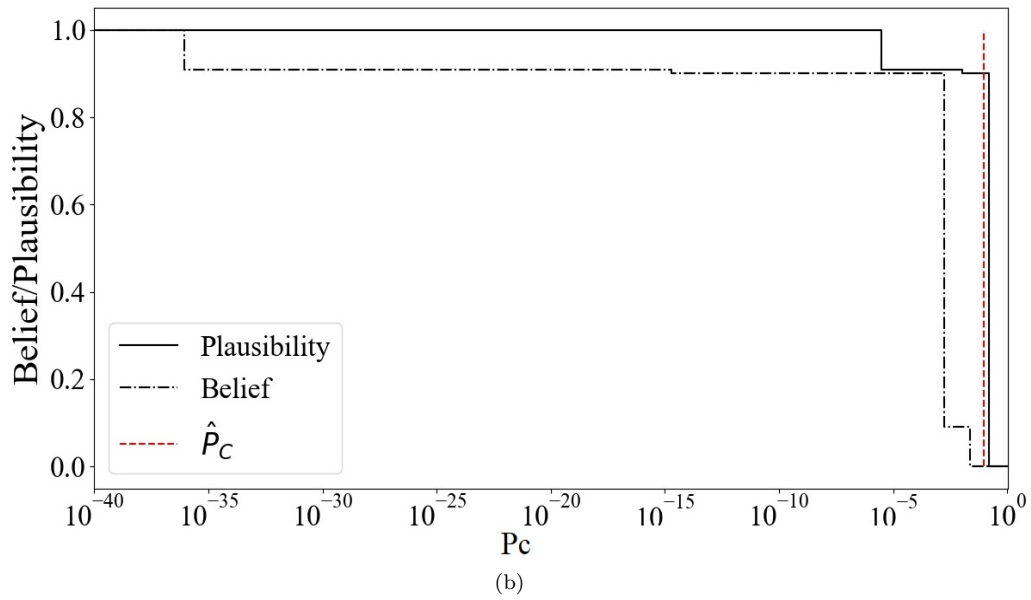
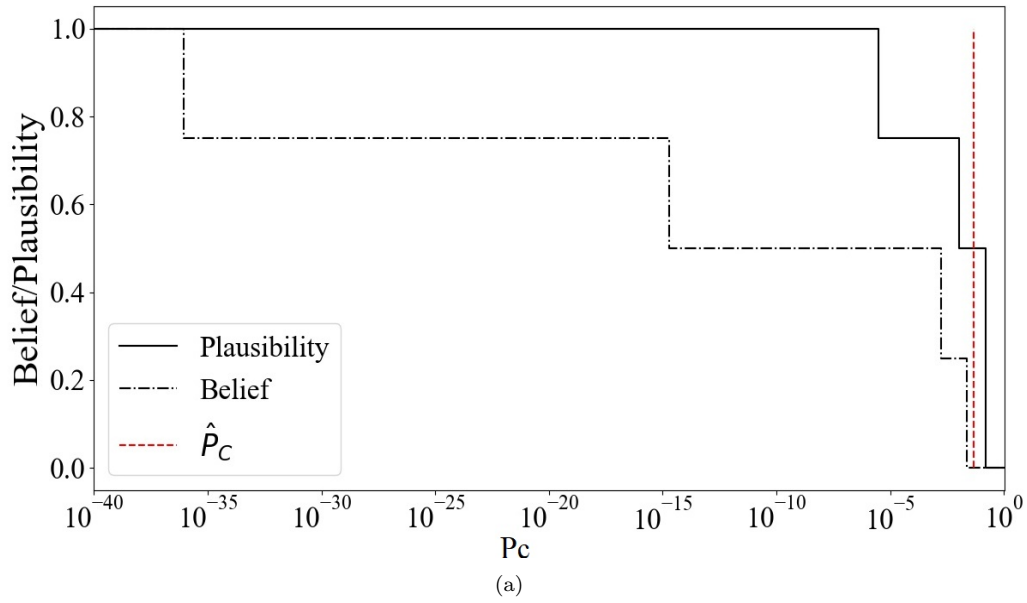


Figure 3: Belief (dash line) and Plausibility (solid line) curves for the geometry in 2 and \hat{P}_C for the combined ellipse (vertical line): (a) equally reliable sources; (b) Source 1 more reliable than Source 2.

former threshold is considered to be low risk while everything above the latter threshold is considered high risk and anything in between deserves further investigations. The Space Debris Office (SDO) of ESA, also considers the time to the TCA and perform a CAM if the value of the P_C is higher than a given threshold a number of days before the TCA [24]. All these models and criteria always assume that the P_C is affected exclusively by aleatory uncertainty.

3.2. Evidence-based Classification Criteria

As explained in the previous section, a classification based solely on P_C triggers actions when the Probability of Collision is found to be higher or lower than a given threshold. However, no information is given on the correctness of P_C . Thus the operator might react to false positives or do nothing in the case of false negatives. Furthermore, unknown cases, that require further observations, are identified only by the thresholds and not by an actual quantification of the degree of ignorance on the probability of a collision.

Consider now the case in which one can quantify the Bel and Pl that the P_C is above a given threshold P_{C0} . In this case three additional pieces of information are available to the operator: the value of Bel , the value of Pl and the gap between the two. A large gap between Pl and Bel implies that there is a lack of information on the calculation of P_C . If both Pl and Bel are low then there is little evidence that the value of P_C is to be trusted. Finally if Bel is low one should be wary that the evidence fully supporting the value of the P_C is low.

Starting from these considerations one can devise a classification that allows operators to differentiate between high-risk, low-risk and uncertain events based on the degree of confidence in the correctness of the P_C . In Table 2 we propose a possible classification exploiting Bel and Pl (category *Criterion 1*).

In the category *Criterion 1* of Table 2 we define three classes: High-Risk (HR), Low-Risk (LR) and Not-Known (NK). Within the last class we further group events in two subclasses. Events classified as subclass 1 in the NK class are those with a low degree of confidence for the value of P_C , the difference between Plausibility and Belief is too high and there is too much uncertainty on the correctness of the value to make a decision. The response to this situation is to collect more information to reduce the knowledge gap. Subclass 2 in the NK class implies a low confidence level in the correctness of the value of P_C . In this case the knowledge gap is low and the evidence

Table 2: Classification criteria according to the proposed three different Event-based approaches: Criterion 1, Criterion 2 and Criterion 3.

CRITERION 1				
P_C value	Pl-Bel gap	Plausibility/Belief	Class	Subclass
$P_C \geq P_{C0}$	$Pl - Bel > \Delta$	-	NK	1
	$Pl - Bel \leq \Delta$	$Pl < Pl_0$	NK	2
		$Bel \geq Bel_0$	HR	-
$P_C < P_{C0}$	$Pl - Bel > \Delta$	-	NK	1
	$Pl - Bel \leq \Delta$	$Pl < Pl_0$	NK	2
		$Bel \geq Bel_0$	LR	-

CRITERION 2			
	P_C for $Bel = Bel_0$	Degree of confidence at P_{C0}	Class
	$P_{Cb} \geq P_{C0}$	-	HR
	$P_{Cb} < P_{C0}$	$Pl(P_{C0}) - Bel(P_{C0}) \leq \Delta$	LR
		$Pl(P_{C0}) - Bel(P_{C0}) > \Delta$	NK

CRITERION 3			
Time to TCA	P_C for $Bel = Bel_0$	Degree of confidence at P_{C0}	Class
$t_{TCA} < T1$	$P_{Cb} \geq P_{C0}$	-	1
	$P_{Cb} < P_{C0}$	$Pl(P_{C0}) - Bel(P_{C0}) \leq \Delta$	5
		$Pl(P_{C0}) - Bel(P_{C0}) > \Delta$	1
$T1 \leq t_{TCA} < T2$	$P_{Cb} \geq P_{C0}$	-	2
	$P_{Cb} < P_{C0}$	$Pl(P_{C0}) - Bel(P_{C0}) \leq \Delta$	5
		$Pl(P_{C0}) - Bel(P_{C0}) > \Delta$	3
$T2 \leq t_{TCA}$	$P_{Cb} \geq P_{C0}$	-	2
	$P_{Cb} < P_{C0}$	$Pl(P_{C0}) - Bel(P_{C0}) \leq \Delta$	4
		$Pl(P_{C0}) - Bel(P_{C0}) > \Delta$	3

supporting the correctness of the value of P_C is also low, thus the operator can consider the value of P_C to be not credible. As a consequence a decision cannot be made with that value of P_C .

3.2.1. Criteria Comparison

In this section we use a database of virtual conjunction geometries with different combinations of miss distances and standard deviations to compare different classification methods. We consider a classification method solely based on \hat{P}_C , a classification that computes Bel and Pl at \hat{P}_C , a classification that looks at the values of Bel and Pl for $P_C < P_{C0}$ and $P_C \geq P_{C0}$ and an extended classification approach with additional criteria.

Five different scenarios have been identified depending on the information provided by the two sources, as explain in Section 2 . Each scenario corresponds to a particular set of geometries. Each of the sets of geometries includes several representative individual encounter cases. Table 3 shows the intervals for each variable. On each of the sets of geometries, the first row refers to the intervals provided by Source 1 and the second row the intervals of Source 2. Geometry 1 represents cases where the families of ellipses provided by both sources tightly overlap each other and both overlap HBR. In this scenario both sources are in agreement and support the fact that a collision might occur with high probability. Geometry 2 are cases where the two set of ellipses tightly overlap each other but do not overlap HBR. In this scenario both sources are in agreement and support the fact that a collision is unlikely to occur. In Geometry 3, both sources are disjoint but one overlaps HBR. In this scenario there is a high degree of uncertainty on the correctness of the prediction. In Geometry 4 both are disjoint and none overlaps HBR. In this scenario joint distribution at 3σ is assumed not to intersect the HBR thus although the sources are in disagreement the conclusion is that a collision is unlikely to happen. Finally, Geometry 5 includes cases where the ellipses are disjoint but the joint distribution at 3σ intersects the HBR. In this scenario the sources are in disagreement but the conclusion is that a collision is highly likely to occur because \hat{P}_C is high. The key difference between Geometry 5 and 3 is that in Geometry 5 the high value of \hat{P}_C is an artifact because both sources think that no collision is going to happen.

Geometry 1 and 2 present low conflict between the sources and can be expected to provide high Bel at \hat{P}_C with little gap between Pl and Bel . Geometry 1 will have $\hat{P}_C > P_{C0}$ and Geometry 2 will have $\hat{P}_C < P_{C0}$. Geometry 3 includes conflict cases, but depend on the distribution of confidence among

Table 3: Bounds for the upper and lower values of the uncertainty intervals for of $[\mu_x, \mu_y]$ and $[\sigma_x, \sigma_y]$ for each of the set of geometries. The first row represents the limits for Source 1 and the second row for Source 2.

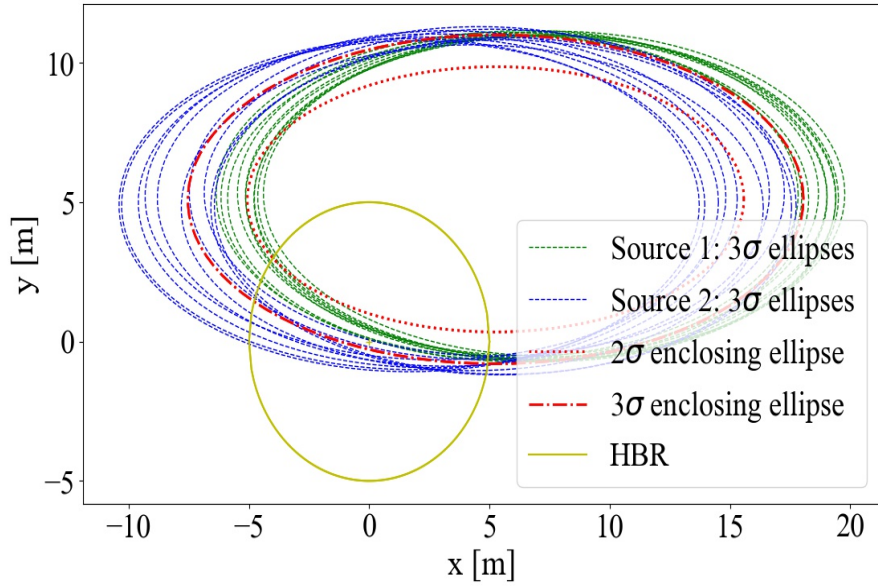
Geometry		μ_x [m]	μ_y [m]	σ_x [m]	σ_y [m]
Geo. 1	Ellipses overlapping	[0,10]	[0,7]	[0.1,4]	[0.2,2]
	Both overlapping HBR	[0,10]	[0,7]	[0.1,4]	[0.2,2]
Geo. 2	Ellipses overlapping	[750,1000]	[-100,100]	[10,25]	[10,18]
	None overlapping HBR	[750,1000]	[-100,100]	[10,25]	[10,18]
Geo. 3	Ellipses not overlapping	[0,10]	[0,7]	[0.1,4]	[0.2,2]
	One overlapping HBR	[750,1000]	[-100,100]	[10,25]	[10,18]
Geo. 4	Ellipses not overlapping	[750,1000]	[200,500]	[10,25]	[10,18]
	None overlapping HBR	[500,800]	[-300,0]	[10,25]	[10,18]
Geo. 5	Ellipses not overlapping	[30,200]	[30,200]	[3,12]	[3,12]
	Both overlapping HBR	[-200,-30]	[-200,-30]	[3,12]	[3,12]

the sources, they can present high values of Bel at \hat{P}_C (Low Risk or High Risk) or big gaps between Pl and Bel (Not Known due to uncertainty).

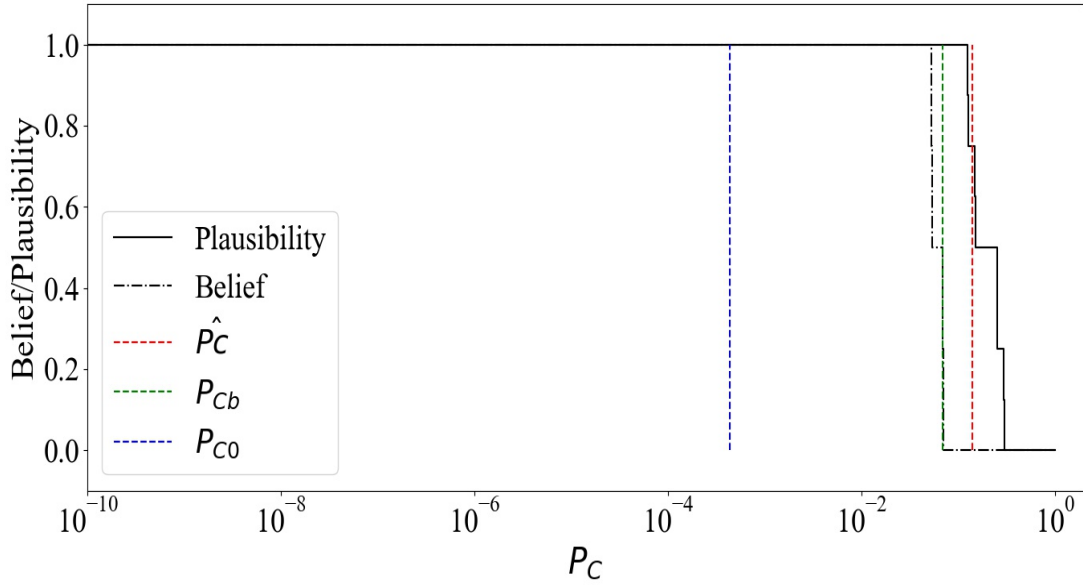
A total of 3000 sample geometries have been created, 600 samples for each type of geometry. Within this 600 samples per geometry, one third (200 samples) corresponds to cases where both sources are equally reliable, another 200 samples to cases where Source 1 is 9 times more reliable than Source 2 and the another remaining third of the samples to cases where Source 1 is nine times less reliable than Source 2.

In Fig. 4 one can see one example of the geometry of an encounter event of the family *Geo. 1* with the associated Bel and Pl curves. Examples of the other families can be found in Appendix A. It is now possible to compare the outcome of the different classification methods.

As indicated before only one threshold on Probability of Collision, P_{C0} is considered, in order to allow for a simpler classification, using \hat{P}_C only, between Collision and No-Collision cases. The values of the thresholds used in this example can be found in Table 4.



(a)



(b)

Figure 4: Representative scenario of a single case of family *Geo*.
 1: (a) encounter geometry, (b) the associated *Bel* and *Pl* curves.

Table 4: Classification thresholds for Evidence-based Criteria

Parameter	Value	Criterion
Probability of Collision (P_{C0})	4.4×10^{-4}	1, 2, 3
Degree of confidence (Δ)	0.3	1, 2, 3
Belief (Bel_0)	0.5	1, 2, 3
Plausibility (Pl_0)	0.5	1
Lower time threshold (T1)	2 days	3
Upper time threshold (T2)	4 days	3

Fig. 5 shows two histograms comparing a classification based on \hat{P}_C only and a classification that accounts for Bel and Pl at \hat{P}_C . In Fig. 5a cases classified as Collisions are divided in actual High Risk events (black), Not Known events category 1 (blue) and Not Known events category 2 (light blue). In the same figure cases classified as No Collision are further divided in actual Low-Risk events (red), Not Known events class 1 (orange), and Not Known events class 2 (pink). Fig. 5b shows that the use of Bel and Pl provides a better classification as it moves all false positive and false negative into a Not Known class.

On the other hand the Not Known class contains a number of cases that would reasonably correspond to actual Collisions or No Collision but are classified as Not Known because of the way \hat{P}_C is computed. In other words the use of a joint distribution encompassing the distributions provided by the two sources leads to a significant degree of uncertainty that leads to the need to acquire further, more precise, information. For the numerical values of each of the bars in Fig. 5, broken down by Geometry families, go to Table 13 in Appendix A.

Consider now the value of Probability of Collision P_{Cb} at which Bel switches from $Bel \geq Bel_0$ to $Bel < Bel_0$. If P_{Cb} is to the right of P_{C0} the operator has enough evidence that any value of P_C between P_{C0} and P_{Cb} is credible and, therefore, the event has to be classified as High-Risk. If P_{Cb} is lower than P_{C0} then one has to check the $Pl - Bel$ gap at P_{C0} and beyond. If the gap at P_{C0} is high then there is some evidence that supports the fact that the event might be high risk but there is not enough information to make a decision. On the contrary if the gap is small one can consider the event low-risk. We can now introduce the new classification *Criterion 2* as defined in Table 2.

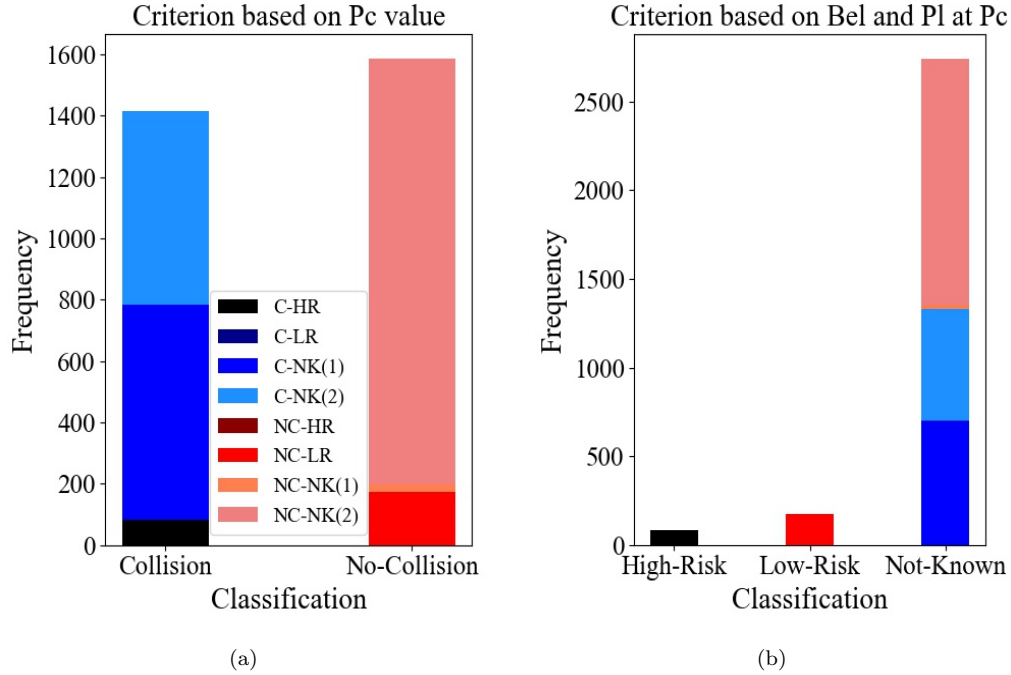


Figure 5: Histograms comparing criterion using only value of P_C (a) and E-b Criterion 1 (b). Each color represents the fraction of samples from each bin on Histogram (a) that moves to one of the bins on Histogram (b).

Using the samples of the previous example, Fig. 6 shows two histogram comparing the Probabilistic criteria based on \hat{P}_C Fig. 6a and the Evidence-based Criterion 2 Fig. 6b, using the thresholds indicated in Table 4. It can be seen that this criterion present much fewer cases classified as Not-Known that the previous one. Moreover, events classified as High-Risk and Low-Risk are free of false confidence. It is worth noting that some events previously classified as Collision are now in the Low-Risk bins, which suggests that there are some events where there is lack of support to the value \hat{P}_C . For the numerical values of each of the bars in Fig. 6, broken down by Geometry families, see Table 14 in Appendix A.

All events that are still classified as Not-Known would suggest two actions: perform a CAM anyway or acquire more information. The decision is often dictated by the available time and resources. If the time to impact is included one could further refine the classification and make a decision based on the

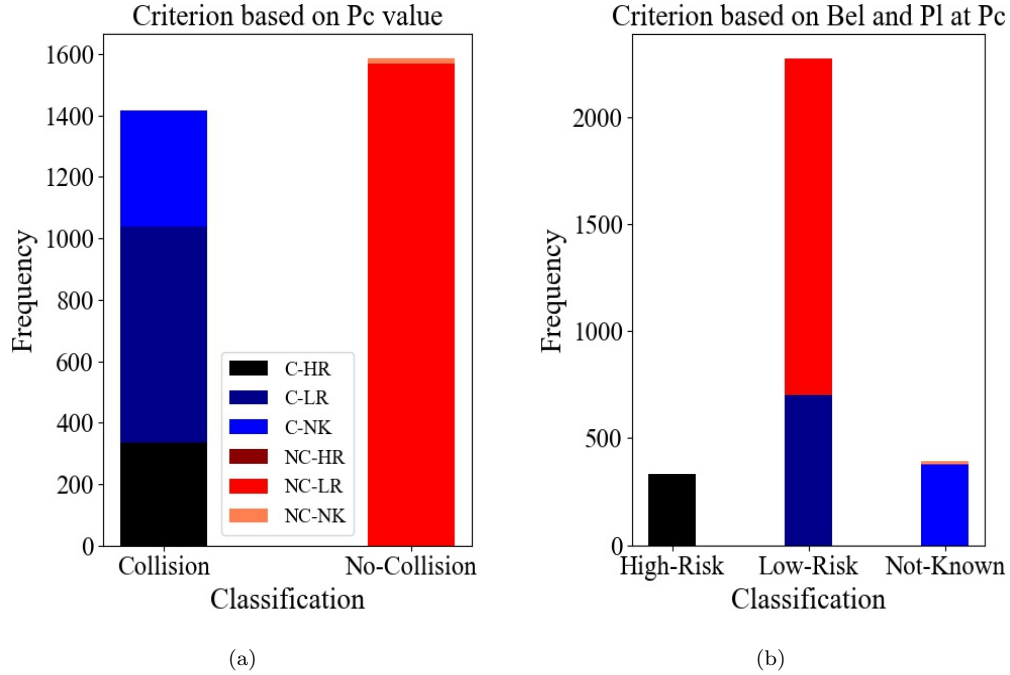


Figure 6: Histograms comparing criterion using only value of P_C (a) and E-b Criterion 2 (b). Each color represents the fraction of samples from each bin on Histogram (a) that moves to one of the bins on Histogram (b).

available time to obtain better quality data.

3.3. Extended Classification Criteria

In order to improve the decision making process, it is desirable to include other parameters on top of Bel and Pl at a given P_C . The goal is to improve the confidence in the decisions of the operators. In particular decision support is required in all those cases that are classified as Not-Known.

To this end, in this section we extend the classification process by including the Time to Closest Approach (TCA). Two thresholds have been considered, dividing events in: long-term, mid-term and instantaneous. Category *Criterion 3* in Table 2 summarises the classes and classification criteria. Events in Class 1 are High-Risk and require the execution of a CAM. Events in Class 2 are High-Risk but there is time to collect better information, either to improve the manoeuvre planning or re-assess the level of risk. Events in

Class 3 are not supported by enough evidence but there is sufficient time to acquire more information. Events in Class 4 are Low-Risk and the time to the encounter is high enough to acquire further information. Events in Class 5 are Low-Risk but there is little time to acquire further information and the best decision is to take no action.

Events with high level of uncertainty but close to the TCA are classified as Class 1, choosing a conservative approach from the point of view of safety. Further analyses can include other considerations, as a comparison between the risk of the event and the inherent risk of executing a manoeuvre or the cost of the manoeuvre against the cost of the satellite pondered by the risk of the encounter.

Fig. 7 shows the histograms comparing the probabilistic classification with this last Evidence-based approach using the same sample as the previous examples and the thresholds stated in Table 4. In order to account for the time to the TCA, three different times to the TCA have been associated to each sample, creating three event per each one. These three times belonged to one of the three time bins (long-term, mid-term and instantaneous). For the numerical values of each of the bars in Fig. 7, broken down by Geometry families, see Table 15 in Appendix A.

In the next section, an Intelligent Event Classification System is presented for automatically performing the classification presented in this Section.

4 . Intelligent Event Classification System

This section introduces an Intelligent Event Classification System (IECS) that exploits the Evidence-based Classification Criteria presented in previous sections to provide decision making support to operators. The IECS uses Machine Learning (ML) to predict the classes of risk of the events given a close encounter geometry parameters. The main advantage of such a system is that it can provide a fast an automatic classification of conjunction events from observations.

4 .1. Classification Systems

Two intelligent event classification systems are proposed: *System 1* takes $Pl(P_{C0}) - Bel(P_{C0}) = \Delta(P_{C0})$, time to the TCA, t_{TCA} , and P_{Cb} as inputs and returns the 5 risk classes discussed in previous sections; *System 2* takes the upper and lower bounds of each of the intervals of each variable $[\mu_x, \mu_y]$ and $[\sigma_x, \sigma_y]$, the corresponding *bpa*, the time t_{TCA} to the TCA and returns

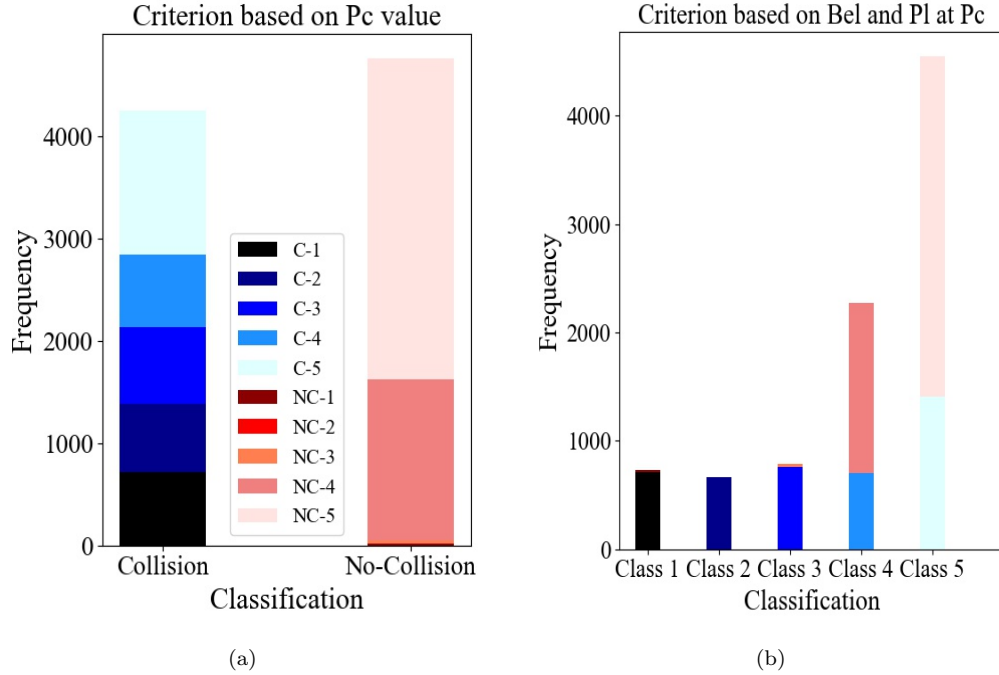


Figure 7: Histograms comparing criterion using only value of P_C (a) and E-b Criterion 3 (b). Each color represents the fraction of samples from each bin on Histogram (a) that moves to one of the bins on Histogram (b).

the 5 classes of risk. Thus the second system bypasses the computation of Bel and Pl , which is a time-consuming step, especially when several sources provide information. In practice the intervals for $[\mu_x, \mu_y]$ and $[\sigma_x, \sigma_y]$ would come from a propagation of the orbital parameters at a time $t < t_{TCA}$. In this paper we assume that we already computed propagated intervals. See [13] for an example of fast propagation of the orbital parameters of pieces of debris.

4.1.1. Synthetic Data-sets

In order to test the two classification systems, we generated two data-sets of synthetic data representing a variety of conjunction scenarios: data-set DB_1 and data-set DB_2 . DB_1 contains 9,000 different conjunction scenarios built using the intervals in Table 3. Each set of geometries contributes with 1,800 samples where the encounter geometry is provided by two Sources

of information in the form of intervals for each of the uncertain variables and associated bpa_s . These 1,800 samples are distributed in groups of 600 samples into each of the three times bins indicated in Section 3.3: immediate encounters ($t_{TCA} < 2$ days), medium-term encounters ($2 \text{ days} \leq t_{TCA} < 4$ days) and long-term encounters ($t_{TCA} \geq 4$ days), where t_{TCA} is the time to the TCA. Each of these groups are composed of three subgroups of 200 samples: in one subgroup the sources are equally reliable ($bpa_1 = bpa_2 = 0.5$), in a second sub-group Source 1 has an associated $bpa_1 = 0.9$ and Source 2 an associated $bpa_2 = 0.1$, and in the third sub-group, Source 1 has an associated $bpa_1 = 0.1$, and Source 2 and associated $bpa_2 = 0.9$.

Data-set DB_2 includes other 19,800 samples, out of which 9,000 have been obtained similarly to the samples in DB_1 and 10,800 feature an extended set of geometries divided in 9 sub-sets characterised by the intervals in Table 5. For each of the 9 sub-sets half of the samples have both sources equally reliable (see bpa_s defined for DB_1), and the other half, with Source 1 more reliable than Source 2. Within each half three different encounter geometries have been generate by changing the time to the TCA, t_{TCA} . The rest of the samples are the same as in DB_1 . In total, DB_2 is made of 28,800 samples.

For each data-set, we used 80% of the samples for training and 20% for testing (Test Set). We labelled DB_{11} and DB_{12} the training and test data-sets from DB_1 , and DB_{21} and DB_{22} the training and test data-sets from DB_2 . The training sub-set was further divided into 80% for the purely training samples (Training Set) and 20% for validation samples (Validation Set) using during the training process to avoid overfitting the system.

4.2. Machine Learning Techniques

Different supervised ML techniques have been tested to assess which one is best at predicting the classes of risk. The techniques tested in this paper are: Artificial Neural Networks (ANN) [25], Random Forest (RF) [26], K-Nearest Neighbours (KNN) [25] and Support Vector Machine (SVM) [27].

The **Artificial Neural Network** trained in this paper contains only one hidden layer. Since it is a classification problem, the output layer contains five nodes, one per class, indicating the probability of the output of being one of them. The final output corresponds to the Class presenting a higher probability. Ideally, when inputs of a class k are provided, all nodes in the output layer should take value 0 but node k that should take value 1. The only hyperparameter modified in this study was the number of neurons in the hidden layer: 10, 20, 50, 100. Each neuron posses a hyperbolic tangent

activation function. The optimiser is based on the Levenberg-Marquardt method and the loss function is the Multi-Class Cross-Entropy. Each configuration has been trained iteratively 10 times with different initialization of the parameters (weights and bias of the neurons). The ANN providing

Table 5: Bounds for the upper and lower values of the uncertainty intervals for of $[\mu_x, \mu_y]$ and $[\sigma_x, \sigma_y]$ for each of the set of geometries used for creating DB_2 . The first row represents the limits for Source 1 and the second row for Source 2.

Geometry		μ_x [m]	μ_y [m]	σ_x [m]	σ_y [m]
Geo. 1	Low μ	[0,4]	[0,4]	[0.1,5]	[0.1,5]
	Low σ	[2,20]	[2,20]	[0.1,5]	[0.1,5]
Geo. 2	Low μ	[0,4]	[0,4]	[10,20]	[10,20]
	High σ	[2,20]	[2,20]	[10,20]	[10,20]
Geo. 3	Low μ	[0,4]	[0,4]	[0.1,20]	[0.1,20]
	Wide σ	[2,20]	[2,20]	[0.1,20]	[0.1,20]
Geo. 4	High μ	[100,2000]	[100,2000]	[10,100]	[10,100]
	Low σ	[100,2000]	[100,2000]	[10,100]	[10,100]
Geo. 5	High μ	[100,2000]	[100,2000]	[500,5000]	[500,5000]
	High σ	[100,2000]	[100,2000]	[500,5000]	[500,5000]
Geo. 6	High μ	[100,2000]	[100,2000]	[10,5000]	[10,5000]
	Wide σ	[100,2000]	[100,2000]	[10,5000]	[10,5000]
Geo. 7	Low μ Low σ	[0,5]	[0,5]	[0.1,5]	[0.1,5]
	High μ Low σ	[100,500]	[100,500]	[10,100]	[10,100]
Geo. 8	Low μ High σ	[0,5]	[0,5]	[10,20]	[10,20]
	High μ Low σ	[100,500]	[100,500]	[10,100]	[10,100]
Geo. 9	Low μ Low σ	[0,5]	[0,5]	[0.1,5]	[0.1,5]
	High μ High σ	[100,500]	[100,500]	[500,2000]	[500,2000]

the lowest value of the loss function on the Validation Set has been selected as the best model. The ANN was implemented using the MATLAB *Deep Learning* Toolbox. [28]

Random Forest is an ensemble method that combines several independent Decision Trees during the training step, feeding each of them with different subsets of the training set. The predicted class is the mode of the output of every single tree. Random Forests overcome the overfitting problem usually faced by Decision Trees.

In this paper, a few RFs configurations have been trained, by modifying the number of trees in the forest, *n_estimators*, (50, 100, 200, 400), the maximum depth of the tree, *max_depth*, ('None', 50, 100), the minimum number of samples required to be at a leaf node, *min_samples_leaf*, (1, 10^{-4} , 10^{-7}), the minimum number of samples to split a node, *min_samples_split*, (2, 20), and the number of features to consider when looking for the best split, *max_features*, ('auto', 0.5, 'log2'). All the other parameters were kept equal to the default values in the Scikit library of Python [29]. The model with the combination of hyperparameters that provides the best results over the Validation Set is saved as the best model.

The **K-Nearest Neighbours** algorithm predicts the output of a given sample by comparing its proximity with the training samples. Once an input vector is given, the closeness to the other points is computed (it can be Euclidean distance or any other metric) and the K nearest points are selected. The class of the sample is selected as the mode of those points. The selection of the optimum K depends on the problem.

In this paper, different KNN models were trained, each one with a different number of neighbors, *n_neighbors*, K (1,3,5), weight function, *weights* ('uniform' and 'distance') and leaf size, *leaf_size*, (30,60). All the other hyperparameters were kept equal to the default values in the Scikit library of Python [30]. The model with the combination of hyperparameters that provides the higher accuracy over the Validation Set is saved as the best model.

Support Vector Machine algorithms classify samples from different classes by building the hyperplane that splits the classes of the training point, maximising the distance to the nearest point on each class. Kernels can be used to define the problem in a higher dimensional space where the division may be easier. In this paper, only the decision function of shape, *decision_function_shape*, ('ovo' or 'ovr') available in the Scikit library on SVM in Python [31] were tested. The best model is the one which higher

accuracy over the Validation Set.

The accuracy of the best model of each classification method was tested over the Test Set of simulated conjunction events. The overall accuracy of the system over all the samples and the precision and recall by class are provided as the score to compare methods and systems. Accuracy is defined as the percentage of samples correctly predicted over all the samples, precision is the percentage of the samples correctly predicted from a certain class over the total number of samples predicted on this class, and recall is the percentage of samples correctly predicted on a certain class over the total number of samples in that class. The results are presented in the next section.

4.3. Classification Results

Table 6 shows the results using the best settings of each ML technique for *System 1*. It can be seen that all ML methods provide good results and an accurate classification. RF and ANN show slightly better results, with accuracy levels shy of 100%. The hyper-parameters selected for the best model of each technique can be found in Appendix A.

Note, however, that the Test Set contains only 5 representative geometries because it is taken from DB_1 . Thus we put the same ML models to the test taking DB_{22} as Test Set. The result can be found in Table 7. It can be

Table 6: Overall accuracy and Precision and Recall by classes. *System 1* trained with DB_{11} and tested with DB_{12} . The best technique is underlined.

Method	Total Acc.		Class 1	Class 2	Class 3	Class 4	Class 5
ANN	99.9	Prec.	98.8	99.8	100	100	99.9
		Rec.	100	99.6	99.4	99.8	100
<u>RF</u>	<u>99.9</u>	Prec.	<u>100</u>	<u>100</u>	<u>99.4</u>	<u>100</u>	<u>100</u>
		Rec.	<u>98.8</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>
KNN	97.0	Prec.	88.9	93.4	100	99.8	99.8
		Rec.	98.8	99.6	93.8	96.3	97.7
SVM	99.2	Prec.	98.7	100	100	99.2	99.0
		Rec.	96.3	99.1	93.8	99.6	100

Table 7: Overall accuracy and Precision and Recall by classes. *System 1* trained with DB_{11} and tested with DB_{22} . The best technique is underlined.

Method	Total Acc.		Class 1	Class 2	Class 3	Class 4	Class 5
ANN	95.6	Prec.	91.3	84.4	100	99.7	99.8
		Rec.	73.4	95.0	99.1	99.8	100
<u>RF</u>	<u>99.9</u>	Prec.	<u>99.6</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>
		Rec.	<u>100</u>	<u>99.8</u>	<u>99.7</u>	<u>100</u>	<u>100</u>
KNN	93.6	Prec.	90.7	90.3	96.0	96.7	95.5
		Rec.	91.6	85.9	92.6	96.0	96.6
SVM	85.2	Prec.	96.6	99.6	92.2	81.8	81.7
		Rec.	63.3	50.0	91.3	99.4	100

seen that the performances of all the techniques, if slightly worse, are in general over 90% accuracy, even close to 100% for some techniques as RF. Apart of the levels of accuracy, it also demonstrates that different geometries correspond to similar classifications using the *Bel* and *Pl* curves.

Table 8 presents the performance of all ML techniques on classification System 2, trained with training set DB_{11} and tested on data-set DB_{12} .

When using classification System 2 the performance of all ML methods degrades due to the wider set of input parameters and the more complex relationship between input parameters and risk classes. This drop is specially remarkable for KNN and SVM. In any case, both the accuracy of ANN and RF remains relatively high, with values above 80%.

Table 9 shows the prediction accuracy of the ML algorithms using classification System 2, DB_{11} for training and DB_{22} for testing. It can be seen, that opposite to what happened in Table 7 where accuracy levels remained high, here they fall for all ML algorithms. The main reason is the extended range of the input parameters combined with the more nonlinear link between input and output underneath System 2.

More complex ML models, for examples deeper RF or ANN with more Hidden Layers, might improve the results. However, the problem here is

Table 8: Overall accuracy and Precision and Recall by classes.
System 2 trained with DB_{11} and tested with DB_{12} . The best technique is underlined.

Method	Total Acc.		Class 1	Class 2	Class 3	Class 4	Class 5
<u>ANN</u>	<u>84.6</u>	Prec.	<u>48.7</u>	<u>75.2</u>	<u>90.4</u>	<u>99.3</u>	<u>99.5</u>
		Rec.	<u>97.5</u>	<u>98.6</u>	<u>81.5</u>	<u>84.4</u>	<u>84.2</u>
RF	84.4	Prec.	48.7	73.7	89.4	99.5	99.5
		Rec.	97.5	98.6	78.4	84.4	84.4
KNN	58.3	Prec.	44.1	68.0	73.4	100	65.0
		Rec.	79.0	84.8	51.2	77.1	86.5
SVM	51.3	Prec.	0.0	43.4	0.0	0.0	64.1
		Rec.	0.0	84.4	0.0	0.0	89.0

Table 9: Overall accuracy and Precision and Recall by classes.
System 2 trained with data-set DB_{11} and tested with data-set DB_{22} . The best technique is underlined.

Method	Total Acc.		Class 1	Class 2	Class 3	Class 4	Class 5
ANN	63.6	Prec.	77.4	59.9	15.9	77.4	86.4
		Rec.	51.7	56.4	60.4	67.6	68.9
<u>RF</u>	<u>77.8</u>	Prec.	<u>83.5</u>	<u>70.5</u>	<u>57.7</u>	<u>78.9</u>	<u>80.1</u>
		Rec.	<u>62.8</u>	<u>47.3</u>	<u>47.9</u>	<u>93.0</u>	<u>92.9</u>
KNN	58.9	Prec.	73.2	58.6	30.4	50.0	60.6
		Rec.	62.8	47.9	47.9	93.0	92.9
SVM	53.3	Prec.	0.0	41.9	0.0	0.0	60.7
		Rec.	0.0	87.1	0.0	0.0	87.8

Table 10: Overall accuracy and Precision and Recall by classes.
System 2 trained with training data-set DB_{21} and tested with data-set DB_{22} . The best technique is underlined.

Method	Total Acc.		Class 1	Class 2	Class 3	Class 4	Class 5
ANN	92.4	Prec.	88.0	82.1	75.0	99.1	99.1
		Rec.	98.7	96.3	71.4	91.8	91.9
<u>RF</u>	<u>93.3</u>	Prec.	<u>88.4</u>	<u>83.5</u>	<u>82.9</u>	<u>99.2</u>	<u>99.2</u>
		Rec.	<u>98.7</u>	<u>97.4</u>	<u>77.7</u>	<u>92.4</u>	<u>92.4</u>
KNN	61.9	Prec.	73.7	62.1	61.5	66.7	60.0
		Rec.	59.4	78.2	43.8	0.0	88.5
SVM	53.8	Prec.	0.0	46.0	0.0	0.0	57.6
		Rec.	0.0	46.0	0.0	0.0	57.6

more the limited range of the training set than the ML algorithms. A more sensible approach would be using a bigger and, especially, wider training set. Table 10 shows the results of the best model of all ML algorithms using classification System 2, training data-set DB_{21} and test data-set DB_{22} . It can be seen that results improve due to the extended training data-set.

Table 11, shows the training time for the best settings of each ML method. The running time during testing are included in Table 12, for each of the cases studied in this paper. The machine used for the training and prediction of all the models has been an Intel Core i7-3520M CPU @ 2.90GHz.

From the comparison between Table 11 and Table 12 one can see that training is the most expensive step. The training time grows with the size of the dataset, as it the difference in training time between DB_{11} and DB_{21} demonstrates. Among the different methods, the ANNs requires the longest training time, which is due to the more complex structure and greater number of hyperparameters. The complexity of ANN also explains the longer training time in the case of System 2.

Nevertheless, from the point of view of the automation of operations, training times are not as relevant as prediction times. Furthermore the initial training is expected to be the most time demanding while all following

updates to improve predictions with new verified data are expected to be faster.

Table 12 shows that prediction times are range from tens of seconds to a few seconds to generate predictions over the whole test set.

Table 11: Training time (in seconds) for the best settings of each method applied to each *System*.

Method	System 1		System 2	
	DB ₁₁	DB ₂₁	DB ₁₁	DB ₂₁
ANN	54.378	1151.576	429.119	9899.855
RF	0.481	0.494	0.577	7.543
KNN	0.997	0.256	0.146	0.795
SVM	0.809	7.987	1.668	19.730

Table 12: Running time (in seconds) for prediction over Test Set for the best settings of each method.

Method	System 1		System 2		
	Trained in DB ₁₁		Trained in DB ₁₁		Trained in DB ₂₁
	Val. DB ₁₂	Val. DB ₂₂	Val. DB ₁₂	Val. DB ₂₂	Val. DB ₂₂
ANN	1.275	1.296	1.221	1.554	1.523
RF	0.045	0.084	0.318	0.220	0.555
KNN	0.122	0.282	0.227	0.615	0.905
SVM	0.144	0.321	0.401	1.032	3.702

5 . Conclusions

In this paper, we proposed a new approach to improve the classification of conjunction events. The new approach accounts for epistemic uncertainty and mitigates the issue with the dilution of probability affecting the pure Probability of Collision.

It was demonstrated that by using Dempster-Shafer Theory of Evidence it is possible to construct classification criteria that improve the accuracy of the classification and correctly classify events for which additional information is required. It was also demonstrated how extending the classification with the

time to collision in combination with Evidence Theory can further improve the classification and provide decision support to operators.

This new evidence-based classification system was then made automatic by training a Machine Learning system. It was demonstrated that both Random Forest and ANN algorithms can provide a reliable classification of conjunction events with very high probability. In particular, it was demonstrated that both RF and ANN can provide a correct prediction starting from observations and associated credibility.

The classification and ML systems presented in this paper were tested on synthetic data and assuming two separate sources of information providing a single confidence interval each. Furthermore, it was assumed that the Probability of Collision could be computed with a simple two dimensional integral on the impact plane of the target and that the uncertainty was Gaussian in nature and unimodal.

These are fairly restrictive assumptions that apply in many cases but not all of the cases. Future work is dedicated to remove these assumptions and have a more general calculation of the probability of collision. We are also working on an extended Intelligent Event Classification System that includes the automatic allocation of manoeuvres. We envisage also an extension of the classification criteria to include more thresholds as it is currently done in practice.

Last but not least the use of real data is essential to properly test the classification system. In particular, Evidence Theory allows for a subjective quantification of the credibility of the source. On the other hand with the same theory one can accommodate also a more precise quantification of the credibility that comes from the characterisation of sensors and data sources. It is, therefore, important to transition from the conceptual work presented in this paper to the actual quantification of the epistemic uncertainty in the observations.

Acknowledgment

The authors would like to express their appreciation to the Royal Aeronautical Society for having funded their attendance to the 2nd International Conference on Space Situational Awareness, ICSSA-2020, and the University of Strathclyde for financing the research.

References

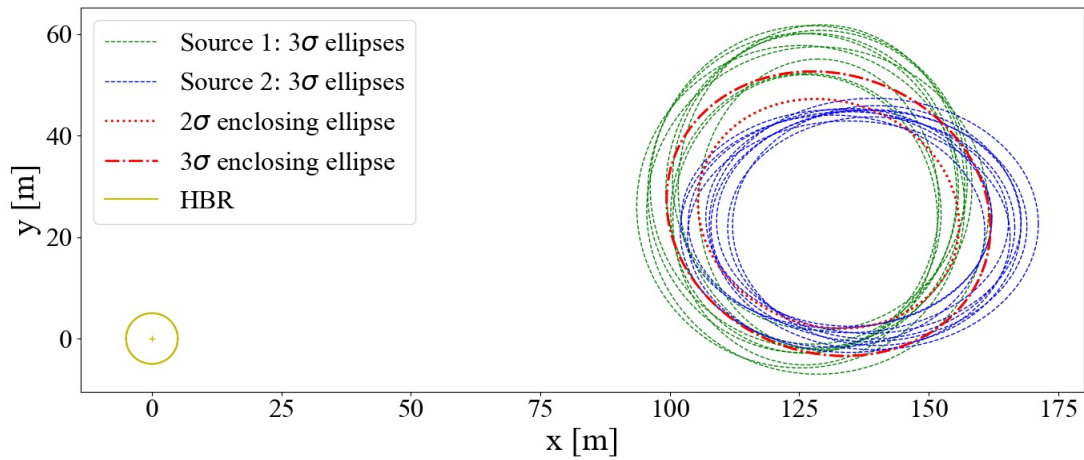
- [1] G. Peterson, M. Sorge, and W. Ailor. “Space Traffic Management in the age of New Space”. In: *Center for Space Policy and Strategy. The Aerospace Corporation* (Apr. 2018).
- [2] H. Lewis et al. “Sensitivity of the space debris environment to large constellations and small satellites”. In: *Journal of the British Interplanetary Society* (2017), pp. 105–117.
- [3] C. Contant-Jorgenson, P. Lála, and K.U. Schrogl. *Cosmic Study on Space Traffic Management*. International Academy of Astronautics, 2006. ISBN: 2-9516787-5-4.
- [4] T. Muelhaupt et al. “Space traffic management in the new space era”. In: *The Journal of Space Safety Engineering* 6 (2 June 2019), pp. 80–87.
- [5] V. Benjamin Bastida et al. “CREAM - ESA’s Proposal for Collision Risk Estimation and Automated Mitigation”. In: *1st International Orbital Debris Conference, Sugar Land, Texas, USA* (Dec. 2019).
- [6] A. Mashiku et al. “Predicting Satellite Close Approaches in the Context of Artificial Intelligence”. In: *AAS/AIAA Astrodynamics Specialist Conference, Portland, Maine, United States* (Aug. 2019).
- [7] S. Nag et al. “System Autonomy for Space Traffic Management”. In: *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*. 2018.
- [8] C. Ramirez-Atencia et al. “New Artificial Intelligence approaches for future UAV Ground Control Stations”. In: *2017 IEEE Congress on Evolutionary Computation (CEC), San Sebastian, Spain* (June 2017), pp. 2775–2782. DOI: 10.1109/CEC.2017.7969645.
- [9] R. Barhydt and K. Krishnamurthy. “Design of a Multi-mode Flight Deck Decision Support System for Airborne Conflict Management”. In: *2004 International Conference on Human Computer Interaction in Aeronautics, Toulouse, France* (Sept. 2004).
- [10] H. Peng and X. Bai. “Exploring Capability of Support Vector Machine for Improving Satellite Orbit Prediction Accuracy”. In: *Journal of Aerospace Information Systems* 15.6 (2018), pp. 366–381.

- [11] H. Peng and X. Bai. “Artificial Neural Network–Based Machine Learning Approach to Improve Orbit Prediction Accuracy”. In: *Journal of Spacecraft and Rockets* 55.5 (2018), pp. 1248–1260.
- [12] M. Vasile et al. “Artificial Intelligence in Support to Space Traffic Management”. In: *68th International Astronautical Congress, Adelaide, Australia* (Sept. 2017).
- [13] L. Sanchez, M. Vasile, and E. Minisci. “AI to support decision making in collision risk assessment”. In: *70th International Astronautical Congress, Washington DC, USA* (Sept. 2019).
- [14] R. Abay et al. “Open Source Collision Avoidance Maneuver Planning Tool”. In: *68th International Astronautical Congress, Adelaide, Australia* (Sept. 2017).
- [15] R. Patera. “General Method for Calculating Satellite Collision Probability”. In: *Journal of Guidance Control and Dynamics - J GUID CONTROL DYNAM* 24 (July 2001), pp. 716–722.
- [16] S. Alfano. “A numerical implementation of spherical object collision probability”. In: *Journal of the Astronautical Sciences* 53 (Jan. 2005), pp. 103–109.
- [17] S. Alfano. “Review of conjunction probability methods for short-term encounters”. In: *Advances in the Astronautical Sciences* 127 (Jan. 2007), pp. 719–746.
- [18] R. Serra et al. “Fast and Accurate Computation of Orbital Collision Probability for Short-Term Encounters”. In: *Journal of Guidance, Control, and Dynamics* 39 (Jan. 2016), pp. 1–13.
- [19] S. Alfano and D. Oltrogge. “Probability of Collision: Valuation, variability, visualization, and validity”. In: *Acta Astronautica* 148 (Apr. 2018), pp. 201–316.
- [20] M. Balch, R. Martin, and S. Ferson. “Satellite conjunction analysis and the false confidence theorem”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 475 (July 2019).
- [21] G. Shafer. *A mathematical theory of evidence*. 1st ed. Princeton, NJ: Princeton University Press, 1976, Apr. 1976, p. 314. ISBN: 9780691100425.
- [22] K. Sentz and S. Ferson. *Combination of Evidence in Dempster-Shafer Theory*. Sandia National Laboratories, Jan. 2002. DOI: 10.2172/800792.

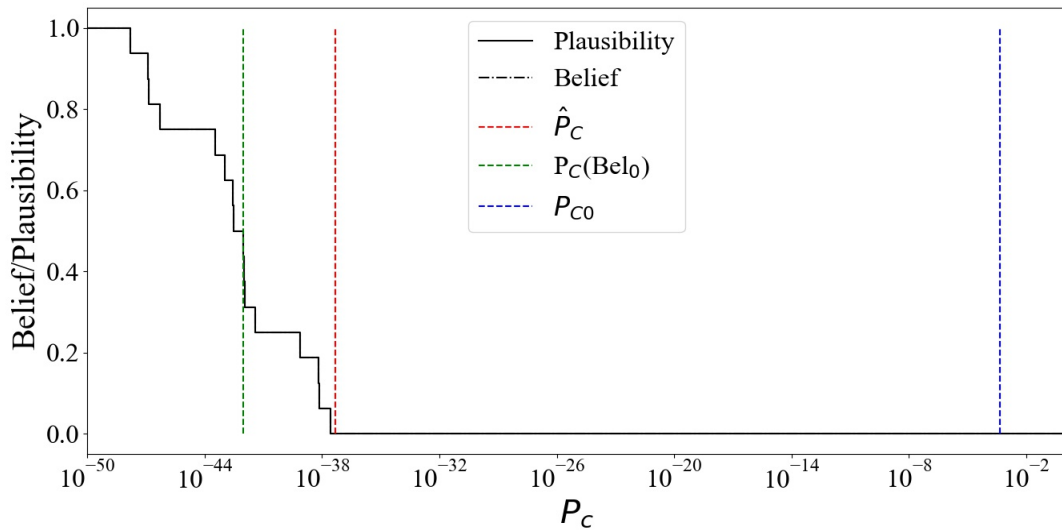
- [23] L. Newman et al. “Evolution and Implementation of the NASA Robotic Conjunction Assessment Risk Analysis Concept of Operations”. In: *Advanced Maui Optical and Space Surveillance Technologies Conference, Maui, Hawaii, USA* (Sept. 2014).
- [24] K. Merz et al. “Current Collision Avoidance service by ESA’s Space Debris Office”. In: *Proc. 7th European Conference on Space Debris, Darmstadt, Germany* (Apr. 2017).
- [25] M. Kubat. *An Introduction to Machine Learning*. 2nd ed. Springer, 2017. ISBN: 978-3-319-63912-3. DOI: 10.1007/978-3-319-63913-0.
- [26] L. Breiman. “Random Forest”. In: *Machine Learning* 45 (1 Oct. 2001), pp. 5–32. DOI: 10.1023/A:1010933404324.
- [27] C. Cortes and N. Vladimir. “Support-vector networks”. In: *Machine Learning* 20 (3 1995), pp. 273–297. DOI: 10.1007/BF00994018.
- [28] *MATLAB Deep Learning Toolbox*. <https://uk.mathworks.com/help/deeplearning/index.html>.
- [29] *Python Scikit learn library. RFF Ensemble Classifier*. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>.
- [30] *Python Scikit learn library. KNN Classifier*. <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>.
- [31] *Python Scikit learn library. SVM Classifier (SVC)*. <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC>.

Appendix A

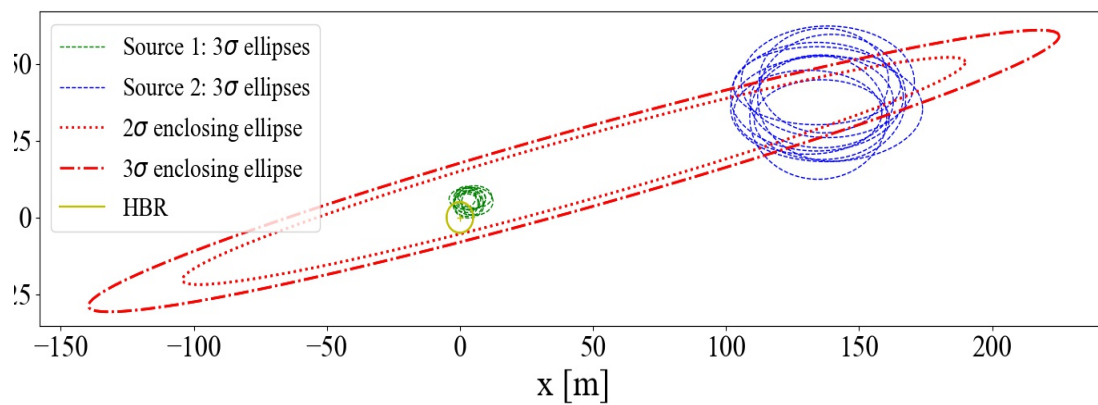
Figure 8 shows representative examples of conjunction events for each of the families of geometries described in Table 3, but the set *Geo. 1*, that is shown in Figure 4.



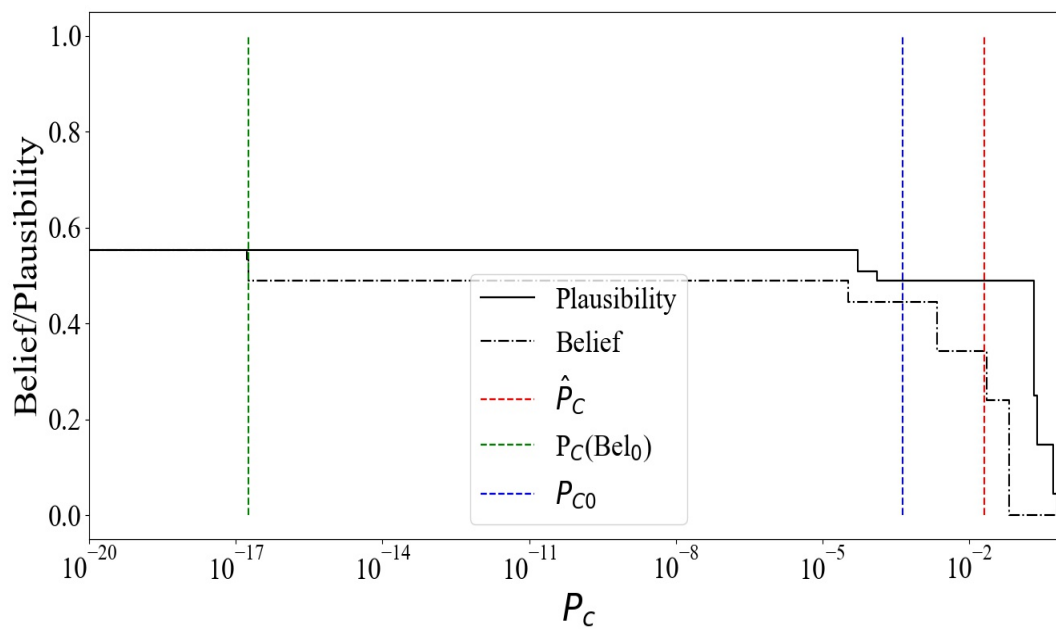
(a)



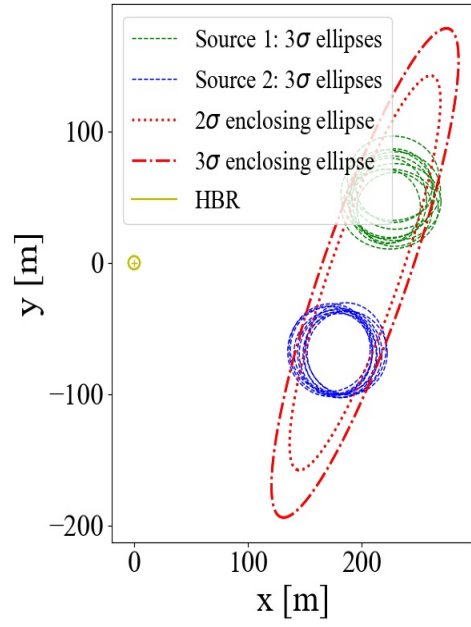
(b)



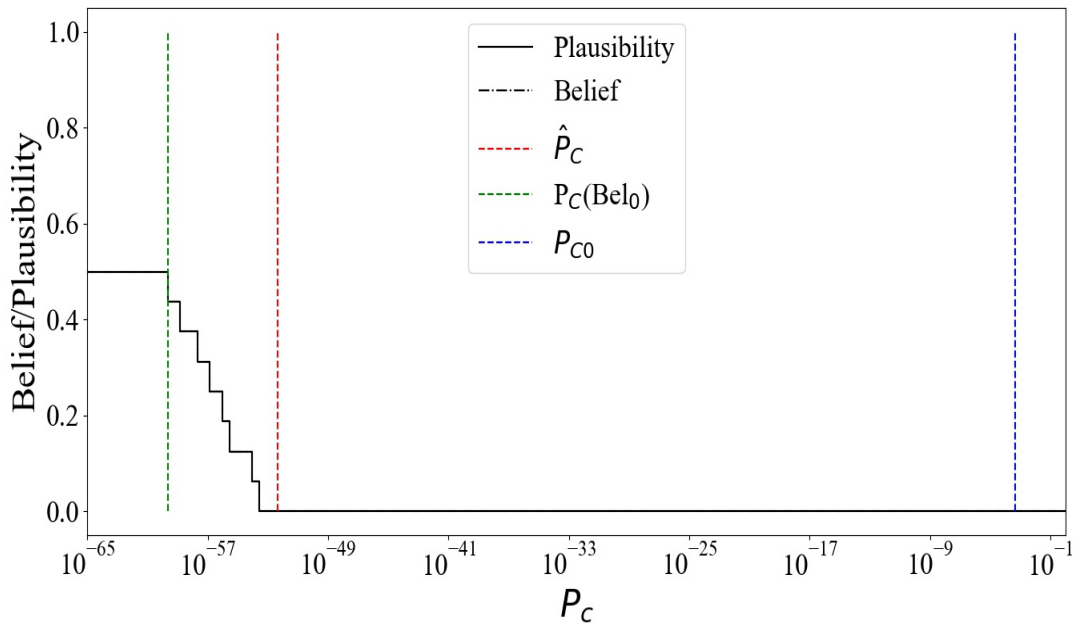
(c)



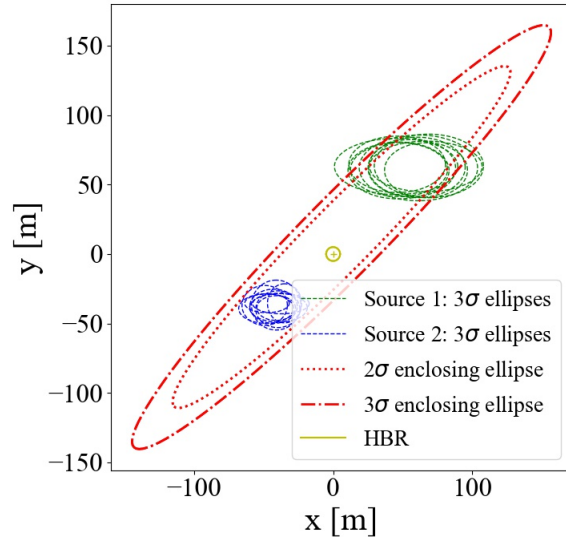
(d)



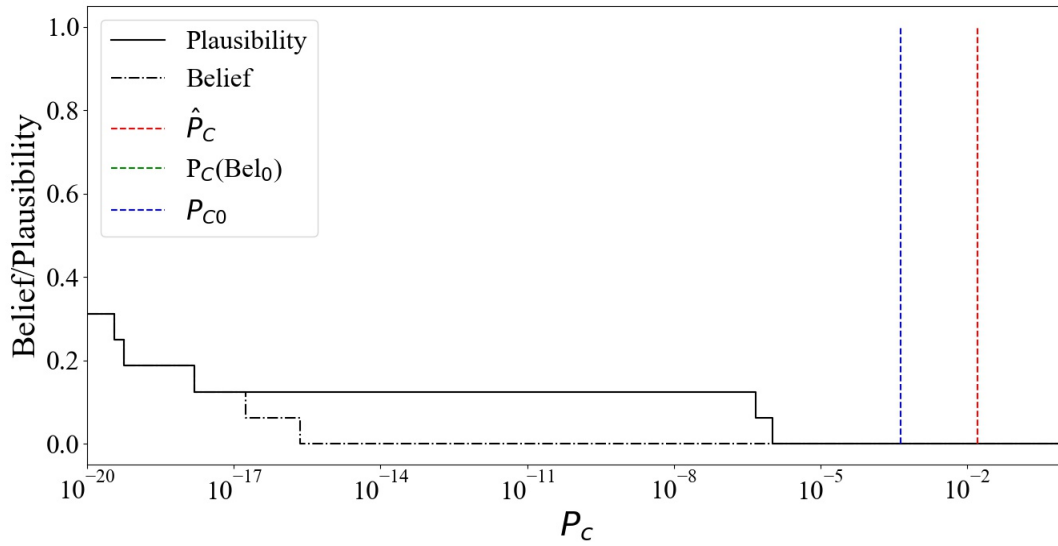
(e)



(f)



(g)



(h)

Figure 8: Representative scenarios of each geometry detailed in Table 3, but *Geo. 1*, with the associated *Bel* and *Pl* curves. (a) and (b) *Geo. 2*, (c) and (d) *Geo. 3*, (e) and (f) *Geo. 4*, (g) and (h) *Geo. 5*

Tables 13, 14 and 15 details the outcome shown in Figure 5, 6 and 7, respectively, where the Evidence-based criteria presented in Section 3 (Criterion 1, Criterion 2 and Criterion 3, respectively) are compare with a purely Probabilistic classification criterion. Results are broken down by geometry families.

Table 13: Classification distribution (in percentage) by sets using criterion based only on P_C value and using E-b Criterion 1.

Set	Number samples	\hat{P}_C based		Evidence based		
		Col. %	NoCol. %	HR %	LR %	NtKn. %
Total	3000	47.2	52.8	2.8	5.8	91.4
Geo. 1	600	96.3	3.7	0.0	0.0	100
Geo. 2	600	0.0	100	0.0	21.8	78.2
Geo. 3	600	66.7	33.3	14.0	0.0	86.0
Geo. 4	600	0.0	100	0.0	7.3	92.7
Geo. 5	600	72.8	27.2	0.0	0.0	100

Table 14: Classification distribution (in percentage) by sets using criterion based only on P_C value and using E-b Criterion 2.

Set	Number samples	\hat{P}_C based		Evidence based		
		Col. %	NoCol. %	HR %	LR %	NtKn. %
Total	3000	47.2	52.8	11.1	75.7	13.1
Geo. 1	600	96.3	3.7	39.5	10.3	50.2
Geo. 2	600	0.0	100	0.0	100	0.0
Geo. 3	600	66.7	33.3	16.2	68.3	15.5
Geo. 4	600	0.0	100	0.0	100	0.0
Geo. 5	600	72.8	27.2	0.0	100	0.0

Table 15: Classification distribution (in percentage) by sets using criterion based only on P_C value and using E-b Criterion 3.

Set	Number samples	\hat{P}_C based		Evidence based				
		Col. %	NoCol. %	C1 %	C2 %	C3 %	C4 %	C5 %
Total	9000	47.2	52.8	8.1	7.4	8.8	25.2	50.5
Geo. 1	1800	96.3	3.7	29.9	23.6	34.4	3.4	6.9
Geo. 2	1800	0.0	100	0.0	0.0	0.0	33.3	66.7
Geo. 3	1800	66.7	33.3	10.6	10.8	10.3	22.7	45.6
Geo. 4	1800	0.0	100	0.0	0.0	0.0	33.3	66.7
Geo. 5	1800	72.8	27.2	0.0	0.0	0.0	33.3	66.7

Table 16 shows the hyper-parameters chosen for the best model of each techniques on each of the four training scenarios detailed in Section 4 .

Table 16: Hyper-parameters of the best model of each techniques on each of the fours training scenarios.

ANN				
	System 1		System 2	
	DB₁₁	DB₂₁	DB₁₁	DB₂₁
Number of neurons in hidden layer	100	100	50	100
RF				
	System 1		System 2	
	DB₁₁	DB₂₁	DB₁₁	DB₂₁
Number of trees	50	100	200	400
Maximum depth of tress	'None'	'None'	'None'	'None'
Min. number samples at leaf node	1	1	1	1
Min. number samples to split node	20	2	2	2
Number features when looking for best split	'auto'	'auto'	'auto'	'auto'
KNN				
	System 1		System 2	
	DB₁₁	DB₂₁	DB₁₁	DB₂₁
Number of neighbors	1	1	20	20
Weight function	'uniform'	'uniform'	'uniform'	'uniform'
Leaf size	30	30	30	30
SVM				
	System 1		System 2	
	DB₁₁	DB₂₁	DB₁₁	DB₂₁
Decision function of shape	'ovo'	'ovo'	'ovo'	'ovo'