Dealing with Uncertainty in Hybrid Conflict: A Novel Approach and Model for Uncertainty Quantification in Intelligence Analysis

Thomas Powell **Serena Oggero** **Joris Westerveld** **Emma Schook**
Netherlands Organisation for Applied Scientific Research (TNO)
TNO Defence, Safety & Security

tom.powell@tno.nl

ABSTRACT

Uncertainty is a central concept in hybrid conflict. Much of hybrid conflict is covert, deniable in nature, and conducted in the grey zone between normal state-to-state relations and armed conflict. Moreover, signals of hybrid conflict can arise from a vast multiplicity of open and covert sources collected over an extended period of time, and possessing different levels of reliability. Together, these factors pose a major challenge for decision-making in hybrid conflict: Dealing with elevated levels of uncertainty demands innovative solutions in intelligence analysis and assessment. A promising approach in this regard is the explicit estimation of uncertainty. In this research, we drew on knowledge of statistics, intelligence analysis and AI to propose a novel approach and develop a statistical model for the quantification and systematic estimation of uncertainty. The model accounted for several important elements of uncertainty in intelligence analysis: source reliability, information credibility, probability language, and timeliness. We tested our approach using labelled and simulated data and discussed the opportunities and challenges for automating this process using AI and data science. By doing so, this research takes a step towards intelligent analytical tooling that mitigates the challenges of uncertainty in decision-making for hybrid conflict.

1.0 INTRODUCTION

Uncertainty is an essential concept in intelligence analysis. Almost every intelligence assessment made should be expressed in terms of uncertainty. This is because assessments either attempt to make inferences from incomplete or often ambiguous data, or try to predict future events (Mandel & Irwin, 2020). Several characteristics of hybrid conflict[1] increase uncertainty in intelligence analysis and assessment. Much of hybrid conflict is covert, deniable in nature, and conducted in the grey zone between normal state-to-state relations and armed conflict, leading to weak and often incomplete signals of emerging threats (Monaghan, Cullen & Wegge, 2019). Moreover, signals of hybrid conflict can arise from a vast multiplicity of open and covert sources collected over an extended period of time, and possessing different levels of reliability. These issues are amplified by increasing digitization which has led to an explosion of open source information, which is increasingly subject to mis- or disinformation (Treverton, 2021).

These characteristics of hybrid conflict present a number of challenges to intelligence analysts: increased uncertainty, ambiguity, elusiveness, activities below detection thresholds, information overload and an explosion in digital sources (Cullen, 2018). These challenges have in turn intensified calls from the intelligence community for the systematic and quantitative estimation of uncertainty. Innovative methods are needed to deal with these unprecedented levels of uncertainty and ambiguity in intelligence analysis of hybrid conflict. This in turn would support decision makers in devising policy responses to complex hybrid threats. These calls are echoed by empirical research showing the potential for reducing analytical bias, improving collaboration, fostering analytical transparency, and paving the way towards a (semi)automatized intelligence analysis process (TR-SAS-114, 2020)

Although qualitative standards exist across (inter)national intelligence organisations and methods have been proposed to improve uncertainty estimation, to the best of our knowledge no intelligence organisations

[1] For definitions see EU, 2018, NATO, 2019 and at the Dutch level, NCTV, 2019.

employ a systematic method of uncertainty estimation (Friedman & Zeckhauser, 2012). In other words, uncertainty estimates occur more or less implicitly "in the head of the analyst". Moreover, uncertainty is most often expressed in a qualitative manner (e.g., 'highly likely'), rather than quantitatively (e.g., '75% likely'). Although some reluctance exists towards the quantification of uncertainty – concerns involve an illusory sense of concreteness (with "hard numbers") to judgements that are a 'best estimate' – numerical values have the potential to mitigate language barriers, resolve semantic differences in uncertainty vocabularies and encourage analyst accountability (Dhami & Mandel, 2020).

This research seeks to address these gaps and present an approach to systematically and quantitively estimate and express uncertainty. This work is novel in that it proposes and tests a method to extract uncertainty information from intelligence reporting itself rather than heavily relying on analyst judgements (e.g., Lesot, Pichon & Delavallade, 2013; Schum & Morris, 2007). As such, our approach also aims for more objective and reproducible estimates of uncertainty. Another major contribution is that we consider the steps, opportunities and challenges involved in automating this process (using techniques from Data Science and Artificial Intelligence). After all, systematically estimating uncertainty by hand would dramatically increase the analyst's workload. By doing so this work takes a step towards faster, more systematic and objective uncertainty judgments that mitigate the challenges of decision making support in hybrid conflict.

1.1 Different types of uncertainty in intelligence analysis

We define estimating uncertainty as an evaluative process that determines the quality of the information upon which intelligence assessments are made (e.g., Friedman & Zeckhauser, 2012; Lesot & d'Allones, 2017; Mandel & Irwin, 2020;). This evaluation relies on different properties of the available information, including: *source reliability*, *information credibility* (in this paper also referred to *confirmation*), *probability language*, and *timeliness* (e.g., Lesot & d'Allones, 2017). The result of this evaluation process is commonly communicated using two terms: *probability* and *confidence* (Dhami & Mandel, 2020). These concepts and their operationalisation in this study are shown in Table 1 and explained in sections 1.2 and 1.3 below. Despite attempts at standardisation, accurately estimating and communicating uncertainty remains a serious challenge that is intensified in the context of hybrid conflict.

Table 1 – Concepts used in this research for estimating uncertainty

Term and Definition	Operationalisation (how it is quantified)	Example
Source reliability. Confidence in a source based on past performance.	Numeric rating system (from 0 to 1) based on Admiralty Code source reliability scale and quantified as shown in Appendix B.	Information from a *usually reliable* source = *0.85.* (see Appendix B)
Information credibility, or confirmation. Extent to which new reporting supports or opposes an intelligence hypothesis.	Score reflecting information that supports (labelled 1) or opposes (labelled -1) an intelligence hypothesis.	*An incoming piece of information indicates a weapon is present in location X = 1*
Probability language. Verbal qualifiers of a statement.	Words such as *possibly*, *improbable*, *highly likely*, *believed to be*, placed on a numerical scale (from 0 to 1).	Sources *believe* there was a weapon at location X *= 0.5.* (see Appendix C)
Timeliness. The age of a piece of information.	Time reduction in months (between 0 and 6 months).	*3 months ago* a weapon was observed. (see Section 2.2 for further explanation)

1.2 Estimating uncertainty

The information evaluation grading system presented in Allied intelligence doctrine is known as the Admiralty Code or NATO System (Hanson, 2015). Under the Admiralty Code, information is assessed on two dimensions: source reliability and information credibility. *Source reliability* is based on "confidence" in a given source, based on past performance. *Information credibility* (equivalent to *information confirmation* in Table 1) reflects the extent to which new information conforms to previous reporting (JDP-2, 2012). Users are instructed to consider these components independently and to rate them on two separate scales, see Appendix A. The resulting rating is expressed using the corresponding alphanumeric code (e.g., *probably true* information from a *usually reliable* source is rated B2).

There are several issues associated with the continued use of the Admiralty Code. Firstly, semantic issues in *source reliability* judgements. A "usually reliable" source has a "history of valid information most of the time" (JDP-2, 2012). However, one analyst may assign "usually reliable" to sources that provide valid information 60% of the time, whilst another might interpret the same term to mean >80% of the time (Irwin & Mandel, 2020). We attempt to resolve this type of semantic issue by quantifying the source reliability terms in the Admiralty Code. We do this based on studies that ask analysts to quantitatively estimate verbal expressions of source reliability (on a scale from 0 to 1, e.g., Samet, 1975; TR-SAS-114, 2020). This resulted in a range of estimates, and we select a score in the mid-point of the range for each individual term, as summarised in Appendix B.

Secondly, open-sources pose an challenge to the application of these standards. Given that any citizen with an internet connection can function as a sensor, along with the increased relevance of disinformation in today's information ecosystem, estimating the accuracy of the abundance and variety of open source information has never been more important. Fortunately, several websites exist that collect crowdsourced ratings of the reliability of open source outlets and articles. In this study, we used three websites[2] as a reference for determining *source reliability* scores for the range of open sources used in our experiment. We then translated these website ratings into the standardised scores shown in Appendix B.

A third challenge is that expressions of probability (*probability language*) in open-source reporting do not adhere to a standardised terminology. Intelligence reports, both single- and multi-source, use a standard dictionary of words to express the likelihood that an event has happened or will happen. The Probability Yardstick (UK DI, 2018), even links probability language with a numerical range (Probable or likely equals 55-70%). This is not the case for open-sources that use a far broader and unstandardized vocabulary. We attempted to address this issue by developing an expanded dictionary which, based on the reviewed literature (summarised in Irwin & Mandel, 2020), standardises and quantifies a fuller range of probability language, as shown in Appendix C.

A final drawback of the existing Admiralty Code standards is that reliability and credibility scores, once assigned, remain static in time. Instructions are lacking for grading pieces of information that are first confirmed (i.e., an information credibility score of "confirmed", or 1) and then later disconfirmed ("improbable", or 5). Moreover, assessments of reliability and credibility do not account for situations in which *timeliness* is an important factor. For instance a reported location of an anti-aircraft weapon may be less reliable months or weeks later since the weapon may have been used or moved elsewhere. We account for these issues in two ways. First, by using a probabilistic approach for combining uncertainty information that updates on the basis of new incoming information. Second, we apply a mathematical (decay) model that reduces the quality of information over time. These are explained in more detail in section 2.2.

[2] *www.adfontesmedia.com*, *www.mediabiasfactcheck.com* and *www.allsides.com/media-bias/media-bias-ratings*

1.3 Communicating uncertainty

Expressing uncertainty in intelligence assessments involves the communication of the estimated uncertainty in a way that can be understood by decision makers. The expression of uncertainty in intelligence analysis can be broken down into two concepts: *probability* and *confidence* (defined below). The challenges in expressing uncertainty are not as numerous as for estimating uncertainty, and mainly revolve around semantic interpretation and cross-national differences in existing standards (Dhami & Mandel, 2020). However, the difficulty in applying these standards should not be underestimated. For instance, Friedman and Zeckhauser (2012) found that analysts and policy makers confounded probability judgements with judgements of confidence. We adopt definitions of probability and confidence from US Defence Intelligence Agency standards (DIA, 2015, see also Lesot & d'Allonnes, 2017).

- *Probability* refers to the likelihood that an event or development has happened or will happen (i.e., 'how likely is it that a weapon is in location X?').
- *Confidence* refers to the perceived strength an analyst places in his or her analysis, based on gaps and assumptions in the analysis process, as well as the number, quality and diversity of sources (e.g., 'based on a limited partially contradictory reporting with substantial information gaps').

By combining the different properties of uncertainty and expressing them in these well-known terms, we hope to make the results as meaningful as possible to intelligence practitioners. Figure 1 illustrates the way in which the properties used to estimate uncertainty are combined to express uncertainty in this study.

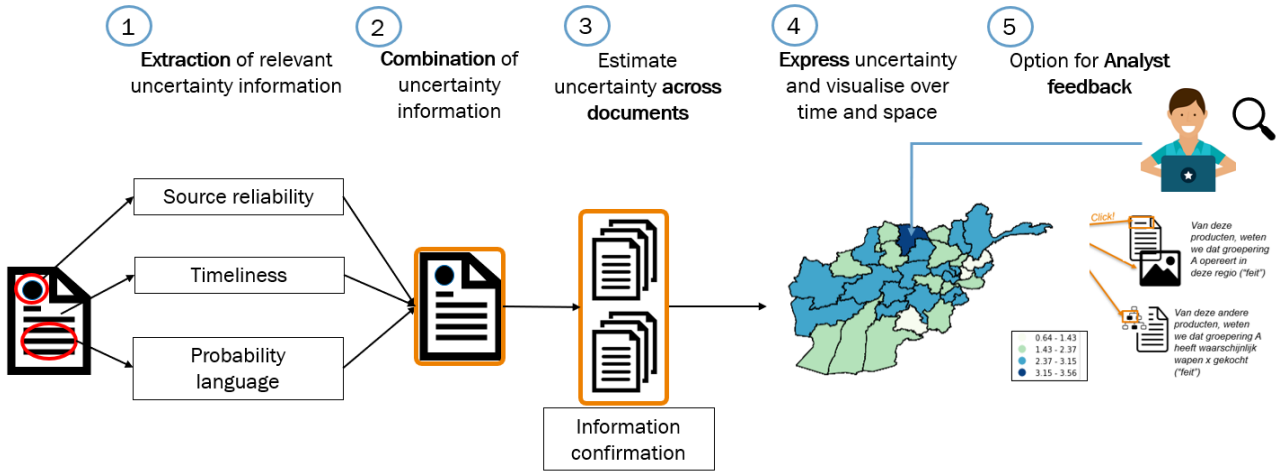

Figure 1 – Uncertainty modelling in this research as a semi-automated analytical process

2.0 METHOD

2.1 Link to existing literatures and a novel approach to uncertainty modelling

In order to express uncertainty in their assessments, intelligence analysts need to weigh up the different determinants of information quality (Section 1.2) and communicate them in terms of probability and confidence (Section 1.3). This mental transformation of uncertainty estimates into a verbal expression is an extremely challenging task, involving conditional probability assessments about numerous pieces of information derived from sources with differing capabilities over an extended time frame. This research proposes an analytical approach to support this process, building on existing literature.

Previous research has made steps towards tackling this issue using Bayesian methods (or probabilistic

approaches more generally). As new information becomes available, Bayesian networks can be updated coherently; that is, respecting the principles of probability theory, such as unitarity, additivity, and non-negativity (Karvetski, Olson, Mandel & Twardy, 2013). For instance, McNaught and Sutovsky (2012) proposed using a Bayesian network as a computer assisted framework to facilitate evidence management and the fusion of information of varying quality. While they suggest that such methods may help analysts explore uncertain situations and overcome cognitive biases, they warn that if input parameters are not known to a "reasonable degree" – something which is very plausible for some lesser known open-sources – then use of these models could generate inaccuracy (Irwin & Mandel, 2020). Therefore, even if a Bayesian approach is adopted, the rule remains: "garbage in, garbage out".

In line with this, we argue that the complete automation of uncertainty estimation is undesirable (see Section 4.2 and Lesot & d'Allones, 2017). An analyst needs to be able to investigate and modify the multiple relevant inputs, amending automatically assigned uncertainty estimates where necessary, and be able to revise his or her own edits in light of new information if needed. In line with this, Lesot et al. (2013) proposed a semi-automated method for evaluating information derived from textual documents (see also Schum & Morris, 2007). Their method involves identifying pieces of information relevant to the requirement at hand, and then attaching an independent level of confidence to each piece of information. These manually assigned ratings are then combined probabilistically to calculate an overall degree of confidence derived from all available information.

Lesot et al's (2013) approach includes some automation of information extraction and fusion from relatively reliable sources. The key differences between the present study and Lesot et al. (2013) is that we propose a method in which uncertainty information could be *automatically* extracted from text and can later be modified by an analyst (see Section 4.2), rather than uncertainty ratings being assigned manually. Moreover, this study accounts for more determinants of uncertainty (including information confirmation and timeliness) and a wider range of (open and covert) sources of varying reliabilities.

Our overall approach to the estimation and expression of uncertainty is illustrated in Figure 1. It comprises five steps.

1. Information bits relevant to the intelligence question introduced are extracted, including the determinants of uncertainty: *source reliability*, *timeliness* and *probability language.*
2. We combine these separate determinants of uncertainty for each piece of information.
3. We account for the extent to which these determinants provide confirming or conflicting information across all pieces of information. This is achieved with a probabilistic model (described in Section 2.2).
4. Uncertainty is expressed in terms of probability and confidence and is visualised over time and geographical space (see Section 3).
5. Options for collecting and incorporating analyst feedback should be provided. We discuss the need for analyst feedback in this semi-automated approach in Section 4).

Note that the main aim of this paper is to present this novel approach and the lessons learned from testing it using experiments on realistic data. The approach is built on a mathematical foundation and the experiments are conducted without yet being incorporated into a (semi-)automated process.

We tested our approach using a case study revolving around a specific intelligence question: *where are anti-aircraft weapons located in a named area of interest (NAI).* The NAI selected for this case study was Afghanistan. Although on the face of it this example does not seem directly relevant to hybrid conflict, intelligence questions regarding the presence of absence of capabilities at certain locations are certainly relevant to hybrid threats. Take the gradual increase of Chinese military presence in the South China Sea, or Russian covert military presence prior to the annexation of Crimea in 2014. Understanding the projection of influence in such regions is certainly relevant to hybrid threats. The covert and gradual nature of such

activities only heightens the need to explicitly express uncertainty in assessments of them. Another reason for choosing Afghanistan as a case study was because it provides an abundance of reporting (from both open and covert sources) for experimental testing of our approach.

2.2 Model

In this section the mathematical model underpinning our novel approach will be briefly explained. The model aims to estimate the *probability distribution* of whether anti-aircraft weaponry is present or absent based on information extracted from various sources. In particular, in the following we will refer to "*information bits*" to mean portions or snippets of reported information relevant for the intelligence assessment. How we operationalised information bits is explained in Section 2.3.

The *Beta distribution* is a typical choice for a probability distribution for our case study (Evans, Hastings & Peacock, 2000). The Beta distribution allows us to indicate the probability of whether a weapon is present or absent, and to additionally indicate a confidence range. In our case this is based on the source reliability, probability and timeliness scores, and the proportion of information bits that state anti-aircraft weapons were present and absent. The model is illustrated conceptually in Figure 2. The blue boxes show our key concepts (summarized in Table 1) and how these relate to the mathematical variables used to estimate the probability distribution and its parameters, shown by the green boxes. A brief explanation of these parameters is given below and a fuller description of the mathematical model can be read in Appendix D.

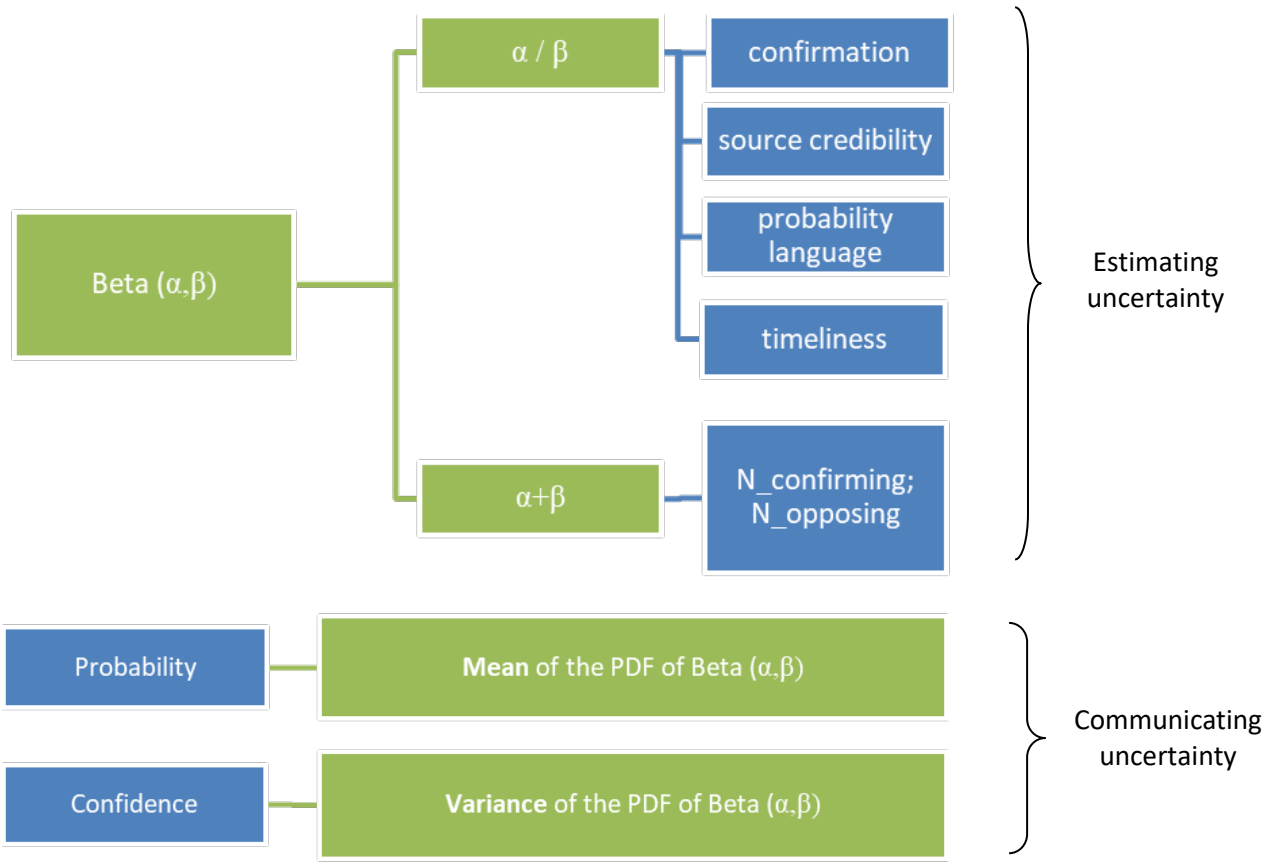

Figure 2 – How the concepts related to uncertainty (blue boxes, see Table 1) used in this study relate to the variables used to build the mathematical model (green boxes)

The **Beta distribution** has two shape parameters called α and β. In the context of our intelligence question, we specified values of α and β such that:

- the ratio α/β corresponds to an indication of how strongly all pieces of information support a given (intelligence) hypothesis[3].
- α and β increase (and therefore the variance decreases) if more information about the same statement is available.

As described in Table 1, **timeliness** can influence the accuracy of information. For our experiments, we choose a time reduction value of 6 months, meaning that sources older than 6 months are seen as considerably less 'valuable' than younger sources.

Based on the information bits (or 'observations' in probabilistic terms) the model estimates the probability density function (**PDF**) of the Beta distribution. The ***mean*** of the PDF describes the total probability that a hypothesis is true (i.e., that a weapon is present). The ***variance*** of the PDF expresses how certain the previous statement is, given the processed information.

In the following tables, Table 2 and Table 3, the numerical workings of the model are illustrated, on the base of a small subsample of the dataset. For more information concerning how the dataset was collected, see Section 2.3.

Table 2 - Example of a subsample of data used as input to the model.

Information bit ID #	Source Reliability	Probability Language	Information Confirmation	Date	Province
1	0,75 (usually reliable)	0,5 (none)	-1	20201201	Kabul
2	0,35 (not usually reliable)	0,5 (possible)	1	20210101	Kabul
3	0,85 (completely reliable)	0,3 (unlikely	1	20210201	Kabul

Table 2 shows an example of the model parameters from an illustrative subsample of the dataset. The first information bit contained no *probability language*, and is assigned a value of 0.5, which will have no strengthening or weakening influence on the total probability score. We can also observe that the *source reliability* of bit 2 is fairly low, while it is high for bits number 1 and 3. This illustrates the diversity of types of sources an intelligence analyst may deal with. The dates show that when the experiments were conducted, all information bits are recent. This means that the *time reduction factor* will not have a strong effect on the weight of the information bits.

Table 3 lists the results of the model applied to the data from Table 2. Note that each row in this table calculates the total probability and its confidence iteratively, meaning that when a new information bit is added, the total set of available information bits are taken into account and updated. In the first row, the probability is based only on the information bit 1; in the second row, the probability is based on the information reported in bit 1 and bit 2; in the last row, it is the cumulative result of all three information bits. Based on the outcome of this small example, one may draw the following conclusion: *the probability that there is an anti-aircraft weapon present in Kabul is higher than random chance (65%). However, this statement has a range of uncertainty that should not be ignored (0.26).*

[3] This indication can be interpreted as a "likelihood ratio" and can be seen as the ratio between confirming evidence (e.g., in our case, evidence of the hypothesis "anti-aircraft weapons are present in the NAI in Afghanistan") and opposing evidence (e.g., evidence of the hypothesis "there are no anti-aircraft weapons located in the NAI in Afghanistan");

Table 3 - Results from the model using the subsample of data in Table 2. Probability and confidence are iteratively calculated (i.e. third row takes all three bits into account).

Information bit ID #	Probability (mean)	Confidence (variance)
1	0.37	0.23
2	0.46	0.31
3	0.65	0.26

2.3 Experimental datasets and design

To test the mathematical model, study its limitations, applicability and opportunities for automatization, we conducted two experiments. We rely on a *"real" dataset* and a *simulated dataset*.

2.3.1 "Real" dataset

The "real" dataset was obtained by manually extracting information bits from several intelligence reports and open-source articles. This was done for two reasons: (1) to learn lessons about the feasibility of automating the collation process; and (2) to analyse the data as illustration of the utility of the mathematical model. Most of the open-source reports and articles were collected for this experiment by a partner organisation for a scenario related to the Afghan conflict. We also enriched the dataset with open source articles collected online. The open-sources include news articles (from The Guardian and CNN), technical reports (such as "Small Arms Survey 2014" and a report on *Airspace of Afghanistan* from the European Union Aviation Safety Agency).

The intelligence reports were created specifically for the purpose of this research. They were modelled following the structure of scenario inject reporting obtained from an exercise conducted in 2018, but modified to contain unclassified information and to relate to the Afghan conflict. In this way, we created reports that resemble true intelligent reports. In total we used seven reports resembling covert reports: 1 Imagery Intelligence (IMINT) report, 2 Open Source Intelligence (OSINT) reports, 2 Human Intelligence (HUMINT) reports and 2 Electronic Warfare (EW) reports. In total the dataset consists of 18 reports and articles, from which we only used information from text and not images.

All documents in the experiment dataset have individually been read through and "collated" by two of the authors of this report into a logbook. The collation in this exercise meant extraction of separate bits of information concerning a specific intelligence question: *Where are anti-aircraft weapons located in Afghanistan?* An information bit includes a report of the presence or absence of a weapon, (ideally) a specific location (district or province were the minimum needed, sometimes only the country is provided) and the date of the observation (ideally specific, but may be a window of time). Several other types of metadata were also captured in the logbook, and the full logbook can be requested from the authors. The most important information captured in the logbook is listed in Table 4. The entries that were used to conduct the modelling are highlighted in blue. The final real dataset contained 24 information bits extracted from the 18 manually collated reports. The information spanned a large timeframe from 2010 to 2020.

2.3.2 Simulated dataset

The simulated dataset is synthetically generated directly in numerical form, meaning that no information sources are used but values are pseudo-randomly generated. The goal was to test the mathematical formula with large amounts of information and over a more recent time period that allows testing of the time reduction factor.

In total 500 information bits have been pseudo-randomly generated. Roughly 30% of the 500 bits reported

that a weapon is present (1), and 70% that no weapons are present (-1). The source reliability scores and probability language were randomly generated within the range of possible values shown in Table 1. Together, this meant that 22 of the 34 provinces of Afghanistan had a low probability that weapons were present (probability lower than 0.5) and 12 provinces had a high probability that weapons were present (greater than 0.5).

For each information bit, a date of reporting is also randomly generated. All information bits spanned a time range from January 2020 to December 2020 (12 months). Each information bit was assigned a time frame of 6 months. Therefore, with this experiment we can use the time reduction factor to calculate the probability that anti-aircraft weapons are present per province per month.

Table 4 – Logbook entries, with a description and an example for each entry. The variables relating to uncertainty and/or used to compute the final probability are highlighted in blue.

Logbook entry	Description ***Example***
BIT	Is (in general) the sentence concerning the mention of weapons and sometimes also the date and/or location in brackets if these are mention in another part of the source. *"Within half an hour of the first attack (…on 20070530, in Helmand province, Afghanistan…), two Apache helicopter gunships were fired on by what the pilots thought was a missile. The helicopters were not damaged."*
Report type	Various types of reports are read. In this case study simply noted as open or covert sources. *Open-Source*
Source type	Type of the source. For covert reports: HUMINT, IMINT, SIGINT, etc. For open-source reports: News media, academic article, blog, think tank report, Industry report, etc. *News media*
Source name	Name of the source. *CNN*
Information date	Date associated with the BIT of information. *20070530*
Location province	Province-level location, which concerns the information in the bit, if present. *Helmand*
Weapon	Name of the weapon *Missile / MANPAD / anti-aircraft missile*
Weapon presence	Is a weapon reported as present (1) or absent (-1)? This is used as input to the mathematical model. *Present.* *1*
Actor	Name of the group of person in possession of the weapon. *Taliban*
Source reliability	Overall assessment of the reliability of the source, for which the Admiralty Code is used. *B (usually reliable)*
Probability language	Probability/likelihood regarding the whole information BIT, expressed through language. *None in this case ("thought was" specifically refers to the type of weapon used)*
Coder Comment	If desired, comment on report or coding decisions. Including specific comments about reliability coding decision. *CNN scores "most reliable for news" according to https://www.adfontesmedia.com/*

3.0 RESULTS

By analysing the real data, we can see how the model would perform in a scenario of limited (24) information bits. With the simulated data, we can examine the outcomes of the model with an abundance of information bits and including the time reduction factor.

3.1 Real data results

The dates of the reports ranged from 2010 to 2020, spanning a too large period of time to be relevant for the time reduction factor which we chose for the model, which applies a reduction to bits of information within a year of the present date. Due to this, the time reduction factor was not included in the model.

After running the mathematical model, the results can be visualized as in Figure 3. The colour assigned to each province represents the probability that a weapon is present in a province. The colour bar shows the probability score ranging from 0 to 1. Blue equals low probability, red equals high probability, white corresponds to equal chance of presence or absence (this is the case if there is no information). The results show that 11 of the 34 provinces have a fairly high probability that an anti-aircraft weapon is present. The other 23 provinces have no information available (0.5).

As well as the probability scores, we can also look at the derived confidence scores for each province. These are shown in the distributions on the right side of Figure 3. The mean probability score between, for instance, the provinces of Uruzgan (0.99) and Nangarhar (0.97) are similar. The confidence (variance) scores, however are somewhat different between Uruzgan (0.04) and Nangarhar (0.07). This means that the model is more certain of the assessment in Uruzgan than in Nangarhar.

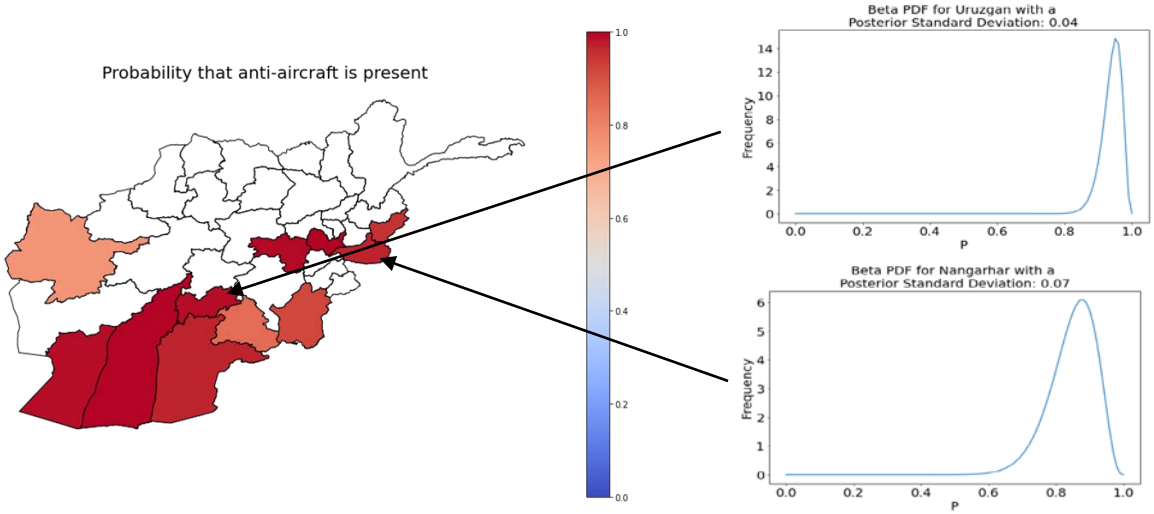

Figure 3 – Map (left) showing probability that anti-aircraft weaponry is present in a given province, for the real data experiment. Dark-red = high probability; dark-blue = low probability; white = equal chance of presence or absence. The probability distribution functions (right) show the confidence (deviation) scores (i.e., width of distribution) for Uruzgan and Nagarhar province.

3.2 Simulated data results

In Figure 4, two different images are shown with the results from the model for May 2020 (month 5) and November 2020 (month 11). With the passage of time, more bits are included in the model. This means that while in May 2020 the probabilities are generally around initial value of 0.5 (midrange of the colour scales), in November 2020 probability scores become more extreme for many regions. This shows that while older

bits become less relevant, as many more bits are added, the probability scores become higher to indicate presence (dark red) or lower to indicate absence (dark blue). Figure 4 shows that there is a very high probability anti-aircraft weaponry is present in Jawzjan province, and a low probability for Hirat.

In a similar way to the real data experiment, looking at the confidence scores from our simulated data (figure not shown due to space limitations) provides valuable information about how certain one can be of this probability assessment. To provide an example: even though the probability that anti-aircraft weapons are present in Baghlan is high, the confidence in this assessment is relatively low.

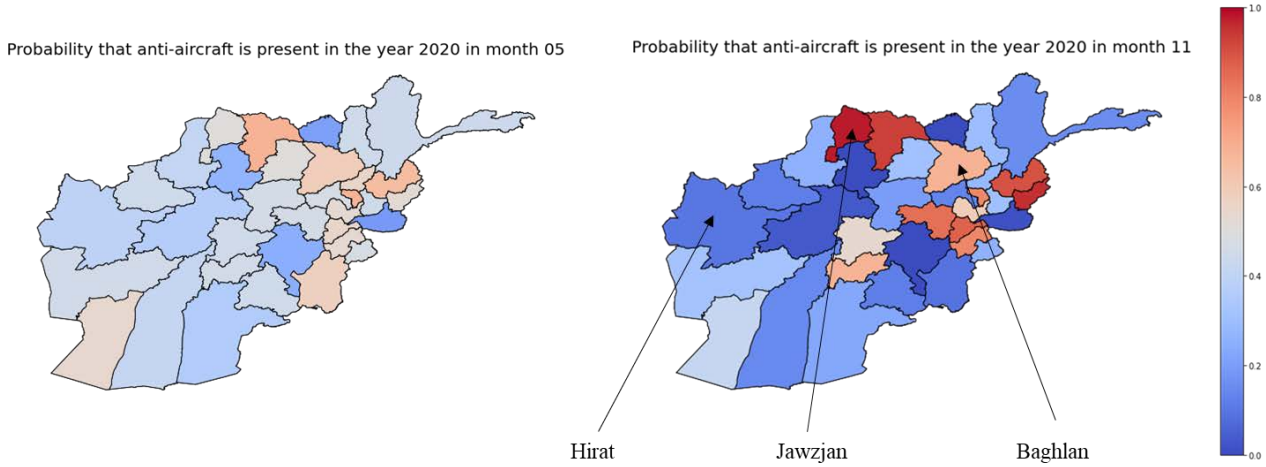

Figure 4 – Map of probability scores for the simulated data experiment for May and November 2020

4.0 DISCUSSION AND CONCLUSION

In this section, we consider some challenges for our approach, outline the opportunities for automation of it using data science and AI, and state our conclusion.

4.1 Challenges of our approach

There are inevitably a number of challenges and limitations of the proposed approach to uncertainty modelling. One is that the use of a probabilistic model runs the risk of washing out weak signals reported by a single source but that might be especially important to the intelligence question. A way to mitigate this would be to highlight such risks to the analyst and give them the opportunity to manually (and subjectively) alter the weighting of an information bit in the probabilistic model. This fits with the idea to include a feedback-loop with analysts using this model in a semi-automated approach (see next section 4.2).

Secondly, the accurate determination of source reliability remains a challenge. This is especially so for open sources that are not included in the websites we used in our study. Moreover, reliability estimates for second- or third-hand information – for instance, when an all-source report or OSINT report fuses information from other sources which are not quoted – is beyond the scope of the present research. Thirdly, regarding probability language: linking a term in our expanded probability dictionary (e.g., “believed to be”) to the standardised terms of the intelligence community (e.g., “probable”) remains in large part subjective. Future research should collect quantitative probability estimates of this expanded dictionary so that this linking can be done on the basis of empirical evidence.

Finally, the intelligence question addressed in this paper is a descriptive one concerning the location of

weapons in Afghanistan (Pherson & Heuer, 2020). For other types of intelligence questions specific choices regarding our model parameters should be made. For instance, intelligence questions about hybrid conflict are often explanatory or predictive in nature. These types of questions will inevitably involve a greater degree of uncertainty since they are more complex and involve forecasts about future events.

4.2 Opportunities for automation

In our proposed approach there are a number of interesting opportunities for automation that could help in objectifying and speeding up the process. Firstly, automatic extraction and quantification of probability language in covert sources can quite easily be implemented through algorithms from the natural language processing (NLP) field, supported by a standard scheme or dictionary. For non-covert / open-sources, where no standard scheme or dictionary is used to convey this information, semantic similarity and word/concept embeddings involving vocabulary from the covert scheme can be added to NLP modules.

Secondly, the time reduction factor can be automatically estimated through the model, once time-related data is extracted from the report – which can also be automatized through word-search or NLP techniques. Thirdly, concerning the reliability of source, the several (openly available) databases concerning trustworthiness of media sources used in this research could be automatically scraped and matched to the source extracted from the report's metadata, again through word-search or NLP techniques.

Importantly, we also envisage a number of feedback loops between the automated process and the analyst (Lesot et al., 2013). For instance, the analyst could support in determining source reliability where this cannot be automatically matched in available databases. Another example is that an analyst should determine the appropriate 'aging' of information to ensure the time reduction factor is appropriate to the context. To facilitate these steps towards automation, future research should seek end-user feedback from analysts (e.g., task analysis), and also seek a qualitative analyst-determined uncertainty assessment up-front which can be used as a 'ground-truth' comparison for the automated assessment of uncertainty.

4.3 Conclusion

In this research we drew on knowledge of statistical modelling and intelligence analysis to propose a novel approach and develop a statistical model for the quantification and systematic estimation of uncertainty. This type of intelligence support is important for understanding hybrid conflict due to the many challenges it poses to the intelligence analyst – increased uncertainty, ambiguity, elusiveness, below detection thresholds, information overload and digitalization (Cullen, 2018). Although more research is needed to further operationalise and automate our approach and model, this research takes a step towards intelligent analytical tooling that mitigates the above challenges and makes the estimation of uncertainty systematic and quantifiable. This, in turn, will provide much needed decision-support to policy makers responsible for formulating responses to hybrid conflict.

REFERENCES

[1] Cullen, P. (2018). Hybrid threats as a new 'wicked problem' for early warning. Strategic Analysis, Number 8. Helsinki, Finland: The European Centre of Excellence for Countering Hybrid Threats.

[2] Dhami, M. K. & Mandel, D. R. (2020). "UK and US policies for communicating probability in intelligence analysis: A review," in *Assessment and Communication of Uncertainty in Intelligence to Support Decision-Making*, TR-SAS-114, NATO Science & Technology Organisation, pp. 17.1-17.9.

[3] European Union (2018). A Europe that protects: Countering hybrid threats. Brussels: EEAS. Available

via eeas.europa.eu/sites/eeas/files/hybrid_threats_en_final.pdf. Accessed on 2nd of September, 2021.

[4] Evans, M., Hastings, N., & Peacock, B. (2000). "Beta Distribution." Ch. 5 in *Statistical Distributions*. New York: Wiley.

[5] Friedman, J. A. & Zeckhauser, R. (2012). Assessing uncertainty in intelligence, *Intelligence and National Security*, vol. 27, no. 6, pp. 824-847.

[6] Hanson, J. M. (2015). The Admiralty Code: A cognitive tool for self-directed learning. *International Journal of Learning, Teaching and Educational Research, 14*(1): 97-115.

[7] Irwin, D. & Mandel, D. R. (2020) "Standards for evaluating source reliability and information credibility," in *Assessment and Communication of Uncertainty in Intelligence to Support Decision-Making*, TR-SAS-114, NATO Science & Technology Organisation, pp. 7.1-7.13.

[8] Joint Doctrine Publication - 2 (2012). *Intelligence.* Netherlands Ministry of Defence, The Hague.

[9] Karvetski, C.W., Olson, K. C., Mandel, D.R., & Twardy, C. R. (2013). Probabilistic coherence weighting for optimizing expert forecasts. *Decision Analysis*, vol. 10, no. 4, pp. 305-326, 2013.

[10] Lesot, M. J. & d'Allonnes, A. R. (2017). "Information quality and uncertainty," in *Uncertainty Modelling*, ffhal-01615344, Springer, pp. 135-146.

[11] Lesot, M. J., Pichon, F. & Delavallade, T. (2013) "Quantitative information evaluation: Modeling and experimental evaluation.," in *Information Evaluation*, Wiley, 2013, pp. 187-230.

[12] Mandel, D. R. & Irwin, D. (2020). Uncertainty, Intelligence, and National Security Decisionmaking. *International Journal of Intelligence and CounterIntelligence*, https://doi.org/10.1080/08850607.2020.1809056, pp. 1-25, 2020.

[13] McNaught, K. & Sutovsky, P. (2019). "Representing variable source credibility in intelligence analysis with Bayesian networks," in Proceedings of the Australian Security and Intelligence Conference, Perth, Western Australia.

[14] Monaghan, S., Cullen, P., Wegge, N. (2019). MCDC Countering hybrid warfare project: Countering hybrid warfare. MCDC.

[15] NATO (2019). *NATO's response to hybrid threats.* Available via https://www.nato.int/cps/en/natohq/topics_156338.htm. Accessed on August 11th, 2021.

[16] NCTV (2019). *Chimaera: An analysis of the 'hybrid threat' phenomenon*. The Hague: National Coordinator for Security and Terrorism. Available via https://english.nctv.nl/documents/publications/2019/09/05/analysis-of-the-%E2%80%98hybrid-threat%E2%80%99-phenomenon. Accessed on 12th of August, 2021.

[17] Office of the Director of National Intelligence (2015). Intelligence Community Directive ICD 203: Analytic Standards.

[18] Pherson, R.H., Heuer Jr., R. J. (2020). Structured analytic techniques for intelligence analysis. Los Angeles, CA, USA: CQ Press.

[19] Samet, M. (1975). Quantitative interpretation of two qualitative scales used to rate military intelligence.

Human Factors, vol. 17, no. 2, pp. 192-202.

[20] Schum, D. A. & Morris, J. R. (2007). Assessing the competence and credibility of human sources of intelligence evidence: contributions from law and probability. *Law, Probability & Risk*, vol. 6, no. 1-4, pp. 247-274.

[21] Treverton, G.F. (2021). An American view: Hybrid threats and intelligence. In: Weissmann, M., Nilsson, N., Palmertz, B. & Thunholm, P. (Eds.), Hybrid Warfare: Security and Assymetric Conflict in International Relations (pp. 36-45). London: I.B. Tauris.

[22] TR-SAS-114 (2020). Assessment and Communication of Uncertainty in Intelligence to Support Decision-Making. *NATO Science & Technology Organisation*.

[23] UK Defence Intelligence (2018). *Professional Head of Intelligence Assessment Probability Yardstick*. London, UK.

Appendices

APPENDIX A: JDP-2 (UK, 2011) SOURCE RELIABILITY AND INFORMATION CREDIBILITY SCALES

Reliability of the Collection Capability		Credibility of the Information	
A	Completely reliable	**1**	Completely credible
B	Usually reliable	**2**	Probably true
C	Fairly reliable	**3**	Possibly true
D	Not usually reliable	**4**	Doubtful
E	Unreliable	**5**	Improbable
F	Reliability cannot be judged	**6**	Truth cannot be judged

APPENDIX B: QUANTIFICATION OF SOURCE RELIABILITY SCORES FROM THE ADMIRALTY CODE

Scores range from 0 to 1. The ranges given are derived from the reviewed literature (e.g., Mandel, 2018; Samet, 1975). *Note: 0.5 is chosen since is the mid-point of the probability range in our model, meaning that information from a source neither makes a conclusion more (> 0.5) or less likely (< 0.5).

Source reliability code	**Description**	**Range from literature**	**Reliability Score in this paper**
A - Completely reliable	No doubt of authenticity, trustworthiness, or competency; has a history of complete reliability	0.65-0.99	0.85
B - Usually reliable	Minor doubt about authenticity, trustworthiness, or competency; has a history of valid information most of the time	0.55-0.90	0.75
C - Fairly reliable	Doubt of authenticity, trustworthiness, or competency but has provided valid information in the past	0.40-0.80	0.55
D - Not usually reliable	Significant doubt about authenticity, trustworthiness, or competency but has provided valid information in the past	0.15-0.70	0.35
E - Unreliable	Lacking in authenticity, trustworthiness, and competency; history of invalid information	0.05-0.53	0.15
F - Cannot be judged	No basis exists for evaluating the reliability of the source	No rating given.	0.5*

APPENDIX C: EXPANDED AND QUANTIFIED PROBABILITY LANGUAGE DICTIONARY USED IN THIS STUDY

The selected scores range from 0 to 1 and are derived from the reviewed standards in the intelligence community (summarised in Mandel & Irwin, 2020). *Note: 0.5 was assigned to no language since is the mid-point of the probability range in our model, meaning that the information neither makes a conclusion more (> 0.5) or less likely (< 0.5).

Standard terminology	**Expanded dictionary**	**Probability Range from literature**	**Probability Score in this paper**
No doubt	Undoubtedly; Beyond doubt; Beyond question; Undeniable	Not included	0.99
Almost certain	Surely	Not included	0.95
Highly likely	Highly probable; Very likely; Have little doubt	>0.9	0.9
Likely	Probable; Believed to be; Foresee; Expect	0.6-0.9	0.7
Even chance	Possible; May/Might; Potentially; Could/Can; Perhaps; Unsure; Thought was; Not known/Unable to assess; [No language]*	0.4-0.6	0.5
Unlikely	Improbable; Doubtful; Not likely; Do not expect	0.1-0.4	0.3
Highly unlikely	(Very) doubtful; Very unlikely; Highly improbable; Outside chance	<0.1	0.1
No chance	Will not	Not included	0.01

APPENDIX D: DETAILED EXPLANATION OF THE MATHEMATICAL MODEL

Typically, whether anti-aircraft weaponry is present or absent can be expressed as a probability p of a binomial distribution. If p is high, presence is likely, if p is low, presence is unlikely and if p is about 0.5, no definite answer can be given. However, it is also useful to include a confidence range based on the amount of observations and the reliability of this information. For example, there is a difference between absence of information (which might lead to an estimate for p of 0.5) and the presence of contradictory information (which also lead to an estimate of 0.5). In the latter situation, we are more convinced that an estimate close to 0 or 1 is unlikely, therefore the confidence range is smaller. Therefore, the model not only aims to estimate the probability p of whether anti-aircraft weaponry is present or absent based on information extracted from various (data and information) sources, but also its distribution. In particular, in the following we will refer to "*information bits*" to mean portions or snippets of extracted information (in the present research this is always text) relevant for the intelligence assessment.

A typical choice for a probability distribution in a scenario like this, in which the distribution of the

parameter p of the binomial distribution (representing the probability of a hypothesis being true) is determined, is the *Beta* distribution [42] [43]. This allows us to indicate the probability of whether a weapon is present or absent, and to additionally indicate a confidence range, which is based on previous observations (previously collected information bits). Simply said, it takes into account the history of reporting by taking the proportions of information bits that claimed that anti-aircraft weapons were present and absent. Note that the effectiveness and reliability of this model is dependent on the information (quality) that is used.

The Beta Distribution

The Beta distribution is a mathematical function largely adopted in statistics:

$$Beta(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

in which $\Gamma(x)$ is the gamma function.

To work with "events", or "observations" of a given process under study, in probabilistic terms, it is useful to work with the probability density function (PDF) of the Beta distribution. This is given by:

$$f(p|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

where p is the probability for that event to occur, hence a proportion between 0 and 1.

The *Beta* distribution has two shape parameters called α and β. Typically, these are associated with the number of successes and the number of failures, when the process under study is a hypothetical experiment whose possible events can either be successes or failures. How this relates to an intelligence analysis will be discussed in the next section.

The figure below shows an example of several Beta distributions with the same ratio of α/β. The *mean* of the PDF is given by $\frac{\alpha}{\alpha+\beta}$, therefore, as long as the ratio α/β is constant, also the mean (indicated by the purple line or the peak of the distribution) remains constant (in this figure equal to 0.5). However, the distributions shown have a different *variance*, seen as the "narrowness" of the peak around the mean. As both α and β increase, the shape of the distribution changes and the peak around the mean of the distribution gets narrower (from red, to blue, to green).

We can understand this as follows: an increase of α and β means an increase in available information, and it reflects in a lower variance; the higher α and β are, the lower the variance.

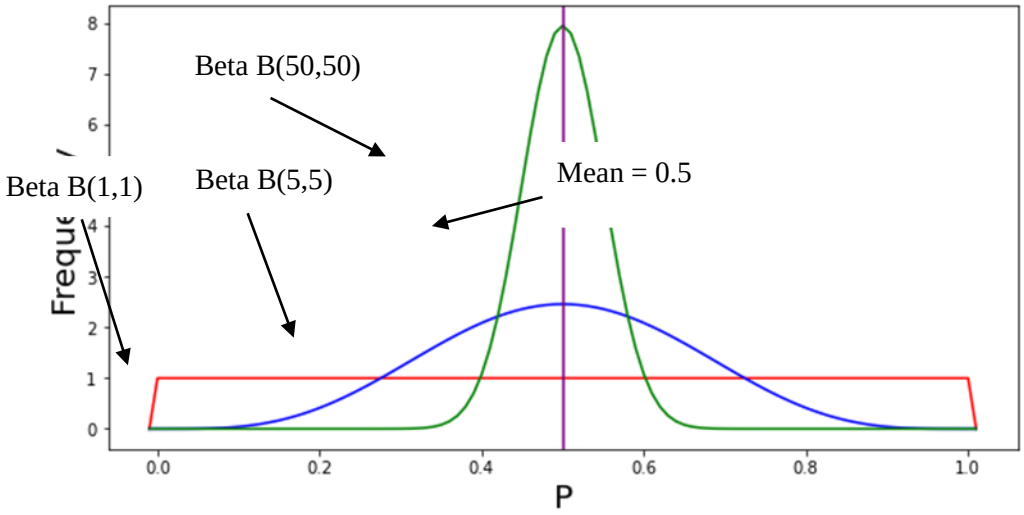

Figure: Examples of the probability density function (PDF, here frequency f) of the Beta distribution, with different α and β values.

The Beta Distribution applied to intelligence analysis

In the intelligence process, bits of information cannot be simply counted as success or failure, because their information values differ as well as their reliability and timeliness. Therefore, the suggestion is to choose appropriate values of α and β such that:

- the ratio α/β corresponds to an indication of how strongly all pieces of information support a given (intelligence) hypothesis[4].
- α and β increase (and therefore the variance decreases) if more information about the same statement is available.

Typically the likelihood ratio for one bit of information is expressed in a so called Bayes Factor (BF). In our case this value starts with a value of 100 (a score of 100 is in general seen as strong evidence [40]). In our case the likelihood ratio gets weakened by other variables (the source reliability, the probability language and a time reduction factor)[5]. So the Bayes Factor is expressed as:

$$\alpha/\beta = BF = 100^{(confirmation\ *\ (source\ reliability * probability\ language * time\ reduction\ factor))}$$

Where $source\ reliability$, $probability\ language$ and (information) $confirmation$ are assigned numerical values as explained in Table . The $time\ reduction\ factor$ reflects how old the source is and how much this decay function should influence the Bayes Factor; this will be explained in Section **Error! Reference source not found.**.

Note that all values smaller than 1 decrease the value of BF, hence decrease the likelihood of a certain evidence.

[4] This can be interpreted as a "likelihood ratio" and can be seen as the ratio between confirming evidence (e.g., in our case, evidence of the hypothesis "anti-aircraft weapons are present in the NAI in Afghanistan") and opposing evidence (e.g., evidence of the hypothesis "there are no anti-aircraft weapons located in the NAI in Afghanistan");

[5] Hence BF covers a range from 0 to 100

The Bayes Factor is determined for each bit of information. If more pieces of information are available, their Bayes Factors can be multiplied. Hence:

$$(\alpha/\beta)_{total} = \prod_i BF_i$$

Now that the ratio α/β has been determined, the values of α and β themselves have to be determined. A way to do this is by looking at the total number of confirming pieces of information relative to the number of opposing pieces of information: if the number of confirming pieces of information outweighs the number of opposing pieces of information, the conclusion is more likely to be certain. The same holds the other way around. In our example it is the ratio between the number of times information bits said that anti-aircraft weaponry is present versus the number of times information bits said that anti-aircraft weaponry is absent. This can be written as:

$$\beta = \max\left(\frac{N_{confirming}+1}{N_{opposing}+1}, \frac{N_{oppposing}+1}{N_{confirming}+1}\right)$$

With $N_{confirming}$ the number of confirming pieces of information (anti-aircraft weapon is present) and $N_{opposing}$ the number of opposing pieces of information (anti-aircraft weapon absent). We add 1 to each value so there is no possibility to divide by zero. Furthermore, the relative added value of an extra piece of information decreases when more information is already available. Both ratios are used as the situation in which more opposing information is available, is comparable to the situation in which more confirming information is available. The maximum value of the two ratios is selected to make sure that the strongest set of evidence determines the value of β.

Given the value of the ratio α/β and the value of β, the value of α can be calculated:

$$\alpha = \beta \prod_i BF_i$$

In this way, the calculated α and β still represent confirming and opposing information respectively.

Time reduction factor

One of the weakening variables of the likelihood ratio is time, because time can influence the accuracy of information. Therefore, a time reduction factor has been implemented in the formula; this factor expresses how "valuable" newer sources are with respect to older ones. The time reduction factor formula is expressed with:

$$time\ reduction\ factor\ = Exp^{\frac{-(t_{now} * ln0.1\ )}{t_{decay}}}$$

In which t_{now} is the time of the observation (in our case measured in months) and is a variable that describes how fast the value of a bit of information decays in time. In this case the older the source is (in months) the less valuable it becomes. for the experiment we conduct, we choose a standard value of t_{decay} of 6 months; this places a cut-off at 6 months, meaning that sources older than 6 months are seen as considerably less valuable than younger sources. This cut-off point should be chosen by analysts on the basis of their experience and the context of operation. Lastly the value 0.1 also weighs down the final *p* value, so that at

the chosen t_{decay} the value only weights for 0.1 (10%, e.g. the red line in figure below).

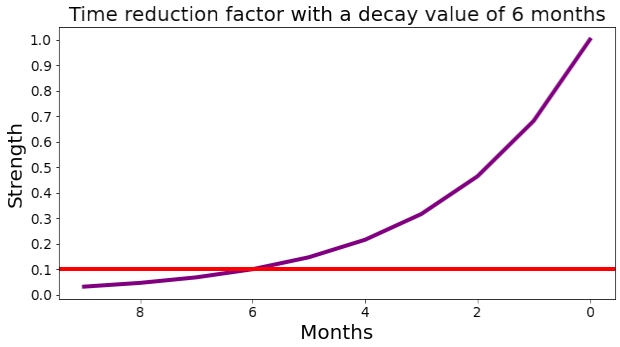

Figure: illustration of the chosen function for the time reduction factor (on the vertical axis) versus "age" of the observation (on the horizontal axis, or months between current moment t=0 and moment of observation). This shows how 'fast' the value of information decreases with aging of information.

Total calculation of probability and confidence

Once the values of α and β themselves have been determined as illustrated in Section **Error! Reference source not found.**, we can calculate the final estimate for the probability score. This describes the total probability p_{tot}, given the processed information, that a hypothesis is true, and is now given by the *mean* of the PDF:

$$p_{tot} = mean(p|\alpha,\beta) = \frac{\alpha}{\alpha+\beta}$$

We also calculate[6] the final *variance* score, which explains how certain the previous statement is, given the processed information. The variance is calculated by using:

$$var(p|\alpha,\beta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

[6] The mathematical model and experiments are available on Gitlab if access is given. The experiment.ipynb file can be opened using Python in Jupyter notebook or Jupyter lab.

In order to calculate the Beta distribution for our experiments, we have used the a python package called Scipy.