Dissecting uncertainty handling techniques: Illustration on maritime anomaly detection

ANNE-LAURE JOUSSELME
GIULIANA PALLOTTA

Detecting and classifying anomalies for Maritime Situation Awareness highly benefits from the combination of multiple sources, correlating their output for detecting inconsistencies in vessels' behaviour. Adequate uncertainty representation and processing are crucial for this higher-level task where the operator analyses information in conjunction with background knowledge and context. This paper addresses the problem of performance criteria identification and definition for information fusion systems in their ability to handle uncertainty. In addition to the classical algorithmic performances such as accuracy, computational cost or timeliness, other aspects such as the interpretation, simplicity or expressiveness need to be considered in the design of the technique for uncertainty management for an improved synergy between the human and the system. The Uncertainty Representation and Reasoning Evaluation Framework (URREF) ontology aims at connecting these criteria to other uncertainty-related concepts. In this paper, we dissect six classical Uncertainty Representation and Reasoning Techniques (URRTs) in their basic form framed into three uncertainty models of probability, belief functions and fuzzy sets, and addressing a fusion problem for maritime anomaly detection. We introduce the *Uncertainty Supports* as a means to capture what is the carrier of uncertainty and distinguish between three types of supports, that are single variables, sets of variables and uncertainty representations. The latter type indeed captures second-order uncertainty. The different URRTs are qualitatively evaluated according to their expressiveness along the uncertainty supports, and quantitatively evaluated according their accuracy and conclusiveness (uncertainty and imprecision) when processing real AIS data with pseudo-synthetic anomalies. This study illustrates a possible use of the URREF for the assessment and comparison of uncertainty handling methods in fusion systems. The framework provides solid basic foundations for a formal assessment to guide further development and implementation of fusion schemes, as well as for the definition of associated criteria and measures of performance.

Manuscript received April 27, 2018; revised March 5, 2019; released for publication April 17, 2019.

Refereeing of this contribution was handled by Pieter de Villiers.

Authors' addresses: A-L. Jousselme, NATO STO Centre for Maritime Research and Experimentation, La Spezia, Italy (E-mail: anne-laure.jousselme@cmre.nato.int). G. Pallotta, Lawrence Livermore National Laboratory, Livermore, California, USA (E-mail: pallotta2@llnl.gov).

The authors wish to thank the NATO Allied Command Transformation (NATO-ACT) for supporting the CMRE project on Data Knowledge and Operational Effectiveness (DKOE).

This paper is an extension of a preliminary version presented in [24].

1557-6418/18/$17.00 © 2018 JAIF

I. INTRODUCTION

In the field of Maritime Situation Awareness (MSA), detecting and classifying vessels' abnormal behaviour is a challenging and crucial task at the core of the compilation of the maritime picture [32, 31]. It requires not only the extraction of relevant contextual patterns-of-life information shaped for instance as maritime routes or loitering areas [42], but also the real time monitoring of the maritime traffic by a set of sensors mixing co-operative self-identification systems (such as the Automatic Identification System (AIS)) and non-cooperative systems such as coastal radars or satellite imagery, to overcome the possible spoofing of the AIS signal [44]. In many cases, intelligence information is of great help to refine and guide the search in the huge amount of data to be processed, filtered and analysed.

In order to take informed decisions, the operator needs to get good quality information. Furthermore, he/she needs to understand additional characteristics of the provided information, including for instance, how that information has been obtained, processed, or what was the context of its creation. In particular, understanding how an anomaly detector came up with an alert is of great importance to the Vessel Traffic System (VTS) operator. More specifically, the operator would benefit from knowing which were the reference data used, which were the sources processed, if the information and associated uncertainty were obtained in objective or subjective manner, whether the decision process considered the sources' quality and how, if the contextual information was considered in the decision, what was the meaning of numerical output values expressing uncertainty, and what was the underlying logical reasoning providing the answer. Second-order information quality may also be highly valuable. For example, probability maps about possible threats could be supplemented by uncertainty assessments about the validity of the probability values, represented as intervals or error estimations on algorithms performance. The benefit of including these different information quality dimensions is twofold: on the one hand, they increase the operator's situation awareness and, on the other hand, they improve trust in the use of the system.

To characterise the outputs provided to operators by some information system, the standard performance criteria of algorithms such as precision, accuracy, False Alarm Rate (FAR), Area Under the Receiver Operating Characteristic (ROC) curve (AUC), timeliness or computational cost [1, 33, 11] may not be sufficient and should be complemented by others to cover the interaction of humans and systems. For instance, some criteria such as explanation, adaptability, simplicity, expressiveness could be considered as well. The Evaluation of Techniques for Uncertainty Representation (ETUR) working group of the International Society of Information Fusion (ISIF) addresses since

2011 the definition and articulation of assessment criteria for uncertainty models and frameworks, uncertainty types, uncertainty derivation, uncertainty nature [8]. Outcomes of this work provide guidance for the selection and design of adequate tools for reasoning support, uncertainty traceability and understandability (e.g., [5, 10, 43]). It is also a first step towards some standardisation of the characterisation and assessment of uncertainty management techniques and, by extent, of fusion schemes.

Evaluating or comparing uncertainty *calculi* in the absolute is not trivial task because these make different fundamental assumptions about the nature and interpretation of uncertainty they aim at representing or processing (see for instance [29, 19]). Fundamental and global formal evaluation and analysis have been in particular presented in [48, 55, 56, 17, 29, 52] and more recently in [19]. For instance, probability, possibility and fuzzy set theory are non comparable since they are appropriate to deal with different types of uncertainty. Rather than competitors, they appear to be "complementary theories of uncertainty that utilise distinct types of uncertainty for expressing deficient information" [29]. Belief functions [47, 54] are "[...] aimed directly at modeling incomplete evidence, but certainly not incomplete knowledge," and designed to handle singular uncertainty [47, 19]. Fusion rules have their own meaning and application constraints as well. While being an updating rule, Bayes' rule is also widely used for fusion purposes (e.g., [34, 7]). However, Bayes' rule is not applicable in case of probable knowledge, unanticipated knowledge and introspective knowledge [15]. In Shafer's view, Dempster's rule is specifically dedicated to combine uncertain and imprecise singular information, such as testimonies. Dempster's rule should also be applied only to independent and reliable sources [47, 53]. It appears thus that rather than competitors the different models for uncertainty representation and associated reasoning schemes are dedicated to different problems and different types of information. As a step toward a formal analysis of uncertainty representation and reasoning techniques, the work presented in this paper aims at bringing the comparisons and descriptions of the classical uncertainty models under the Uncertainty Representation and Reasoning Evaluation Framework (URREF).

In this paper, we compare six (6) different approaches (hereafter called Uncertainty Representation and Reasoning Techniques, URRTs) to combine pieces of information from a set of heterogeneous sources (hard and soft) as the core of a maritime anomaly detector for route deviation. In complement to comparative analyses (e.g., [26, 2]), this paper identifies additional comparison elements which may have an impact on the behaviour (and performances) of the fusion schemes. The maritime anomaly detection problem is first introduced in Section II, covering route extraction and route association problems together with some associated uncertainty-related challenges. In Section III, we briefly review the current state of the URREF ontology and introduce the *uncertainty support* as part of possible refinement of the Expressiveness criterion. Six URRTs are introduced in Section IV as alternative basic fusion schemes to solve the above defined problem, with an emphasis on the uncertainty representation. The six URRTs are compared in Section V in a qualitative way regarding their expressiveness (relatively to their uncertainty support and imperfection type captured) but also in a quantitative manner through more classical but complementary quality criteria, processing a real AIS dataset augmented with pseudo-synthetic anomalies. We conclude in Section VI on future work and further challenges to be addressed in the coming years by the Evaluation of Techniques for Uncertainty Representation working group.

II. MARITIME ANOMALY DETECTION

We illustrate the discussed methods via a real-world example of maritime anomaly detection. Although a unique definition of anomalies in the maritime domain is not available yet, we here use the term "maritime anomaly" to indicate a *deviating* behaviour from traffic normalcy, which we learn from spatio-temporal data of ships at sea. More specifically, the analysis of traffic spatio-temporal data streams provided by the AIS, a cooperative self-reporting system allows detecting and characterising inconsistencies or ambiguities, which can be ultimately transformed into usable and actionable knowledge [40]. We briefly introduce in this section the problem of route extraction, which builds the normalcy models, in our case, traffic normalcy, in such a way that these models can be further exploited for anomaly detection. We then introduce the problem of associating a vessel to a pre-defined route selected from the extracted system of routes which represents the traffic normalcy functional to the anomaly detection. We conclude this section with some uncertainty challenges related to the way we represent the maritime routes which affects the maritime anomaly detection.

A. Route extraction

The Traffic Route Extraction and Anomaly Detection (TREAD) tool presented in [42] implements an unsupervised classification approach which we here use to derive a dictionary of the maritime traffic routes by processing spatio-temporal data streams from terrestrial and satellite AIS receivers. The analysis and synthesis of the activity at sea as patterns of life is referred to as *maritime routes* and summarises the normal maritime traffic over a given period of time, a given area and a specified set of employed sensors (or sources). The AAP-6-2014 NATO glossary of terms defines a route as "*The prescribed course to be travelled from a specific point of origin to a specific destination.*" A

TABLE I
Examples of source statements expressed by ϕ about different vessel attributes.

Attribute i	ϕ	Type of statement
SOG (knots)	10.3	Single measurement (precise and certain)
Type [Type1 Type2 Type3 Type4 Others]	[0.2 0.1 0 0.7 0]	Probability vector
Size	"Big vessel"	Natural language

TREAD *route* is then defined by a starting point and an ending point, together with a subset of intermediate waypoints, describing a physical path on a portion of the sea. If the area under surveillance is captured by a big enough bounding box, the route starting and ending points are the centroids of stationary areas, either coastal areas such as ports, either offshore areas such as islands, either offshore platforms, or open-sea areas such as fishing areas. The TREAD algorithm first reconstructs the single-vessel trajectories by linking the vessels' contacts and then clusters the trajectories followed by vessels into groups having the same starting and ending points. Each of these clusters represents a maritime route. The average path along a route is called *synthetic route*. The basic uncertainty around this path is computed using the trajectories of all the vessels which transited along that route in the given time window.

While only temporal streams of positional information is processed to extract the set of maritime routes, they can be further characterised by additional attributes representing the traffic of vessels composing it, such as speed, type or heading distributions. The associated uncertainty characterisation of the route along these attributes can be more or less complex, ranging from simple average values, to added variance parameters, to histograms, to estimated complete probability distributions, to sets of distributions (see Section II-C). The maritime traffic, and thus the set of routes, may be influenced by meteorological conditions (some areas may be avoided), season, economical context (ships may decide their destination based on the current stock market linked to their cargo) or areas of conflict. Also, in order to derive the average path (i.e., synthetic route) from the route cluster an extent parameter is included in the TREAD algorithm which allows adjusting the search range radius dynamically, thus enhancing the computation of intermediate waypoints while still avoiding issues such as land crossing.

The set of routes summarises thus some kinematic patterns of life of vessels over a given period of time and region, possibly layered by specific vessel types (e.g. fishing vessels, tankers, passenger vessels). This synthetic information and associated uncertainty characterises part of the context or background knowledge for the problem of route association and detection of anomalies at sea. It provides a *reference* or *normalcy* against which the current vessel contacts will be compared, and the anomalies detected.

B. The route association problem

A route deviation detector is to be designed to help the Vessel Traffic System (VTS) operator to (1) associate vessels to existing routes (and possibly predict their destination), and (2) detect abnormal behaviours to be further investigated.

We consider a vessel V observed by a series of heterogeneous sources $\mathcal{S} = \{s_1, \ldots, s_N\}$ such as a coastal radar and its associated tracker, a SAR (Synthetic Aperture Radar) image with associated either ATR (Automatic Target Recognition) algorithm or a human analyst, a visible camera operated by a human analyst, the AIS information sent by the vessel itself or some intelligence source. Let $\mathcal{A}$ be the set of features of interest, either observed and thus about which information is either provided by or extracted by some sources of $\mathcal{S}$, or to be inferred. For solving our problem of route association, we consider attributes such as the position (latitude, longitude), Course Over Ground (COG), Speed Over Ground (SOG), Type, Length, and also the maritime route followed by the vessel. Let denote by $\mathcal{A}$ the set of features of interest, by $\mathcal{X}$ the set of uncertain variables corresponding to features of $\mathcal{A}$, by X_i the variable of $\mathcal{X}$ corresponding to feature $i \in \mathcal{A}$ and by U a subset of variables of $\mathcal{X}$. We further denote by Ω_i the domain of definition of X_i containing the set of its possible values, by $x_i \in \Omega_i$ a singleton of Ω_i and by $A_i \subseteq \Omega_i$ a subset of Ω_i. Let Ω be the corresponding space, defined as the Cartesian product of the Ω_i corresponding to vessel features of interest at a given timestamp t. Also, $\mathbf{x}_t = \{\phi(X_i, s, t)\}_{(i \in \mathcal{A}, s \in \mathcal{S})}$ denotes a set of information items jointly provided by some sources from $\mathcal{S}$ about some features in $\mathcal{A}$ at a specific instant in time t. This notation of information item encompasses the general case where sources provide some uncertainty about their statement and thus ϕ denotes a source statement either as a single measurement (precise and certain), either as a probability vector (expressing some uncertainty interpreted as provided by the source itself), either as a natural language expression (possibly vague), etc. Table I lists some examples. In the specific case of precise and certain measurements defined over a scale of real numbers, $\mathbf{x}_t$ would simply be a vector of real values of Ω. For the purpose of the discussion in this paper, we consider that each feature estimation is provided by a single source (while in general several sources may provide information about the same feature). Moreover, we focus on the fusion of all (singular) observations obtained at the same instant in time t. Thus for the sake of simplicity of the exposure, the index t and s will be omitted, and information items will be denoted simply by $\phi(X_i)$ or ϕ_i. Uncertainty about state transitions $\mathbf{x}_t \rightarrow \mathbf{x}_{t+1}$ will be considered in further extension of this work.

Let $\Omega_R = \{R_0, R_1, \dots, R_K\}$ be the finite set of possible routes followed by the vessel V for the given area of interest, where R_k for $k = 1, \dots, K$ is a pre-computed route and R_0 represents "none of the K routes": R_k, for $k = 1, \dots, K$, is the label to be output by the fusion process corresponding to the event "The vessel V follows route R_k" and R_0 is a rejection class corresponding to the vessel following no specific pre-computed route. This class gathers the events of "The vessel is physically off-route," "The vessel is in the reverse traffic on the route," "The speed is not compatible with the route followed," "The type of the vessel is not compatible with the route followed," representing some Maritime Situational Indicators of possible interest to the VTS operator. In the following, we consider a quite simple reasoning scheme according to which an anomaly is detected based on a joint assessment (fusion) of the 5 features of Position, COG, SOG, Length, Type provided by the AIS report of the vessel and describing the route. Other said, the behaviour of a vessel V is detected as being abnormal if *the set* of its estimated features is *not compatible* with any existing route. Compared to [40], the nature of the anomaly will not be specified. However, identifying the features which contribute the most to the disbelief toward any of the routes would provide information about the nature of the anomaly.

For convenience, we partition Ω into the observation space, say Ω_o and decision space Ω_R. The fusion scheme to be designed aims thus at establishing a mapping $\Psi : \Omega_o \rightarrow \Omega_R$ such that $\hat{R} = \Psi(\mathbf{x})$ is the route label assigned to V represented by $\mathbf{x}$ (at time t). The underlying reasoning is that any observed feature at t combined with possible background knowledge contributes to a global belief (disbelief) that V is following a pre-established route from Ω_R. Indeed, if all the observed (measured) features match the corresponding feature values of a specific existing route, then the corresponding route label is assigned to the vessel. If some "inconsistency" or "conflict" exists between the set of observed features and the routes features (e.g. if the distance between $\mathbf{x}$ and each of the R_k is too high, or if the set of compatible routes according to the speed does not match the equivalent set according to the type) then V is assigned to no route and an anomaly is reported (label R_0).

The same set of pieces of information would then be used for two purposes:

(1) Associating a vessel to route, under the assumption that the sources are reliable and
(2) detecting anomalies, under the assumption that an inconsistency among the set of estimated features would reveal a possible behaviour of interest.

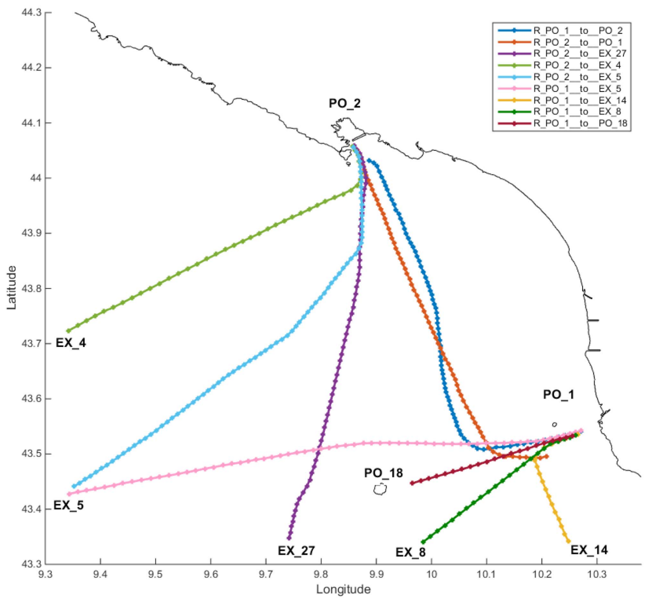


Fig. 1. Historical route prototypes extracted via the TREAD algorithm [42] in the area between La Spezia and Livorno, Italy, from AIS data (Jan 1–Feb 20 2013).

However, on the one hand, information is inherently imperfect (incomplete or imprecise, uncertain, gradual, granular [19]—See Section III-B) and on the other hand inconsistencies may arise either from sources limitations (e.g. gaps in or weak coverage of sensors, limited reasoning abilities, storage limitations, false detections or identifications), and lack of reliability in general) or malevolent behaviour of the vessel such as deception. The appropriate detection and identification of anomalies highly relies on the technique for fusing the different pieces of information and detecting inconsistencies, which include the handling of uncertainty.

C. Uncertainty in Maritime Anomaly Detection

Figure 1 illustrates the set of maritime routes previously extracted with TREAD algorithm [42] from a large number of AIS contacts for the area between La Spezia and Livorno in Italy. The used AIS data are part of a reference dataset published at CMRE [41].

As computed by TREAD, a maritime route is a cluster of vessel detections (positions from AIS contacts) with label R_k and identifies the geographical area where vessels have been observed travelling between a predefined entry point and exit point in the past temporal window. From this set of contacts (cluster) several synthetic representations can be extracted more or less complex, more or less rich, more or less precise. As an example, each route is represented in a synthetic way by a series of *intermediate* waypoints with associated average headings. Additional features characterising the traffic can be further extracted such as the distribution of speed, length and type of vessels traveling on this route. As a matter of fact, routes are, by nature, uncertain objects and the characterisation and representation of their uncertainty is of primary importance for

TABLE II
Dictionary of routes and examples of simple associated uncertainty representations

Route	Label	Synthetic route $\mathbf{r}^{(k)}$	Traffic information statistics			
	$R-Origin-to-Destin$	POSITION $\pm$ width [km]	$\overline{COG}\pm$STD	SOG [KNOTS]	LENGTH [M]	TYPE [FREQUENCY]
R_1	R_PO_1__to__PO_2	$\{WP\}^{(1)} \pm 2.77$	$297 \pm 85°$	$\mathcal{N}(10,4)$	[80 : 100];[260 : 300]	[0.8 0.1 0 0 0.1]
R_2	R_PO_2__to__PO_1	$\{WP\}^{(2)} \pm 3.01$	$140 \pm 49°$	$\mathcal{N}(11,4)$	[0 : 130];[250 : 300]	[0.37 0.13 0 0 0.5]
R_3	R_PO_2__to__EX_27	$\{WP\}^{(3)} \pm 1.19$	$185 \pm 11°$	$\mathcal{N}(12,2)$	[120 : 250]	[0.75 0 0 0 0.25]
R_4	R_PO_2__to__EX_4	$\{WP\}^{(4)} \pm 2.86$	$221 \pm 31°$	$\mathcal{N}(15,4)$	[100 : 200];[260 : 350]	[0.97 0 0 0 0.03]
R_5	R_PO_2__to__EX_5	$\{WP\}^{(5)} \pm 5.08$	$209 \pm 19°$	$\mathcal{N}(11,1)$	[100 : 300];[200 : 210]	[1 0 0 0 0]
R_6	R_PO_1__to__EX_5	$\{WP\}^{(6)} \pm 1.91$	$255 \pm 18°$	$\mathcal{N}(13,4)$	[0 : 25];[110 : 300]	[0.82 0.09 0 0.09 0]
R_7	R_PO_1__to__EX_14	$\{WP\}^{(7)} \pm 1.25$	$210 \pm 90°$	$(\mathcal{N}(10,2);\mathcal{N}(18,2))$	[50 : 100];[120 : 200]	[0.93 0 0 0.07 0]
R_8	R_PO_1__to__EX_8	$\{WP\}^{(8)} \pm 0.86$	$225 \pm 14°$	$(\mathcal{N}(11,2);\mathcal{N}(19,2))$	[100 : 150];[190 : 240]	[0.38 0.24 0 0.38 0]
R_9	R_PO_1__to__PO_18	$\{WP\}^{(9)} \pm 0.98$	$244 \pm 21°$	$\mathcal{N}(11,3)$	*not reported*	[0 0 0 1 0]

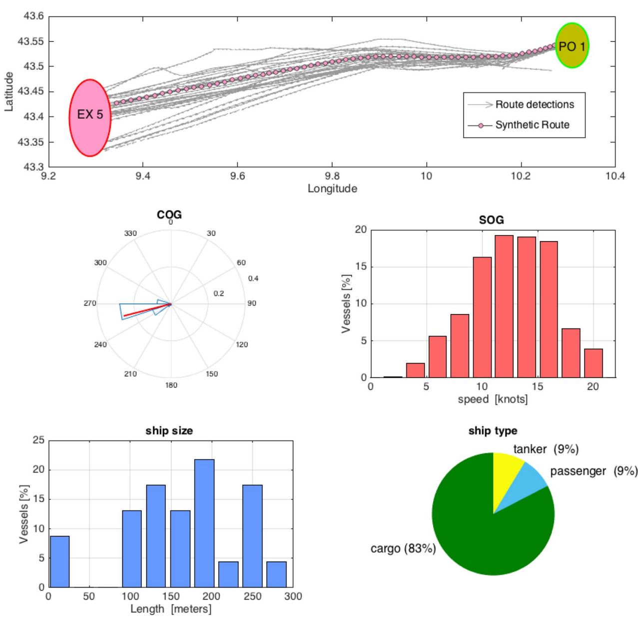


Fig. 2. An example of multi-dimensional uncertainty representation for Route R_6 with label R_PO_1__to__EX_5 reported in Table II.

a proper use of this information for the anomaly detection task. Figure 2 gives an example of how some dimensions of uncertainty for a specific route can be represented: top panel—the geographical displacements of vessel positions with respect to the synthetic (average) route; middle left panel—distribution of COGs; middle right panel—distribution of SOGs; bottom left panel—distribution of ship length; middle right panel—frequency of types of the ships which transited along that route in the given time window).

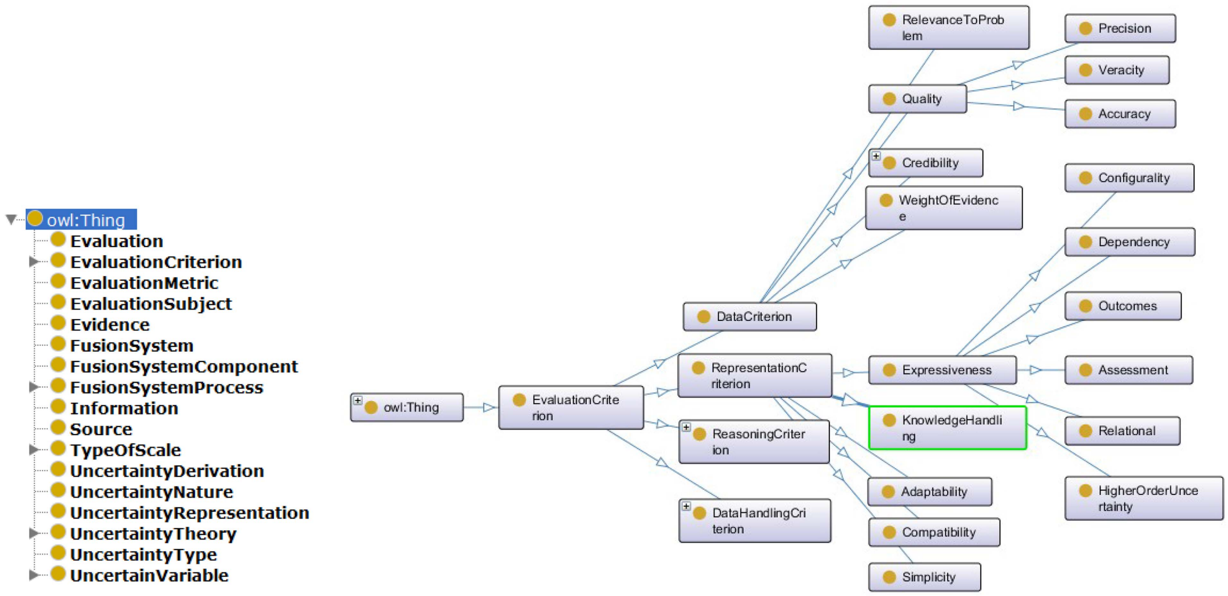
Table II lists several examples of simple uncertain representations for the different routes in the derived dictionary and illustrates how this multi-dimensional uncertainty of the routes can be encoded in a compact way.[1]

For instance, each route R_k may be represented by a prototype $\mathbf{r}^{(k)}$ corresponding to the mean or most frequent trajectory of the cluster. Those features are *precise* and *certain* values to which some *imprecision* or *uncertainty* can be added for a richer representation, based on the statistical information from the raw AIS messages which contain many additional fields of interest. The route width $w^{(k)}$ is defined as the maximum of the distances of each route point (i.e., vessel positions associated to the route) to the closest waypoint on the synthetic route. It defines an area where the transited vessels have been observed in the past.

The statistics extracted from the raw AIS dataset may serve two purposes: On the one hand, they can be used as the basic ingredient for the generic uncertainty representation captured by the route objects and, on the other hand, they are possibly transferred to express some uncertainty about new singular measurements. The histograms of the different features (SOG, COG, Length, Type) can be further interpreted as likelihood functions $p(X_i = x \mid R_k)$ (see Section IV-C) and approximated by different models. For instance, the distribution of the speed variable X_S for Route R_1 can defined by the couple $(\bar{s}_1; \sigma_1^{(s)})$ representing the mean and standard deviation of speed values estimated on the training dataset used to build R_1. With the additional assumption of a Gaussian (normal) model, these two parameters would completely define one estimation of a probability distribution for X_S. A Mixture-of-Gaussian (MoG) model could be used for the conditional likelihoods of the SPEED and LENGTH for instance, as well as more sophisticated techniques of joint density estimation, or models of dynamics of vessels, considering as well the interaction between speed and position (e.g., [46, 38]).

However in some cases, the amount of data (e.g., number of trajectories) building the cluster may not be large enough to estimate reliable distributions and considering second-order uncertainty could be appropriate (see Sections IV-E and IV-F). Also some AIS fields, especially the ones entered manually, are often missing or miss-spelled. For instance, the destination may not be specified or may not be valid, the Estimated Time of Arrival (ETA) may not be updated. The positional and

[1] The field 'TYPE' in Table II corresponds to the following encoding: [T1 T2 T3 T4 T5]=[Cargo Tanker Fishing Passenger Others]



(a) (b)

Fig. 3. Excerpt of the URREF ontology for the EVALUATIONCRITERION classes. UNCERTAINTYTYPE can be used to refine the EXPRESSIVENESS criterion. Displayed with the Protégé software [39]. The full and last version of the ontology is available at eturwg.c4i.gmu.edu/files/ontologies/URREF.owl. (a) Top-level concepts of the URREF ontology. (b) URREF EVALUATIONCRITERION class with subclasses. EXPRESSIVENESS is a subclass of UNCERTAINTYREPRESENTATIONCRITERION. PRECISION and ACCURACY are subclasses of DATAQUALITYCRITERION.

kinematic information being automatically sent is more reliable but can suffer from incompleteness to due a lack of coverage of the AIS receivers resulting in missing reports for a certain period of time. The non-reception of the AIS signal may arise as well from an intentional manipulation, either simply to conceal some activity either legal (e.g., fishing) or illegal (e.g., smuggling), or to keep hidden from pirates. Finally, the AIS signal can be spoofed for instance shifting the positional information to another area, or by modifying the MMSI or IMO identifier of the vessel for instance [44]. Previous studies have demonstrated that roughly 5% of AIS data is generally inconsistent (see e.g. [35]).

The consideration of these different imperfections of information is crucial in the design of maritime anomaly detection solutions. However, it requires a prior proper understanding of the origins of uncertainty, of the kinds of imperfection, of the type of information (be it relevant to a population of situations or to a single one—generic or singular) to provide a meaningful solution and to properly interpret the estimates output by the algorithms and made available to the user. In the following we provide a brief overview of the URREF ontology which aims at capturing assessment criteria on the one hand, and relevant uncertainty-related concepts that impact the solution assessment on the other hand.

III. THE UNCERTAINTY REPRESENTATION AND REASONING EVALUATION FRAMEWORK (URREF)

The URREF ontology [8] identifies, defines and links uncertainty-related concepts which come into play when evaluating the uncertainty representation and reasoning approaches underlying information fusion schemes. As the work is still on-going and some elements are currently under discussion within the ETUR group, this section only provides a partial description of the ontology focusing on the concepts relevant to this paper. The reader is referred to the ETUR working group collaboration website for an up-to-date description of the URREF ontology.[2]

The top level concept THING (see Figure 3(a)) contains concepts such as UNCERTAINTYNATURE (epistemic vs aleatory), UNCERTAINTYTYPE, UNCERTAINTYTHEORY (mathematical framework), UNCERTAINTYDERIVATION (objective vs subjective), SOURCE (of information), EVALUATIONSUBJECT and associated EVALUATIONCRITERION. The EVALUATIONCRITERION class is further split into DATACRITERIA, DATAHANDLINGCRITERION, REPRESENTATIONCRITERION and REASONINGCRITERION classes (see Figure 3(b)).

A. Evaluation subjects

Evaluation subjects are the elements composing the URRT which assessment through the URREF is meaningful [25]. An evaluation subject is any item which can be compared and evaluated through the URREF ontology according to a series of corresponding criteria. The uncertainty representation process (that we denote here by h) corresponds to the abstraction process of modelling [9] and aims at capturing the uncertainty (i.e. imperfection) arising from (in particular but not only):

[2]eturwg.c4i.gmu.edu/files/ontologies/URREF.owl

- the measurements, including the links between the variables, the mapping from the measurement space to the decision space, and finally the uncertainty over the decision space, including the route definition (i.e., the normalcy definition);
- the source quality, either provided by the source itself (i.e., self-confidence) which expresses some doubt about the estimated value or testimony provided, or estimated by the algorithm designer (or user) based on past experience with the source (i.e., reliability). If we relate reliability to the ability of the source to consistently provide correct outputs, then self-confidence and reliability differ in the sense that the source may have a low confidence in its declaration (singular information) while being still highly reliable (generic information), or being highly confident while being always wrong (low reliability).

The uncertainty representation is assessed by the REPRESENTATIONCRITERION of the URREF ontology.

The fusion method builds a series of uncertainty functions over the space Ω, that we split for convenience between the measurement and decision spaces, i.e. $\Omega = \Omega_o \times \Omega_R$, and involves at least one instance of the following elements: (1) a combination function ρ acting over (possibly some subsets of) Ω, (2) a mapping function g from Ω_o to Ω_R, and (3) a decision mapping l from an uncertainty function over Ω_R to a singleton of Ω_R.

The Atomic Decision Procedure (ADP) underlying the fusion method Ψ is thus composed of the elements $\{h,g,\rho,l\}$. The scheme Ψ before decision (l), outputs an uncertainty function over Ω_R representing some belief degrees we may have at time t regarding the different hypotheses of Ω_R, based on a set of pieces of information (either singular measurements received by the sources at t or generic information extracted from historical data or background knowledge and can formally be denoted as:

$$\Psi(\phi(U,S)) = \phi(X_R,\Psi) \tag{1}$$

where $U \subseteq \mathcal{X}$, $S \subseteq \mathcal{S}$ and ϕ is an information item provided by S over U. Equation (1) expresses that Ψ processes some pieces of information defined over a subset U of variables from $\mathcal{X}$, provided by a subset S of sources $\mathcal{S}$ and including some uncertainty, and outputs another piece of information defined over Ω_R, then provided by ψ as a source. As we will illustrate in Section IV, the order of the elements of Ψ is not fixed, since the fusion operation ρ can be performed within different subsets of Ω or Ω_o (e.g., URRT#1, URRT#2, URRT#3) or solely within Ω_R (e.g., URRT#4, URRT#6), the fusion can occur after the decision step (e.g., majority vote in classifier combination), etc. The reasoning scheme is assessed by the REASONINGCRITERIA of the URREF ontology.

B. Information deficiencies

In the current state of the URREF ontology, some information quality dimensions are covered by the UNCERTAINTYTYPE class (Ambiguity, Incompleteness, Vagueness, Randomness, Inconsistency). Alternative categorisations of information deficiencies could be considered instead, such as either Smets' structured thesaurus of imperfection of information [51], either Klir and Yuan's typology [30], or the typology of defects of information of Dubois and Prade [19]. In this paper, we will refer to the later one, and following the authors we will distinguish between the four information defects of *incompleteness* (or yet imprecision), *uncertainty*, *graduality*, and *granularity*.

- *Imprecision*—Refers a set of possible values, regardless how they have been obtained: The bigger the size of the set, the higher the imprecision. It represents the inability of the source to provide a single value or to discriminate between several values. Imprecision is interpreted as a type of incompleteness as it arises from a lack of information. For instance, the statement "The vessel is following either route R_1 or R_2" is imprecise and provides only incomplete information not allowing to answer the question "What is the route followed by the vessel?".
- *Uncertainty*—Arises when an agent does not know (or partially knows) if a proposition is true or false. It can be expressed by a degree (or a set of degrees) of confidence assigned to a specific (or set of) value(s) to be "true." Its nature can either correspond to a lack of knowledge (epistemic uncertainty) or to the variability of an underlying process (aleatory uncertainty). When assigned by the source itself it corresponds to "self-confidence." Uncertainty can also be expressed at the output of the fusion process itself with an equivalent interpretation, meaning that the fusion process does not provide a maximal confidence toward its output. For instance, the probability distribution over the set of possible types of vessels C_i as output by some classifier can be interpreted as an uncertainty expression, i.e., expressing a set of (normalised) degrees of confidence in the truth of the proposition "The vessel is of class C_i."
- *Graduality (or gradualness)*—Arises usually from linguistic expressions and induces propositions with some possible degrees of truth (i.e., non Boolean). That kind of imperfection allows a proposition to be more or less true or false. For instance, "The vessel is fast" is a gradual information item, using the *gradual predicate* "fast," and is typically represented by fuzzy sets. As we will illustrate later, "on-route" can be considered as a gradual predicate making maritime routes ill-defined objects.
- *Granularity*—Refers to the support over which the proposition is defined, i.e. to the set of pre-established possible values. Granularity refers to the partition granules used in the definition of a set. For instance,

the set $\Omega_1 = \{$FISHING VESSEL; NOT FISHING VESSEL$\}$ describing exhaustively the types of vessels has a rougher granularity than the set $\Omega_2 = \{$FISHING VESSEL; CARGOS; TANKERS; OTHERS$\}$ covering also exhaustively the possible types of vessels. The change of granularity is done through the operations of refinement or coarsening (see for instance [47]).

Not a defect *per se*, we also consider the dimension of "trueness" (vs falseness):

- *Trueness*—It is considered here as the criterion relating a piece of information (either input or output) to truth or to some reference value. It is defined in [22] as the "closeness of agreement between the expectation of a test result or a measurement result and a true value." The notion of trueness covers two different aspects that are *how close* the results are to truth, especially in case on measurements on continuous scales or *how frequently* the results correspond to truth, especially in case of nominal scales such as output of classifiers.

Usually, on the one hand, imprecision (or precision) and uncertainty (or certainty) are opposed [51]: "I'm certain that the speed of the vessel is between 3 and 6 knots" (Imprecise but certain statement) versus "I'm not certain that the speed of the vessel is 5 knots" (Precise but uncertain statement). On the other hand, precision and trueness are often associated in performance assessment of systems, and are gathered under the term *accuracy* in ISO 5725 [22], referring to a series of independent tests. The way these information deficiencies relate to the concepts of UNCERTAINTYTYPE, UNCERTAINTYDERIVATION and DATACRITERIA is still under discussion within the ETUR working group and is not addressed in this paper.

These five deficiencies (or imperfections) introduced above will be used in the following to characterise both input and output information of the fusion method. We will denote in the following by η the imperfection to be captured by the uncertainty representation process h.

C. Type of information

Following [19], we distinguish between *generic* and *singular* information. Generic information refers to a population of situations such as statistical models, physical rules, logical rules or commonsense knowledge. It is a synthesis of previous knowledge. Singular information is about the current state of the world such as an observation, a testimony or a sensor measurement. This distinction is similar to the one sometimes made between knowledge and evidence: According to Pearl (as cited in [16]) knowledge is understood as "judgments about the general tendency of things to happen," whereas evidence refers to the description of a specific situation."

Therefore, as a matter of convention in this paper, the notions of data, knowledge, evidence and information are all covered by the single term *information*. This is driven only by the need to avoid confusion between the terms and by no means to deny any existence of distinction between these notions. Consequently, "incomplete knowledge," "uncertain evidence," "erroneous data," etc, are all covered by the general term "*imperfect information*."

Moreover, we reserve the term *uncertainty* to the definition introduced in Section III-B. Indeed, *uncertainty* may be used sometimes abusively to cover the different types of imperfection (or information defects) as they all induce some uncertainty in the decision maker's mind. Note that *uncertainty* is also considered as the dual of *information* as classically understood in the field of Generalised Information Theory (GIT) [28]: To some increase of information corresponds equivalent reduction of uncertainty, as captured for instance by Shannon entropy measure. Hence, instead of uncertainty, we rather use the general terms of *imperfect information*, *imperfection*, *information defects*, *information deficiencies*, uncertainty being one of them.

D. Uncertainty theory

The UNCERTAINTYTHEORY class contains the mathematical theories for the representing and reasoning with uncertainty. It typically includes, but is not limited to, probability theory, fuzzy set theory, possibility theory, belief function theory, rough set theory, imprecise probability theory (see [19] for a survey). In the following, we will consider the three mathematical frameworks of probabilities, belief functions and fuzzy sets. Although geometry is not traditionally considered as an uncertainty theory, contrary to probabilities, belief functions or fuzzy sets, we also provide in this section the description of distance measures together with some justifications for its consideration.

Let us denote by Ω a set of hypotheses which could correspond either to the joint space ($\Omega_o \times \Omega_R$), either to the measurement space only Ω_o, either to the decision space only Ω_R or to any other subset of it.

A Probability Mass Function (PMF) p satisfies the following properties:

(p.1) $p : \Omega \to [0; 1]$
(p.2) $\sum_{x \in \Omega} p(x) = 1$

A probability measure P satisfies the following properties:

(P.1) $P : 2^{\Omega} \to [0; 1]$
(P.2) $P(\emptyset) = 0$ and $P(\Omega) = 1$
(P.3) $P(A) = \sum_{x \in A} p(x), \quad \forall A \subseteq \Omega$ and p the PMF
(P.4) $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$

We have that $P(\{x\}) = p(x)$. The additivity property (P.4) constrains in particular $P(A) + P(\bar{A}) = 1$, if $\bar{A}$ denotes the negation (or complement) of A, i.e., $\bar{A} = \Omega \backslash A$.

A state of complete ignorance about the value of x is usually represented by a uniform distribution over Ω such that $P(x) = 1/|\Omega|$, for all $x \in \Omega$, where $|.|$ denotes the cardinality. The additivity property is what distinguishes probability measures from other non-additive measures such as belief functions.

Dempster-Shafer theory, or evidence theory, or belief function theory [12, 47], is often described as an extension of probability theory in which the axiom of additivity is relaxed for an axiom of sub-additivity on belief functions. In other words, the underlying distribution of a belief function is no longer defined over the singletons of Ω but rather over its powerset 2^Ω.

A Basic Probability (or Belief) Assignment (BPA or BBA) is a function m such that

(m.1) $m : 2^\Omega \rightarrow [0;1]$
(m.2) $\sum_{A\subseteq\Omega} m(A) = 1$
(m.3) Closed-world assumption: $m(\emptyset) = 0$ OR Open-world assumption: $m(\emptyset) \neq 0$

A belief function is a function Bel such that:

(Bel.1) $\text{Bel} : 2^\Omega \rightarrow [0;1]$
(Bel.2) $\text{Bel}(\emptyset) = 0$ and $\text{Bel}(\Omega) = 1$
(Bel.3) $\text{Bel}(A) = \sum_{B\subseteq A} m(B), \quad \forall A \subseteq \Omega.$
(Bel.4) $\text{Bel}(A\cup B) \leq \text{Bel}(A) + \text{Bel}(B)$ for all $A, B \subseteq \Omega$ such that $A \cap B = \emptyset$

A plausibility function is a function Pl such that:

(Pl.1) $\text{Pl} : 2^\Omega \rightarrow [0;1]$
(Pl.2) $\text{Pl}(\emptyset) = 0$ and $\text{Bel}(\Omega) = 1$
(Pl.3) $\text{Pl}(A) = \sum_{B\cap A\neq\emptyset} m(B), \quad \forall A \subseteq \Omega.$
(Pl.4) $\text{Pl}(A\cup B) \geq \text{Pl}(A) + \text{Pl}(B)$ for all $A, B \subseteq \Omega$ such that $A \cap B = \emptyset$

The belief function Bel and plausibility function Pl are thus respectively sub-additive ($\text{Bel}(A) + \text{Bel}(\bar{A}) \leq 1$) and super-additive ($\text{Pl}(A) + \text{Pl}(\bar{A}) \geq 1$). The uncertainty functions Bel and Pl are dual of each others ($\text{Bel}(A) = 1 - \text{Pl}(\bar{A})$) and can be interpreted (under Dempster's statistical view [12]) as respectively lower and upper bounds of an (unknown) probability of A: $\text{Bel}(A) \leq P(A) \leq \text{Pl}(A)$, $\forall A \subseteq \Omega$. The open-world assumption [54] relaxes the exhaustivity of the original Dempster-Shafer model, allowing the empty set to have a non-null mass. That means that other hypotheses than the ones initially considered in Ω can actually be true. It is interesting in our practical case of route association as this empty set would then act as a rejection class for "off-route" vessels (see Section IV-F). Evidence theory "includes extensions of probabilistic notions (conditioning, marginalisation) and set-theoretic notions (intersection, union, inclusion, etc.)" [13]. The conjunctive rule is based on the intersection between sets (see (10)). A non-null mass to the empty set denotes thus a conflict (or inconsistency) between the two belief functions combined and may be interpreted as an indicator to an anomaly. A state of complete ignorance is represented by the vacuous BPA $m(\Omega) = 1$ (or equivalently by $[\text{Bel}(A);\text{Pl}(A)] = [0;1]$ for all $A \subseteq \Omega$, $A \neq \emptyset$ and $A \neq \Omega$), which is distinct from the uniform distribution.

A fuzzy set μ satisfies the following properties [57]:

(f.1) $\mu : \Omega \rightarrow [0;1]$
(f.2) $\max_x \mu(x) = 1$
(f.3) $\mu(A\cup B) = \max(\mu(A), \mu(B))$
(f.4) $\mu(A\cap B) = \min(\mu(A), \mu(B))$

Compared to probabilities and belief functions which define degrees of belief regarding the occurrence (or truth) of an event, being itself either true or false, fuzzy sets define degrees of truth for events which are thus allowed to be more or less true.

Geometric distances are not an uncertainty model *per se*. However, they are at the basis of the computation of trueness, precision or accuracy in measurement data (e.g. [22]) which all convey notions of uncertainty. Moreover, pattern matching techniques (see Sections IV-A and IV-B) rely on distances computation. Finally, uncertainty may be derived from distance measures as the farther to a route the vessel, the higher our uncertainty that it follows that route (see Section IV-D). For these reasons we include here the basic properties of distance measures.

A (metric) distance function d satisfies the following properties:

(d.1) $d : \Omega \times \Omega \rightarrow [0;1]$
(d.2) $0 \leq d(x_1, x_2) \leq 1$
(d.3) $d(x_1, x_2) = d(x_2, x_1)$
(d.4) $d(x, x) = 0$
(d.5) $d(x_1, x_2) = 0 \Rightarrow x_1 = x_2$
(d.6) $d(x_1, x_2) \leq d(x_1, x_3) + d(x_3, x_2)$

All these properties define *metric distances*, but relaxing some of them lead to weaker forms of distances such as pseudo-metrics or semi-metrics. The properties of the functions introduced here correspond to some desirable behaviours of the uncertainty handling models within the fusion method to be designed. One of the tasks of the designer is to identify and select the uncertainty representation together with the associated mathematical framework in order to meet the requirements of the expected underlying logic of the method. To sum up, and referring to the basic information quality dimensions identified in Section III-B, probabilities convey the notion uncertainty only, belief functions convey both uncertainty and imprecision, while fuzzy sets convey the notion of graduality which can be assessed using distance measures.

E. Uncertainty supports

In order to refine the assessment of uncertainty representations, we introduce the concept of *uncertainty support* as an item about which some uncertainty (or

TABLE III
Examples of pieces of information for different uncertainty supports and generic and singular information

Uncertainty Support		Generic uncertainty ("0")	Singular uncertainty ("t")
(us.1)	X_i	$\eta(X_i, s, 0)$ *Uncertainty about the type of vessels deduced from past AIS records*	$\eta(X_i, s, t)$ *Uncertainty about the type of a specific vessel, as provided by an ATR classifying a SAR imagery*
	X_R	$\eta(X_R, s, 0)$ *Prior uncertainty about the routes followed*	$\eta(X_R, s, t)$ *Uncertainty about the route followed by the vessel at t*
(us.2)	(X_i, X_j)	$\eta((X_i, X_j), s, 0)$ *Uncertainty linking type and speed of vessels, in general*	$\eta((X_i, X_j), s, t)$ *Uncertainty about the type of the vessel given the current speed*
	(X_i, X_R)	$\eta((X_i, X_R), s, 0)$ *Uncertainty linking the speed and the route*	$\eta((X_i, X_R), s, t)$ *Uncertainty about the route followed by a specific vessel at t given its speed*
(us.3)	$\eta(., 0)$	$\eta(\eta(., s_2, 0), s_1, 0)$ *Uncertainty about the prior distribution over the routes as lower and upper bounds*	$\eta(\eta(., s_2, 0), s_1, t)$ *Uncertainty at t about the routes previously extracted*
	$\eta(., t)$	$\eta(\eta(., s_2, t), s_1, 0)$ *Uncertainty about the source s_2 declaration provided at t (e.g., prior reliability)*	$\eta(\eta(., s_2, t), s_1, t)$ *Uncertainty about the current source s_2 statement itself including some uncertainty*

imperfection in general) needs to be captured and represented (in other words, what "we are uncertain about") and distinguish between:

(us.1) *Individual states of the world* as represented by any single variable of $\mathcal{X}$

(us.2) *links between states* as represented by subsets of variables from $\mathcal{X}$

(us.3) *uncertainty expression* η over the above supports (us.1) or (us.2).

Supports (us.1) are a special case of (us.2). The supports of type (us.3) correspond to abstract states covering for instance uncertainty or imprecision about a probability distribution, about a probabilistic model linking several variables, etc. The joint distribution of length and types of vessels can be itself the support of some uncertainty or imprecision since its estimation may not reflect the real distribution (due to a lack of data for instance). This is a HigherOrderUncertainty (i.e. second-order uncertainty), which a subclass of Expressiveness criteria captured in the URREF ontology under the RepresentationCriterion class (see Figure 3(b)). Advantages of considering second-order uncertainty are for instance discussed in [45, 34, 7, 2]. One of the purposes of the URREF is to analyse and capture these features of second-order uncertainty.

Table III lists examples of uncertainty supports (for both generic and singular information) together with the notation and meaning.

The examples of uncertainty supports provided in Table III are for two variables only, although these cover any subsets of variables. To distinguish between generic and singular information, we will use the indexes 0 and t respectively to the corresponding uncertainty supports. Moreover, we assign the symbol of the information source s from which the imperfection has to be captured. For instance:

$$\eta(X_T, \text{AIS dataset}, 0)$$

denotes the imperfection of the type of vessels observed in the past pertaining to the AIS dataset of interest.

In Section V-A, the URRTs will be compared according to their ability to capture the different imperfection types of our problem at hand as exemplified by Table III.

F. Evaluation criteria

We will focus in Section V-A on the Expressiveness criterion of the RepresentationCriterion class of the URREF ontology. Expressiveness is defined as the power of an uncertainty representation technique to convey relevant aspects of a given fusion problem [8]. The uncertainty supports are a "relevant aspect" of the problem as they are able to convey the idea of Dependency (between variables), Higher-order uncertainty, (source) Self-confidence and extend to the source's reliability. Note that this assessment along the expressiveness criterion is not be ordinal in the sense that the methods are not be ordered according to their expressiveness. It is rather a comparative assessment where the methods are characterised according to their expressiveness. Instead of establishing some ranking of the URRTs, the expressiveness assessment is aimed at improving the understanding of the semantics of the different approaches. We further expand the Expressiveness criterion to cover the ability of the URRTs to capture the different types of imperfection as defined in Section III-B.[3] Figure 3(b) displays the EvaluationCriterion class split into RepresentationCriterion, ReasoningCriterion, DataCriterion and DataHandlingCriterion. Expressiveness is a subclass of RepresentationCriterion, having itself other subclasses such as HigherOrderUncertainty or Dependency.

Additionally in Section V-B, we also assess the URRTs globally on their outputs through the DataCriterionQuality. Notions of Trueness (or Falseness), Imprecision and Uncertainty are quantitatively evaluated when the fusion scheme solving the route association problem is implemented processing real AIS data.

IV. UNCERTAINTY REPRESENTATION AND REASONING TECHNIQUES

Six different uncertainty representation and reasoning techniques (URRTs) are presented below, as six instantiations of the fusion scheme Ψ, to be further assessed through the URREF. The URRTs presented here are very basic and simple schemes far less complete than the ones reported in the literature addressing the problems of maritime anomaly detection or route association. However, this deliberate simple exposure is aimed at "*dissecting*" the underlying uncertainty representation and reasoning, as a first step for comparison and improved understanding.

A. URRT#1: Pattern matching—Euclidean

Intuitively, the closer the observed vessel under consideration is to the centroid of the routes, the more likely it is to belong to the route. A pattern matching approach captures this basic reasoning. Prototype matching differs from template matching (such as 1-nearest neighbour) in a way that a perfect match is not expected. It provides better flexibility and allows some tolerance to handle uncertainty. A standard pattern (prototype) matching approach computes the Euclidean distances between $\mathbf{x}$ and each of the routes of Ω_R as:

$$d^{(E)}(\mathbf{x}, R_k) = \sqrt{(\mathbf{x} - \mathbf{r}^{(k)})'(\mathbf{x} - \mathbf{r}^{(k)})} = \sqrt{\sum_{i \in \mathcal{A}} (x_i - r_i^{(k)})^2} \quad (2)$$

where $\mathbf{r}^{(k)}$ is the prototype corresponding to route R_k (see Table II), defined in the feature space Ω and $\mathbf{x}'$ is the transpose vector of $\mathbf{x}$. The ith components of $\mathbf{x}$ and $\mathbf{r}^{(k)}$ are denoted by x_i and $r_i^{(k)}$ respectively. The quantity $(d_i^{(k)})^2 = (x_i - r_i^{(k)})^2$ can be interpreted as an inverse *degree of match* of the observation x_i to the equivalent prototypical element of R_k, that we denote as $r_i^{(k)}$: The lower the square distance, the higher the degree of membership of the vessel to that route. Let us define by $\mu_i^{(k)}$ the degree of membership of $\mathbf{x}$ to R_k according its feature x_i. Then, adopting a similarity view of fuzzy sets [3, 18], $\mu_i^{(k)}$ can be defined through $d_i^{(k)}$ as, for instance:

$$\mu_i^{(k)} = \exp(-(d_i^{(k)})^2)$$

which tends toward 0 whenever the distance tends toward infinity and equals to 1 if the distance is null. Equation (2) can then be written as:

$$d^{(E)}(\mathbf{x}, R_k) = \sqrt{-\sum_{i \in \mathcal{A}} \log(\mu_i^{(k)})} \quad (3)$$

where $\mu_i^{(k)} \in]0;1]$ is a normalised degree of membership. Eq. (3) is a bisymmetrical continuous strictly monotonous mean [6]. The fusion operator in (2) is a sum (disjunction) which averages local dissimilarities with R_k along the different features. It acts as a compromise between min (conjunctive) and max (disjunctive) operators.

We then consider the following decision rule:

$$\hat{R} = \begin{cases} \arg\min_k d^{(E)}(\mathbf{x}, R_k) & \text{if } d^{(E)}(\mathbf{x}, R_k) < \epsilon_1 \\ R_0 & \text{otherwise} \end{cases} \quad (4)$$

where ϵ_1 is a threshold to be set according to the operator's needs or expectations, representing some tolerance over the global distance over the 5 features. In practice, ϵ_1 can be deduced from some aggregation of the individual thresholds ϵ_1^i for each feature. This decision rule allows some imprecision in the decision space as it can lead to a set of possible routes, without identifying a single one. An anomaly is detected if it does not match any route. Many anomaly detection approaches are based on distances computation as an implementation of the notion of "closeness to normalcy" (e.g. [11]). Semantic distances can also be used to assess the different meanings between attributes (e.g. [4]).

B. URRT#2: Pattern matching—Mahalanobis

A modified version of the Euclidean pattern (prototype) matching scheme is obtained by using the Mahalanobis distance:

$$d^{(M)}(\mathbf{x}, R_k) = \sqrt{(\mathbf{x} - \mathbf{r}^{(k)})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mathbf{r}^{(k)})} \quad (5)$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of the random vector $\mathbf{X}$ associated to $\mathbf{x}$, whose coordinates are r.v. X_is. The superscript $^{-1}$ denotes the inverse matrix. The element $\sigma_{i,j}$ of $\boldsymbol{\Sigma}$ is the covariance of X_i and X_j defined as $E(X_i, X_j) - E(X_i)E(X_j)$ where E is the expectation operator such that $E(X) = \sum x p(X = x)$ for a discrete random variable X. The same decision rule (4) than for the Euclidean pattern matching is used. However, another threshold ϵ_2 must be used instead of ϵ_1, based on the covariance matrix.

As in (2), the fusion operator in (5) is a disjunction but including weights which would discount the local individual dissimilarities relatively to the variance of their corresponding feature, and pairs of errors relatively to their covariance.

[3]Note that this link between Expressiveness and UncertaintyType is not currently implemented in the URREF ontology and is at a stage of proposal for inclusion.

The Euclidean and Mahalanobis distances in (2) and (5) are well suited to features defined over numerical and continuous scales while they reduce to logical AND for nominal variables such as the type. Better suitable distance measures are usually used based on the aggregation of individual for each feature, possibly using different definitions than the square difference (e.g. [21]). Other distances such as the log-normal probability density (e.g. [1]) would account for the routes statistics as well. Mahalanobis distance is used in [36] to associate vessel tracks to maritime routes.

C. URRT#3: Probability-based—Bayesian

In the standard Bayesian approach to fusion, the function $p(\mathbf{X} = \mathbf{x} \mid R_k)$ represents the likelihood of observing a specific set of values $\mathbf{x}$ on a given route R_k, and is usually derived from past observations used to compute the routes. The different observations are combined following Bayes' rule:

$$P(R_k \mid \mathbf{x}) \propto p(R_k) \prod_{i \in \mathcal{A}} p(x_i \mid R_k), \quad \forall R_k \in \Omega_R \tag{6}$$

under the assumption of *independent and identically distributed observations*. $p(R_k)$ is some prior probability that the vessel follows a specific route. The resulting posterior probability $P(R_k \mid \mathbf{x})$ represents some belief that the route followed by the vessel of interest is R_k given that we currently observe $\mathbf{x}$. A normalisation factor ensures that a probability distribution is obtained. Equation (6) is known as Naïve Bayes model in classification. This combination rule (6) can be written using the individual posterior probabilities as $P(R_k \mid \mathbf{x}) \propto p(R_k)^{-(|\Omega_R|-1)} \prod_i p(R_k \mid x_i) p(x_i)$. The decision rule is the Maximum A Posteriori (MAP) probability:

$$\hat{R} = \begin{cases} \arg\max_k p(R_k \mid \mathbf{x}) & \text{if } p(R_k \mid \mathbf{x}) > \epsilon_3 \\ R_0 & \text{otherwise} \end{cases} \tag{7}$$

where ϵ_3 is a threshold: if the posterior probability is too uniformally distributed among the routes, then no clear matching is detected and an anomaly is returned. The Bayesian reasoning scheme is at the basis of the Bayesian network approach proposed for instance in [23].

The fusion operator is a conjunctive operator, i.e. the product of individual likelihoods. It has the property of decreasing very fast to 0 as the number of features to be combined increases. Also, the result is exactly 0 if only one likelihood is null.

D. URRT#4: Probability-based—Non-Bayesian

In a still probabilistic but non-Bayesian approach, each measured feature is considered providing some evidence about the membership of $\mathbf{x}$ to a given route R_k. For instance, $p_s(R_k) = p(R_k \mid x_s)$ is the contribution of the speed observation to the membership of the vessel V to R_k and is interpreted as the probability that V belongs to R_k given (or according to) the estimated speed. Then, the observations are aggregated by a weighted sum as:

$$p(R_k \mid \mathbf{x}) = \sum_{i \in \mathcal{A}} \alpha_i p(R_k \mid x_i), \quad \forall R_k \in \Omega_R \tag{8}$$

where $\alpha_i \in [0,1]$ is a weight reflecting either the confidence in the soft decision values computed by the individual sources, and possibly be deduced from $p(x_i)$, or the relevance of the features to the fusion problem (for instance, the position and heading may be given a higher weight than the type). This rule is derived in [27] from (6) under the assumption of uniform $p(R_k)$. Contrary to the Bayesian approach, the posteriors are combined. The decision rule is then (7).

The fusion operator is a disjunctive operator, as in (2) and (5), but probabilities are combined rather than distances.

E. URRT#5: Transferable Belief Model (TBM) model-based

The reasoning scheme considered here is the one proposed in [45, 14] within the Transferable Belief Model (TBM) framework and making use of the Generalised Bayes Theorem (GBT) [50] as the combination rule, given by the following plausibility measure for a subset of routes A:

$$\text{Pl}(A \mid \mathbf{x}) = 1 - \prod_{R_k \in A} (1 - \text{Pl}(\mathbf{x} \mid R_k)), \quad \forall A \subseteq \Omega_R \tag{9}$$

where $\text{Pl}(A) = \sum_{A \cap B \neq \emptyset} m(B)$ is the plausibility of $A \subseteq \Omega_R$, with m being a Basic Belief Assignment (see Section III-D). $\text{Pl}(A \mid \mathbf{x})$ is the conditional plausibility of A and is interpreted as the maximum confidence that can be assigned to A (i.e., that the route followed belongs to the subset A) given that $\mathbf{x}$ has been observed. As proposed in [45], $\text{Pl}(\mathbf{x} \mid R_k)$ is the least committed plausibility function corresponding to the probabilistic likelihood function considered as the pignistic probability. For a BBA m, the pignistic probability [49] is defined for any singleton of Ω_R as $\text{BetP}(R_k) = \sum_{R_k \in A} m(A)/|A|$. As introduced in Section III-D, pairs of plausibility and belief values can be interpreted as intervals over the probability of any subset of routes $A \subseteq \Omega_R$. However, if we restrict to singletons only, (9) reduces to the product of plausibility under the independence assumption. The decision rule requires then two steps: (a) the transformation of the Pl measure into a probability distribution over Ω_R (e.g. the pignistic probability) such that (b) the MAP rule (7) can be applied (with the appropriate threshold).

The fusion operator is again a conjunctive operator with similar properties than the ones described in Section IV-C.

F. URRT#6: Belief functions—Database query

Similarly to the probabilistic non-Bayesian URRT#4, each observed feature x_i of $\mathbf{x}$ is assumed to provide some evidence about route R_k being followed by V. The uncertainty is modeled by belief functions rather than probabilities. Each observation x_i is regarded as a query to Ω_R such that only the items (i.e. routes) satisfying the associated criterion are retrieved, to form a set of possible routes A_i according to x_i. A_i is the subset of routes satisfying the query x_i:

$$A_i = \{R \in \Omega_R \mid x_i \in \Omega_i\}$$

For instance, A_1 is the set of routes compliant with a measured speed of 5 knots. The multivalued mapping between the observation space Ω and decision space Ω_R assigns to any singleton of Ω a subset of Ω_R. Let us consider that some singular information about the observation x_i under the form of a probability, provided for instance by a classifier: $p^{(T)}(X_T = \text{Cargo}) = 0.4$ is the probability that the observed vessel is a Cargo type, as estimated based on current observations. Let $\mathbf{p}_T = [0.4\ 0.3\ 0\ 0.3\ 0]'$ be the uncertainty of the classifier (source) expressed as a probability distribution about the type of the vessel. This uncertainty is transferred to the corresponding subsets of Ω_R previously defined by the multivalued mapping, defining thus a BBA m_i over Ω_R, where the numerical weight $m_i(A_i) = p_i(x_i)$ is interpreted as the degree of belief that can be assigned to A_i and to none other subset of A_i. Then, $A_{\text{Cargo}} = (R_1, R_2, R_3, R_4, R_5)$ is the set of routes possibly followed by cargo vessels and is assigned a weight of 0.4. Equivalently, $A_{\text{Tanker}} = (R_2, R_3, R_5)$ and $m(A_{\text{Tanker}}) = 0.3$ and $A_{\text{Passenger}} = (R_2, R_5)$ and $m(A_{\text{Passenger}}) = 0.3$. This multivalued mapping does not induce a probability distribution over Ω_R but a BBA.

The resulting BBA m over Ω_R is obtained by combining the individual contributions of each feature by the conjunctive rule, where weights are assigned to conjunctions of sets of routes A_i and A_j:

$$m(A) = \sum_{A_i \cap A_j = A} m_i(A_i) m_j(A_j), \quad \forall A \subseteq \Omega_R \tag{10}$$

The rule (10) defines a conjunctive fusion based on the intersection between sets. The decision rule is similar to (7) but considers the conflict measure as a criterion for anomaly:

$$\hat{R} = \begin{cases} \arg\max_k \text{BetP}(R_k) & \text{if } m(\emptyset) < \beta \\ R_0 & \text{otherwise} \end{cases} \tag{11}$$

where BetP is the pignistic transformation of m. The quantity $m(\emptyset)$ is the BBA of the empty set after combination and represents the global weight of conflict between all the sources (or features).

V. ASSESSMENT OF URRTS

We now characterise the different approaches previously described through the URREF and its associated ontology, Expressiveness, in Section V-A and output Quality criteria in Section V-B.

A. Expressiveness assessment

Table IV summarises the comparative description of the 6 URRTs presented in Section IV as candidate solutions to the same problem of maritime route detection. The expressiveness of the URRTs relatively to different uncertainty supports identified in Section III-E is first assessed in a binary way, so that an empty cell means that the technique (as actually defined in the previous section) does not account for the uncertainty on the corresponding support. The types of imperfection (*graduality*, *uncertainty*, *imprecision*) are mentioned in case the URR technique captures them, together with the corresponding notation. The *granularity* is kept constant for all the methods and is just reported as the list of possible values for all variables in the first rows. In the third part of the table the reasoning schemes are compared along their respective uncertainty representation, marginalisation, decision elements.

1) URRTs analysis:

The uncertainty supports introduced in Section III-E are mentioned for each method in Table IV. We thus refer the reader to Table IV for details on the uncertainty supports about the URRTs analysis.

URRT#1—We observe that the standard pattern matching approach (URRT#1) does not account for many uncertainty supports: The route representation is considered as precise and certain since the prototypes are defined by single values (either the mean, or the mode for the type); the dependency between variables is not considered, nor is the possible links between routes; sources' uncertainty (or self-confidence) about their singular declaration at t is not considered; sources' reliability is not represented, nor is any second-order uncertainty. URRT#1 captures a single imperfection type as a notion of *graduality* through a distance measure, the route prototype being considered as a reference: the distance to route can be interpreted as a degree of membership of $\mathbf{x}$ to R_k. From this generic information, a singular imperfection is further derived as $\eta(X_R, t)$ combining with the observation of the vessel at t. The fusion is performed through the distance definition by a sum operator acting as an average of inverse of similarities along the different features of $\mathcal{A}$: The higher the local similarities, the lower the global distance and the higher the membership of $\mathbf{x}$ to R_k.

URRT#2—The extension of URRT#1 using the Mahalanobis distance as described by URRT#2, accounts for both the spread of the routes along the different features (through the individual standard deviations σ_is) and the dependency between variables (through the covariances $\sigma_{i,j}$s). The variance can be interpreted as a measure of *imprecision* regarding X_i. The covariance describes how the variables vary with each other, measures the dependency between them, and expresses then some statistical *uncertainty* on the link between X_i and X_j. Compared to URRT#1, URRT#2 considers some imperfection about the reference objects (the routes). Still,

TABLE VI
Expressiveness comparison of Uncertainty Representation and Reasoning Techniques based on Uncertainty Support.

	URRT Name	URRT#1	URRT#2	URRT#3	URRT#4	URRT#5	URRT#6
	Mathematical framework	Geometry	Geometry, Statistics	Probability, Bayesian	Probability, non-Bayesian	Evidence theory, TBM	Evidence theory
Domains	Ω_1-Latitude	$\{[43.2; 43.4], \ldots,]44; 44.2]\}$					
	Ω_2-Longitude	$\{[9.3; 9.5], \ldots,]10.1; 10.3]\}$					
	Ω_3-Speed	$\{[2.0; 3.0], \ldots,]33; 34]\}$					
	Ω_4-Heading	$\{[0; 20], \ldots,]340; 360]\}$					
	Ω_5-Type	{Cargo; Tanker; Fishing; Passenger; Other}					
	Ω_R-Route	$\{R_0, R_1, \ldots, R_8\}$					
	UNCERTAINTY REPRESENTATION (EXPRESSIVENESS)						
Uncertainty supports (see Table III)	X_i		Imprecision $\eta(X_i^0) = E(\{X_i^0\}^2)$				Uncertainty $\eta(X_i^t) = p_i(X_i^t)$
	X_R			Uncertainty $\eta(X_R^0) = p(R_k)$			Uncertainty + Imprecision $\eta(X_R^t) = m_i(A)$
	(X_i, X_j)		Uncertainty $\eta(X_i^0, X_j^0) = E(X_i^0, X_j^0)$				
	(X_i, X_R)	Graduality $\eta(X_i^0, X_R^0) = \mu_i^{(k)}$	Graduality $\eta(X_i^0, X_R^0) = \mu_i^{(k)}$	Uncertainty $\eta(X_i^0, X_R^0) = p(x_i\|R_k)$	Uncertainty $\eta(X_i^0, X_R^0) = p(x_i\|R_k)$	Uncertainty $\eta(X_i^0, X_R^0) = \mathrm{Pl}(x_i\|R_k)$	Imprecision $\eta(X_i^0, X_R^0) = A \subseteq \Omega_R$
	η^t				Uncertainty $\eta(\eta^t) = \omega_i$		
	η^0					Imprecision $\eta(\eta^0) = [\mathrm{Bel}_i(A); \mathrm{Pl}_i(A)]$	Imprecision $\eta(\eta^0) = [\mathrm{Bel}_i(A); \mathrm{Pl}_i(A)]$
	REASONING						
ADP elements	h	$\eta(X_i^0, X_R^0)$	$\eta(X_i^0, X_R^0)$ $\eta(X_i^0)$ $\eta(X_i^0, X_j^0)$	$\eta(X_i^0, X_R^0)$ $\eta(X_R^0)$	$\eta(X_i^0, X_R^0)$ $\eta(\eta^t)$	$\eta(X_i^0, X_R^0)$ $\eta(\eta^0)$	$\eta(X_i^0, X_R^0)$ $\eta(X_i^t)$ $\eta(X_R^t)$ $\eta(\eta^0)$
	ρ	Ω, $\sqrt{\sum_i (.)^2}$	Ω, $\sqrt{\sum_{i,j} \sigma_{i,j} (.)^2}$	Ω, $\prod_i (.)$	Ω_R, $\sum_i w_i(.)$	Ω_R, $\prod_i (.)$	Ω_R, $(\cap, \sum, \prod_i)$
	$\downarrow \eta^t(X_R)$	Graduality $d^{(E)}(\mathbf{x}, \mathbf{r}^{(k)})$	Graduality $d^{(M)}(\mathbf{x}, \mathbf{r}^{(k)})$	Uncertainty $p(R_k\|\mathbf{x})$	Uncertainty $p_{\mathcal{A}}(R_k)$	Uncertainty + Imprecision $\mathrm{Pl}(R_k\|\mathbf{x})$	Uncertainty + Imprecision $m_{\mathcal{A}}(A)$, $A \subseteq \Theta$
	l	$\min d + \epsilon_1$	$\min d + \epsilon_2$	$\max p(R_k\|x_i) + \epsilon_3$	$\max p_{\mathcal{A}}(R_k) + \epsilon_4$	$\max \mathrm{BetP}(R_k) + \epsilon_5$	$\max \mathrm{BetP}(R_k) + \epsilon_6$ on $m(\varnothing)$

there is no consideration of singular uncertainty about the observations at t, excepted the *graduality* measured by the distance to the prototype route.

URRT#3—As in URRT#1, the independence assumption between variables applies to the Bayesian approach presented in URRT#3. No consideration for either the source's reliability nor self-confidence and the measurement itself is assumed both certain and precise by the source. Rather *uncertainty* is considered over the mapping between Ω and Ω_R where the likelihoods $p(x_i \mid R_k)$ describe how likely it is to obtain some specific measurement given that the vessel follows route R_k. Prior *uncertainty* about routes is explicitly considered by $p(R_k)$ which could be based on other contextual information such as meteorological or seasonal. The fusion is done through a product operator which has the drawback of decreasing very rapidly to 0 once one of the likelihoods is very low. This rule is named "severe" for that reason [27], since it is very sensible to one source's negative opinion. The product is a conjunctive operator (corresponding to a logical AND) making the underlying assumption either that all the measurements are correct, or that all the sources are reliable. Although the independence assumption between features is in our case wrong, this naive Bayesian fusion rule is however shown to provide good (accurate) results. This can be explained by the randomness of likelihood estimates, the low variance mitigating the obvious bias [20]. Including the source's reliability about measurements is a direct extension of URRT#3 (see for instance [34]), as well as considering the dependencies between variables. The final assessment $\eta(X_R, t)$ expresses some *uncertainty* degree that the vessel is actually following route R_k.

URRT#4—In the probabilistic non-Bayesian approach of URRT#4, the individual probabilities are assumed to provide local belief degrees toward each route. They are summed up to give a global belief so that the higher the belief degree according to each feature, the higher the global belief. URRT#4 does not consider the dependency between features. However, some notion of source's *reliability* can be captured by the weights ω_i that can be derived from some likelihood measures extracted from a confusion matrix. This expresses some *second-order uncertainty* about the source's declaration at t. The combination rule is a disjunction (logical OR) and is known to be less sensitive to estimation errors (unreliable sources), and to single source's opinion [27] making the approach more robust. This is a more cautious rule to be used in case of less reliable sources.

URRT#5—URRT#5 may be seen as an extension of URRT#3 within the TBM model, where non-additive functions (i.e., plausibility functions) are used rather than probabilities. The plausibility function $\mathrm{Pl}(\mathbf{x} \mid R_k)$ models some *imprecision* about the (assumed precise but unknown) likelihood function $p(\mathbf{x} \mid R_k)$ (itself capturing some *uncertainty*) used in URRT#3. Equation (9) is obtained under the assumption of a vacuous prior on Ω_R, meaning that no prior uncertainty on routes is considered. The output of the GTB expressed by $\eta(X_R, t)$ being also a plausibility function, assigns plausibility values to *subsets* of routes and captures thus some *imprecision* over Ω_R. $\eta(X_R,t)$ defines then *second-order uncertainty* by means of a couple belief-plausibility measure expressing some uncertainty about the posterior event $(R_k \mid \mathbf{x})$. This second-order uncertainty is not considered in the traditional Bayesian approach where the probability estimations are considered certain. Other equivalent approaches exist framed into imprecise probability or robust Bayesian frameworks.

URRT#6—In URRT#6, the *uncertainty* output by the sources about the measurement provided at t is considered. Rather than a single (precise and certain) measure, each source outputs a probability distribution over the set of values of their respective feature which induces as many multivalued mappings over Ω_R when querying the dictionary of routes. The multivalued mappings define some *imprecision* over the set of routes, since to a single value in Ω_i corresponds a subset A of Ω_R. The prior imperfection on the links within Ω or between Ω_i and Ω_R is characterised as sets of routes (imprecision) satisfying some criteria about the features. This *imprecision* is further combined with the singular uncertainty of the source at time t defining the resulting BBA $\eta_i(X_R,t)$. The main characteristic of this scheme is to deal with subsets of routes, in a qualitative way, with an additional quantification. The explicit notion of *conflict* is a way to detect inconsistencies between the subsets of routes compatible with each feature. The fusion is performed through a conjunctive rule, assuming the independence between sources as well as totally reliable sources.

2) Interpretation:

The type of imperfection handled by URRT#1 and URRT#2 is graduality meaning that the route is considered as an "ill-defined object," with fuzzy boundaries, to which vessels belong more or less. The distance measure provides an aggregated inverse degree of membership of the vessel to a given route: If the distance is low then the vessel belongs to the route with a high degree of membership. Contrarily, the other methods (URRT#3 to URRT#6) express a "degree of belief" that the vessel is following the route. This is a difference between a binary event (URRT#3 to URRT#6) and a fuzzy event (URRT#1 and URRT#2). This semantic aspect highlights the need for a clear semantics for the concept of *maritime route*, whether it means either "following a specific path and thus ending in a specific destination" (binary event) or "being positioned on a portion of the sea with ill-defined boundaries" (fuzzy event).

3) Enrichment of basic URRTs:

Each of the URRT above could be enriched to account for more uncertainty supports. As examples only, the reliability of the sources is classically considered in URRT#6 by introducing discounting (or reinforcement) operations for belief functions such as described in [37]. Also, the reasoning scheme of Equation (6) in URRT#3 can be enriched by considering the reliability of the sensors in providing accurate measurements, and introducing factors $p(Z_i \mid X_i)$ where Z_i is the measurement provided by the source while the true value was X_i, as proposed in [34] for instance. URRT#3 can be easily implemented as a Bayesian network (e.g., [23]) where the dependency between variables is considered. A Bayesian network has the advantage of a better *transparency* in the reasoning for the user, which could also be an interesting assessment criterion to be considered in the URREF ontology. Moreover, the *computational cost* is improved by local computations.

B. Output quality assessment

The qualitative analysis above is now complemented by a quantitative analysis based on more standard criteria. We provide below a series of possible criteria for quantitative assessment of the six URRTs discussed in this paper, that we implement to discover abnormal behaviours of vessels within a real AIS dataset complemented by pseudo-synthetic anomalies.

1) Output criteria:

We consider the output quality criteria of TRUENESS (or falseness), PRECISION (or imprecision) and CERTAINTY (or uncertainty). The Trueness notion captures how correct the results are after decision. To measure this criterion we use the standard F_β-score (or measure), classically defined as:

$$\mathrm{Tru}(\Psi) = F_\beta(\Psi) = \frac{(1+\beta^2)TP}{(1+\beta^2)TP + \beta^2 FN + FP} \quad (12)$$

where $\beta \in [0;1]$ is a parameter weighting the two types of errors, TP, FN and FP are the number of true positives, false negatives and false positives respectively, N is the number of negative samples and P is the number of positive samples.

The IMPRECISION and UNCERTAINTY are assessed *before* the final decision (labelling to a single route) is taken, and quantify how much the URRT is non-specific and uncertain before the labelling, respectively. They are assessed through the Hartley measure and Shannon entropy[4]:

$$\mathrm{Imp}(\Psi) = \frac{1}{\log_2(|\Omega_R|)}\log_2(|A|) \quad (13)$$

$$\mathrm{Unc}(\Psi) = -\frac{1}{\log_2(|\Omega_R|)}\sum_{R\in\Omega_R} p(R)\log_2(p(R)) \quad (14)$$

where $|.|$ denotes the cardinality of sets and p is the probability distribution over the set of routes before decision is taken. The equations above are normalised versions of the measures. In (13), A is the set of compatible routes according to the corresponding decision criteria.

[4]In (14), the distances in URRT#1 and URRT#2 are transformed into probability distributions over the set of routes, with thus a different meaning. Shannon entropy may not be an adequate measure in this case.

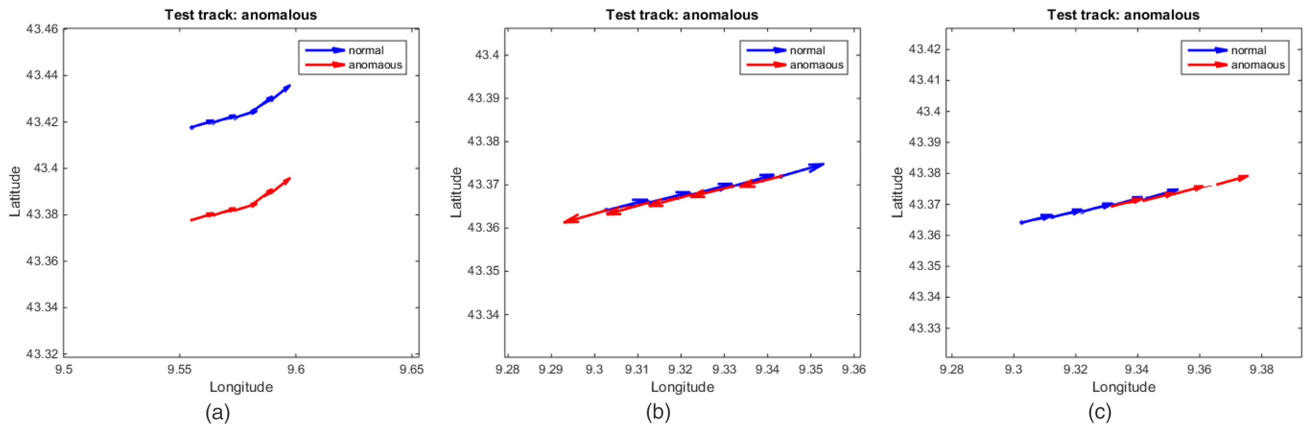


Fig. 4. Examples of simulated anomalies starting from the real data in [41]: the blue track is the normal track derived from the trajectories belonging to the subset of 8 routes, the red one is the synthetic anomalous track, reproducing a specific anomalous behaviour. (a) Positional anomaly: shifted track. (b) Directional anomaly: reverse flow track. (c) Kinematic anomaly: high speed track.

2) Dataset of anomalous tracklets:

The six URRTs are tested on a reference data set of AIS data developed at CMRE. The tracklet dataset consists of raw positional data collected for research purposes via the ground-based Automatic Identification System (AIS) receiver located in Castellana (La Spezia–Italy) owned by CMRE. The dataset contains the reports of the vessels equipped with AIS transponders, which were transiting over a section of the Northern Tyrrhenian Sea framing La Spezia harbour during the time period which goes from January 1st through February 20th 2013. The dataset contains both real tracklet data (labelled as "normal tracklets") and pseudo-synthetic tracklet data (labelled as "anomalous tracklets"). The original Castellana dataset [41] which has to be considered the source of the current dataset, is in the form of terrestrial AIS (T-AIS).

Two classes are considered: Class R_1 corresponds to normal trajectory segments and Class R_0 corresponds to anomalous trajectory segments. The normal trajectory segment of each evaluation trajectory is constructed by first selecting a random tracklet from the set of normal evaluation trajectories of a given length of 5 consecutive points: 95 tracklets are extracted from the system of pre-computed routes. Each route is decomposed into single-vessel trajectories and then further divided into tracklets of 5 consecutive points. The anomalous trajectory segment of each evaluation trajectory is constructed by first selecting a random tracklet from the set of normal evaluation trajectories of a given length of 5 points, replicating it and then altering its features. More specifically, a total of 275 anomalous tracklets were generated as follows:

- *Positional anomalies:* 80 Off-route tracks were created by shifting either the LONGITUDE or LATITUDE sequence (of a given magnitude);
- *Directional anomalies:* 108 high-speed tracks were created by increasing the initial instant speed of the track and by using a Near-Constant-Velocity Model to derive the new coordinates (LONGITUDE, LATITUDE), given the observed reported course;
- *Kinematic anomalies:* 87 opposite-flow tracks were created by changing the initial heading of the track and by using a Near-Constant-Velocity Model to derive the new coordinates (LONGITUDE, LATITUDE), given the observed reported speed SOG.

Figures 4 shows examples of the three types of simulated anomalous tracklets. As the traffic normalcy, we considered a subset of 8 routes as displayed in Figure 1.

3) Results and discussion:

We present here results of anomaly detection, thus considering two classes only, R_0 the class of anomalous tracklets containing three kinds of anomalies as described above and R_1, the class of normal tracklets belonging to the subset of 8 routes. Figure 5 displays the output quality results on a spider (radar) graph, with the three criteria of TRUENESS ($F1$-score), UNCERTAINTY (reverse entropy) and IMPRECISION (reverse non-specificity). The best method is the one covering the widest area in the graph. The ranges of the criteria are indicated in brackets. The TRUENESS criterion as measured by F_1 aggregates the TP and FN and hides thus the contribution of each corresponding type of errors. Table V expands the criterion of TRUENESS by displaying additional measures to the TPR, as the TNR, the F_1, F_2 and $F_{0.5}$ measures. While F_1 assigns equal weights to false negatives and true positives, F_2 gives more emphasis on false negatives and $F_{0.5}$ attenuates the influence of false negatives.

Through these criteria and associated performance measures, we observe that URRT#1 provides excellent results in terms of TRUENESS and PRECISION. That means that URRT#1 was able to correctly detect the anomalies and the on-route vessels. Moreover, before decision the set of compatible routes was minimum (a singleton). However, the entropy was quite high meaning some UNCERTAINTY before decision. The extension of

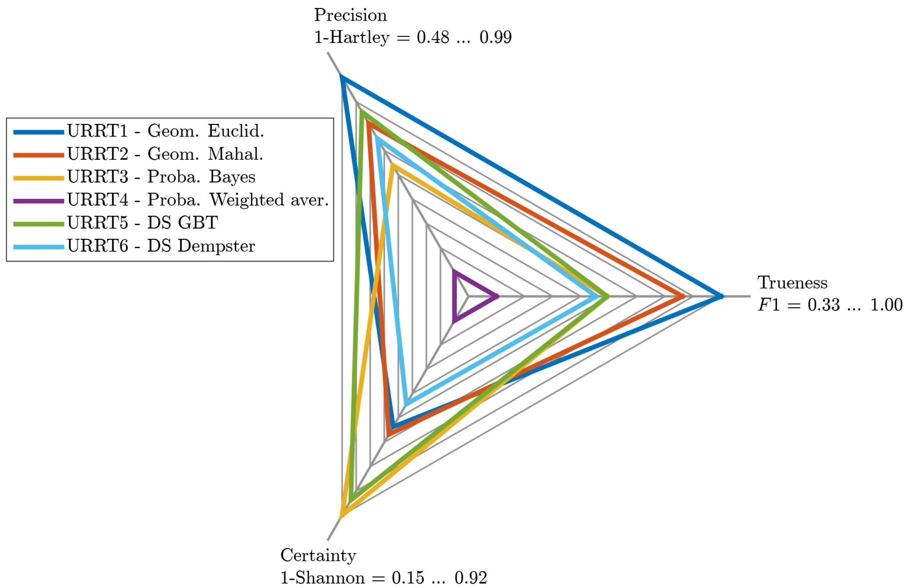


Fig. 5. Spider graph of three output quality criteria for the six URRTs.

TABLE V
Trueness measures for the six URRTs.

	TPR	TNR	F_1	F_2	$F_{0.5}$
URRT#1	1.00	1.00	1.00	1.00	1.00
URRT#2	1.00	0.91	0.88	0.95	0.83
URRT#3	1.00	0.64	0.66	0.83	0.55
URRT#4	0.21	0.97	0.33	0.25	0.48
URRT#5	1.00	0.64	0.66	0.83	0.55
URRT#6	0.77	0.76	0.62	0.70	0.56

URRT#1 to the Mahalanobis distance provides slightly lower results in terms of Trueness though, especially regarding the TNR (some anomalies have been missed) as we can see in Table V. Indeed, it appears that considering the dependency between the attributes in the observation space, although more correct than the naive independence assumption under URRT#1, leads to a slight decrease in the performances. In both URRT#1 and URRT#2, the uncertainty representation is based on the distance of the tracklet to the routes, computed by a Hausdorff distance. If the set of points of the tracklet belongs to the set of points of the routes, then the distance will be very low, or null.

The Bayesian approach URRT#3 and its evidential extension URRT#5 provide similar performance results. Compared to the pattern matching approaches, the TPR is still maximum while the TNR is only 60%. However the Uncertainty is lower meaning that the decisions could be taken with a quite high confidence. However, combined with the low TNR, this is not a desirable behaviour as this apparent confidence of the algorithm may be miss-interpreted by the decision maker. These two approaches use the likelihoods extracted from the routes' statistics as a basis for uncertainty representation. No probability distribution estimation method was applied and the likelihoods were simply extracted from the histograms. The evidential approach based on the Generalised Bayes Theorem (URRT#5) uses plausibility functions instead of probabilistic likelihoods and allows by that to account for some Imprecision on the probability distributions. It is particularly interesting when the amount of data available does not guarantee a reliable estimation of the probability distribution. Indeed, as illustrated in Table II, some routes are built upon only a few trajectories and their uncertainty may be better represented by lower and upper bounds of unknown probability distributions (as provided by belief and plausibility measures respectively) or simply by crisp intervals.

The weighted average of probabilities (URRT#4) provides the worse results along the three criteria, while the TNR is actually better than most of the other approaches. From Table V, it appears that the bad performance of URRT#4 is mainly due to a very low TPR (around 20%). That means that on-route vessels are seldom detected and wrongly detected as anomalies instead. The disjunctive operator (+) averages the posterior probabilities and a very low probability along one feature (denoting an anomaly) would be diluted among other higher probabilities. It would thus be more difficult to detect the directional and kinematic anomalies. As mentioned previously, the disjunctive operator is a rather cautious fusion operator, more suited to a consensus. We should not however conclude that URRT#4 is not a good approach, as its strength is to be robust to errors and unreliable sources, something that was not reflected in our dataset.

The evidential approach using the conjunctive rule (URRT#6) provides mediocre TPR and TNR while this pair of values is actually better than all the approaches expected the pattern matching ones. The Uncertainty and Imprecision are both quite high meaning that the decision was taken with still a high hesitation. URRT#6 is the only method which rejection criterion is based on a measure of conflict (here Dempster's conflict). The conflict is represented by the empty set between subsets of routes. The core of the reasoning relies thus on the intersection of the subsets of routes compatible with the features. In case this intersection is empty, no route is actually detected as compatible and the tracklet is classified as abnormal. The BBA was set to represent the uncertainty originating from the source's quality, which acts as a discounting over the categorical BBA of the set of compatible routes. However, in case the source expresses some (lack of) SELF-CONFIDENCE about its declaration, this singular uncertainty could be considered as well with this approach.

Finally, note that all the URRTs but the URRT#6 rely on *generic imperfection* only. URRT#6 is the only approach (again, as currently implemented) which accounts for the uncertainty expressed at the current instant in time t. All the other approaches rely on uncertainty, imprecision or graduality derived from past observations.

REMARK The results presented here should be read as an instantiation of the exploitation of the URREF mainly, as the application of such techniques to maritime anomaly detection requires deeper work. In particular, the synthetic anomaly generation technique may have a high impact on the results. The fact that the technique essentially shifts tracklets from their original position in the feature space may explain why the pattern matching approaches provide better results.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we dissected six (6) uncertainty and reasoning techniques (URRTs) to information fusion and proposed detailed description and comparison in their ability to handle uncertainty, in representation and fusion. We selected a variety of classical and simple schemes from (or adapted from) the literature which are all good candidates to solve the two problems of maritime route association and anomaly detection. We introduced the *uncertainty support* as an element conveying uncertainty, which allowed to make clearer which uncertainty is actually captured in the different reasoning schemes. We distinguished between uncertainty over individual variables or links between them, as well as second-order uncertainty. We framed our discussion within the Uncertainty Representation and Reasoning Evaluation Framework (URREF) and illustrated that considered jointly with the type of information either generic (from historical data or prior knowledge) or singular (at the time of the observation), the uncertainty support concept covers some elements of EXPRESSIVENESS of the URREF ontology (DEPENDENCY, HIGH-ORDER UNCERTAINTY, SELF-CONFIDENCE) and could expand to other criteria such as RELIABILITY.

The implementation of the URRTs to detect anomalies of a real AIS dataset allowed us to illustrate that the expressiveness criterion should not be assessed in isolation and that it is the joint assessment of the various criteria that makes the URREF powerful. Indeed for instance, a lack of expressiveness about the dependency between variables may still provide a good overall accuracy of the algorithm through some natural balance process.

Rather than identifying a "winner" approach, the comparison between the URRTs presented herein aimed at highlighting the *differences* and possible *complementarity* in uncertainty representation and reasoning. The approaches have been kept simple for a clearer understanding and in future works we will build upon this thin characterisation of the basic techniques together with the quality of the data available, taking advantage of the diversity of the different approaches, to design an efficient algorithm with easily interpretable results for detecting the anomalies at sea.

REFERENCES

[1] B. Auslander, K. M. Gupta, and D. W. Aha
A comparative evaluation of anomaly detection algorithms for maritime video surveillance.
In *Proc. SPIE, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense X*, volume 8019, Orlando, Florida, USA, May 2011.

[2] A. Benavoli and B. Ristic
Classification with imprecise likelihoods: A comparison of TBM, random set and imprecise probability approach.
In *Proc. of the 14th Int. Conf. on Information Fusion*, 2011.

[3] T. Bilgiç and I. B. Türkşen
Measurement of membership functions: Theoretical and empirical work.
Handbook of Fuzzy Sets and Systems, 1 Fundamentals of Fuzzy Sets, 1995.

[4] E. Blasch, E. Dorion, P. Valin, and E. Bossé
Ontology alignment using relative entropy for semantic uncertainty analysis.
In *Proceedings of the IEEE National Aerospace and Electronics Conference (NAECON)*, pages 140–148, Fairborn, OH, 14–16 July 2010.

[5] E. Blasch, A. Josang, J. Dezert, P. C. G. Costa, K. Laskey, and A.-L. Jousselme
URREF self-confidence in information fusion trust.
In *Proc. of the International conference on Information Fusion*, Salamanca, Spain, July 2014.

[6] I. Bloch
Information combination operators for data fusion: A comparative review with classification.
IEEE Transactions on Systems, Man and Cybernetics—Part A: Systems and Humans, 26(1):52–67, 1996.

[7] S. Challa and D. Koks
Bayesian and Dempster-Shafer fusion.
Sādhanā, 29:145–176, April 2004.

[8] P. C. G. Costa, K. Laskey, E. Blasch, and A.-L. Jousselme
Towards Unbiased Evaluation of Uncertainty Reasoning: The URREF Ontology.
In *Proc. of the 15th Int. Conf. on Information Fusion*, Singapore, 2012.

[9] J. P. de Villiers, A.-L. Jousselme, A. de Waal, G. Pavlin, K. Laskey, E. Blasch, and P. Costa
Uncertainty evaluation of data and information fusion within the context of the decision loop.
In *Proc. of the 19th Int. Conf. on Information Fusion*, Heidelberg, Germany, July 2016.

[10] P. de Villiers, G. Pavlin, P. Costa, K. Laskey, and A.-L. Jousselme
A URREF interpretation of Bayesian network information fusion.
In *Proc. of the International Conference on Information Fusion*, Salamanca, Spain, July 2014.

[11] G. K. D. de Vries and M. van Someren
Machine learning for vessel trajectories using compression, alignments and domain knowledge.
Expert Systems with Applications, 39(18):13426–13439, Dec. 2012.

[12] A. Dempster
Upper and lower probabilities induced by multivalued mapping.
The Annals of Mathematical Statistics, 38:325–339, April 1967.

[13] T. Denœux
Introduction to belief functions.
Fourth School on Belief Functions and their Applications, 2017. URL https://www.hds.utc.fr/ tdenoeux/dokuwiki/_media/en/lecture1.pdf.

[14] T. Denoeux and P. Smets
Classification using belief functions: Relationship between case-based and model-based approaches.
IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics, 36(6):1395–1406, 2006.

[15] P. Diaconis and S. Zabell
Some alternative to Bayes's rule.
In B. Grofman and G. Owen, editors, *Proc. of the Second University of California, Irvine, Conference on Political Economy*, pages 25–38, 1986.

[16] D. Dubois and H. Prade
Evidence, knowledge, and belief functions.
Int. Journal of Approximate Reasoning, 6:295–319, 1992.

[17] D. Dubois and H. Prade
Fuzzy sets and probability: Misunderstandings, bridges and gaps.
In *Proc. of the 2nd IEEE Int. Conf. on Fuzzy Systems*, pages 1059–1068, 1993.

[18] D. Dubois and H. Prade
The three semantics of fuzzy sets.
Fuzzy Sets and Systems, 90:141–150, 1997.

[19] D. Dubois and H. Prade
Formal representations of uncertainty,
volume Decision-making—Concepts and Methods, chapter 3, pages 85–156. ISTE, London, UK & Wiley, Hoboken, N.J. USA, 2009. Invited paper.

[20] J. H. Friedman
On bias, variance, 0/1-loss, and the curse-of-dimensionality.
Data Mining and Knowledge Discovery, 1:55–77, 1997.

[21] K. M. Gupta, D. W. Aha, and P. Moore
Case-based collective inference for maritime object classification.
In *Proceedings of the Eighth International Conference on Case-Based Reasoning*, pages 443–449, Seattle, WA, 2009. Springer.

[22] ISO 5725
Accuracy (trueness and precision) of measurement methods and results—part 1: Introduction and basic principles.
Technical report, ISO International Standardization, 2011. Available at standardsproposals.bsigroup.com/Home/getPDF/830.

[23] F. Johansson and G. Falkman
Detection of vessel anomalies—a Bayesian network approach.
In *Proc. of the 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, pages 395–400, Melbourne, Qld., Dec. 2007.

[24] A.-L. Jousselme and G. Pallotta
Dissecting uncertainty-based fusion techniques for maritime anomaly detection.
In *Proc. of the 18th Int. Conf. on Information Fusion*, Washington, D.C., USA, July 2015.

[25] A.-L. Joussselme and P. Maupin
A brief survey of comparative elements for *uncertainty calculi* and decision procedures assessment.
In *Proc. of the 15th Int. conf. on Information Fusion*, 2012. Panel *Uncertainty Evaluation: Current Status and Major Challenges*.

[26] A. Karlsson, R. Johansson, and S. F. Andler
Characterization and empirical evaluation of Bayesian and credal combination operators.
Journal of Advances in Information Fusion, 2011.

[27] J. Kittler
Combining classifiers: A theoretical framework.
Pattern Analysis and Applications, 1(18–27), 1998.

[28] G. J. Klir
Generalized information theory.
Fuzzy Sets and Systems, 40:127–142, 1991.

[29] G. J. Klir
Probability-possibility transformations: A comparison.
International Journal of General Systems, 21:291–310, 1992.

[30] G. J. Klir and B. Yuan
Fuzzy Sets and Fuzzy Logic: Theory and Applications.
Prentice Hall International, Upper Saddle River, NJ, 1995.

[31] R. O. Lane, D. A. Nevell, S. D. Hayward, and T. W. Beaney
Maritime anomaly detection and threat assessment.
In *Proc. of the 13th Int. Conference on Information Fusion*, Edinburgh, UK, 2010.

[32] R. Laxhammar
Anomaly detection for sea surveillance.
In *Proc. of the Int. Conference on Information Fusion*, Firenze, Italy, July 2008.

[33] R. Laxhammar
Anomaly Detection in Trajectory Data for Surveillance Applications.
PhD thesis, School of Science and Technology at Örebro University, 2011.

[34] H. Leung and J. Wu
Bayesian and Dempster-Shafer target identification for radar surveillance.
IEEE Transactions on Aerospace and Electronic Systems, 36(2):432–447, 2000.

[35] B. Liu, E. N. de Souza, S. Matwin, and M. Sydow
Knowledge-based clustering of ship trajectories using density-based approach.
In *proc. of the IEEE International Conference on Big Data (Big Data)*, pages 603–608, Oct 2014. doi: 10.1109/BigData.2014.7004281.

[36] F. Mazzarella, M. Vespe, and C. Santamaria
SAR ship detection and self-reporting data fusion based on traffic knowledge.
12:1685–1689, 2015.

[37] D. Mercier, B. Quost, and T. Denœux
Refined modeling of sensor reliability in the belief function framework using contextual discounting.
Information Fusion, 9:246–258, 2008.

[38] L. Millefiori, P. Braca, K. Bryan, and P. Willett
Modeling vessel kinematics using a stochastic mean-reverting process for long-term prediction.
IEEE Transactions on Aerospace and Electronic Systems, 52(5):2313–2330, October 2016.

[39] M. Musen
The Protégé project: A look back and a look forward.
AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4):4–12, June 2015. URL DOI: 10.1145/2557001.25757003.

[40] G. Pallotta and A.-L. Jousselme
Data-driven detection and context-based classification of maritime anomalies.
In *Proceedings of the 18th International Conference on Information Fusion*, Washington, D. C. (USA), July 2015.

[41] G. Pallotta and M. Vespe
Vessel traffic dataset from ground-based AIS receiver Castellana: description and use.
Technical Report CMRE-DA-2014-001, NATO STO CMRE, 2014. NATO UNCLASSIFIED.

[42] G. Pallotta, M. Vespe, and K. Bryan
Vessel pattern knowledge discovery from AIS data—A framework for anomaly detection and route prediction.
Entropy, 5(6):2218–2245, 2013.

[43] G. Pavlin, A.-L. Jousselme, P. J. de Villiers, P. C. G. Costa, and P. de Oude
Towards the Rational Development and Evaluation of Complex Fusion Systems: A URREF-Driven Approach.
In *Proc. of the 21st Int. Conf. on Information Fusion*, pages 679–687, Cambridge, UK, 2018.

[44] C. Ray, R. Gallen, C. Iphar, A. Napoli, and A. Bouju
DeAIS project: Detection of AIS spoofing and resulting risks.
In *OCEANS 2015—Genova*, pages 1–6, May 2015. doi: 10.1109/OCEANS-Genova.2015.7271729.

[45] B. Ristic and P. Smets
Target classification approach based on the belief function theory.
IEEE Transactions on Aerospace and Electronic Systems, 41(2):574–583, 2005.

[46] B. Ristic, B. La Scala, M. Morelande, and N. Gordon
Statistical analysis of motion patterns in AIS Data: Anomaly detection and motion prediction.
In *Proc. of the 11th Conference on Information Fusion (FUSION)*, Cologne, Germany, 2008.

[47] G. Shafer
A Mathematical Theory of Evidence.
Princeton University Press, 1976.

[48] G. Shafer and A. Tversky
Languages and designs for probability judgment.
Cognitive Science, 9:309–339, 1985.

[49] P. Smets
Constructing the pignistic probability function in a context of uncertainty.
Uncertainty in Artificial Intelligence, 5:29–39, 1990. Elsevier Science Publishers.

[50] P. Smets
Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem.
International Journal of Approximate Reasoning, 9:1–35, 1993.

[51] P. Smets
Imperfect information: Imprecision—uncertainty.
In A. Motro and P. Smets, editors, *Uncertainty Management in Information Systems. From Needs to Solutions*, pages 225–254. Kluwer Academic Publishers, 1997.

[52] P. Smets
Probability, possibility, belief: Which and where?
In P. Smets, editor, *Quantified Representation of Uncertainty & Imprecision*, volume 1, pages 1–24. Kluwer, Doordrecht, 1998.

[53] P. Smets
Analyzing the combination of conflicting belief functions.
Information Fusion, 8:387–412, 2007.

[54] P. Smets and R. Kennes
The transferable belief model.
Artificial Intelligence, 66:191–234, 1994.

[55] L. Sombé
Reasoning under Incomplete Information in Artificial Intelligence,
volume 5 of *Int. J. of Intelligent Systems (Special Issue)*. John Wiley and Sons, INC., 1990.

[56] P. Walley
Measures of uncertainty in expert systems.
Artificial Intelligence, 83:1–58, 1996.

[57] L. A. Zadeh
Fuzzy sets.
Information and Control, 8:338–353, 1965.

Anne-Laure Jousselme received her PhD degree from the Electrical Engineering Department of Laval University in Quebec City (Canada) and the Institut National Polytechnique de Grenoble (France) in 1997. Formely with Defense Research and Development Canada (DRDC), she is now with the NATO STO Centre for Maritime Research and Experimentation (CMRE) in La Spezia (Italy), where she conducts research activities on reasoning under uncertainty, high-level and hard & soft information fusion, information quality assessment and serious gaming applied to maritime situational awareness and anomaly detection. She is area editor of the International Journal of Approximate Reasoning and associate editor of the Perspectives on Information Fusion magazine. She is a member of the Boards of Directors of the International Society of Information Fusion (ISIF) where she serves as VP membership and of the Belief Functions and Applications Society (BFAS) where she serves as Secretary. She serves on program committees of the International Conference of Information Fusion (FUSION) and the International Conference on Belief Functions (BELIEF). She was Tutorial Chair of FUSION 2007 in Quebec City (CA), International Co-chair of FUSION 2015 in Washington and Technical Co-chair of FUSION 2019 in Ottawa (CA). She was general Chair of the Canadian Tracking and Fusion Conference (CTFG) in 2014 in Ottawa (CA) and Local Organizer of the International Conference of Scalable Uncertainty Management (SUM) in 2015 in Quebec City (CA).

Giuliana Pallotta is a data scientist at Lawrence Livermore National Laboratory (LLNL), Computational Engineering Division, where she is working on diverse projects relevant to national security (e.g., climate, global security, evolving networks). Prior to joining LLNL as a staff member, she was a Research Scientist in maritime big data analytics at NATO Science and Technology Organization CMRE during 2012–2015. She is a Fulbright Scholar, recipient of a 2013 IEEE JPLC Information Fusion Best Paper Award for her work on pattern recognition and anomaly detection from vessel motion data. In 2011 she was a post-doctoral Research Associate at the University of Washington, Seattle–US. She received her PhD in Applied Statistics in 2009. Her research interests are in the areas of Bayesian inference, Statistical modeling, Big Data, Machine Learning, Anomaly Detection, Situational Awareness and Information Fusion for surveillance and security applications.