DRDC-RDDC-2016-P124

A URREF interpretation of Bayesian network information fusion

J. P. de Villiers*† G. Pavlin‡ P. Costa§ K. Laskey§ A.-L. Jousselme¶

* CSIR, Pretoria, South Africa, jdvilliers1@csir.co.za
†University of Pretoria, Pretoria, South Africa, pieter.devilliers@up.ac.za
‡D-CIS Lab, Thales Research and Technology, Delft, The Netherlands, gregor.pavlin@d-cis.nl
§George Mason University, Fairfax, VA, USA, [pcosta,klaskey]@gmu.edu
¶DRDC-Valcartier, Quebec, Canada, anne-laure.jousselme@drdc-rddc.gc.ca

Abstract*—In order for the uncertainty representation and reasoning evaluation framework (URREF) ontology for the evaluation of information fusion systems to have maximum value, it must be generally applicable irrespective of the application, uncertainty representation, reasoning scheme or data format. Since the URREF ontology is still an evolving framework, it is the focus of ongoing refinement through the efforts of the Evaluation of Techniques for Uncertainty Representation Working Group (ETURWG). Recent efforts by the authors to apply the URREF definitions to the evaluation of Bayesian network (BN) fusion systems have identified a need to translate and map the terminology of the URREF to the Bayesian network paradigm. The BN-to-URREF mapping is addressed in this paper within the context of the atomic decision procedure (ADP) and the latest view of the URREF ontology. The atomic decision procedure describes the information fusion process from input to output and consists of information sources (ADP-1), the interpretation and processing of this information and the qualification of the uncertainty it contains (ADP-2), the fusion of and reasoning with this uncertain information (ADP-3) and finally the decision scheme and output information (ADP-4). The URREF evaluation of the BN information fusion process allows for evaluation according to (1) representation criteria which relate to ADP-2 and evaluate modeling and model parameterization, (2) reasoning criteria which relate to ADP-3 and evaluate the reasoning scheme, which in the case of a BN entails the computation of marginal probability densities over hypothesis variables and (3) data criteria which relate to ADP-1 and evaluate the sources and the information generated by said sources, together with the qualification and representation of the uncertainty inherent to the information. However, the current URREF framework does not make provision for (4) decision criteria, which relate to ADP-4 and evaluates the decision scheme. BNs do not address decisions *per se*, but a BN often forms the inference component of a decision system. In this case, the problem is to identify the maximum utility or minimum loss decision.

I. Introduction

The improvement of the different aspects of an information fusion system (IFS) requires objective and unbiased evaluation. As such, in order to perform level 4 (Process Refinement) functions of the Data Fusion Information Group (DFIG) data fusion model [1], methods for evaluating fusion system performance are required. Initial efforts to establish a framework for the evaluation of IF systems commenced with [2] and [3], with consolidation later in [4]. These efforts rely on measures of effectiveness (MOEs) [5], measures of performance (MOPs) and measures of force effectiveness (MOFEs). The authors of [6] identify quality of service (QOS) [5] and quality of information (QOI) [7], [8] as important MOEs, as well as their impact on decision making [9],[5]. The author of [10] considers the lack of practical application of performance measures in the literature. Challenges that were identified include the lack of ground truth, the difficulty in capturing and consolidating the different facets of fusion performance, the need to adapt performance measures to an application or situation, and theoretical assumptions that are violated in practice.

The link between uncertainly representation and performance evaluation is discussed in [6]. The application of fusion level metrics such as timeliness, accuracy and confidence are directly influenced by the uncertainty representation (mathematical framework) and reasoning procedure, referred to as uncertainty *calculi* in [11]. There is a general view that the fusion process consists of the generation of information (the *Input step*), the representation of knowledge and uncertainty, the combination of information, and reasoning under uncertainty (the *Representation and Reasoning step*) and the reporting of results of the reasoning process along with some characterisation of uncertainty of the reported information (the *Output step*). This flow of information is further refined and termed the *Atomic Decision Procedure* (ADP) in [11] and [12]. The ADP assists in characterising the objects that will be under evaluation and, for the purpose of this paper, will include a) the performance of the fusion system, b) the sources of information at the input of the fusion system, c) the information itself at the input of the fusion system and finally d) the representation of knowledge and uncertainty in the fusion system.

The evaluation of any system that fuses information from several heterogeneous sources is a challenging exercise. This statement is highlighted by the ongoing efforts of Evaluation of Techniques for Uncertainty Representation Working Group (ETURWG). Bayesian networks (BNs) can be viewed as a versatile fusion technology because BNs can fuse heterogeneous information from domain experts as well as incorporate measured data. The objectives of this paper are both to list common BN evaluation techniques that are available in the literature, and, where possible, to link them to the latest (and constantly evolving) views on the uncertainty representation and reasoning evaluation framework (URREF) ontology [6] and current efforts of the ETURWG. In order to make the

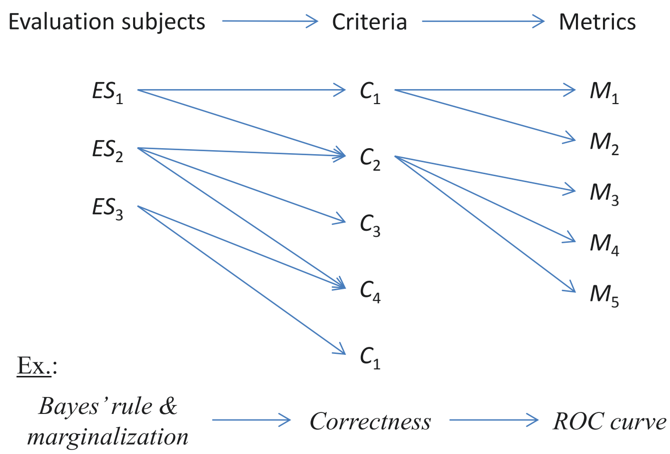

Fig. 1. The figure depicts the distinction between evaluation subjects, criteria and metrics. As an example, Bayes' rule and marginalisation is the subject of evaluation, with "Correctness" the criterion and the receiver operating characteristic (ROC) the metric.

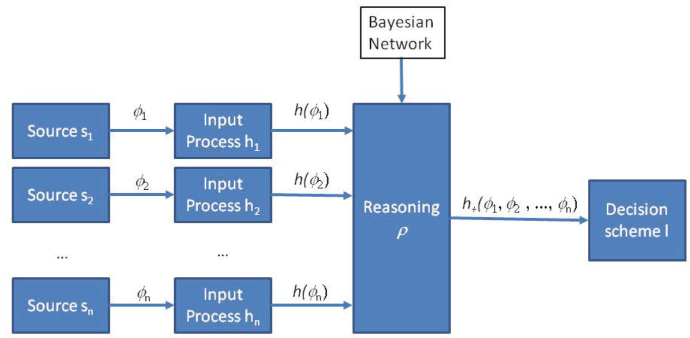

Fig. 3. The ADP process in BN fusion

link, existing BN evaluation techniques need to be linked to the appropriate objects of URREF ontology. One of the most complete and recent expositions of evaluation techniques for BNs is found in [13]. Techniques are classified into four categories: methods to evaluate criteria such as sensitivity and influence, model prediction performance, uncertainty in posterior probability distributions, and alternative posterior probability distributions.

The authors distinguish between *evaluation subjects* (ES), that are assessed according to *criteria* through specific *metrics* (see Figure 1). For a given evaluation subject ES, a series of associated criteria C_ks are defined. In some cases, two distinct subjects may be evaluated according to the same criterion. For instance, a reasoning scheme ES_1 and an uncertainty representation ES_2 may be assessed according the criterion of *consistency*, with however different definitions and associated metrics.

The work in this paper attempts to address the adaptation of URREF criteria and apply them to several evaluation subjects of BN fusion and to further demonstrate a conceptual framework for evaluation in a practical problem.

II. Rhino poaching case study

In order to map the URREF criteria to BN information fusion, a rhino poaching case study is used. The BN models the problem of rhino poaching and attempts to capture the causal elements (see Figure 2). An existing rhino poaching model [14] is used to illustrate the mapping. This model relates to a particular geographically bounded area or cell. Thus consider a map of a reserve or game park partitioned into $j = 1, \ldots, M$ areas where the jth area or cell is denoted by G_j. Furthermore, consider a Bayesian network with $N = 15$ nodes. The variable "$Season$" determines the presence of drinkable water (denoted by the "$Water$" variable) and vegetation which is suitable for rhino consumption (denoted by the "$Vegetation$" variable). The state of the "$Water$" and "$Vegetation$" variable determine the presence of rhinos in a cell. The presence of a rhino (denoted by "$Rhino\_present$") contributes to the presence of a poacher (denoted by "$Poacher\_present$"), where a rhino is sighted by a ranger (denoted by "$Rhino\_sighted$") and whether a tagged rhino returns radio collar a track (denoted by "$Rhino\_track$") in the cell. The time of day ("$TimeOfDay$"), weather condition (denoted by "$Weather$") and moon phase (denoted by "$Moon$") determine the vulnerability of the rhino. The presence of a ranger (denoted by "$Ranger\_present$"), the presence of a poacher and the vulnerability of the rhinos determine whether a poaching event will happen, which in the end will result in a poaching report if discovered (denoted by "$Poaching\_report$"). Finally the presence of a poacher will determine if a tourist or ranger sees the poacher. This variable is denoted by ("$Possible\_poacher\_sighting$").

III. The Atomic Decision Process

In order to systematically identify evaluation aspects and facilitate mapping to URREF, BNs need to be analysed in the context of an Atomic Decision Procedure (ADP) [11], [12]. The ADP structure implies multiple elements that can be evaluated [12], such as (ADP-1) sources of information s_i and input information ϕ_i, (ADP-2) uncertainty representation process h and uncertain information $h(\phi_i)$, (ADP-3) reasoning ρ and combined information $h_+(\phi_1, \phi_2, \ldots, \phi_N)$ and (ADP-4) decision scheme l and output information. Figure 3 depicts the elements of the ADP.

A. ADP-1: Sources and Input Information

Bayesian networks are often used in problems that involve many different types of sources. For the rhino poaching case study, sources and the input information they produce include the vegetation map (denoted by s_1), the water map (s_2), ranger reports (s_3), the rhino tracking system (s_4), blue-force tracking reports on ranger presence (s_8), and other discrete inputs such as time of day (s_5), season (s_6) and weather(s_7).

B. ADP-2: Input Process and Uncertain Information

This section describes the processes that translate the outputs of various sources, such as sensor signals and reports,

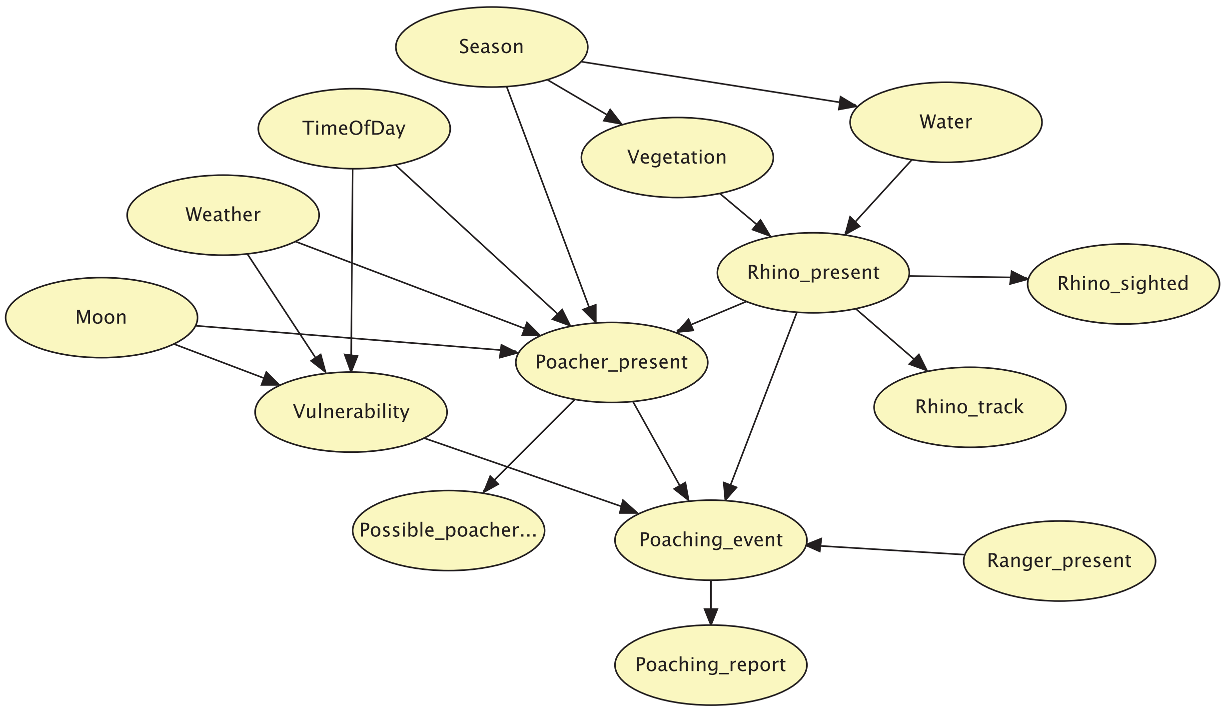

Fig. 2. The rhino poaching model

to the inputs of the fusion system. An input process h_i translates output ϕ_i of source s_i to an instantiated mathematical representation $h_i(\phi_i)$. In Bayesian networks, each output ϕ_i is associated with a variable A_i. The input process h_i transforms output ϕ_i to $h(\phi_i)$, which is an assignment of evidence to the states of the corresponding variable A_i. The uncertainty regarding ϕ_i is introduced in two ways.

1) Variable A_i corresponding to output ϕ_i is assigned hard evidence $h(\phi_i)$ which states with certainty that a specific state of A_i has materialized. The uncertainty associated with ϕ_i is captured by the the conditional probability table (CPT) describing the relations between A_i and the variables that directly influence it. Thus, in case of hard evidence, the characteristic of the source s_i is represented via the CPT that is part of the overall domain model. Such a CPT represents a sensor model for the source corresponding to variable A_i.
2) Variable A_i is assigned uncertain evidence $h(\phi_i)$ which specifies some weighting over the states of A_i. There exist different forms of uncertain evidence [15], such as likelihood evidence [16], [17] (also called virtual evidence), and soft evidence. Virtual evidence assigns likelihoods to the states of A_i, and reflects the uncertainty about a reported observation (the observation is uncertain). Soft evidence assigns a probability distribution over the states of A_i, and reflects the uncertainty of the underlying variable which is observed (the event is uncertain). In case of uncertain evidence, the output ϕ_i carries information about the uncertainty of the observations or event for virtual and soft evidence respectively. Like in the case of hard evidence, additional uncertainty is introduced by the CPT relating variable A_i to its direct causes. Sources of such uncertain evidence have capability to estimate distributions/likelihoods over states of variables. For example, in the modular fusion approach introduced in [18], messages exchanged between different fusion modules can be viewed as uncertain evidence.

The following list illustrates the input processes by using the example model. The sources, associated variables, their ranges and their links to other variables of the rhino poaching example problem are presented in Table I.

1) In the case of the vegetation map s_1, the output ϕ_1 is a map datum read out corresponding to area G_j. If the map s_1 indicates vegetation suitable for rhino consumption within area G_j, then the input process h_i produces hard evidence $h(\phi_1)$ instantiating the corresponding variable $Vegetation$ to $available$. If no suitable vegetation is indicated on the map, $h(\phi_1)$ instantiates $Vegetation$ to $not\_available$. The water map corresponding to source s_2 is used in a similar manner.
2) A ranger s_3 can report the presence of rhinos in area G_j. Ranger provides output ϕ_3 by submitting a form, where the presence/absence of rhinos at location G_j is indicated by a check box. The state of this element is translated by h_3 to hard evidence $h(\phi_3)$ instantiating Boolean variable $Rhino\_sighted$ either to $true$ or $false$; i.e. Note, output ϕ_3 does not carry any information about the source quality/performance, although the source is noisy; the ranger can make mistakes and confuse other animals for rhinos or miss rhinos. The source noise is in this case captured

by the CPT relating variables $Rhino\_present$ and $Rhino\_sighted$. This CPT describes the chance that a ranger will correctly report the presence or absence of a rhino if asked whether rhinos are present at G_j. This CPT is a sensor model for the reporting process by rangers.

3) The rhino tracking system s_4 provides output ϕ_4, a report indicating the chance that the rhinos are in area G_j. Report ϕ_4 is assumed to carry estimated posterior distributions over the potential locations of rhinos, that could be produced with the help of an arbitrary tracking approach, such as for example Kalman filters and Particle filters. The input process h_4 computes $h(\phi_4)$ by integrating the posterior distribution delivered by ϕ_4 within area G_j. $h(\phi_4)$ assigns Boolean variable $Rhino\_tracked$ a distribution over its states. This signal can be considered as uncertain evidence.
4) There are also other types of inputs which correspond to simple instantiations of discrete random variables in the example model, such as the time of the day (morning, afternoon, evening and night) represented by multi state discrete random variable $TimeOfDay$, and season (spring, summer, autumn, winter) represented by multi state discrete random variable $Seasons$. Also weather is considered as a phenomenon that can easily be observed and represented by a multi-state discrete random variable $Weather$. These variables have corresponding sources $\{s_5, \ldots, s_7\}$ with outputs $\{\phi_5, \ldots, \phi_7\}$. The Input processes for these variables produce hard evidence $\{h(\phi_5), \ldots, h(\phi_7)\}$.
5) The presence of rangers is assumed to be easily observable either through a blue-force tracking system or through a reporting system s_8. Such a report ϕ_8 could be translated by h_8 to hard evidence $h(\phi_8)$ instantiating binary variable $Ranger\_present$. If the tracking approach is noisy, then $h(\phi_8)$ is uncertain evidence.

C. ADP-3: Reasoning and Combined Information

Reasoning is the centrepiece of the fusion system. In this process, states of hidden variables representing hypotheses are estimated. Reasoning with Bayesian networks is based on three distinctive elements.

1) A domain model in form of a Bayesian network. Mathematically, the model used by the algorithm is given as a collection of local distributions: CPTs that describe the uncertainty of the relations between the non-root variables, together with marginal probability distributions over the states of the root variables. The product of these local distributions encodes a joint probability distribution over all variables in the BN. The graphical representation captures merely the qualitative knowledge about the relationships between the relevant variables; quantitative information about the strength of the relationships is given by the local distributions.
2) Instantiated variables which capture the evidence obtained via signals from different sources.
3) A generic, domain independent inference process supported by the inference engine. This inference process produces an exact or approximate representation of the joint posterior distribution of the variables in the BN conditional on the evidence. Exact solutions are typically a variant of the sum-product algorithm, such as Λ-Π algorithm [16], Junction tree algorithm [19], etc. There are also approximate methods, such as Loopy propagation [20], particle filters, etc., which typically are applied when exact computation would be intractable.

The reasoning process has two types of inputs associated with uncertainties, (i) the uncertain inputs resulting from the Input processes and (ii) the domain models. The uncertain inputs and the domain model are used by the inference process that is based on some belief propagation algorithm that produces uncertain output information. The domain models capture prior knowledge about the domains obtained through analysis of historical data or expert introspection (expert interviews, workshops). The domain models can be viewed as a program that instructs the generic algorithm to carry out the right operations, dependent on the domain model and the states of instantiated variables (i.e. evidence), represented by ϵ.

The reasoning process computes marginal posterior probability distributions $P(H_k|\epsilon)$ over the states of hidden variables of interest H_k, i.e. the so-called hypothesis variables, conditional on the observed evidence. If desired, it is also possible to compute other marginalizations of the joint posterior distribution, such as the joint posterior distribution of two hypotheses. The reasoning process takes into account the prior knowledge about the dependencies between the variables and their strength encoded in the corresponding CPTs.

Uncertainty is introduced into the reasoning processes via local distributions and soft evidence, as they describe uncertain relations between various phenomena. The uncertainties of the relations are captured by conditional probabilities that describe the chance that various states of random variables will materialize as a function of their conditioning variables (parents in the BN graph). For example, as shown in Figure 2, the probability distribution of the "$Rhino\_present$" variable is assessed conditional on the values of the "$Vegetation$" and "$Water$" variables.

D. ADP-4: Decision Scheme and Output Information

The outputs of a reasoning process are used (i) in some decision making process or (ii) as inputs to another fusion process. Various decision making schemes I are possible. One common approach is maximum $aposteriori$ (MAP), in which the state with the highest posterior probability is assumed to occur. Note that the joint MAP for several variables is typically not the same as the individual MAP for each variable taken individually. Alternative schemes could be based on simple thresholding; or the posteriors over the alternative states can be used for ranking. The output information of such a process could be an indicator of the decision, an ordered list of the first n highest ranked hypotheses, etc. Finally, the approach with the strongest Bayesian justification is the Bayes risk approach. In this approach, a $loss function$ $L(d, x)$ is specified to measure the cost of taking decision d when the true state of variable X is x. For continuous variables, $L(d, x)$ is typically a function of the distance between d and x, such as the quadratic loss

TABLE I. A TABLE DEPICTING INFORMATION SOURCES, ASSOCIATED VARIABLES, THEIR RANGES AND THEIR LINKS TO OTHER VARIABLES IN THE BN OF THE RHINO POACHING EXAMPLE PROBLEM.

Source	**Variable name**	**Range**	**Links**
Vegetation Map	$a_1 = Vegetation$	$X_{a_1} = \{available, not\_available\}$	$Season$, $Rhino\_present$
Water Map	$a_2 = Water$	$X_{a_2} = \{available, not\_available\}$	$Season$, $Rhino\_present$
Ranger report	$a_3 = Rhino\_sighted$	$X_{a_3} = \{true, false\}$	$Rhino\_present$
Rhino tracking system	$a_4 = Rhino\_track$	$X_{a_4} = \{in\_area, not\_in\_area\}$	$Rhino\_present$
Time of day	$a_5 = TimeOfDay$	$X_{a_5} = \{morning, noon, evening, night\}$	$Poacher\_present, Vulnerability$
Season	$a_6 = Season$	$X_{a_6} = \{spring, summer, autumn, winter\}$	$Poacher\_present, Water$
Weather	$a_7 = Weather$	$X_{a_7} = \{favourable, inclement\}$	$Poacher\_present, Vulnerability$
Ranger blue force tracking system	$a_8 = Ranger\_present$	$X_{a_8} = \{in\_area, not\_in\_area\}$	$Poaching\_event$

function; for categorical variables, $L(d, x)$ measures the cost of confusing x with d. The Bayes decision is to choose the state d that minimizes the expected loss $E_X[L(d, X)]$, where the expectation is taken over the posterior distribution of X given the evidence.

E. Primary Evaluation of the ADP and its Context

The ADP assists with the explicit formulation of where uncertainties enter within the overall fusion process and also provides a context for the evaluation of uncertainty representations and their use in fusion processes. The main objects under evaluation (or evaluation subjects) are undoubtedly in the reasoning part of the ADP (labeled ADP-3), whereas the other parts of the ADP can be viewed as providing a context for such evaluation; that is, they are boundary conditions of the primary evaluation. The suitability of the representation and the reasoning schemes will depend on various given elements, such as sources, the resulting signals as well as the chosen decision making processes and compatible outputs of the fusion system. These elements result from user requirements regarding functionality, performance, operational constraints and the costs of implementation/maintenance. When available, previously developed models for similar problems can be invaluable tools for model specification. There are also commonly used modeling patterns (e.g., noisy-OR, divorcing) that can employed where applicable (c.f., [21])

For example, in the targeted domain of Figure 3, the authors are dealing with very different correlated sources and related phenomena. Thus, in order to keep inference and construction of such systems tractable, a systematic description of the correlations between the relevant variables is required. Moreover, BNs are suitable in the targeted domain as no explicit representation and handling of reliability of sources is required. The authors assume a simple Bayesian decision making process or ranking based on the computed likelihood of poaching.

IV. BN VALIDATION AND EVALUATION UNDER URREF

Fusion system evaluation criteria form part of the URREF Ontology (version 2) [22]. As currently defined, the URREF ontology has four criteria groups. The first relates to criteria associated with the handing of data (labeled $DataHandlingCriterion$). These include *Data interpretation* and *Traceability* as evaluation criteria. The second group (labeled $RepresentationCriterion$) relates to how the problem domain and data are represented through the criteria of *Knowledge handling*, *Simplicity*, *Expressiveness*, *Adaptability* and *Compatibility*. The third group (labeled $ReasoningCriterion$) relates to how well the fusion system reasons. Here, the evaluation criteria include *Correctness*, *Consistency*, *Performance* (*Throughput* and *Timeliness*), *Computational cost* and *Scalability*. The fourth and final evaluation criteria group (labeled $DataCriterion$) consists of criteria relating to the fidelity of the data (both input and output, since the output of one fusion process may form the input of another). It seems here it also relates to some aspects of the source of the data. Such criteria include the *Weight of evidence* of the data (*Sensitivity*), *Relevance to the problem*, the *Quality* of the data (*Veracity*, *Precision* and *Accuracy*), and *Credibility* (*Observational sensitivity*, *Objectivity* and *Self confidence*). In this section we attempt to map these criteria to Bayesian networks as fusion systems.

A. Evaluation according to the Representation criteria

The representation criteria in the URREF ontology relate to ADP-2, or uncertainty representation. In a Bayesian network, uncertainty is represented by a joint probability distribution over the uncertain variables. The internal model is parameterized by local probability distributions, which may be a elicited from human subject-matter experts or may be informed by historical data. These representation criterion (*Knowledge handling*, which includes *Simplicity, Expressiveness, Adaptability and Compatibility*) relate to how straightforward it is to specify the graph and parameterize the model for competent engineers without deep knowledge of Bayesian theory; whether the formalism is sufficiently expressive to capture kinds of uncertainty needed to address the problem at hand; the ability of the model to adapt to new knowledge, changes in input sources, or new output configurations; and how compatible the model is with data standards respectively.

BNs have seen wide application to many problems of reasoning under uncertainty. Their popularity stems in great part from their ability to represent complex uncertain dependence relationships such as those shown in Figure 2 in a parsimoniously parameterized and computationally tractable manner. For example, to specify a fully general joint probability distribution on the variables in Figure 2 would require 262144 parameters; whereas an equivalent factorized representation of the joint probability distribution over the same set of variables can be defined by only 400 parameters, if the direct dependencies depicted in Figure 2 are considered. This vast reduction in the complexity of specification of the joint probability distribution makes both expert elicitation and learning from data feasible. In addition, the directed graph naturally represents cause and effect relationships and allows modeling the effects of external interventions to set the value of a variable (c.f., [23]). Furthermore, BNs allow combining knowledge from theory (e.g., laws of physics; models from economic theory) with informed expert judgement and data. Further, easily accessible software tools support both model specification and inference. The modular nature of the local distributions supports model maintenance and adaptation. Finally, extensions such as multi-entity Bayesian networks and probabilistic ontologies provide a semantically rich extension to Bayesian networks [24].

BNs have been criticized for their inability to represent incomplete knowledge (i.e., unknown local distributions), and belief functions have been proposed as an alternative representation It has been argued in response that sensitivity analysis and higher-order probability (parameter uncertainty) can address this issue [25] . Another issue is the representation of fuzzy knowledge. For example, one might argue that vulnerability is a matter of degree, and even precise knowledge of all the measurable factors contributing to vulnerability is insufficient to give a precise definition of vulnerability. Fuzzy logic has been proposed as an alternative representation. There have also been formulations proposed for how to model fuzzy knowledge from the Bayesian perspective. To our knowledge, there have been few direct comparisons of alternative formulations on the same problem. It is our hope that the URREF will encourage such comparisons.

The modeling and reasoning process, through the casting of sensor, target, and environmental conditions and their causal relations into a BN, can be likened to operational conditional modeling in Chapter 13 of [26]. In the rhino poaching example sensors correspond to sightings and radio collars, targets correspond to rhinos and poachers, and environmental conditions correspond to weather, vegetation, season, time of day, etc. The BN can only reason based on visible data (weather, season, time of day, rhino sightings etc.) since the true states that are needed to perform a complete analysis are latent (hidden). This creates uncertainty of the sources, the information they produce and the performance of the reasoning process. Conditioning reduces uncertainty, hence any information that is relevant will assist the fusion process. For example, the inclusion of variables representing the price of rhino horns and a market to sell may assist in creating an a-priori estimate of the chance of a poacher being present.

B. *Evaluation according to Reasoning criteria*

Bayesian network prediction performance can be evaluated when cases are available for which outcomes are known. Common metrics include error rates and confusion matrices, receiver operating characteristics, and several performance indices such as the Bayes information criterion (BIC) and true skill statistic [13]. In the language of [11] and [12] these metrics probably evaluate the "reasoning step" or ADP-3 of the atomic decision procedure - element number. In other words, how well does the BN predict certain outcomes in comparison to real-world data from the underlying process being modelled? In this case the criteria are most likely the *Consistency* and *Correctness* of the fusion algorithm, which are both reasoning criteria as defined in the URREF ontology version 2.

Furthermore, the aforementioned metrics can be used, together with a loss function which encodes the risk attitude of the decision maker, and corresponds to the "decision step" of the atomic decision procedure (ADP-4) in [11]. Here it seems that a suitable corresponding evaluation criterion does not exist in the URREF.

Model sensitivity relates to the extent to which specific variables influence the posterior probability distribution of the variable of interest. This relates to the relevance of specific variables for making inferences about the posterior variable of interest. The metrics which can evaluate model sensitivity include variance reduction, entropy reduction, case file simulation and scenario evaluation [13]. This is related to the evaluation of the choice of variables of the so called "universe of discourse", which forms the first construct of the elementary atomic decision procedure (ADP-1) in [11]. In this case the criteria are the *Expressiveness*, *Simplicity* and *Compatibility* of the BN model and can be classified as the evidence related criteria.

C. *Evaluation according to the Data criteria*

In the case of BNs, the information sources can be sensor data, human observers or other heterogeneous sources. While the current URREF considers only sources feeding the fusion system at runtime, we propose to explicitly discuss also sources that enter fusion via the modeling process. Namely, sources of information influence fusion using BNs in three ways, through human experts specifying the causal relations between variables, by humans or data informing or defining the model's CPTs, and by sources of information providing evidence (observations). The latter is covered by the current version of the URREF. However, also the experts constructing BN components and databases providing data for parameter learning are design time sources s_i that should be characterized. In analogy to the ADP, the design time sources produce outputs ϕ_i which are translated via a design/learning process h_i to mathematically rigorous models $h(\phi_i)$, i.e. the encoding of their beliefs. Clearly, neither the design time sources s_i nor the learning processes are perfect.

Uncertainty thus enters the BN in three forms: model structure uncertainty (causality uncertainty), model parameter uncertainty (uncertainty in the CPTs) and observational uncertainty (uncertain evidence). It is foreseeable that these uncertainties can be evaluated according to the criteria of *Veracity, Objectivity, Observational sensitivity* and *Self confidence*. An interesting attempt to model expert knowledge imprecision can be found in [27]. The discretization of inherently continuous variables may also lead to a loss of observational sensitivity and it may be necessary to quantify this in some way.

Moreover, also the inputs and outputs of a fusion system should be evaluated according to the remaining Data criteria. Here *Accuracy, Precision and Veracity* are relevant criteria. Uncertainty in the posterior probability distribution of the variable of interest can be characterised by metrics which measure its dispersion (and its reduction in the presence of mounting evidence). Such metrics include Bayesian credible intervals, a posterior probability certainty index with a certainty envelope, and a new adaptation of the Gini coefficient and the Lorenz curve [13]. The object under evaluation for these metrics seems to be a subjective metric of the uncertainty at the output of the reasoning step (after fusion). The reason this measurement may be subjective, is because this measurement of output uncertainty is based on the internal model of uncertainty and not as such the true uncertainties of the underlying process being modelled.

D. Information itself (Data criteria)

The criteria that are discussed in this section refer to the the pieces of information just as they leave the information sources, and hence *before* they enter the fusion system and are translated into a specific uncertainty or knowledge representation. Here the metrics are *Relevance, Weight of evidence, Conclusiveness, Veracity, Dissonance, Ambiguity, Completeness, Confidence* by source S, *Accuracy, Precision and Interpretation*. It may be worth demonstrating the application of these criteria using the rhino poaching example in this paper. Consider for example the "vegetation map" information source. If one considers an area or cell G_j on the map of a game reserve, the vegetation map may give a discrete or continuous index for the suitability of consumption by a rhino as a value between 0 and 100.

When the criteria of *Relevance* and *Weight of evidence* are considered, this must be within the context of giving a final answer, such as the chance that a rhino will be poached in that cell in the next 24 hours. In other words the question may be stated as a) Is the vegetation index at all relevant? and b) If so, how relevant compared other causal factors? This again relates to a sensitivity analysis, but *before* the signal ϕ_1 is translated to the boolean variable a_1. Hence, traditional BN sensitivity analyses will not suffice, and the relevance and weight of evidence need to be ascertained for the signal ϕ_1 and *not* its corresponding BN variable $a_1 = Vegetation$.

The criterion of *Veracity* is not relevant for this example, but in other cases will relate to the truthfulness of the piece of information ϕ_n. The criterion of *Dissonance* implies a clash between different sources, and here two maps of the same area may give different vegetation suitability indices. *Ambiguity* may not be relevant for this example, but in other cases a piece of information may imply more that one underlying explanation. *Completeness* in the vegetation map example may related to uncharted areas, or a map in which the vegetation composition was not sampled finely enough. The criterion of *Confidence* may be relevant if the point on the map queried is far from a sample point. Hence the confidence of vegetation suitability index may be low. If there was probabilistic uncertainty in the vegetation suitability index, then the distance from the true value would indicate the *Accuracy* and the amount of variability of the vegetation suitability index around true value would indicate the *Precision*. If the map was to be read by a person using a gray scale, with him/her being aware that 0 maps to white and 100 maps to black, one may have different map readers interpreting the blackness of the measured point differently. Here, the *Interpretation* criterion may be relevant.

V. Concluding Remarks

It is clear that a systematic and general approach is needed for the evaluation of uncertainty representation and reasoning. Hence, it is necessary to clearly define at which points uncertainty enter into the reasoning and representation processes of an information fusion system, how information flows and how it is transformed within the fusion system. To this end, the ADP formalises these aspects. The casting of BN information fusion into the ADP framework assists greatly in facilitating an understanding of the objects that are under evaluation as well as which URREF criteria are relevant in which parts of the information fusion processing and decision chain.

Several evaluation criteria are absent in traditional BN evaluation and validation methods when compared the URREF ontology. By the same measure, there may be aspects of BN evaluation that are not currently addressed by the URREF ontology. The authors thus propose the following tasks to address these shortcomings:

1) Carefully identify the purpose of a BN in a specific fusion setting. What are the objectives of the BN? Is it used to gain understanding about the problem domain as well as the effect of interventions? Is it used to fuse data from heterogeneous sources? Is it used to fill in missing information? Is is used to combine prior information and data in a causal way such as to reduce uncertainty? Is it used to mitigate the over fitting of data?
2) Once the purpose of the the BN has been characterised, the *relevant* URREF criteria should be identified. Corresponding metrics that evaluate these criteria should be identified, and where they do not exist, they should be developed.
3) If there are important objectives to be achieved by a BN that are not addressed by the URREF framework, they should be included in the ontology and corresponding metrics should be identified or developed.

It is important to concede that the although URREF should lend itself to be applied generally, the user should realise that only the URREF criteria that are relevant to the objectives of the fusion system (in this case the BN) should be retained.

References

[1] E. Blasch, "Sensor, user, mission (sum) resource management and their interaction with level 2/3 fusion," in *Information Fusion, 2006 9th International Conference on*, July 2006, pp. 1–4.

[2] E. Waltz and J. Llinas, *Multisensor Data Fusion Systems*. Artech House, 1990, ch. System modelling and performance evaluation.

[3] R. DeWitt, "Principles for testing a data fusion system," PSR (Pacific Sierra Research, Inc), Internal report, March 1998.

[4] H. D. Liggins M.E. and J. Llinas, *Handbook of Multisensor Data Fusion*, 2nd ed. CRC Press, 2009, ch. Assessing the performance of multisensor fusion processes, pp. 655–675.

[5] E. Blasch, P. Valin, and E. Bossé, "Measures of effectiveness for high-level fusion." in *FUSION*. IEEE, 2010, pp. 1–8.

[6] P. Costa, K. Laskey, E. Blasch, and A. L. Jousselme, "Towards unbiased evaluation of uncertainty reasoning: The URREF ontology," in *Information Fusion (FUSION), 2012 15th International Conference on*, July 2012, pp. 2301–2308.

[7] M. Johnson and K. Chang, "Quality of information for data fusion in net centric publish and subscribe architectures," in *Information Fusion, 2005 8th International Conference on*, vol. 2, July 2005, pp. 8 pp.–.

[8] B. Yu and K. Sycara, "Learning the quality of sensor data in distributed decision fusion," in *Information Fusion, 2006 9th International Conference on*, July 2006, pp. 1–8.

[9] G. Rogova and E. Bossé, "Information quality in information fusion," in *Information Fusion (FUSION), 2010 13th Conference on*, July 2010, pp. 1–8.

[10] J. van Laere, "Challenges for if performance evaluation in practice," in *Information Fusion, 2009. FUSION '09. 12th International Conference on*, July 2009, pp. 866–873.

[11] A. Jousselme 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.

[12] A. L. Jousselme. (2014, January) Objects under evaluation (evaluation subjects). [Online]. Available: http://eturwg.c4i.gmu.edu/?q=t8

[13] B. G. Marcot, "Metrics for evaluating performance and uncertainty of Bayesian network models," *Ecological Modelling*, vol. 230, no. 0, pp. 50 – 62, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304380012000245

[14] H. Koen, J. de Villiers, A. de Waal, G. Pavlin, F. Mignet, P. de Oude, and J. Roodt, "In press: A framework for inferring predictive distributions of rhino poaching events through causal modelling," in *Information Fusion (FUSION), 2014 17th International Conference on*. ISIF, July 2014.

[15] Y. Peng, S. Zhang, and R. Pan, "Bayesian network reasoning with uncertain evidences," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 18, no. 05, pp. 539–564, 2010.

[16] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[17] H. Chan and A. Darwiche, "On the revision of probabilistic beliefs using uncertain evidence," in *IJCAI*, 2003, pp. 99–105.

[18] G. Pavlin, P. de Oude, M. Maris, J. Nunnink, and T. Hood, "A multi-agent systems approach to distributed bayesian information fusion," *Information Fusion*, vol. 11, no. 3, pp. 267–282, 2010.

[19] F. V. Jensen, *Bayesian Networks and Decision Graphs*. Springer-Verlag, 2001.

[20] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," *CoRR*, vol. abs/1301.6725, 2013.

[21] K. B. Korb and A. E. Nicholson, *Bayesian artificial intelligence*, 2nd ed. Boca Raton, FL: CRC Press, 2011.

[22] P. Costa. URREF ontology - version 2.0. Evaluation of Technologies for Uncertainty Representation Working Group (ETURWG). [Online]. Available: http://eturwg.c4i.gmu.edu/?q=node/43

[23] J. Pearl, *Causality: models, reasoning, and inference*, 2nd ed. Cambridge, U.K.; New York: Cambridge University Press, 2009.

[24] R. Carvalho, P. Costa, K. Laskey, and K. Chang, "PROGNOS: predictive situational awareness with probabilistic ontologies," in *2010 13th Conference on Information Fusion (FUSION)*, Jul. 2010, pp. 1–8.

[25] P. P. Shenoy, "Valuation-Based Systems: A framework for managing uncertainty in expert systems," in *Fuzzy Logic for the Management of Uncertainty*, L. A. Zadeh and J. Kacprzyk, Eds. John Wiley and Sons, 1992, pp. 83–104.

[26] E. Blasch, E. Bosse, and D. Lambert, *High-Level Information Fusion Management and Systems Design*, ser. Artech House intelligence and information operations series. Artech House, 2012. [Online]. Available: http://books.google.co.za/books?id=rMG0Ks9gst0C

[27] S. L. Rinderknecht, M. E. Borsuk, and P. Reichert, "Bridging uncertain and ambiguous knowledge with imprecise probabilities," *Environ. Model. Softw.*, vol. 36, pp. 122–130, Oct. 2012. [Online]. Available: http://dx.doi.org/10.1016/j.envsoft.2011.07.022