From tactical picture to situation assessment evaluation: A CUAS illustration

Anne-Laure Jousselme and Benjamin Pannetier
CS Research Lab, CS Group - France
Email:{anne-laure.jousselme,benjamin.pannetier}@csgroup.eu

***Abstract*—Evaluating information fusion algorithms and systems is instrumental to the proper prediction of error, to the rational improvement of solutions and *in fine* to the acceptance of solutions by end-users. Performance criteria define general semantics for a desirable behavior of the systems, while corresponding metrics implement that semantics for computable quality. Evaluation of the first levels of the JDL model of data fusion (detection and individual object assessment) is classically measured with objective and more or less standardized metrics. Evaluation of higher levels of processing such as situation assessment is less formalized as the evaluation criteria seat somewhere between the tactical picture quality and the decision-maker situation awareness. In this paper, we propose a formalization of the situation assessment problem, which bridges level 1 and 2 of the JDL model. Secondly, we define a global measure of quality encompassing the criteria of completeness, accuracy, clarity, which can be applied to both level 1 and level 2. We illustrate the metrics on a Counter-Unmanned-Aerial System (CUAS) scenario, comparing two uncertainty handling methods for a threat assessment solution through a Dynamic Bayesian Network. We finally conclude and sketch ideas for future steps of this research.**

***Index Terms*—Uncertainty evaluation; tactical picture; quality metrics; situation assessment; threat assessment; CUAS.**

I. Introduction

The evaluation of information fusion algorithms and systems enables the proper prediction of error, the rational improvement of solutions and *in fine* the acceptance of solutions by end-users. While performance criteria define general semantics for a desirable behavior of the systems, corresponding metrics implement that semantics for a computable quality [1].

The first levels (0 and 1) of the Joint Directors of Laboratories (JDL) model of data fusion [2] aim at building the Tactical Picture (TP), focusing on individual objects assessment, with classical quality criteria such as completeness, accuracy, clarity, continuity, timeliness and commonality [3]. Level 2 processing aims at a rather global assessment, focusing on either relationships between individual objects and their environment (context) or relationships between objects themselves. The situation assessment (SA) process is aiming at providing the decision-maker with the necessary understanding of the situation for an informed decision. Compared to levels 0 and 1, the criteria for situation assessment quality are less documented (*e.g.*, [4], [5]) as they lay somewhere between the tactical picture quality and the decision-maker situation awareness (SAW) quality, for which methods such as the Situation Awareness Global Assessment Technique (SAGAT) [6] are used. The Activities of Interest Score (AoI) proposed by [7], measures how well the system identifies relevant activities. It captures in an objective way quality dimensions of confidence, purity, cost Utility and timeliness. Confidence is a measure of how well the system detects the true activities and is composed of three factors of recall, precision and fragmentation, referring to some ground truth information, and close to the objective standard measures used for classification at level 1. Purity characterizes the quality of the activities detection and characterization, in a similar way than clarity for a TP.

The objective evaluation of situation assessment quality is thus a challenge as both aspects of the TP and SAW should be considered. The goal of this paper is not to solve that problem but rather to set up some bases bridging the levels in terms of evaluation. Such an harmonization through the fusion levels is enabled by, on the one hand, a list of generic criteria of completeness, accuracy, clarity, continuity, timeliness and commonality, classically accepted for the TP quality and extended to cover SA quality.

In Section II, we propose a formalization which bridges lower levels of processing to level 2, and introduce the "semantic elevation" step as a means to bridge individual object assessment and situation assessment. Section III defines metrics for the TP quality as well as for the SA quality, capturing accuracy, clarity and completeness criteria. For illustration purposes, we implement in Section IV a solution for a Counter-Unmanned Aircraft System (CUAS) scenario by means of a Dynamic Bayesian Network (DBN). We display some results of evaluation comparing two uncertainty representations (using either hard or soft evidence) and their impact on both the TP quality and the global SA quality. Conclusions are drawn in Section V, and future work is discussed.

II. Formalization

We consider a scene delimited by a volume of time and space, within which a set of M relevant objects evolve. Each object is characterized by a set of N relevant features. Several agents a (or sources) estimate the objects features at different instants in time k.

A. Tactical Picture quality

The quality of the TP can be characterized along six criteria [3], in accordance with standard documents (*e.g.*, [8]). For a given situation $\mathbf{s}$ setting in a scene defined by a volume of space and time:

1) COMPLETENESS is the degree to which objects of $\mathbf{s}$ are detected, tracked, recognized, identified;
2) ACCURACY is the degree to which the estimated attributes values for each object of $\mathbf{s}$ agree with the corresponding true values;
3) CLARITY is the degree to which the TP is free from ambiguous values;
4) CONTINUITY is the degree to which the object attributes are maintained in time;
5) TIMELINESS is a degree of the latency in the TP compilation process;
6) COMMONALITY in a multi-agent context, is the degree to which the local TP held by each agent contains similar or consistent information.

Two distinct evaluation phases can be distinguished: (1) laboratory evaluation, during which ground truth is available and thus accuracy and completeness can be assessed; (2) operational evaluation, when the system is running (accuracy is difficulty assessed at that stage, unless some user feedback can be considered).

In this paper, we will focus on the three dimensions of accuracy, clarity and completeness, while the extension to other criteria is left for future work.

B. Level 1: Object assessment

We denote by $\mathbf{x}_k^{(m)}$ the feature vector representing object m at time k and by $x_{k,n}^{(m)}$ the feature n's value for that object at k[1]:

$$\mathbf{x}_k^{(m)} = \left[x_{k,1}^{(m)} \dots x_{k,N}^{(m)}\right]'$$

The problem modeling step assigns a random variable $X_n^{(m)}$ to each feature of each object, hence $N \times M$ variables. We denote by $\mathbb{X}$ that set of basic variables. The configuration space is defined as [9]:

$$\Omega = \bigtimes_{m\in\mathcal{M};n\in\mathcal{N};j} \left(X_n^{(m)} = x_n^j\right) \tag{1}$$

Let furthermore denote by $\mathcal{X}_n^{(m)}$ the universal set of $X_n^{(m)}$. Random variables are defined over either nominal (categorical) or numerical scales, and categorized accordingly. Typical numerical features are the position, speed, acceleration of the object, but also the heading or the distance to a point of interest to name a few. Nominal features are typically the type of the object (its class or category), and features describing its behavior (*e.g.*, intent, capability or threat nature). One challenge in defining global quality metrics along the different criteria is to properly account for the different nature of features.

We denote by $\hat{\mathbf{t}}$, the tactical picture gathering all estimated relevant features of detected objects, and by Acc$(\hat{\mathbf{t}})$, Clar$(\hat{\mathbf{t}})$, Comp$(\hat{\mathbf{t}})$ and Qual$(\hat{\mathbf{t}})$ measures of corresponding accuracy, clarity, completeness and global quality of $\hat{\mathbf{t}}$, respectively.

Desirable properties for Qual$(\hat{\mathbf{t}})$, encompassing the three dimensions of accuracy, clarity and completeness are the following:

(Q_1) Boundedness: $0 \leq \text{Qual}(\hat{\mathbf{t}}) \leq 1$;
(Q_2) Maximum: Qual$(\hat{\mathbf{t}}) = 1$ if and only if Comp$(\hat{\mathbf{t}}) = 1$, Acc$\left(\hat{\mathbf{t}}\right) = 1$ and Clar$\left(\hat{\mathbf{t}}\right) = 1$, meaning that all objects have been detected, correctly characterized, and true values are estimated with no ambiguity;
(Q_3) Minimum: Qual$(\hat{\mathbf{t}}) = 0$ if either Comp$(\hat{\mathbf{t}}) = 0$, Acc$\left(\hat{\mathbf{t}}\right) = 0$ or Clar$\left(\hat{\mathbf{t}}\right) = 0$, meaning that either no object has been detected or none of the detected objects have been perfectly characterized (accuracy and clarity);
(Q_4) Monotonicity 1: For fixed values of Acc$(\hat{\mathbf{t}})$ and Clar$(\hat{\mathbf{t}})$, Qual$(\hat{\mathbf{t}})$ increases with Comp$(\hat{\mathbf{t}})$;
(Q_5) Monotonicity 2: For a fixed value of Comp$(\hat{\mathbf{t}})$, Qual$(\hat{\mathbf{t}})$ increases with Acc$(\hat{\mathbf{t}})$ and Clar$(\hat{\mathbf{t}})$.

We will derive in Section III a measure of global quality satisfying these properties, aggregating measures of completeness, accuracy and clarity building up from features (distinguishing between numerical and nominal features) to individual objects and finally to picture.

C. Level 2: Situation assessment

Situation Assessment involves inferring relationships between objects, or between objects and their environment, to recognize or classify situations and provide the decision-maker with an appropriate interpretation and understanding to enable an informed decision [6], [10]. Situations are abstract or complex objects, involving several individual (physical) objects, and can be characterized by situation features defined upon objects' relationships of different nature. If we denote by $\mathcal{R}$ the set of relevant relationships between objects of $\mathcal{M}$, $\mathcal{R}$ can be represented by an adjacency tensor such that:

$$R_{m,l,r} = \begin{cases} 1 \text{ if } m \text{ and } l \text{ are linked by } r \\ 0 \text{ else} \end{cases} \tag{2}$$

with $r \in \mathcal{R}$ is a relationship and $(m,l) \in \mathcal{M}^2$ is a pair of objects. If r is an equivalence relation (*i.e.*, symmetric, transitive and reflexive), then r defines a partition of $\mathcal{M}$.

Let $\mathcal{M}$ denote the set of detected objects (at level 0) and by $\mathcal{P}(\mathcal{M})$ its power set. For the sake of completeness, $\mathcal{M}^+$ includes objects of interest m as well as the object environment e, closing the world with "everything which does not belong to the situation"[2].

We denote by $\mathbf{s}^{(i)}$ a situation which relevant features $s_n^{(i)}$ are themselves defined over subsets of features of subsets of objects (instead of individual objects). At the situation level, we may prefer the term "uncertain variable" to capture the possible epistemic nature of uncertainty, due to a lack of knowledge rather than the randomness of the phenomenon (more frequent at levels 0 and 1). Each situation feature is

[1] To lighten the notations, we will omit the subscript k and superscript (m) whenever irrelevant, keeping in mind that metrics (as features) are defined for each instant k, and each object m.

[2] The notion of environment could be named "context", encompassing all exogenous variables impacting the situation but not being part of it (*e.g.*, [11])

associated with an uncertain variable S_n with universe $\mathcal{S}_n$. Compared to individual objects' variables, situation variables are generally nominal and often binary.

Situation features bear thus on elements A of $\mathcal{P}(\mathcal{M})$ such that $|A| \geq 2$. For instance, $s^{(A)}$ could be the number of objects in A, could capture a rendez-vous between two objects, or if a group of objects exhibits some coherency in behavior (see Section IV). Equivalently to physical objects, a situation at time k can then be described by a finite set of relevant features bearing on subsets of objects:

$$\mathbf{s}_k = \left[s_{k,1}^{(A_1)} \dots s_{k,N}^{(A_N)} \right]'$$

for N subsets of objects of interest. Generally, $\hat{s}_n^{(A_j)}$ denotes the estimation of feature n for the group of objects A_j, and $s_n^{*,(A_k)}$ is its true value. Then, $\hat{\mathbf{s}}_k$ is the estimated situation at k.

D. Semantic elevation

At level 0 and 1 of the process, individual objects features $x_{k,n}^{(m)}$ are essentially directly extracted from sensors' output, and numerical. To enable some reasoning at level 2 and assess the situation, additional variables need to be added, either completely new or derived from the level 1 ones. If $\mathbb{X}$ denotes a set of variables representing the fusion problem at level 1, and $\$$ denotes a set of variables at level 2, we define a *semantic elevation* to produce level 2 variables as follows:

$$\sigma : \qquad 2^{\mathbb{X}} \rightarrow \$ \tag{3}$$

$$\sigma\left((X_1, \mathcal{X}_1) \dots, (X_j, \mathcal{X}_j)\right) \mapsto (S, \mathcal{S}) \tag{4}$$

where $(X, \mathcal{X})$ denotes an uncertain variable and its universe. The semantic elevation can create new variables with either different meanings from the ones represented in $\mathbb{X}$, or with the same meaning but defined on a different universe, for a more coarse-grained one. We identify several kinds of such mappings:

(σ_1) Relationship between objects:
$\sigma_1\left((X_1^{(1)}, \mathcal{X}_1), (X_1^{(2)}, \mathcal{X}_1)\right) \mapsto (S, \mathcal{S})$

(σ_2) Relationship between object and environment:
$\sigma_2\left((X_1^{(1)}, \mathcal{X}_1), (X^{(e)}, \mathcal{X}_e)\right) \mapsto (S, \mathcal{S})$

(σ_3) Relationship between features:
$\sigma_3\left((X_1^{(1)}, \mathcal{X}_1), (X_2^{(1)}, \mathcal{X}_2)\right) \mapsto (S, \mathcal{S})$

(σ_4) Coarsening: $\sigma_4(X, \mathcal{X}) \mapsto (S, \mathcal{S})$ with $\mathcal{S} = \rho(\mathcal{X})$, ρ being a coarsening function.

As an example of (σ_1), if $X_p^{(m)}$ denotes the position of two objects m, $m = 1, 2$, with possible values defined by $\mathcal{X}_p$ then may want to define $S_r^{(12)}$ to denote a rendez-vous (expressing a specific kind of relationship) between the two objects, with universe $\mathcal{S}_r$, such that:

$$\sigma_1\left((X_p^{(1)}, \mathcal{X}_p), (X_p^{(1)}, \mathcal{X}_p)\right) = \begin{cases} S_r^{(12)} = |X_p^{(1)} - X_p^{(2)}| < \tau \\ \mathcal{S}_r = \{0; 1\} \end{cases} \tag{5}$$

with $\tau \in \mathbb{R}^{*+}$ a threshold below which the rendez-vous is assumed. Another variable could include the speed of the two vessels, to define that rendez-vous, being a mix of mapping of kind (σ_1) and (σ_3). As an example of (σ_4): if $X_T^{(m)}$ denotes the type of object m, with possible values $\mathcal{X}_T = \{$Drone; Plane; Bird; Other Air; Truck; Car; Pedestrian; Other Ground$\}$ we define:

$$\sigma_4(X_T^{(m)}, \mathcal{X}_T) = \begin{cases} S_T^{(m)} = X_T^{(m)} \\ \mathcal{S}_T = \{\text{Drone}; \text{Not drone}\} \end{cases} \tag{6}$$

This example of mapping from $\mathcal{X}_T$ to $\mathcal{S}_T$ is a coarsening as each element of $\mathcal{X}_T$ is mapped onto a single element of $\mathcal{S}_T$. In case $\mathcal{S}_T$ is defined with fuzzy concepts, fuzzy membership functions will be used for the mapping. For instance, the mapping of $\mathcal{X}_{rcs}$ to $\mathcal{S}_S = \{\text{Small}; \text{Medium}; \text{Large}; \text{X-Large}; \text{XX-Large}\}$, *i.e.*, the Radar Cross Section (RCS) to the size would be defined by three membership functions $\mu_{sm}(x)$, $\mu_{me}(s)$,$\mu_{la}(x)$, $\mu_{xla}(x)$ and $\mu_{xxla}(x)$, with $x \in \mathcal{X}_{rcs}$ (see Section IV).

The semantic elevation step enables (1) setting some relevant granularity for the proposed solution as guided by operational needs (σ_4), (2) reasoning over relationships between objects (σ_1) and their environment (σ_2), and (3) defining more meaningful variables (σ_3). Following the JDL, the level 2 variables are produced by (σ_1) and (σ_2) mappings, while (σ_3) and (σ_4) mappings can still produce level 1 variables (for individual objects assessment). To identify that transition from level 1 to level 2, we introduce a "level 1.5" at which individual objects are still assessed, but with a higher semantic level as produced by mappings (σ_3) and (σ_4). We will illustrate that semantic transition in Section IV, on a threat assessment problem for CUAS.

E. Uncertainty representation

Let $\mathbb{X} \cup \$$ be the set of uncertain variables representing the relevant features for our problem, either at the object or situation level. Each $X \in \mathbb{X}$ is defined over a universal set $\mathcal{X}$. At an instant in time k, the result of the inference process after fusing the different pieces of information issued from sensors or sources in general, can be captured by an uncertainty function on the joint space $\mathcal{X}^{(N)} = \bigtimes_{X \in \mathbb{X}} \mathcal{X}$. Let denote by f such an uncertainty function which could be a probability function, or other non-additive functions such as plausibility or possibility function.

III. Measures through the levels

We now propose a global measure for the instantaneous Tactical Picture Quality (TPQ), considering the accuracy, clarity and completeness dimensions, that can be used for Situation Assessment Quality (SAQ).

Let denote by $\mathbf{x}$ either the tactical picture $\hat{\mathbf{t}}$ or the result of the situation assessment $\hat{\mathbf{s}}$. Let further denote by $\text{Acc}(\hat{\mathbf{x}})$, $\text{Clar}(\hat{\mathbf{x}})$, $\text{Comp}(\hat{\mathbf{x}})$ and $\text{Qual}(\hat{\mathbf{x}})$ respectively measures of corresponding accuracy, clarity, completeness and global quality.

A. Accuracy

As introduced in Section II, a measure of accuracy quantifies how "close" to truth[3] an estimation is. Hence, the estimation $\hat{x}_n$ of feature n value (of object m by agent a) is accurate if it is conform to its true value, x_n^*. We define the bounded measure for local error as:

$$\delta_{a,b}\left(\hat{x}, x^*\right) = \frac{1}{2} + \frac{1}{2}\tanh\left[\frac{b}{a}\left(|\hat{x} - x^*| - a\right)\right] \tag{7}$$

where a and b are two parameters, such that:

- $a \in \mathbb{R}^+$ controls the tolerance regarding what is accepted as "true", especially for numerical features;
- $b \in \mathbb{R}^{*+}$ controls the slope of the decreasing function from 1 to 0. Whenever $b \to +\infty$, (7) is $H_e(x - a)$, the Heavyside function centered on $x = a$.

The $\tanh$ function is a means to get a bounded measure. For an object m, we use the definition of a Minkowski distance to aggregate local errors:

$$d^{(\alpha)}(\hat{\mathbf{x}}, \mathbf{x}^*) = \left(\sum_{n=1}^{N} \delta(\hat{x}_n, x_n^*)^\alpha\right)^{\frac{1}{\alpha}} \tag{8}$$

where $\alpha \in \mathbb{R}^{*+}$, corresponding to the definition of a Minkowski distance. $\alpha = 2$ in (8) defines the Euclidean distance. The accuracy of estimation for a given object m at time k is then defined by:

$$\mathrm{Acc}_{a,b,\alpha}\left(\hat{\mathbf{x}}^{(m)}\right) = 1 - \frac{d^{(\alpha)}\left(\hat{\mathbf{x}}^{(m)}, \mathbf{x}^{(m),*}\right)}{N^{\frac{1}{\alpha}}} \tag{9}$$

Then, an accuracy measure for the TP is:

$$\mathrm{Acc}_{a,b,\alpha}(\hat{\mathbf{t}}) = \frac{1}{M}\sum_{m=1}^{M} \mathrm{Acc}_{a,b,\alpha}\left(\hat{\mathbf{x}}^{(m)}\right) \tag{10}$$

$\mathrm{Acc}(\hat{\mathbf{t}})$ is maximum (equal to 1) whenever the accuracy for each object is maximum, and minimum (equal to 0) whenever the accuracy for each object is minimum.

The role of the two parameters a and b is twofold: Firstly, they help handling heterogeneous features, bounding the "error" between 0 and 1 with more or less quantification for numerical features (b parameter) and secondly, they enable bridging level 1 and level 2 assessments by controlling the tolerance on error, or the granularity of estimation (a parameter). For binary variables, and a specific pair of (a, b) values, Eq. (7) gives:

$$\delta_{0.5,50}(\hat{x}, x^*) \approx \delta_{01} \tag{11}$$

the Kronecker index such $\delta_{01}(\hat{x}, x^*) = 0$ if $\hat{x} \neq x^*$ and $\delta_{01}(\hat{x}, x^*) = 1$ if $\hat{x} = x^*$.

Equation (10) is used for measuring SA accuracy, when situational variables are used and with a proper setting of parameters (a, b) as in (11). In this case, either $M = 1$ if we consider a unique situation of interest (as it is the case in our illustration in Section IV) or $M > 1$ represents the number of situations of interest.

[3]Truth is considered here as a reference value.

B. Clarity

Clarity is the degree to which the TP is free from ambiguous values. It thus considers duplicated tracks, but also how confusing the estimation could be. This criterion is independent of the accuracy and can thus be measured without any reference to ground truth. As introduced in Section II, we assume that each estimation is provided by means of an uncertainty function f (a probability distribution, a belief function, a possibility distribution, etc) over the configuration space.

We define here the notion of clarity referring to the dispersion (or spread) of the uncertainty function, which would be defined either on the TP space or the situation space. We base the definition of clarity on the measure of entropy H which quantifies such a notion with the following properties, for random variables X, X_n from a set $\mathbb{X}$:

(H_1) Boundedness: $0 \leq H(X) \leq H_M \neq +\infty$
(H_2) Minimum: $H(X) = 0$ iff $f(x) = 1$ for only one $x \in \mathcal{X}$
(H_3) Subadditivity: $H(\mathbb{X}) \leq \sum_{n=1}^{N} H(X_n)$ for jointly distributed random variables $\mathbb{X}$
(H_4) Additivity: $H(\mathbb{X}) = \sum_{n=1}^{N} H(X_n)$ if variables from $\mathbb{X}$ are independent

For a set $\mathbb{X}$ of N variables X_n describing the estimated state of an object or a situation, we define the clarity as:

$$\mathrm{Clar}(\mathbb{X}) = 1 - \frac{H(\mathbb{X})}{H_M} \tag{12}$$

where $H(\mathbb{X})$ is an entropy measure for the joint distribution of $\mathbb{X}$, and $H_M = \sum_n \log|\mathcal{X}_n|$ its maximum value. For a single variable, $\mathrm{Clar}(X)$ is minimum (equal to 0) when X is uniformly distributed over $\mathcal{X}$ and maximum (equal to 1) whenever the distribution focuses on a singleton. Thanks to the properties of entropy measures we have:

$$\mathrm{Clar}_*(\mathbb{X}) \leq \mathrm{Clar}(\mathbb{X}) \leq \mathrm{Clar}^*(\mathbb{X}) \tag{13}$$

where

$$\mathrm{Clar}_*(\mathbb{X}) = 1 - \frac{\sum_{n=1}^{N} H(X_n)}{H_M} \tag{14}$$

$$\mathrm{Clar}^*(\mathbb{X}) = 1 - \frac{\max_{n=1}^{N} H(X_n)}{H_M} \tag{15}$$

In practice, the distribution $\mathbb{X}$ may be difficulty accessible. The lower and upper bounds are a convenient means to still estimate the clarity of the TP.

The formulation (12) is valid for any uncertainty function f, with H being defined accordingly. For a probability distribution p of $\mathbb{X}$:

$$\mathrm{Clar}(\mathbb{X}) = 1 + \frac{1}{\sum_n \log|\mathcal{X}_n|} \sum_{\mathbf{x} \in \mathcal{X}_1 \times \ldots \mathcal{X}_N} p(\mathbf{x}) \log p(\mathbf{x}) \tag{16}$$

where $p(\mathbf{x}) = p(\mathbb{X} = \mathbf{x})$.

For m objects, a global measure of clarity is then defined by averaging over the m objects:

$$\mathrm{Clar}(\hat{\mathbf{t}}) = \frac{1}{M}\sum_{m=1}^{M} \mathrm{Clar}(\mathbb{X}^{(m)}) \tag{17}$$

where $\mathbb{X}^{(m)}$ is the set of variables describing object m. This measure is maximal (equal to 1) if the clarity of estimation for each object is maximal and minimal (equal to 0) if the clarity for each object is minimal. The clarity for situation $\hat{\mathbf{s}}$ is computed with Eq. (17), with $M = 1$, if we consider a single situation.

C. Global quality

To define the global quality, we first define an aggregate measure for accuracy and clarity as:

$$\mathrm{AC}_{\beta,\alpha,a,b}(\hat{\mathbf{t}}) = \beta \mathrm{Acc}(\hat{\mathbf{t}}) + (1-\beta)\mathrm{Clar}(\hat{\mathbf{t}}) \tag{18}$$

with $\beta \in [0;1]$ and $\mathrm{Acc}(\hat{\mathbf{t}})$ and $\mathrm{Clar}(\hat{\mathbf{t}})$ defined by equations (17) and (10) respectively. This measure is 0 if and only if both the accuracy and clarity are 0, and is equal to 1 if and only if both the accuracy and clarity are 1. In between, the two measures of accuracy and clarity are weighted by β.

A global measure of quality for a tactical picture encompassing the dimensions of accuracy, clarity and completeness, is finally defined as follows:

$$\mathrm{Qual}_{\beta,\alpha,a,b}(\hat{\mathbf{t}}) = \begin{cases} \left(\mathrm{AC}_{\beta,\alpha,a,b}(\hat{\mathbf{t}})\right)^{\gamma(\hat{\mathbf{t}})} & \text{for } \gamma(\hat{\mathbf{t}}) \neq 0 \\ 0 & \text{for } \gamma(\hat{\mathbf{t}}) = 0 \end{cases} \tag{19}$$

where $\gamma(\hat{\mathbf{t}})$ is the completeness of the TP: $\gamma(\hat{\mathbf{t}}) = \frac{|\mathcal{M}|}{|\mathcal{M}^*|}$ is the ratio of detected objects over the total number of real objects of interest. The expression (19) satisfies properties (Q_1) to (Q_5) listed in Section II-B. Additionally, for a given value of $\mathrm{Comp}(\hat{\mathbf{t}})$, if $\beta = 1$, only accuracy and completeness are considered and if $\beta = 1$, only clarity and completeness are considered.

The expression (19) is valid for SA quality, denoted as $\mathrm{Qual}_{\beta,\alpha,a,b}(\hat{\mathbf{s}})$, when situational variables are used and with proper parameters a and b (see discussion in Section III-A). If we are interested in a single situation, $M = 1$ and the completeness is irrelevant.

IV. Illustration: Evaluating a CUAS solution

We can now illustrate the metrics on a scenario dedicated to CUAS, with simulated data. After a brief description of the scenario (Section IV-A), we present in Section IV-B our fusion solution to the CUAS problem, implemented as two Dynamic Bayesian Networks. Some inferred outputs are shown, processing both hard and soft evidences from level 1. Section IV-C is dedicated to the core contribution of this paper, where we illustrate the measures developed in Section III to compare levels 1, 1.5 and 2 processing with either hard and soft evidences.

A. Scenario description

We consider a blue team and a red team whose sensors and areas of interest are depicted in Fig. 1. The objective of the blue team is to protect the plant located in the blue area (#1). Two strategies are scheduled: "stay and hold" (the least favorable) or "lead a counter-attack" (the most efficient) using blue armored vehicles parked in the blue area (#2).

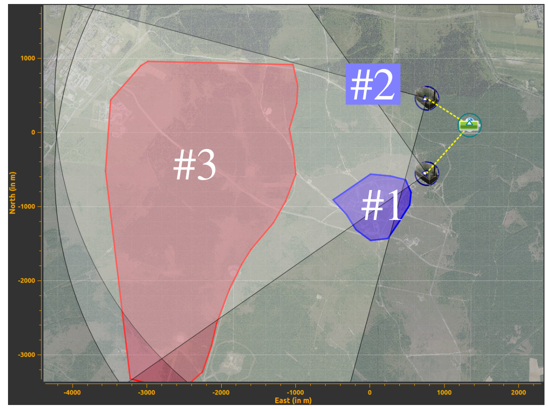

Fig. 1. Blue team's sensors' location and areas of interest for the blue team (#1 and #2) and for the red team (#3)

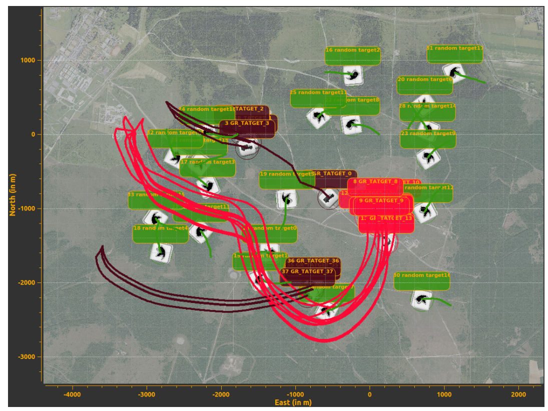

Fig. 2. Ground truth. 1 swarm of 10 drones (in red), 2 sections of 4 trucks (in brown) departing from the red area (#3) heading to the blue area (#2) and 20 random trajectories of birds (in green)

The red team consists of two sections of armored vehicles. Section 1, four armored trucks, is responsible for initiating the main attack on the plant, following the trajectory depicted in brown in Fig. 2. Section 2, also four armored trucks, provides support to Section 1. Additionally, the red team owns a fleet of 10 drones which is aimed at neutralizing the blue defense within the plant. The trajectory of the drone swarm is depicted in Figure 2 by the individual drone trajectories in red. Red drone swarm operates a maneuver to bypass the blue plant #1 and avoid being countered by blue CUAS.

B. Fusion solution for the CUAS problem

1) Level 1: Tracking and classification: To build the TP, the blue command center (CC) is equipped with a counter UAV system with two 3D radars and a central fusion node located within the CC (see Fig. 2). Tracking, classification and threat evaluation tasks rely solely on radar data. The Multiple Target Tracking (MTT) algorithm uses detection from sensors sequentially to initialize, update, or delete estimated tracks. Because we use a Multiple Hypothesis Tracking (MHT) approach, the centralized architecture facilitates straightforward fusion of measurements. An Interacting Multiple Model (IMM) enables leveraging geographical information such as road networks, forested areas, and elevation, and enhancing the precision of estimated track states compared to a MHT with a single motion model [12]. Fusion processing is integrated into our LEXLUTOR platform, specifically designed for evaluating data fusion algorithms [3]. The classifier used in this paper implements a naive Bayesian approach (as described in [13]) to classify tracks. The types of objects to be recognized by the system are listed in the Table I, as the possible states of variable $X_{TS}^{(m)}$. See [12] for details about the classifier.

2) Level 1.5: Individual threat evaluation: The solution to the CUAS problem is implemented through a Bayesian Network depicted in Figure 3. Object features assessment is combined with a threat model. A drone is potentially a threat if it has the opportunity (close distance to the CC), the intent (took off from the Red area), and the capability (its type enables carrying weapons) [14]. This Bayesian Network captures conditional independence in the joint distribution by a directed acyclic graph (DAG). Leveraging these conditional independencies enables the factorization of the joint distribution, thereby facilitating the compact representation of even very large distributions. The observable variables, denoted

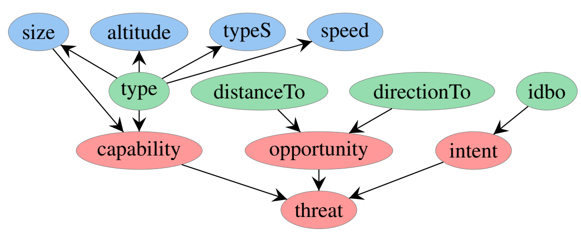

Fig. 3. Dynamic Bayesian Network for individual drone threat assessment

in blue and green in Fig. 3, pertain to level 1 of the JDL model. Each detected object at time k is represented by a set of variables $\mathbb{X}_k^{(m)}$ for m detected objects. Green variables are issued from the semantic elevation step (see Section II-D) and correspond to "level 1.5". The latent variables, depicted in red, correspond also to "level 1.5". These latter variables are denoted as $\mathbb{T}_k$, representing the state at time k for each object.

The observed variables at level 1 or 1.5 are defined according to Table I. The evidence is derived from the attributes and states of the targets. The semantic elevation is implemented by means of a fuzzy discretization (similar to [15]) to obtain the evidences to be propagated into the DBN. Figure 4 displays the corresponding fuzzy memberships for the node "speed". where μ_l, μ_m and μ_h denote the membership functions of low, medium and high velocity, respectively. Parameters of these trapezoidal functions set the granularity of the problem at level 1.5. A coarsening is applied to variables "typeS" and "idbo" (Identification By Origin) casting their original universe into binary ones, (Drone; Not drone) and (Yes; No) respectively.

The transformation from fuzzy membership to probability is defined by equation (20) ($\forall i \in \mathbb{X}_k^{(m)}$):

$$p_i(u) = \frac{\mu_u^{1/\epsilon}}{\sum_{i=1}^{n} \mu_u^{1/\epsilon}} \tag{20}$$

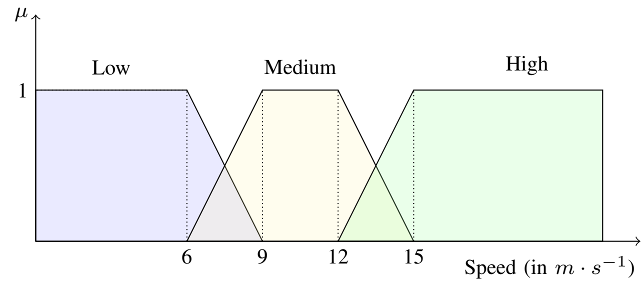

Fig. 4. Speed membership functions

where $u \in \mathcal{X}_{sp}$, and ϵ is a forgetting factor equal to 0.5 in this paper. The probability (20) is used to compute the threat probability with soft evidence in our DBN as illustrated below.

TABLE I
PROBLEM VARIABLES FOR THE DBN OF FIG. 3

Variable	Node	States set $\mathcal{X}_i$ or $\mathcal{S}_j$	States notation
$X_{sz}^{(m)}$	size	Small, Medium, Large, X-Large, XX-Large	sm, me, la, xla, xxla
$X_{sp}^{(m)}$	speed	Low, Medium, High	l,m,h
$X_{a}^{(m)}$	altitude	Low, Medium, High	l,m,h
$X_{TS}^{(m)}$	typeS	Drone, Plane, Bird, Other Air, Truck, Car, Pedestrian, Other Ground	D, P, B, OA, T, C, P, OG
$S_{T}^{(m)}$	type	drone, not drone	D,NoD
$S_{To}^{(m)}$	directionTo	in direction, to be considered, not in direction	InD, TBC, NoInD
$S_{d}^{(m)}$	distanceTo	far, near, closest	f, n, c
$S_{id}^{(m)}$	idbo	yes, no	y, n
$S_{c}^{(m)}$	capacity	yes, no	y, n
$S_{o}^{(m)}$	opportunity	yes, no	y, n
$S_{i}^{(m)}$	intent	yes, no	y, n
$S_{t}^{(m)}$	threat	yes, no	y, n

Figure 5 displays the results of the threat assessment performed by a hard evidence DBN, for one track associated to a drone belonging to the swarm of drones. The probability on the "Opportunity" node (blue curve) increases as the drone gets closer to the area #1. Excepted at time 07h46, the probability decreases because of the drone swarm's maneuver to bypass the area #1 (see Fig. 1) bringing a decrease in probability of the parent node "directionTo". Despite of this variation, the threat probability (red curve) remains high.

Figure 6 displays the threat probability of the same drone track obtained with a soft evidence DBN. We observe that the curves are less noisy and provide a better interpretation on the drone threat. In Section IV-C, we will provide some objective measurement of this intuition.

3) Level 2: Critical Situation evaluation: In accordance to our scenario, we are concerned by a situation in which a group of drones would attack the blue defence and CC. Such a situation would occur if at least 3 drones posing a possible threat to the CC would have a coherent behavior, Drones are considered coherent if meaning they move in close formation. Figure 7 shows the DBN for the situation assessment model. For illustration purposes, we display the threat nodes from three individual targets, as computed by the DBN of Fig. 3.

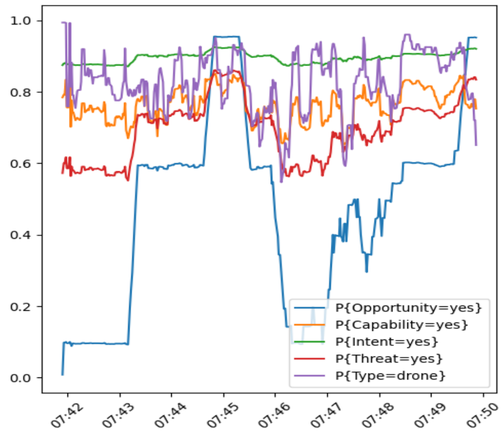

Fig. 5. Individual drone threat probability with hard evidence

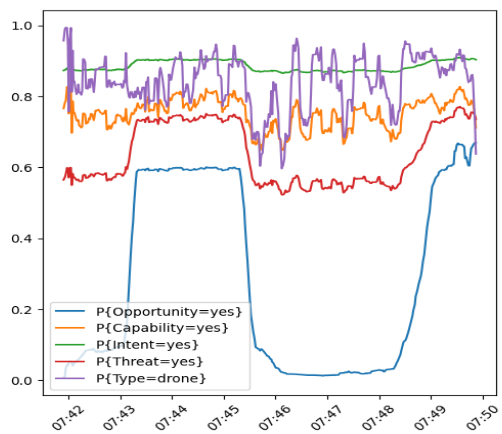

Fig. 6. Individual drone threat probability with soft evidence

Figure 8 displays the threat probabilities for three tracks classified as drone (from level 1.5, node "type") and the critical situation probability (red curve) obtained with soft evidences. We observe that the probability of each drone threat exhibits similar characteristics than the threat probability of individual drones from Figure 6. Additionally, the critical situation probability remains close to 1, except between 7:46 and 7:48, during which the swarm maneuvers to avoid area #1, leading to a decrease in the individual drone threat probability (as the opportunity probability decreases).

C. Evaluating solutions

As an illustration, we present here some results of the TP and SA metrics through levels 1, "1.5" and 2. We compare the two versions of our solution, denoted as "Hard" (when hard evidence is used to feed the DBN at "level 1.5") and "Soft" (when soft evidence is used). We show a subset of quality metrics as accuracy, clarity and global quality.

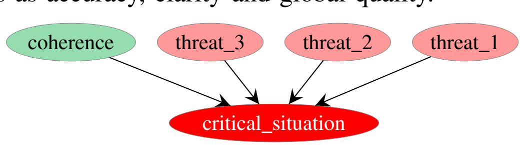

Fig. 7. Critical situation DBN

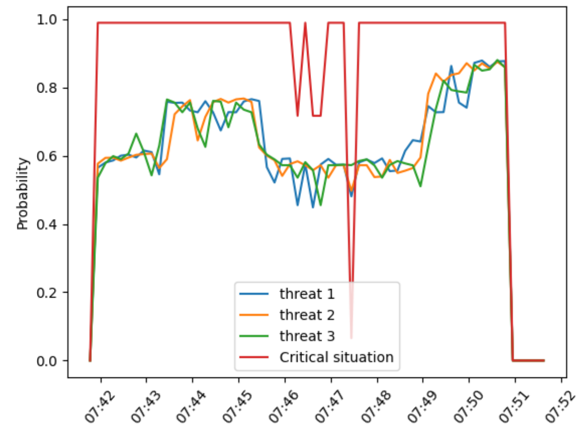

Fig. 8. Critical situation and individual threats probabilities

In Fig. 9, the TP accuracy for the is displayed, comparing levels 1 and 1.5 for the two versions of the solutions. Parameters for accuracy for the level 1 features are set to $\mathbf{a} = [10, 10, 3, 2, 7, 15, 0.5, 10, 7]$ and $\mathbf{b} = [20, 20, 10, 20, 20, 10, 20, 10, 10]$, and $\alpha = 2$. We observe that

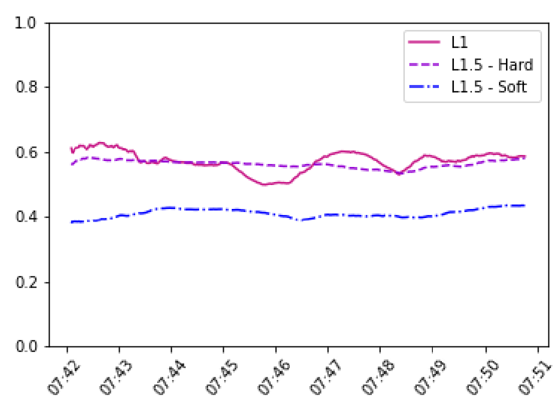

Fig. 9. Tactical Picture accuracy from Level 1 to 1.5

the accuracy at level 1 (computed over numerical features essentially) is less stable than its equivalent measure at "level 1.5" computed on nominal features, while very similar in values. The TP accuracy is however lower with the "Soft" version than with the "Hard" one. The quite low accuracy values (regardless the method) is explained by the high number of birds in the scene, for which the tracker is not adapted. Clarity at "level 1.5" is shown in Fig. 10. Here again, the clarity for the "Soft" version is slightly lower than its "Hard" counterpart. This observation is confirmed by results of Figure 11 where the global quality is displayed.

Figure 12 displays the SA quality of (level 2). We observe a decrease corresponding to the time where the group of drones seems to avoid the area under protection (cf. Fig. 8). Interestingly, the two versions of the solution lead to very similar global performance on the situation assessment, beside the noticeable difference at lower levels. Although

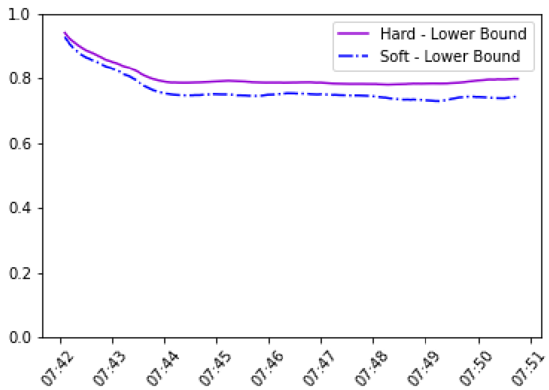

Fig. 10. Tactical Picture clarity for Level 1.5

further studies are needed to draw any conclusion on the performances of solutions, this result shows the ability of the proposed metrics to cover levels 1 and 2 in a consistent way. Furthermore, the global SA evaluation metrics can support the optimal granularity of the problem for possible computational gain Such refinement would be part of level 4 of the JDL.

V. Conclusions

In this paper, we presented an harmonized evaluation of a fusion solution through different levels of the JDL model. We proposed a series of metrics for global evaluation the the Tactical Picture and Situation Assessment quality, encompassing the criteria of accuracy, clarity and completeness. We identified the semantic elevation step as a series of mappings between level 1 and level 2 variables. This semantic transition also highlighted an intermediate level ("level 1.5") where the individual object assessment semantics is increased by means of coarsening and definition of meaningful variables. Metrics are illustrated on a CUAS fusion solution implemented as a Dynamic Bayesian Network. Two uncertainty handling methods are compared, propagating either hard or soft evidence. In future work, we will extend the measures to the other dimensions of timeliness, continuity and commonality. We will also study the impact of the different parameters on the metrics

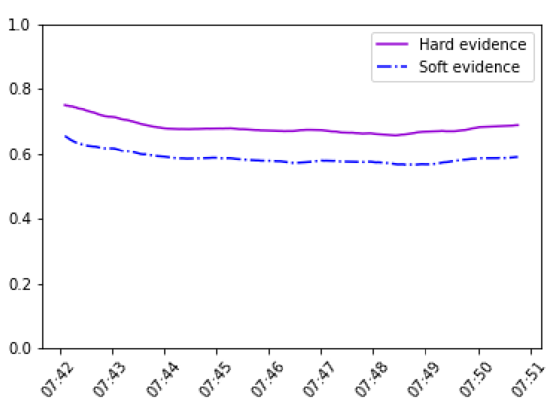

Fig. 11. Tactical Picture quality for Level 1.5

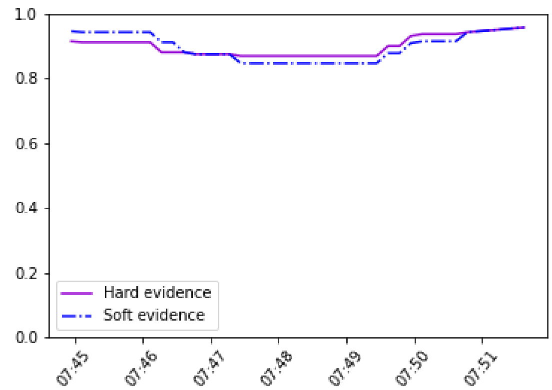

Fig. 12. Global Situation Assessment quality

on the overall result, and will compare with other solutions framed into other uncertainty theories.

References

[1] P. C. G. Costa, A.-L. Jousselme, K. Laskey, E. Blasch, V. Dragos, J. Ziegler, P. de Villiers, and G. Pavlin, "URREF: Uncertainty representation and reasoning evaluation framework for information fusion," *JAIF*, vol. 13, no. 2, pp. 137–157, Dec. 2018.

[2] J. Llinas, C. Bowman, G. Rogova, A. Steinberg, E. Waltz, and F. White, "Revisiting the JDL data fusion model II," in *Proc. of ICIF*, Stockholm, Sweden, June-July 2004, pp. 1218–1230.

[3] B. Pannetier and A.-L. Jousselme, "Evaluation of counter unmanned aerial systems through the levels," in *Proc. of the Int. conference of Information Fusion*, Charleston, South Carolina, Jun. 2023.

[4] T. Nguyen, C. P. Lim, N. D. Nguyen, L. Gordon-Brown, and S. Nahavandi, "A review of situation awareness assessment approaches in aviation environments," *IEEE Systems Journal*, vol. 13, no. 3, pp. 3590–3603, 2019.

[5] E. A. Green, M. E. Armstrong, and J. Mantua, "Scientific measurement of situation awareness in operational testing," *The Journal of Test & Evaluation*, vol. 44, no. 3, September 2023.

[6] M. R. Endsley and D. J. Garland, *Situation Awareness Analysis and Measurement*, M. R. Endsley and D. J. Garland, Eds. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers, 2000.

[7] J. Salerno, "Measuring situation assessment performance throughthe activities of interest score," in *Int. Conference on Information Fusion*, Cologne, Germany, Jul. 2008.

[8] SIAP System Engineering Task Force (SIAP SETF), "Single integrated air picture (SIAP) attributes - 2.0," Tech. Rep. 2003-029, August 2003.

[9] Y. Fischer, A. Reiswich, and J. Beyerer, "Modeling and recognizing situations of interestin surveillance applications," in *IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2014.

[10] A. Steinberg, "A model for situation and threat assessment," in *Information Fusion for Command Support*, ser. Meeting Proceedings RTO-MP-IST-055, Keynote 1, Neuilly-sur-Seine, France, 2006.

[11] P. Kowalski and A.-L. Jousselme, "Context-awareness for information correction and reasoning in evidence theory," *International Journal of Approximate Reasoning*, vol. 153, pp. 29–48, Feb. 2023.

[12] B. Pannetier, J. Dezert, J. Moras, and R. Levy, "Wireless sensor network for tactical situation assessment," *IEEE sensors*, vol. 22, no. 1, 2022.

[13] Y. Bar-Shalom, T. Kirubarajan, and C. Gokberk, "Tracking with classification-aided multiframe data association," *IEEE Trans. on AES*, vol. 41, no. 3, pp. 868–878, 2005.

[14] P. Kowalski, M. Zocholl, and A.-L. Jousselme, "Explaining the impact of source behaviour in evidential reasoning," *Information Fusion*, vol. 81, pp. 41–58, May 2022, SI: Advances in Explainable (XAI) and Responsible (RAI) Artificial Intelligence.

[15] W. Li, W. Zhang, B. Liu, and Y. Guo, "The situation assessment of uavs based on an improved whale optimization bayesian network parameter-learning algorithm," *Drones*, vol. 7, no. 11, 2023.